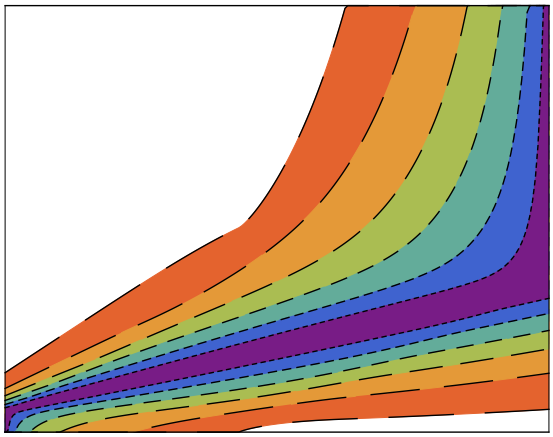
Monotonic Probability Distribution

Characterisation, Measurements under Prior Information, and Application

Dissertationsschrift zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
an der Julius-Maximilians-Universität Würzburg



vorgelegt von

Wolfgang Sans

aus Würzburg

Würzburg 2019

Eingereicht am 24.07.2018

1. Gutachter: Prof. Dr. Rainer Göb

2. Gutachter: Prof. Dr. Sven Knoth

Tag der Disputation: 14.12.2018

Diese Arbeit widme ich
Elart von Collani (1944–2017).
Ein Mentor und Freund.

Abstract

Statistical Procedures for modelling a random phenomenon heavily depend on the choice of a certain family of probability distributions. Frequently, this choice is governed by a good mathematical feasibility, but disregards that some distribution properties may contradict reality. At most, the choosen distribution may be considered as an approximation. The present thesis starts with a construction of distributions, which uses solely available information and yields distributions having greatest uncertainty in the sense of the maximum entropy principle. One of such distributions is the monotonic distribution, which is solely determined by its support and the mean. Although classical frequentist statistics provides estimation procedures which may incorporate prior information, such procedures are rarely considered. A general frequentist scheme for the construction of shortest confidence intervals for distribution parameters under prior information is presented. In particular, the scheme is used for establishing confidence intervals for the mean of the monotonic distribution and compared to classical procedures. Additionally, an approximative procedure for the upper bound of the support of the monotonic distribution is proposed. A core purpose of auditing sampling is the determination of confidence intervals for the mean of zero-inflated populations. The monotonic distribution is used for modelling such a population and is utilised for the procedure of a confidence interval under prior information for the mean. The results are compared to two-sided intervals of Stringer-type. Bayesian statistics feature the ability of incorporating prior information. Bayesian procedures and the here presented frequentist procedure for shortest confidence intervals share some similarities. A comparison with respect to the interval estimation of the mean of the exponential distribution is carried out.

Zusammenfassung

Statistische Verfahren zur Modellierung eines zufälligen Phänomens hängen stark von der Wahl einer bestimmter Familie von Wahrscheinlichkeitsverteilungen ab. Oft wird die Auswahl der Verteilung durch das Vorliegen guter mathematischer Handhabbarkeit bestimmt, dabei aber außer Acht gelassen, dass einige Verteilungseigenschaften gegen die Realität verstoßen können und bestenfalls als Näherung aufgefasst werden können. Die vorgelegte Arbeit beginnt mit einer Konstruktion von Verteilungen, die ausschließlich verfügbare Informationen verwenden und im Sinne des Prinzips der maximalen Entropie die größte Unsicherheit beinhalten. Eine dieser Verteilungen ist die monotone Verteilung, die alleine durch ihren Träger und den Mittelwert festgelegt ist. In der klassischen, frequentistischen Statistik existieren zwar Verfahren zur Schätzung von Verteilungsparametern, die Vorinformationen verarbeiten können, sie finden aber kaum Beachtung. Es wird ein allgemeines frequentistisches Verfahren zur Konstruktion kürzester Konfidenzintervalle für Verteilungsparameter unter Vorinformation vorgestellt. Dieses Verfahren wird zur Herleitung von Konfidenzintervallen für das erste Moment der monotonen Verteilung angewendet, und diese mit klassischen Bereichsschätzern verglichen. Außerdem wird ein approximatives Schätzverfahren für die obere Grenze des Trägers der Monotonen Verteilung vorgeschlagen. Ein Hauptziel der Wirtschaftsprüfung ist die Bestimmung von Konfidenzintervalle für Mittelwerte von Grundgesamtheiten zu bestimmen, die viele Nullen enthalten. Die monotone Verteilung geht in die Modellierung einer solchen Grundgesamtheit und in das Verfahren für ein Konfidenzintervall unter Vorinformation zur Schätzung des Mittelwerts ein. Die Ergebnisse werden mit zweiseitigen Intervallen vom Stringer-Typ verglichen. Die bayessche Statistik zeichnet sich dadurch aus, dass Vorinformationen berücksichtigt werden können. Die bayesschen Verfahren haben eine gewisse Ähnlichkeit zu dem frequentistischen Verfahren der kürzesten Konfidenzintervalle. Ein Vergleich wird bezüglich der Intervallschätzung des Mittelwerts der Exponentialverteilung angestellt.

Contents

1. Introduction

We demand rigidly defined areas of doubt and uncertainty!

Douglas Adams
"The Hitchhiker's Guide to the Galaxy"

The world's evolution is anything but deterministic – it is subject to randomness. Randomness means that any repetition of a given process results in one of a set of different values. In other words, the future development of any process is always uncertain, even if the initial conditions would be exactly known. However, for man, the initial conditions are never known exactly and the resulting ignorance increases the human uncertainty generated by randomness. Thus, human uncertainty about the future has two sources: one is the ignorance about facts from the past, i.e., apart from a certain set of possibilities they are unknown, and the other is the randomness of the future's development, i.e., again apart from a set of possibilities and some knowledge about the structure of randomness the development does not follow deterministic laws.

To be successful, man must make decisions taking into account future developments and, therefore, predictions about the future are necessary. Since humanity exists the art of making predictions was one of the most important issues and many different techniques were developed and applied in the history of mankind. Modern science, which was developed during the last 500 years, represents one of the latest solutions for making predictions. However, modern science is essentially based on the assumption of a deterministic future development and as a consequence scientific predictions never come true and thus may lead to wrong decisions.

To overcome this pure deterministic description of future development, statistics has been developed aiming at describing randomness for making statements about the initial conditions.

A strictly stochastic framework for science has been developed by Jakob Bernuolli

in *Ars conjectandi* [3][1]. Bernoulli's approach was taken up by von Collani in [19] and [20], who derived a stochastic model which describes the uncertain future taking into account both the ignorance about the past and the (structured) randomness of the future. This type of stochastic model enables to deduce prediction procedures yielding in a given situation reliable and accurate predictions based on a unique conceptual framework. Any prediction consists of a future event, i.e., a set of possible outcomes, and its most important properties are the probability, i.e., the degree of certainty, that it actually will occur, and its accuracy, i.e., the size of the predicted event.

Having prediction procedures at hand, von Collani [19, 20] derived measurement, i.e., estimation procedures to reduce the ignorance about initial conditions. Similar to prediction procedures, measurement procedures should meet requirements with respect to reliability, i.e., the probability of yielding correct results, and accuracy, that is, the amount of ignorance reduction.

The aim of this dissertation is twofold: First, the statistical approach with respect to prediction and, especially, measurement procedures is investigated. It is shown, that not all conventional statistical models and the corresponding statistical procedures are suitable to derive reliable and accurate prediction and, consequently, measurement procedures. Second, a stochastic model of uncertainty named Bernoulli-Space is presented which enables reliable and accurate prediction procedures and measurement procedures. One integral part of the Bernoulli-Space is named ignorance, which would be named prior information in conventional statistics. That is, the derived measurement procedures from the Bernoulli-Space take prior information into account. The Bernoulli-Space is developed and applied in the case of monotonic probability distributions.

Monotonic probability distributions defined by a monotonic probability density function[2] (PDF) play in form of the *exponential distribution* an important role in

[1] A short overview of Bernoulli's life and the *Ars conjectandi* can be found in [21].

[2] Usually, the term probability density function (PDF) is used for a continuous probability distribution, whereas in the discrete case, the term probability mass function (PMF) is established. Since both, PDF and PMF, are densities in the mathematical sense, one with respect to the Borel or Lebesgue measure, the other to a counting measure, we use the term PDF. Only if it is necessary, we will distinguish between those terms.

many fields of applications related to the concept of "lifetime" (see [39], pp. 494 et seqq.). The exponential distribution belongs to a continuous random random variable and is characterized by a constant failure rate, i.e., the probability of an instantaneous failure does not change during the entire lifetime. The use of the exponential distribution is especially popular in the following areas of applications:

- In *engineering*, in particular in *reliability engineering* as a lifetime distribution of technical devices and systems: Although, almost all technical devices do not have a constant failure rate over their whole span of life, especially not at the beginning and at the end of their operation time, the exponential distribution is assumed as an approximation for a certain time period between start and end of the lifetime.
- In *queuing theory* as the distributions of the inter-arrival time and the processing time: In this case the exponential distribution is selected because of its simplicity. Examples may be found in the field of planning the capacity of computer servers, call-centers etc.
- In *physics* as the distribution of the time of radioactive decay.
- In *economics* as the time until a default of a mortgage or a debt (this is similar to modelling a lifetime in reliability engineering).
- In *technical simulations*, e.g. simulations of train schedules as the distribution of delay times.

There are two closely connected aspects of the exponential distribution, which obviously conflict with reality, i.e., its infinite support and the so-called memory-less property. The memory-less property may be met approximately in reality. However, an infinite support can in no case be considered as an approximation of a finite one. These unrealistic assumptions are avoided here by admitting a bounded support of the distribution which also excludes the memory-less property.

As in the above exemplary fields of applications, monotonic distributions occur for simple, i.e., not compound processes. A simple process is characterised by the fact that it cannot be divided into two or more different processes. In this sense,

a simple process cannot be split up into a sequence of sub-processes, where the distribution of each later process depends on the outcomes of the preceding processes. Randomness of simple processes may be modelled by uniform probability distributions, by monotonic probability distributions or uni-modal probability distributions.

The dissertation is organized in three parts:

- The first part consists of two chapters and is devoted to the introduction of some general concepts of stochastic theory and the specific properties of monotonic distributions. In Chapter 2 we introduce a general concept of choosing probability distributions according to some qualitative properties. We identify properties which enable us to derive quantitative properties with which the probability distributions may be calculated: the moment equations. Since the same equations may be found in the theory of maximum entropy distributions, we sketch its basic idea which was firstly introduced by Shannon [70]. In Chapter 3 we give an extensive characterisation of the continuous monotonic probability distribution, which helps us to derive measurement and prediction procedures in the following chapters. Some of the presented results were published in the *Economic Quality Control* journal 2003 [67].
- The second part describes in Chapter 4 the classical methods of inference used in statistics. The most commonly used point and set estimators are given together with the corresponding evaluation criterions. A majority of this chapter was published in the *Economic Quality Control* journal 2013 [69]. In Chapter 5 we apply the methods of Chapter 4 to the monotonic probability distribution. It is shown that especially when the support is subject to inference, most of those methods are not suitable at all.
- The third part is devoted to a new way of scientific modelling. In Chapter 6 the concept of the Bernoulli-Space is presented. The Bernoulli-Space fully describes the uncertainty aspect of a real world phenomenon by taking into account the always existant ignorance about the past as well as the randomness with respect to the future. Though, ignorance always exists, it does

not mean that we know nothing at all. The Bernoulli-Space incorporates the existing knowledge and makes it utilisable in stochastic procedures. The derived stochastic measurement and prediction procedures are exemplified by the monotonic probability distribution. The chapter concludes with the possible application of the Bernoulli-Space to statistical distributions. Chapter 7 fully derives the measurement procedure for the expectation $\mathrm{E}[X]$ of a random variable X distributed according to the monotonic probability distribution. The procedure is compared with statistical ones when the approximation with the exponential distribution is considered. The chapter ends with an approximative measurement procedure for the upper bound of the range of variability of X. Chapter 8 presents an application of the afore derived measurement procedure for $\mathrm{E}[X]$ in the field of audit sampling. Since the Bernoulli-Space and the derived procedures seems to have similarities to the Bayesian approach, a comparison appeared to be necessary. Therefore, Chapter 9 compares both concepts by deriving measurement procedures for the expectation of an exponential distributed random variable.

Besides the introduction of the Bernoulli-Space as the model of uncertainty, the dissertation also aims at showing the importance to select a realistic, i.e., bounded support of a probability distribution. Actually, in many applications, particularly when safety is involved, the bounds of the support are of primary interest. Assuming in these cases an infinite support lets the problem to determine a safety bound become more or less insolvable.

All calculations were performed with Mathematica [79].

2. Selection of an Univariate Probability Distribution

An appropriate selection of a probability distribution constitutes a major problem. Often a distribution function is selected which fits well to the observed data. In other cases the availability of statistical methods determines the model. Finally, in many fields of application there are certain traditional practices concerning the selection of probability distributions. Sometimes, more than one of those reasons governs the selection.

This chapter starts with an application of the Weierstrass Approximation Theorem to formulate an approximation theorem for exponential functions $e^{p(x)}$, where $p(x)$ is a polynomial. As those functions are strictly positive, they may serve – after a normalisation – as probability density functions (PDF). It is outlined that the real world's present state d governs an actual probability distribution (random law). Shown by Hausdorff, every univariate PDF on a bounded support is determined by its sequence of moments. This establishes the connection of the present state d and the distribution's moments. Finally, the available qualitative properties of a PDF are translated into quantitative properties in the form of a minimal necessary sequence of moments. This leads to the definition of PDFs denoted as *Minimum Information Representatives* (MIR). Since MIR shares the same form as distributions derived from the Maximum Entropy Principle, the latter is outlined. The three most basic types of PDFs are derived from their respective qualitative properties.

2.1. A Unique Representation

The Weierstrass Approximation Theorem [76] states for every real-valued continuous function $f(x)$ on $[a, b]$ that a polynomial $p(x)$ exists such that

$$|f(x) - p(x)| < \varepsilon \tag{2.1}$$

for all $x \in [a, b]$ and for any given $\varepsilon > 0$. Applied to a probability density function $f_X(x)$ for a random variable X with a bounded and closed range of variability $\mathcal{X} \subset \mathbb{R}$ this means that it can be approximated by a polynomial $p(x)$ to any degree of accuracy. Especially in the discrete case with a finite range of variability it is even possible to give explicitly a polynomial, which exactly fits $f_X(x)$ at the elements of $\mathcal{X}$.

Univariate probability density functions are by definition nonnegative functions on the real line. For every element of the range of variability $\mathcal{X}$ of a random variable X they are strictly positive and zero for any other element[1]. The range of variability $\mathcal{X}$ of X is equal to the support of the related probability density function. Approximating a probability density function with a polynomial may lead to a polynomial with negative values even though the approximation achieves a required accuracy. This problem is solved by using exponential functions for approximation with a polynomial $\widetilde{p}(x)$ as exponent:

$$e^{\widetilde{p}(x)} \tag{2.2}$$

The following theorem is formulated for a continuous function but can be easily applied to a discrete function, as mentioned above.

Theorem 2.1 (Approximating PDF) *Let f be a continuous, positive and bounded function on $[a, b]$ and $\varepsilon > 0$ given, then there exists a polynomial p on $[a, b]$ such that*

$$|f(x) - e^{p(x)}| < \varepsilon \tag{2.3}$$

holds for all $x \in [a, b]$.

Proof of 2.1: As f is a bounded function there exists a $M > 0$ such that $M \geq f(x)$ for all $x \in [a, b]$. Let $\varepsilon > 0$ and $\widetilde{\varepsilon} := \frac{\varepsilon}{M+\varepsilon} > 0$. Define the function $\widetilde{f}$ as

$$\widetilde{f}(x) = \ln f(x) \qquad \text{for all } x \in [a, b]\,. \tag{2.4}$$

[1] Concerning the bounds of $\mathcal{X}$ it may serve useful to include them in $\mathcal{X}$. Since discrete points have zero Borel or Lebesgue measure, this states no problem.

$\widetilde{f}$ is a continuous function on the bounded interval $[a,b]$. The Weierstrass Approximation Theorem implies that for $\widetilde{\varepsilon}$ there exists a polynomial q on $[a,b]$ such that

$$|\widetilde{f}(x) - q(x)| < \widetilde{\varepsilon} \tag{2.5}$$

for all $x \in [a,b]$.
Inequation (2.5) can be written as

$$\widetilde{f}(x) - \widetilde{\varepsilon} \quad < \quad q(x) \quad < \quad \widetilde{f}(x) + \widetilde{\varepsilon} \tag{2.6}$$

$$\Leftrightarrow \quad \underbrace{e^{\widetilde{f}(x)}}_{=f(x)} e^{-\widetilde{\varepsilon}} \quad < \quad e^{q(x)} \quad < \quad \underbrace{e^{\widetilde{f}(x)}}_{=f(x)} e^{\widetilde{\varepsilon}} \tag{2.7}$$

$$\Leftrightarrow \quad e^{-\widetilde{\varepsilon}} \quad < \quad \frac{e^{q(x)}}{f(x)} \quad < \quad e^{\widetilde{\varepsilon}} \tag{2.8}$$

- Left side of inequation (2.8).
  The following relation for the exponential function is well known:

  $$e^x > 1 + x \quad \text{for all } x \in \mathbb{R}\,. \tag{2.9}$$

  Applied to $e^{-\widetilde{\varepsilon}}$ it follows

  $$1 - \widetilde{\varepsilon} < e^{-\widetilde{\varepsilon}} \tag{2.10}$$

  and therefore

  $$1 - \widetilde{\varepsilon} < \frac{e^{q(x)}}{f(x)} \tag{2.11}$$

  This can be transformed to

  $$f(x) - e^{q(x)} < \widetilde{\varepsilon} f(x) \leq \widetilde{\varepsilon} M = \varepsilon \frac{M}{M+\varepsilon} < \varepsilon \tag{2.12}$$

- Right side of inequation (2.8).
  $\varepsilon > 0$ and $M > 0$ yield that $\widetilde{\varepsilon} = \frac{\varepsilon}{M+\varepsilon} < 1$ and with the relation

  $$e^x < \frac{1}{1-x} \qquad \text{for} \quad x < 1 \tag{2.13}$$

  we get

  $$\frac{e^{q(x)}}{f(x)} \quad < \quad e^{\widetilde{\varepsilon}} \quad < \frac{1}{1-\widetilde{\varepsilon}} \tag{2.14}$$

  $$\Leftrightarrow \quad e^{q(x)} - f(x) \quad < \quad \widetilde{\varepsilon} e^{q(x)} \quad \overset{(2.6)}{<} \widetilde{\varepsilon} e^{\widetilde{f}(x)+\widetilde{\varepsilon}} = \widetilde{\varepsilon} e^{\widetilde{\varepsilon}} f(x) < \widetilde{\varepsilon} e^{\widetilde{\varepsilon}} M \tag{2.15}$$

  $$e^{q(x)} - f(x) \quad < \quad \widetilde{\varepsilon} e^{\widetilde{\varepsilon}} M \quad < \frac{\widetilde{\varepsilon}}{1-\widetilde{\varepsilon}} M = \varepsilon \tag{2.16}$$

  $$\Longrightarrow \quad -\varepsilon \quad < \quad f(x) - e^{q(x)} \tag{2.17}$$

Relations (2.12) and (2.17) together and with $p(x) \equiv q(x)$ it follows

$$|f(x) - e^{p(x)}| < \varepsilon \,. \tag{2.18}$$

•

2.1.1. Continuous Approximations

Those values of a random variable representing an aspect of a real random phenomenon, which may actually be observed, form necessarily a finite, that is, discrete set of numbers. This set is called *range of variability of* X denoted by $\mathcal{X}$. With respect to the *probability mass function* f_X, the range of variability $\mathcal{X}$ is equal to the *support of* f_X.

The probability $\mathrm{P}_X(\{x\})$ of each possible outcome x is trivially bounded by 1. The case that a singleton $\{x\}$ has probability 1 represents the degenerated or deterministic case and, therefore, can be excluded from stochastic analysis. For this reason we require that the range of variability $\mathcal{X}$ of a random variable has more than one element.

As every range of variability $\mathcal{X}$ of a random variable X is discrete, any continuous mathematical description constitutes an approximation, notwithstanding that continuous models are generally assumed because they offer mathematical benefits. The question arises how to derive a continuous approximation starting from a discrete model.

The discrete and bounded range of variability $\mathcal{X}$ is a basic property of a random variable X representing a real world aspect. The first step, therefore, is to approximate the discrete range of variability $\mathcal{X}$ by a continuous one denoted by $\widetilde{\mathcal{X}}$. In the univariate case a natural continuous approximation is given by

$$\widetilde{\mathcal{X}} = \{x | \min \mathcal{X} \leq x \leq \max \mathcal{X}\} \tag{2.19}$$

Next the probability distribution function F_X has to be approximated by an absolutely continuous function $F_{\widetilde{X}}$, which should meet some criteria, for instance

$$\begin{aligned} &\text{for } \varepsilon > 0: \ \lim_{\varepsilon \to 0} F_{\widetilde{X}}(\min \mathcal{X} - \varepsilon) = 0 \ , \ \lim_{\varepsilon \to 0} F_{\widetilde{X}}(\max \mathcal{X} + \varepsilon) = 1 \\ &\text{for } \delta > 0: \ \max_{x \in \mathcal{X}} \left| F_X(x) - F_{\widetilde{X}}(x) \right| < \delta \end{aligned} \tag{2.20}$$

As $F_{\widetilde{X}}$ is absolutely continuous its derivative exists and we obtain

$$F_{\widetilde{X}}(x) = \int_{\min \widetilde{\mathcal{X}}}^{x} f_{\widetilde{X}}(y)dy \qquad \text{for} \quad x \in \widetilde{\mathcal{X}} \tag{2.21}$$

Of course, a continuous approximation only makes sense, if the original discrete support consists of a sufficiently large number of elements.

In statistics the actual discrete case is not approximated in the way described above, but a continuous model is selected generally without considering the real situation. For instance, the discrete range of variability is replaced often by the set of real numbers $\mathbb{R}$, and the probability distribution selected is the normal distribution.

2.1.2. Moment Equations

Theorem (2.1) implies that every PDF with bounded support of a random variable X can be expressed as an exponential function

$$f_X(x) \quad = \quad e^{\beta_0+\beta_1 x+\beta_2 x^2+\beta_3 x^3+\ldots} \tag{2.22}$$

The actual probability distribution of a random variable X is determined by *initial conditions*, which describe the present state and situation, respectively, of a real world phenomenon, of which X represents an aspect. For instance, initial conditions may be *current temperature, atmospheric pressure, force of gravitation, motivation and education of workers in a production process etc.* The list of initial conditions could be continued ad infinitum, because no real world phenomenon could be considered to be independent from the rest of the world. Of course, some initial conditions have a bigger impact on the phenomenon of interest than others. Initial conditions describing the state and situation, respectively, of a real world phenomenon are given by *situation based deterministic variables*, denoted with D. One actual situation (i.e., an infinite number of actual aspects) is denoted by d, and $\mathcal{D}$ stands for a set of actual situations.

As $\mathcal{X}$ and the PDF f_X of X are completely determined by the actual situation d we will write $\mathcal{X}(\{d\})$ and $f_{X|\{d\}}$ in the following to express this dependency:

$$f_{X|\{d\}}(x) \quad = \quad e^{\beta_0(d)+\beta_1(d)x+\beta_2(d)x^2+\beta_3(d)x^3+\ldots} \tag{2.23}$$

In the following we call $(\beta_0(d), \beta_1(d), \beta_2(d), \ldots)$ *sequence of exponential coefficients* and write $\beta(d)$:

$$\beta(d) = (\beta_0(d), \beta_1(d), \beta_2(d), \ldots) \tag{2.24}$$

As shown in [34, 35] in the univariate case a PDF $f_{X|\{d\}}$ with bounded support $\mathcal{X}(\{d\})$ of a random variable $X|\{d\}$ is completely determined by the sequence of values of its moments $E\left[X^k|\{d\}\right]$, $k = 0, 1, 2, \ldots$:

$$\begin{aligned} &(\mu_0(d), \mu_1(d), \mu_2(d), \ldots) \\ \text{with} \qquad &\mathrm{P}_X(\mathcal{X}(\{d\})) = \mu_0(d) = 1 \end{aligned} \tag{2.25}$$

The moment $\mu_0(d)$ can be interpreted as the necessary condition on a proper PDF. Since both, the sequence of moments $(\mu_0(d), \mu_1(d), \mu_2(d), \ldots)$ together with $\mathcal{X}(\{d\})$ and the actual situation d completely determine $f_{X|\{d\}}$, we may conclude, that there is an invertible mapping g between them

$$\begin{aligned} g(d) &= (\mu_0(d), \mu_1(d), \mu_2(d), \ldots) \\ d &= g^{-1}(\mu_0(d), \mu_1(d), \mu_2(d), \ldots) \end{aligned} \tag{2.26}$$

In the way that the initial conditions are values of a situation based deterministic variable, we call the sequence of moments *distribution based deterministic variable* and write $d^{(m)}$:

$$d^{(m)} = (\mu_0(d), \mu_1(d), \mu_2(d), \ldots) \tag{2.27}$$

The exponential coefficients $\beta_i(d)$ in (2.23) are obtained as unique solutions of the following system of equations:

$$\mu_i(d) = \int_{\mathcal{X}(\{d\})} x^i e^{\sum_{k=0}^{\infty} \beta_k(d) x^k} dx \qquad \text{for} \quad i = 0, 1, 2, \ldots \tag{2.28}$$

$$\text{with} \qquad \mu_0(d) = 1 \tag{2.29}$$

The system of equations (2.29) provides a mapping h between the sequences of moments $d^{(m)} = (\mu_0(d), \mu_1(d), \mu_2(d), \ldots)$ of a random variable $X|\{d\}$ and the sequences of exponential coefficients $\beta(d) = (\beta_0(d), \beta_1(d), \beta_2(d), \ldots)$ of its PDF:

$$\begin{aligned} &h(d^{(m)}) = h(g(d)) = \beta(g^{-1}(d^{(m)})) = \beta(d) \\ &\qquad \text{has for every value } \beta(d) \text{ exactly one solution } d^{(m)} \end{aligned} \tag{2.30}$$

Although the exponential coefficients $\beta_i(d)$ are determined by the sequence of moments $d^{(m)}$, which are more or less easy accessible, their meanings are not as clear as the meaning of location, shape or scale parameters of many distributions used in classical statistics. Nevertheless, the sequence of exponential coefficients $\beta(d) = (\beta_0(d), \beta_1(d), \beta_2(d), \ldots)$ together with $\mathcal{X}(\{d\})$ also completely determines a probability distribution and, therefore, is a deterministic variable as well, which we call *exponential coefficient based deterministic variable*.

When confusions can be excluded, for the sake of simplicity we will just write β_i and μ_i instead of $\beta_i(d)$ and $\mu_i(d)$, respectively.

2.2. Minimum Information Representative

The range of variability $\mathcal{X}(\{d\})$ and the sequence of values $(\mu_0, \mu_1, \mu_2, \ldots)$ of the moments completely determine the probability distribution (see 2.1.2). Of course, it is generally not possible to get full knowledge of the values of an infinite number of moments. But rather than to try to get knowledge of as many as possible values of the moments, it seems to be more economic to determine the number of moments of which the values should be necessarily known in order to describe the essential – and therefore qualitative – properties of the probability distribution sufficiently well. Also worth mentioning is the fact, that almost all known probability distributions are uniquely determined by a finite number of moments.

To this end the essential properties of a probability distribution have to be identified. In [19] the essential properties are given by the number and type of extremes of the PDF. Besides this, other properties like the range of variability, monotony or symmetry are essential properties, too. The probability distributions which meet the same essential properties E form a family of distributions $\mathcal{P}^{(E)}$.

In this manner 'sufficiently well' means to find a probability distribution of form (2.23) with an exponent of finite degree which meets the essential properties E and is as simple as possible, i.e., the degree of the polynomial should be as small as possible. We name these simplest probability distributions *Minimum Information Representatives* (MIR). Therefore, a MIR is the simplest member of

a family of distributions $\mathcal{P}^{(E)}$. The *information* refers to the essential properties and *minimum* refers to the minimal degree of the exponent with which it is possible to meet the essential properties and, therefore, to the number of necessary moments to obtain the exponential coefficients.

Remark: Consider two different probability distributions f and g. Both satisfy some essential properties E. Thus, both are members of $\mathcal{P}^{(E)}$. Let f be the MIR of $\mathcal{P}^{(E)}$, k be the polynomial degree of the exponent of f, and m be the polynomial degree of the exponent of g. Then, $m > k$ has to hold, since g cannot be the MIR of $\mathcal{P}^{(E)}$ as well. That is, the determination of the exponential coefficient of g would need more moment equations.

Example 2.2 *For illustration consider a continuous random variable $X|\{d\}$ with a bounded range of variability $\mathcal{X}(\{d\}) = \{x \,|\, x_1 \leq x \leq x_N\}$. Assume that it is known, that for the given initial conditions d the PDF $f_{X|\{d\}}$ has exactly one maximum value in $x_M \in X|\{d\}$ and for other values $x \in \mathcal{X}(\{d\})$ the PDF $f_{X|\{d\}}$ decreases with increasing distance to x_M:*

$$\forall\, x \in \mathcal{X}(\{d\}), x \neq x_M \quad : \quad f_{X|\{d\}}(x_M) > f_{X|\{d\}}(x) \tag{2.31}$$

$$\forall y, z \in \mathcal{X}(\{d\}),\, x_1 \leq y < z < x_M \quad : \quad f_{X|\{d\}}(y) \leq f_{X|\{d\}}(z) < f_{X|\{d\}}(x_M) \tag{2.32}$$

$$\forall y, z \in \mathcal{X}(\{d\}),\, x_M < y < z \leq x_N \quad : \quad f_{X|\{d\}}(x_M) > f_{X|\{d\}}(y) \geq f_{X|\{d\}}(z) \tag{2.33}$$

The simplest density function of form (2.23) which reflects the known property has as exponent a polynomial exactly of degree two:

$$f_X(x) = e^{\beta_0 + \beta_1 x + \beta_2 x^2} \tag{2.34}$$

with

$$\beta_2 \neq 0 \qquad \text{and} \qquad \int_{\mathcal{X}(\{d\})} f_X(x)dx = 1 \tag{2.35}$$

A possible way to determine the three exponential coefficients β_0, β_1 and β_2 in (2.34) is to solve three moment equations of (2.28) *and* (2.29)*. To know which three equations to be solved we have to note that the impact of the moments on the probability distribution becomes smaller the higher the order of the moments is. Moreover, moments of lower order may be determined much easily and more*

accurately than those of higher order. Thus, it is reasonable to use the first and second moment for determining the values of β_i, $i = 0, 1, 2$*:*

$$\mu_i = \int_{\mathcal{X}(\{d\})} x^i e^{\beta_0(\mu_0,\mu_1,\mu_2)+\beta_1(\mu_0,\mu_1,\mu_2)x+\beta_2(\mu_0,\mu_1,\mu_2)x^2}\,\mathrm{d}x \quad \textit{for} \quad i = 1, 2 \tag{2.36}$$

$$\textit{and} \quad 1 = \int_{\mathcal{X}(\{d\})} e^{\beta_0(\mu_0,\mu_1,\mu_2)+\beta_1(\mu_0,\mu_1,\mu_2)x+\beta_2(\mu_0,\mu_1,\mu_2)x^2}\,\mathrm{d}x \tag{2.37}$$

Notation: In the following we will write $\mu^{(k)}$ to express the sequence of the first $k+1$ consecutive moments starting with μ_0, i.e., $\mu^{(k)} = (\mu_0, \mu_1, \mu_2, \ldots, \mu_k)$. In the same way $\beta^{(k)}$ stands for $(\beta_0, \beta_1, \beta_2, \ldots, \beta_k)$.

Still, the sequence of values of moments and the range of variability of the random variable $X|\{d\}$, represented by

$$d^{(m)} = (\mu_0(d), \mu_1(d), \mu_2(d), \mu_3(d), \ldots) = (\mu^{(2)}(d), \mu_3(d), \ldots)\,,$$

determine uniquely the PDF. But the Minimum Information Representative (2.34) is only an approximation of the true PDF, therefore, $\beta_i(\mu^{(2)})$, $i = 0, 1, 2$, are different from $\beta_i(d)$, $i = 0, 1, 2$, the first three components of the sequence of exponential coefficients $\beta(d) = (\beta_0(d), \beta_1(d), \beta_2(d), \ldots)$ of the true PDF.

Figure 2.1 shows the different deterministc variables d, $d^{(m)}$ and β, the necessary moments $\mu^{(k)}$ and exponential coefficients $\beta^{(k)}$ of the Minimum Information Representative and the mappings which exist between them.

For instance, with the Minimum Information Representative f_X derived with $\mu^{(k)}$, i.e., the range of variability and the first k moments, it is possible to calculate $E[X^{k+1}]$:

$$E[X^{k+1}] = \int_{\mathcal{X}(\{d\})} x^{k+1} f_X(x)\,\mathrm{d}x\,, \tag{2.38}$$

which in general is different from the true probability distribution's $(k+1)$-th moment.

As the necessary μ_i's, and therefore $\mu^{(k)}$ completely determine the approximation, we write

$$f_{X|\{\mu^{(k)}\}} \tag{2.39}$$

in the following.

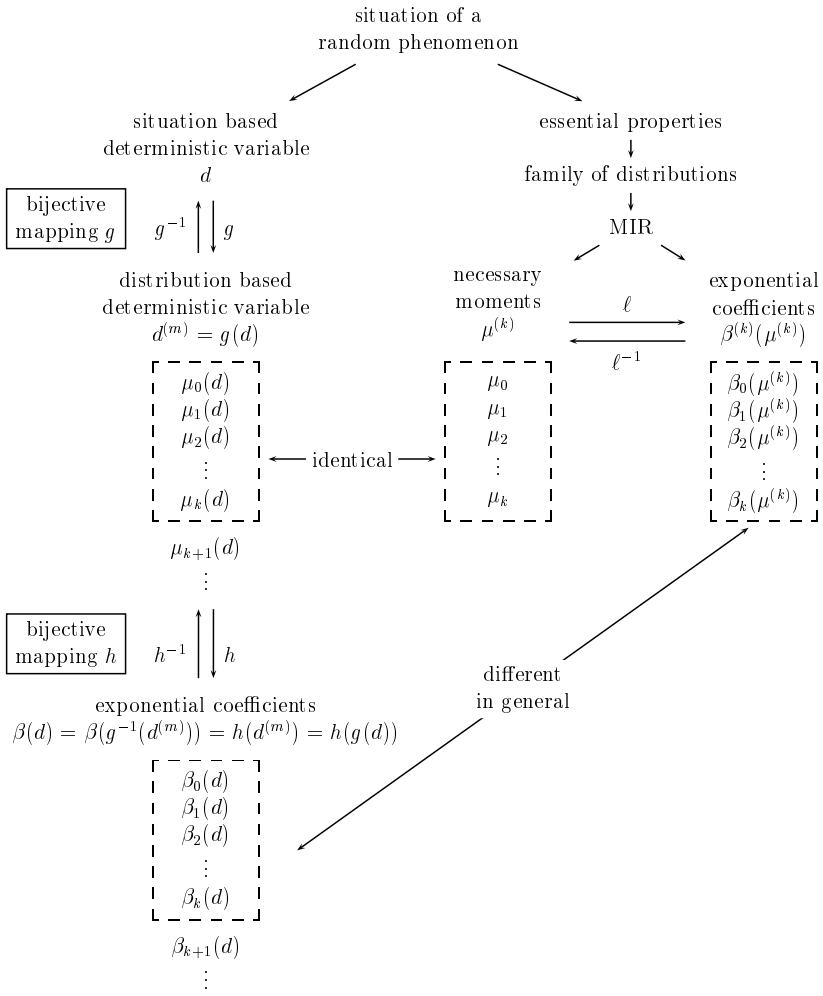


Figure 2.1.: Relations between the different deterministic variables d, $d^{(m)}$ and β, and the necessary moments $\mu^{(k)}$ and exponential coefficients $\beta^{(k)}$ of the Minimum Information Representative.

Example 2.3 *Consider a continuous random variable* $X|\{d\}$ *with the range of variability* $\mathcal{X}(\{d\}) = \{x \,|\, 0 \leq x \leq 1\}$. *Let the true PDF be given by*

$$f_{X|\{d\}}(x) = e^{1.0896748 - 1.5089485x - 2.8528670x^2} \tag{2.40}$$

and, therefore, the value of the first moment $\mu_1(d) = 0.25$.

We can easily check that (2.40) is a monotonic decreasing function on its support $\mathcal{X}(\{d\}) = \{x \,|\, 0 \leq x \leq 1\}$*:*

$$\forall_{y,z \in \mathcal{X}(\{d\})}\ y < z : f_{X|\{d\}}(y) > f_{X|\{d\}}(z) \tag{2.41}$$

If the true probability density function is not known, and only the qualitative property of monotonic decrease is available, we would opt in favour of the monotone distribution family and its simplest Minimum Information Representative

$$f_{X|\{\mu^{(1)}\}}(x) = e^{\beta_0(\mu^{(1)}) + \beta_1(\mu^{(1)})x} \tag{2.42}$$

with $X|\{\mu^{(1)}\}$ *indicating the approximation and* $\mu^{(1)} = (\mu_0, \mu_1)$.

With $\mathcal{X}(\{d\}) = \{x \,|\, 0 \leq x \leq 1\}$, *the value* $\mu_1 = 0.25$ *of the first moment and the two moment equations*

$$1 = \int_0^1 e^{\beta_0(\mu^{(1)}) + \beta_1(\mu^{(1)})x} dx \tag{2.43}$$

$$0.25 = \int_0^1 x e^{\beta_0(\mu^{(1)}) + \beta_1(\mu^{(1)})x} dx \tag{2.44}$$

we obtain the values of the exponential coefficients $\beta_0(\mu^{(1)})$ *and* $\beta_1(\mu^{(1)})$*:*

$$\beta_0(\mu^{(1)}) = 1.3070168 \tag{2.45}$$

$$\beta_1(\mu^{(1)}) = -3.5935120 \tag{2.46}$$

and hence

$$f_{X|\{\mu^{(1)}\}}(x) = e^{1.3070168 - 3.5935120x} \tag{2.47}$$

Figure 2.2 illustrates the difference between the (true but unknown) probability density function $f_{X|\{d\}}$ *and the approximating probabilty density function* $f_{X|\{\mu^{(1)}\}}$.

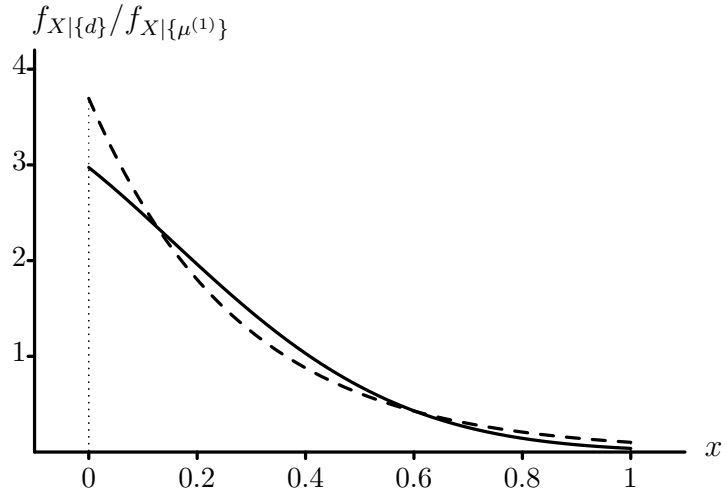
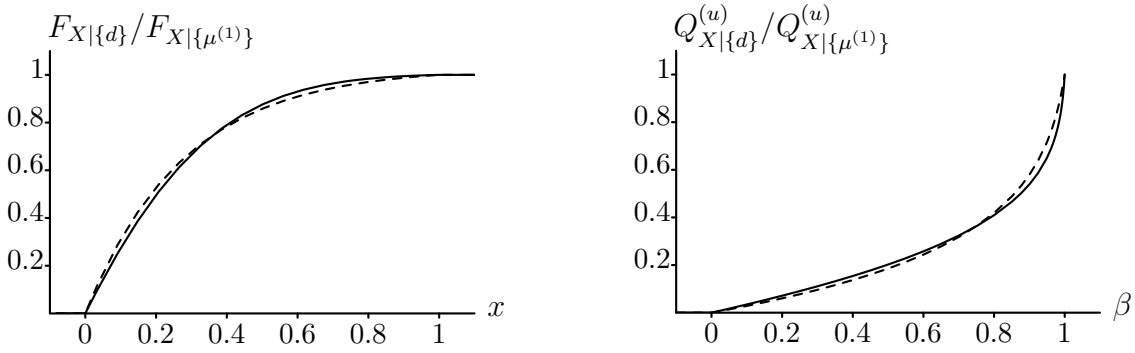


Figure 2.2.: Illustration of the (true but unknown) probability density function $f_{X|\{d\}}$ (solid line) and the approximating probability density function $f_{X|\{\mu^{(1)}\}}$ (dashed line).

Other ways to illustrate the difference between true and approximating probability distribution are, for instance, the probability distribution functions $F_{X|\{d\}}$ and $F_{X|\{\mu^{(1)}\}}$ and the upper quantile functions $Q^{(u)}_{X|\{d\}}(\beta)$ and $Q^{(u)}_{X|\{\mu^{(1)}\}}(\beta)$, $0 < \beta < 1$, as shown in figure 2.3.



(a) Probability distribution functions

(b) Upper quantile functions

Figure 2.3.: The true functions are represented with the solid lines, the approximating functions with the dashed lines.

2.3. MIR and Maximum Entropy Distributions

In information theory and statistical mechanics, the use of available knowledge, i.e., information, and dealing with uncertainty is closely connected to the principle of maximum entropy. In this section this principle is shortly derived. Following this concept, probability distributions arise, which are very similar to the minimum information representatives.

2.3.1. Shannon's Entropy

In 1948 Shannon quantified "uncertainty" generated by randomness by introducing the concept of entropy, initiating the field of information theory [70]. For the discrete case, i.e., for a set of n possible events with occurence probabilities $p_1, p_2, \ldots, p_n$ Shannon derived a measure $H(p_1, \ldots, p_n)$ from three very intuitive requirements, i.e. (see [70], pp. 392-393):

1. H should be continuous in the p_i.
2. If all the p_i are equal, $p_i = \frac{1}{n}$, then H should be a monotonic increasing function of n.[...]
3. If a choice be broken down into two successive choices, the original H should be the weighted sum of the individual values of H.

Shannon showed that the only measure satisfying these three requirements is of form:

$$H = -K \sum_{i=1}^{n} p_i \log p_i \tag{2.48}$$

with a positive constant K.
Shannon called $H = -\sum_{i=1}^{n} p_i \log p_i$ "the *entropy* of the set of probabilities $p_1, \ldots, p_n$" (ibid.), since the same form appears for the entropy in statistical mechanics[2].

[2]Soon after, Jaynes [37] argued, that Shannon's concept and interpretation has so much generality, that it should be re-adopted to statistical mechanics.

2.3.2. Maximum Entropy Principle

Shannon's entropy gave rise to formulate the *Maximum Entropy Principle* (MEP) for selecting approximative probability distribution:

> Select that probability distribution as approximation that meets all known properties of the unknown probability distribution and has maximal entropy.

The entropy function H (2.48) is only well defined for discrete probability distribution, i.e., meets only in this case all requirements. Nevertheless, by

$$\widetilde{H} = -\int_{-\infty}^{+\infty} f(x) \log f(x) \, \mathrm{d}x \tag{2.49}$$

a formal extension for continuous distributions was already defined by Shannon [70] and is often called *differential entropy* (see for example [46]). We have to note, that $\widetilde{H}$ (2.49) does not always meet the requirements, but may be negative or infinite. The derivation of maximum entropy distributions may be done by variational calculus, which we illustrate by the following two simple examples.

Example 2.4 *Let $\mathcal{X}(\{d\}) = \{x \,|\, a \leq x \leq b\} = [a, b]$ be the only known property of the unknown PDF $f(x)$. To derive the distribution with maximum entropy under the constraint $\int_{-\infty}^{+\infty} f(x) \, \mathrm{d}x = \int_a^b f(x) \, \mathrm{d}x = 1$ we define the Lagrangian function L_{λ_0}:*

$$L = \int_a^b f(x) \log f(x) \, \mathrm{d}x - \lambda_0 \left(1 - \int_a^b f(x) \, \mathrm{d}x \right) \tag{2.50}$$

If $f(x)$ maximises the entropy, some small variation $\Delta f(x)$ will yield a variation $\Delta L = 0$ for L :

$$0 = \Delta L = \int_a^b \Delta f(x) \big(\log f(x) + 1 + \lambda_0 \big) \, \mathrm{d}x \,. \tag{2.51}$$

Since the equation has to hold for every small variation $\Delta f(x)$ we get

$$\log f(x) + 1 + \lambda_0 = 0 \tag{2.52}$$

$$f(x) = e^{-\lambda_0 - 1} \tag{2.53}$$

and utilising the constraint yields:

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^b e^{-\lambda_0-1}\,\mathrm{d}x = 1 \tag{2.54}$$

$$e^{-\lambda_0-1}(b-a) = 1 \tag{2.55}$$

$$e^{-\lambda_0-1} = \frac{1}{b-a} = f(x)\,, \tag{2.56}$$

which is a constant PDF, i.e., the uniform distribution on $[a,b]$.
Note, that for $b-a<1$, *i.e.,* $f(x)>1$ *on* $[a,b]$, *we have a negative entropy* $\widetilde{H}(f) = -\log\frac{1}{b-a} < 0$.

Example 2.5 *Let* $\mathcal{X}(\{d\}) = \{x\,|\,0 \le x < +\infty\} = [0,+\infty)$ *and* $\mathrm{E}[X|\{d\}] = \mu_1 > 0$ *be the two known properties of the unknown probability distribution with density* $f(x)$. *That is, we have to derive the maximum entropy distribution under the two constraints* $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = \int_0^{+\infty} f(x)\,\mathrm{d}x = 1$ *and* $\int_{-\infty}^{+\infty} xf(x)\,\mathrm{d}x = \int_0^{+\infty} xf(x)\,\mathrm{d}x = \mu_1$. *After some small calculation we get for* $x \ge 0$

$$f(x) = e^{-\lambda_1 x-\lambda_0-1} \tag{2.57}$$

with the two Lagrangian mutlipliers λ_0 *and* λ_1. *Utilising the constraints yields*

$$e^{-\lambda_0-1} = \lambda_1 \tag{2.58}$$

$$\lambda_1 = \frac{1}{\mu_1} \tag{2.59}$$

$$f(x) = \frac{1}{\mu_1}e^{-x/\mu_1} \tag{2.60}$$

which is the exponential distribution with value μ_1 *of its expectation.*

We note, that the maximum entropy principle uniquely determines probability distributions from a number of constraints, usually on moments, by maximising their entropy. Further examples may be found in [25] where the first and second moments are given. There are various areas of application, in particular: metrology, see [78], [48], [63]; in biology where an overview is provided by [50]; and in survey sampling, see [27].

The difference between a MIR and a MEP distribution refers to the essential properties which are assumed to be known in the case of a MIR distribution. If in the above example the underlying unknown distribution is unimodal, the exponential distribution as the result of a blind application of the maximum entropy principle is clearly misleading.

2.4. Basic Types of Probability Distributions

One main task of modelling a random phenomenon is to derive an appropriate probability distribution. By now we have derived a general representation (2.23) and we have seen how to exploit the qualitative knowledge about a random phenomenon. In this part three basic types of probability distributions will be presented and the knowledge is derived which is at least necessary for selecting an appropriate approximation for each type.

One common (and realistic) property for all types is their bounded support $\mathcal{X}(\{d\})$, which is determined by the initial conditions d. For any realistic situation $\mathcal{X}(\{d\})$ contains a finite number of elements:

$$\mathcal{X}(\{d\}) = \{x_1, x_2, \ldots, x_N\} \tag{2.61}$$

If a continuous approximation of the range of variability is reasonable we set:

$$\mathcal{X}(\{d\}) = \{x \,|\, x_1 \leq x \leq x_N\} \tag{2.62}$$

2.4.1. Constant Distribution

We consider any outcome $x \in \mathcal{X}(\{d\}) = \{x_1, \ldots, x_N\}$ to be equally probable, that is, the Minimum Information Representative coincides with the exact one and has the following probability mass function:

$$f_{X|\{\mu^{(0)}\}}(x) \equiv f_{X|\{d\}}(x) = \frac{1}{|\mathcal{X}(\{d\})|}\mathbb{1}_{\mathcal{X}(\{d\})}(x) = \frac{1}{N}\mathbb{1}_{\mathcal{X}(\{d\})}(x) \tag{2.63}$$

where $\mathbb{1}$ is the indicator function. Since the probability for every element of $\mathcal{X}(\{d\})$ is the same, we talk of a constant probability distribution. If the continuous approximation is reasonable, the resulting density function on $\mathcal{X}(\{d\}) = \{x \,|\, x_1 \leq x \leq x_N\}$ is

$$f_{X|\{\mu^{(0)}\}}(x) = \frac{1}{x_N - x_1}\mathbb{1}_{\mathcal{X}(\{d\})}(x) \tag{2.64}$$

We should note that in the continuous approximation, the density function cannot be taken as an approximation of the probability mass function. The approximation always refers to the probability distribution function $F_{X|\{d\}}$. Figure 2.4 illustrates this fact for the constant distribution.

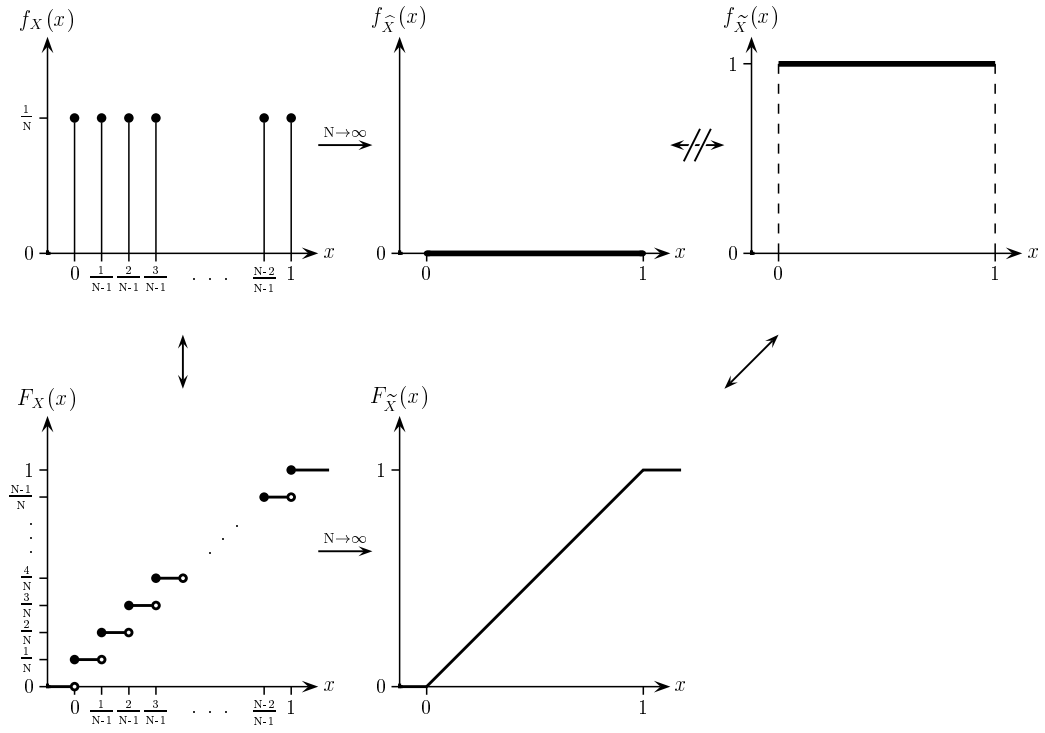


Figure 2.4.: Approximating the true probability mass function f_X by letting the number N of elements in $\mathcal{X}(\{d\})$ converging to infinity would result in $f_{\widehat{X}}$, a function which equals zero and is not a probability density function. The only way for continuous approximation leads via the probability distribution function $F_{X|\{d\}}$ to $F_{\widetilde{X}|\{d\}}$ and then to its density function $f_{\widetilde{X}|\{d\}}$ (which does not represent probabilities).

Remark: The constant probability distribution is better known as uniform distribution, in the continuous case on the interval $[x_1, x_N]$ denoted by $\mathrm{U}(x_1, x_N)$ and in the discrete case on the set $\mathcal{X}(\{d\})$ with N elements denoted by $\mathrm{U}(\mathcal{X}(\{d\}))$.

Example 2.6 *A very popular example for the constant probability distribution is obtained in the case of games of chance, e.g. throwing a dice. The side being on top is represented by the random variable X. Assume the dice has six sides consecutively numbered from 1 to 6, then the range of variability is*

$$\mathcal{X}(\{d_0\}) = \{1, 2, 3, 4, 5, 6\} \tag{2.65}$$

Furthermore, assume the dice being totally symmetric, which is described quantitatively by the value d_0 of a suitable deterministic variable D, then the probability is $\frac{1}{6}$ for every side to be on top.

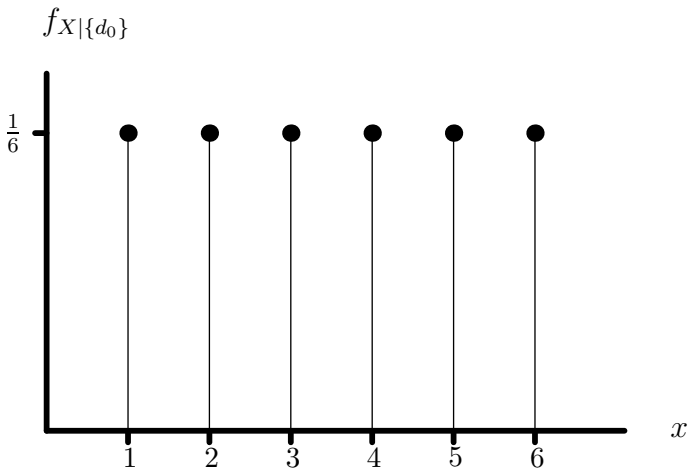


Figure 2.5.: Probability mass function $f_{X|\{d_0\}}$ for throwing a six-sided, symmetric dice.

Example 2.7 *Assume a person wanting to catch a bus at a bus stop, knowing only that there is a bus every 20 minutes but not knowing the exact timetable. The bus timetable d_0 can be described by the time of earliest arrival t_0 and the inter-arrival time Δ. Assume $\Delta = 20$ minutes:*

$$d_0 = \{t \,|\, t = t_0 + \nu 20\,,\ \nu = 0, 1, 2, \ldots\} \tag{2.66}$$

In this way, d_0 represents a part of the initial conditions.

The person's arrival time at the bus stop is assumed to be independent of the arrival time of the bus, and is described by the random variable $Y|\{d_0\}$, which can be considered as a continuous approximation of the discrete observable time. As only the arrival time within a 20 minutes interval between two arriving buses is of interest, we can define a random variable $X|\{d_0\}$ with

$$X|\{d_0\} = (Y|\{d_0\}) - t_0 \bmod 20 \tag{2.67}$$

Then, the range of variability $\mathcal{X}(\{d_0\})$ *of* $X|\{d_0\}$ *is*

$$\mathcal{X}(\{d_0\}) = [0, 20] \tag{2.68}$$

and the density function is the constant one (see also Figure 2.6):

$$f_{X|\{d_0\}}(x) = \begin{cases} \frac{1}{20} & \text{for } 0 \leq x \leq 20 \\ 0 & \text{else} \end{cases} \tag{2.69}$$

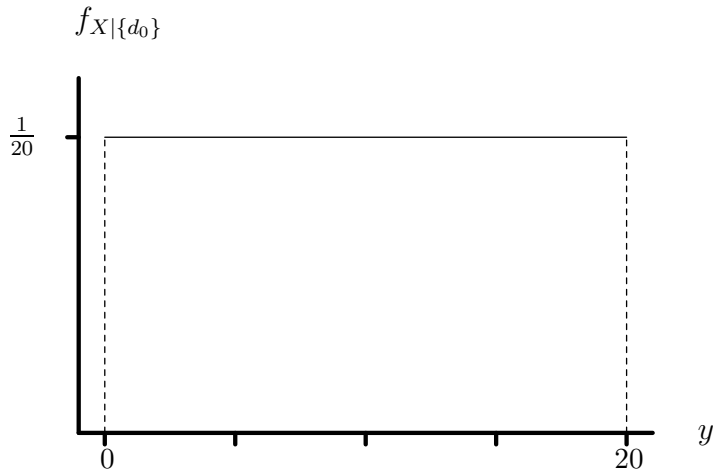


Figure 2.6.: Density function $f_{X|\{d_0\}}$ for the time of arriving at a bus stop in a period of 20 minutes whithout having any preference with respect to a special time.

2.4.2. Monotonic Distribution

Next consider the case of two extrema in the bounds of $\mathcal{X}(\{d\})$ and no inner extremum. This results in the monotonic distribution family, that is, the probability density function $f_{X|\{d\}}$ is

- either monotonous increasing:

$$f_{X|\{d\}}(x) \leq f_{X|\{d\}}(y) \text{ for } x, y \in \mathcal{X}(\{d\}) : x < y \tag{2.70}$$

- or monotonous decreasing:

$$f_{X|\{d\}}(x) \geq f_{X|\{d\}}(y) \text{ for } x, y \in \mathcal{X}(\{d\}) : x < y \tag{2.71}$$

The resulting Minimum Information Representative is given by:

$$f_{X|\{\mu^{(1)}\}}(x) = \begin{cases} e^{\beta_0+\beta_1 x} & \text{for } x \in \mathcal{X}(\{d\}) \\ 0 & \text{else} \end{cases} \tag{2.72}$$

For determining the coefficients β_0 and β_1 the value of the first moment is necessary. Hence, we have to solve the following equations:

- In the discrete case:

$$\begin{aligned} 1 &= \sum_{i=1}^{N} e^{\beta_0+\beta_1 x_i} \\ \mu_1 &= \sum_{i=1}^{N} x_i e^{\beta_0+\beta_1 x_i} \end{aligned} \tag{2.73}$$

- In the continuous case:

$$\begin{aligned} 1 &= \int_{x_1}^{x_N} e^{\beta_0+\beta_1 x} dx \\ \mu_1 &= \int_{x_1}^{x_N} x e^{\beta_0+\beta_1 x} dx \end{aligned} \tag{2.74}$$

Chapter 3 treats in very detail the continuous monotonic distribution, thus, we only note here that for special values of μ_1 equations 2.73 and 2.74, respectively, yield $\beta_1 = 0$. Then the monotonic distribution falls back to the constant distribution.

Example 2.8 *Assume a device for measuring the diameter of steel balls. As the measuring results vary even in the case that the same steel ball is examined several times, the measuring process is represented by a random variable $Y|\{d\}$, where $d = (d_1, d_2, d_3, \ldots)$ with d_1 being the actual value of the diameter. For describing the quality of the measurement device consider the absolute deviation of the measurement from d_1 and define the random variable*

$$X|\{d\} = \left|Y|\{d\} - d_1\right| \tag{2.75}$$

For a reasonable measuring device it should hold that large deviations from the actual value should occur with smaller probability than small deviations. This implies that for the approximating density function of $X|\{d\}$ it should hold

$$f_{X|\{d\}}(x) > f_{X|\{d\}}(y) \text{ for } x < y \tag{2.76}$$

with $x, y \in \mathcal{X}(\{d\})$. *Thus, the density function describes a monotone distribution. Assume that only one steel ball is examined with a diameter of* 1 *cm and that the first moment of the absolute deviation takes the value* $\mu_1 = 0.05$ *cm. The range of variability is known to be* $\mathcal{X}(\{d\}) = [0, 0.2]$. *Solving the moment equations (2.74) we get* $\beta_0 = 2.916455$ *and* $\beta_1 = -17.96756$. *Figure 2.7 shows the resulting density function* $f_{X|\{\mu^{(1)}\}}(x) = e^{2.916455-17.96756x}$.

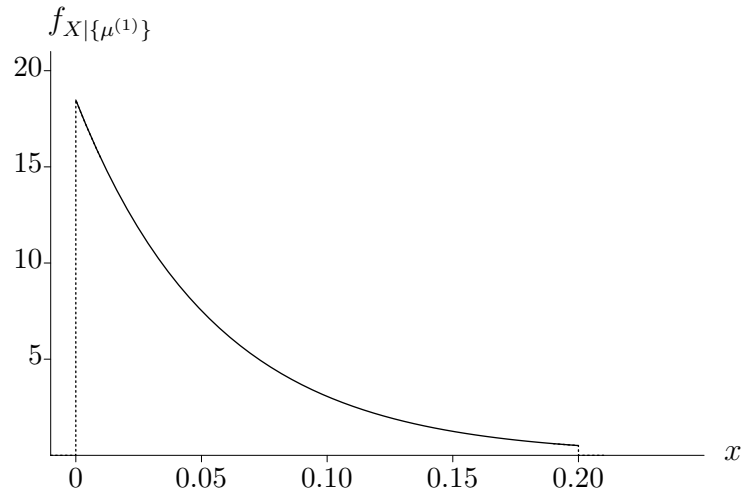


Figure 2.7.: Density function $f_{X|\{\mu^{(1)}\}}$ for the absolute measurement deviation with $\mathcal{X}(\{d\}) = [0, 0.2]$ and $\mu_1 = 0.05$.

2.4.3. Unimodal Distribution

Next, assume a probability density function which is monotone increasing until some $x_M \in \mathcal{X}$, and decreasing from this on. This case defines the unimodal distribution family. In the discrete case that means that one outcome x_M, $x_1 < x_M < x_N$, is most probable and the probability of the others decreases with increasing distance to x_M:

$$\forall\, x \in \mathcal{X}(\{d\}), x \neq x_M \;:\; f_{X|\{d\}}(x_M) > f_{X|\{d\}}(x) \tag{2.77}$$

$$\forall_{y,z \in \mathcal{X}(\{d\})}\, y < z < x_M \;:\; f_{X|\{d\}}(y) \leq f_{X|\{d\}}(z) < f_{X|\{d\}}(x_M) \tag{2.78}$$

$$\forall_{y,z \in \mathcal{X}(\{d\})}\, x_M < y < z \;:\; f_{X|\{d\}}(x_M) > f_{X|\{d\}}(y) \geq f_{X|\{d\}}(z) \tag{2.79}$$

In both, the discrete case and in the case of the continuous approximation we get the resulting Minimum Information Representative:

$$f_{X|\{\mu^{(2)}\}}(x) = \begin{cases} e^{\beta_0+\beta_1 x+\beta_2 x^2} & \text{for } x \in \mathcal{X}(\{d\}) \\ 0 & \text{else} \end{cases} \tag{2.80}$$

To obtain a relative maximum, β_1 and β_2 have to meet some conditions. For the continuous approximation, i.e., the density function $f_{X|\{\mu^{(2)}\}}$ we have[3]:

For a maximum x_M the first derivative of $f_{X|\{\mu^{(2)}\}}$ has to be 0 and the second derivative has to be negative at this point:

$$f'_{X|\{\mu^{(2)}\}}(x_M) = (\beta_1 + 2\beta_2 x_M)e^{\beta_0+\beta_1 x_M+\beta_2 x_M^2} = 0 \tag{2.81}$$

$$\Leftrightarrow x_M = -\frac{\beta_1}{2\beta_2} \tag{2.82}$$

and

$$f''_{X|\{\mu^{(2)}\}}(x_M) = \big((\beta_1 + 2\beta_2 x_M)^2 + 2\beta_2\big)e^{\beta_0+\beta_1 x_M+\beta_2 x_M^2} < 0 \tag{2.83}$$

$$\Leftrightarrow \beta_2 < 0 \tag{2.84}$$

As the maximum has to be an inner point of $\mathcal{X}(\{d\})$ we have the condition

- in the discrete case:
  As $-\frac{\beta_1}{2\beta_2}$ need not to be an element of $\mathcal{X}(\{d\})$, the condition (2.82) may not be fulfilled for any element of $\mathcal{X}(\{d\})$. But since we are seeking an inner maximum, the following condition is necessary
  $$x_1 < -\frac{\beta_1}{2\beta_2} < x_N \tag{2.85}$$
  Even if condition (2.85) is met, the maximum may result to be a boundary maximum. To prevent this the next two conditions

[3] As the discrete probability mass function is in fact a real function defined only at discrete points, it can be treated as a continuous function when determining maxima, minima, monotonicity etc.

have to be fulfilled:

$$f_{X|\{\mu^{(2)}\}}(x_1) < f_{X|\{\mu^{(2)}\}}(x_2) \overset{\beta_2<0}{\Longleftrightarrow} -\frac{\beta_1}{2\beta_2} > \frac{x_1+x_2}{2} \tag{2.86}$$

$$f_{X|\{\mu^{(2)}\}}(x_{N-1}) > f_{X|\{\mu^{(2)}\}}(x_N) \overset{\beta_2<0}{\Longleftrightarrow} -\frac{\beta_1}{2\beta_2} < \frac{x_{N-1}+x_N}{2} \tag{2.87}$$

- in the continuous case:

$$x_1 < -\frac{\beta_1}{2\beta_2} < x_N \tag{2.88}$$

The coefficients β_0, β_1 and β_2 can be obtained as solutions of the following three moment equations:

- In the discrete case:

$$\begin{aligned} 1 &= \sum_{i=1}^{N} e^{\beta_0+\beta_1 x_i+\beta_2 x_i^2} \\ \mu_1 &= \sum_{i=1}^{N} x_i e^{\beta_0+\beta_1 x_i+\beta_2 x_i^2} \\ \mu_2 &= \sum_{i=1}^{N} x_i^2 e^{\beta_0+\beta_1 x_i+\beta_2 x_i^2} \end{aligned} \tag{2.89}$$

- In the continuous case:

$$\begin{aligned} 1 &= \int_{x_1}^{x_N} e^{\beta_0+\beta_1 x+\beta_2 x^2} dx \\ \mu_1 &= \int_{x_1}^{x_N} x e^{\beta_0+\beta_1 x+\beta_2 x^2} dx \\ \mu_2 &= \int_{x_1}^{x_N} x^2 e^{\beta_0+\beta_1 x+\beta_2 x^2} dx \end{aligned} \tag{2.90}$$

Other knowledge can concern the location of the maximum for example which could be exploited instead.

Remark: In the discrete case it is possible that two neighbouring elements x_m and x_{m+1} of $\mathcal{X}(\{d\})$ have both the same maximum value of the probability mass function. Nevertheless, we treat this case as an unimodal distribution.

The unimodal type is divided into two distinguishable types:

- **Symmetric unimodal distribution**

  Obviously, in the case of a symmetric probability density function, the maximum x_M coincides with the midpoint of $\mathcal{X}(\{d\})$:

$$x_M = \frac{x_1 + x_N}{2} \tag{2.91}$$

  Furthermore, the value μ_1 of the first moment also coincides with x_M, and hence the midpoint of $\mathcal{X}(\{d\})$:

$$\mu_1 = \frac{x_1 + x_N}{2} \tag{2.92}$$

  – In the discrete case:

    Symmetry refers both to $\mathcal{X}(\{d\})$ and the values of the probability mass function:

$$x_{i+1} - x_i = x_{N-(i-1)} - x_{N-i} \tag{2.93}$$

$$f_{X|\{\mu^{(2)}\}}(x_i) = f_{X|\{\mu^{(2)}\}}(x_{N-(i-1)}) \tag{2.94}$$

$$\text{for} \quad i = \begin{cases} 1, 2, \ldots, \frac{N-1}{2} & \text{if } N \text{ is odd} \\ 1, 2, \ldots, \frac{N}{2} & \text{if } N \text{ is even} \end{cases} \tag{2.95}$$

    Equation (2.94) may easily be transformed into

$$(x_M =) \; -\frac{\beta_1}{2\beta_2} = \frac{x_{N-(i-1)} - x_i}{2} \tag{2.96}$$

    In fact, equation (2.93) is equivalent to (2.96) and, therefore, actually redundant.

    Remark: If N is even, we always get a probability mass function with two maxima, $x_{\frac{N}{2}}$ and $x_{\frac{N+2}{2}}$, which both have the same probability.

  – In the continuous case:

    The equation

$$-\frac{\beta_1}{2\beta_2} = x_M = \frac{x_1 + x_N}{2} \tag{2.97}$$

    is sufficient for a symmetric density function.

    This equation already yields another representation which is more common when defining a symmetric property of a function:

$$\forall\, \delta,\, 0 \leq \delta \leq \frac{x_1 + x_N}{2} : f_{X|\{\mu^{(2)}\}}(x_M - \delta) = f_{X|\{\mu^{(2)}\}}(x_M + \delta) \tag{2.98}$$

From equations (2.96) and (2.97) we see how the range of variability $\mathcal{X}(\{d\})$ is closely connected with the exponential coefficients β_1 and β_2. Due to the knowledge about the qualitative property of symmetry the effort to obtain the necessary quantitative knowledge, i.e., β_1 and β_2, becomes smaller.

- **Asymmetric unimodal distribution**
  For an asymmetric unimodal distribution the maximum x_M does not coincide with the midpoint of $\mathcal{X}(\{d\})$, and, therefore, also not with the value μ_1 of the first moment:

$$x_M \neq \frac{x_1 + x_N}{2} \tag{2.99}$$

$$\mu_1 \neq \frac{x_1 + x_N}{2} \tag{2.100}$$

If symmetry can be excluded as a qualitative property of a probability distribution, then it is reasonable that it is known whether the maximum is smaller or bigger then the midpoint of $\mathcal{X}(\{d\})$:

$$x_M < \frac{x_1 + x_N}{2} \tag{2.101}$$

or

$$x_M > \frac{x_1 + x_N}{2} \tag{2.102}$$

With equation (2.82), which still holds, this leads to

$$-\frac{\beta_1}{2\beta_2} < \frac{x_1 + x_N}{2} \tag{2.103}$$

or

$$-\frac{\beta_1}{2\beta_2} > \frac{x_1 + x_N}{2} \tag{2.104}$$

The effort to obtain β_1 and β_2 is also reduced, but not as much as in the symmetric case.

Remark: If asymmetry shall be taken into account not only by the condition $\mu_1 \neq x_M$ one has to select at least a cubic polynomial as exponent or

even change to a different class of functions for representing the probability density function. One often assumed class is given by the density function

$$\frac{\gamma}{\alpha}\left(\frac{x-\mu}{\alpha}\right)^{\gamma-1} e^{\left(\frac{x-\mu}{\alpha}\right)^{\gamma}} \mathbb{1}_{[\mu,+\infty)}(x) \quad \text{with} \quad \gamma,\alpha>0\,,$$

which defines the general Weibull distribution. In classical statistics the Weibull distribution is often used as a lifetime distribution. To determine the parameters the most common way is to fit the curve to the data - no qualitative properties of the random phenomenon are considered afore. Besides this its unbounded range of variability also conflicts with reality.

Example 2.9 *Assume the production of steel balls for bearings. One important feature to decide whether a ball is conforming or not is its diameter. For being conforming the actual diameter has to lie in a range given by the specification interval. The process of production can only be adjusted to a target value* d_1*, which itself lies in the specification interval. This target value* d_1 *is one component of the initial conditions* $d = (d_1, d_2, d_3, \ldots)$*. Let* $X|\{d\}$ *be the discrete random variable representing the observable diameter.*

Assume that from theoretical considerations about the process it is known that the probability mass function is unimodal and, therefore we choose as MIR (2.80).

For obtaining the values of β_0*,* β_1 *and* β_2 *we solve the three moment equations:*

$$1 = \sum_{i=1}^{N} e^{\beta_0+\beta_1 x_i+\beta_2 x_i^2} \tag{2.105}$$

$$\mu_1 = \sum_{i=1}^{N} x_i e^{\beta_0+\beta_1 x_i+\beta_2 x_i^2} \tag{2.106}$$

$$\mu_2 = \sum_{i=1}^{N} x_i^2 e^{\beta_0+\beta_1 x_i+\beta_2 x_i^2} \tag{2.107}$$

Assume that the measurement device has a resolution of 0.01 *and that the discrete range of variability is given by*

$$\mathcal{X}(\{d\}) = \{0.92, 0.93, \ldots, 1.04\}\ ,\ N = 13 \tag{2.108}$$

Let the target value $d_1 = 1.00$ *coincide with the value* μ_1 *of the first moment of* $X|\{d\}$. *The variance* σ^2 *of the production process can be considered to be* d_2, *then the second moment* μ_2 *can be directly calculated via* $\mu_2 = \sigma^2 + \mu_1^2$. *Given the value* $\sigma^2 = 0.0004$ *we get* $\mu_2 = 1.0004$. *With this, the values of* β_0, β_1 *and* β_2 *can be calculated:*

$$\beta_0 = -1134.36002 \tag{2.109}$$

$$\beta_1 = 2263.35674 \tag{2.110}$$

$$\beta_2 = -1130.64260 \tag{2.111}$$

Figure 2.8 shows the probability mass function (PMF)

$$f_{X|\{\mu^{(2)}\}}(x) = e^{-1134.36002+2263.35674x-1130.64260x^2} \tag{2.112}$$

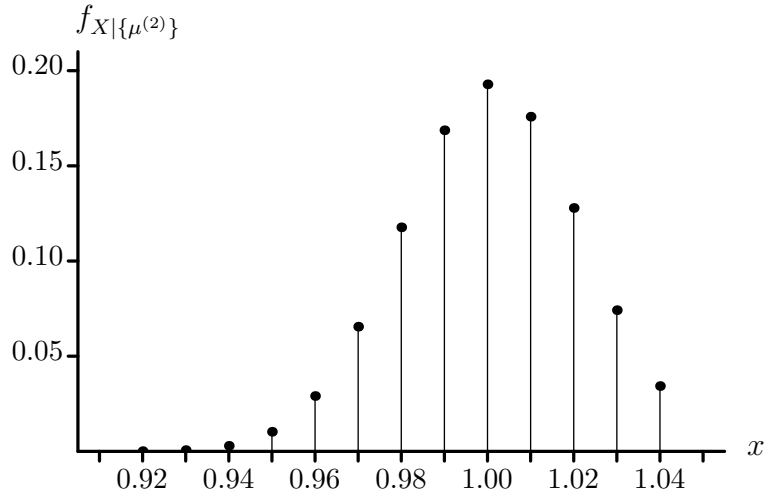


Figure 2.8.: PMF $f_{X|\{\mu^{(2)}\}}(x) = e^{-1134.36002+2263.35674x-1130.64260x^2}$ for the random diameter of steel balls for the discrete range of variability $\mathcal{X}(\{d\}) = \{0.92, 0.93, \ldots, 1.04\}$, $d_1 = \mu_1 = 1.00$ and $d_2 = \sigma^2 = 0.0004$.

3. Monotonic Probability Distributions of Continuous Univariate Random Variables

In [67] the importance of monotonic probability distributions is outlined. Furthermore some properties of the corresponding moments and the n-fold convolution are derived. Finally the goodness of an approximation with the normal distribution is analysed.
In the first two sections of this chapter those results are presented and a far more detailed characterisation is given.

3.1. Characterisations

Let X be a random variable of continuous type with the following essential properties:

- Range of variability: $\mathcal{X}(\{d\}) = \{x \,|\, a \leq x \leq b\} = [a, b]$.
- The density function $f_{X|\{d\}}$ has two boundary extremes and no relative extreme.

According to section 2.4.2 in Chapter 2 the resulting Minimum Information Representative is a member of the monotonic distribution family:

$$f_{X|\{\mu^{(1)}\}}(x) = e^{\beta_0+\beta_1 x}\mathbb{1}_{[a,b]}(x) = \begin{cases} e^{\beta_0+\beta_1 x} & \text{for } x \in \mathcal{X}(\{d\}) = [a, b] \\ 0 & \text{else} \end{cases} \tag{3.1}$$

with coefficients β_0 and β_1 being solutions of the two equations:

$$1 = \int_a^b e^{\beta_0+\beta_1 x}\,\mathrm{d}x \tag{3.2a}$$

$$\mu_1 = \int_a^b x e^{\beta_0+\beta_1 x}\,\mathrm{d}x \tag{3.2b}$$

Equation (3.2a) may be rewritten in the following way:

$$e^{\beta_0} = \frac{1}{\int_a^b e^{\beta_1 x}\,\mathrm{d}x} \tag{3.3}$$

With this, we may reformulate equation (3.2b) to

$$\mu_1 = \int_a^b x e^{\beta_0+\beta_1 x}\,\mathrm{d}x = e^{\beta_0}\int_a^b x e^{\beta_1 x}\,\mathrm{d}x = \frac{\int_a^b x e^{\beta_1 x}\,\mathrm{d}x}{\int_a^b e^{\beta_1 t}\,\mathrm{d}t} \tag{3.4}$$

And finally, the probability density function $f_{X|\{\mu^{(1)}\}}$:

$$f_{X|\{\mu^{(1)}\}}(x) = \frac{e^{\beta_1 x}}{\int_a^b e^{\beta_1 t}\,\mathrm{d}t}\mathbb{1}_{[a,b]}(x) = \begin{cases} \frac{e^{\beta_1 x}}{\int_a^b e^{\beta_1 t}\,\mathrm{d}t} & \text{for } x \in \mathcal{X}(\{d\}) = [a,b] \\ 0 & \text{else} \end{cases} \tag{3.5}$$

Since β_1 is the only distribution parameter besides the support's boundary points a and b we will follow the notation in [67] and obtain:

$$\lambda = \lambda(a,b,\mu_1) = \beta_1 \tag{3.6a}$$

$$\mu_1 = \frac{\int_a^b x e^{\lambda x}\,\mathrm{d}x}{\int_a^b e^{\lambda x}\,\mathrm{d}x} = \frac{\int_a^b x e^{\lambda(a,b,\mu_1)x}\,\mathrm{d}x}{\int_a^b e^{\lambda(a,b,\mu_1)x}\,\mathrm{d}x} \tag{3.6b}$$

$$f_{X|\{(a,b,\mu_1)\}}(x) = \frac{e^{\lambda x}}{\int_a^b e^{\lambda t}\,\mathrm{d}t}\mathbb{1}_{[a,b]}(x) = \frac{e^{\lambda(a,b,\mu_1)x}}{\int_a^b e^{\lambda(a,b,\mu_1)t}\,\mathrm{d}t}\mathbb{1}_{[a,b]}(x) = f_{X|\{\mu^{(1)}\}}(x) \tag{3.6c}$$

Though possible, we have not explicitly solved the appearing integrals throughout the last equations. The cause for this is quite simple: The distribution parameter λ may adopt every value in $\mathbb{R}$, and for $\lambda = 0$ equation (3.6b) yields

$$\mu_1 = \frac{\int_a^b x \, \mathrm{d}x}{\int_a^b \mathrm{d}x} = \frac{\frac{1}{2}(b^2 - a^2)}{b - a} = \frac{a+b}{2} \tag{3.7a}$$

with density function

$$f_{X|\{(a,b,\frac{a+b}{2})\}}(x) = \frac{1}{\int_a^b \mathrm{d}t} \mathbb{1}_{[a,b]}(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) \tag{3.7b}$$

representing the constant (uniform) distribution on the interval $[a, b]$.

Definition 3.1 (Monotonic Distribution) *A random variable X is called monotonic distributed, written $X|\{(a, b, \mu_1)\} \sim Mon(a, b, \mu_1)$, with range of variability $\mathcal{X} = \{x \,|\, a \leq x \leq b\}$, $a, b \in \mathbb{R}$, if its density function equals (3.6c).*

Without proof we can state the following obvious proposition:

Proposition 3.2 (Sign of λ) *Let X be a random variable with range of variability $\mathcal{X} = \{x \,|\, a \leq x \leq b\}$, and with density function (3.6c). Then for λ, μ_1 and $f_{X|\{(a,b,\mu_1)\}}$ one of the following three conditions holds:*

(a) $\lambda < 0 \iff \mu_1 < \frac{a+b}{2}$, *i.e.,* $f_{X|\{(a,b,\mu_1)\}}$ *is strictly decreasing on* $\mathcal{X}$,

(b) $\lambda > 0 \iff \mu_1 > \frac{a+b}{2}$, *i.e.,* $f_{X|\{(a,b,\mu_1)\}}$ *is strictly increasing on* $\mathcal{X}$,

(c) $\lambda = 0 \iff \mu_1 = \frac{a+b}{2}$, *i.e.,* $f_{X|\{(a,b,\mu_1)\}}$ *is constant on* $\mathcal{X}$.

In the two cases (a) and (b) respectively of the above proposition 3.2 we solve the appearing integrals in (3.6b) and (3.6c) and get with $\lambda = \lambda(a, b, \mu_1)$:

$$\mu_1 = \frac{\int_a^b x e^{\lambda x} \, \mathrm{d}x}{\int_a^b e^{\lambda x} \, \mathrm{d}x} = \frac{be^{\lambda b} - ae^{\lambda a}}{e^{\lambda b} - e^{\lambda a}} - \frac{1}{\lambda} \tag{3.8a}$$

$$f_{X|\{(a,b,\mu_1)\}}(x) = \frac{e^{\lambda x}}{\int_a^b e^{\lambda t} \, \mathrm{d}t} \mathbb{1}_{[a,b]}(x) = \frac{\lambda e^{\lambda x}}{e^{\lambda b} - e^{\lambda a}} \mathbb{1}_{[a,b]}(x) \tag{3.8b}$$

Depending on the problem on hand, we will use the implicit terms with the integrals or the explicit terms, whichever suits best.
The distribution function $F_{X|\{(a,b,\mu_1)\}}(x) = \mathrm{P}_{X|\{(a,b,\mu_1)\}}(\{y \,|\, y \leq x\})$ for $\lambda \neq 0$, i.e. $\mu_1 \neq \frac{a+b}{2}$, is

$$F_{X|\{(a,b,\mu_1)\}}(x) = \begin{cases} 0, & x < a \,, \\ \frac{e^{\lambda x} - e^{\lambda a}}{e^{\lambda b} - e^{\lambda a}}, & a \leq x \leq b \,, \\ 1, & x > b \,. \end{cases} \tag{3.9}$$

For $a = 0$ and $b = 1$, i.e., $\mathcal{X} = \{x \,|\, 0 \leq x \leq 1\}$, Figure 3.1 displays the dependency between μ_1 and λ and Figure 3.2 shows some density functions for different values of λ. In the next section a proposition states how the general case $\mathcal{X} = \{x \,|\, a \leq x \leq b\}$ can be traced back to $\mathcal{X} = \{x \,|\, 0 \leq x \leq 1\}$.

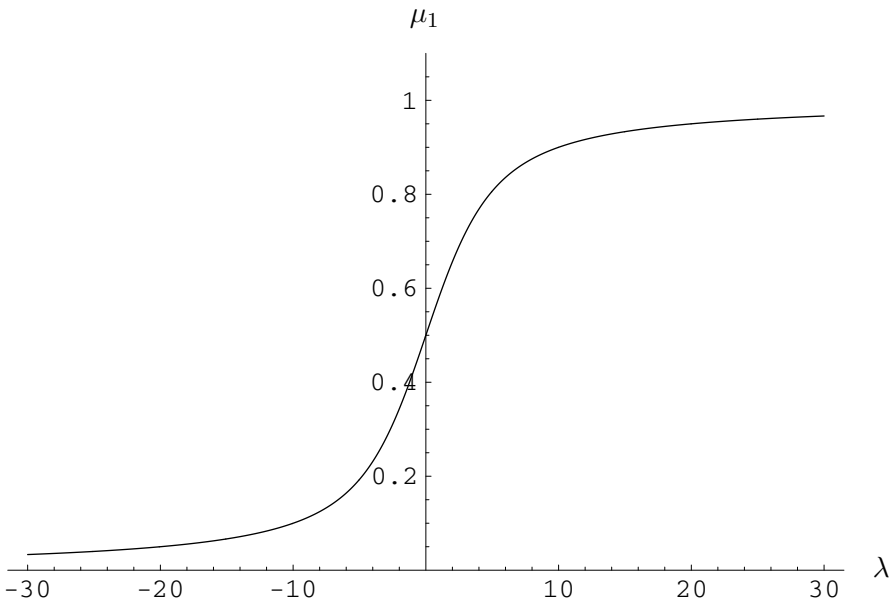


Figure 3.1.: The first moment μ_1 as function of the distribution parameter λ for $\mathcal{X} = \{x \,|\, 0 \leq x \leq 1\}$.

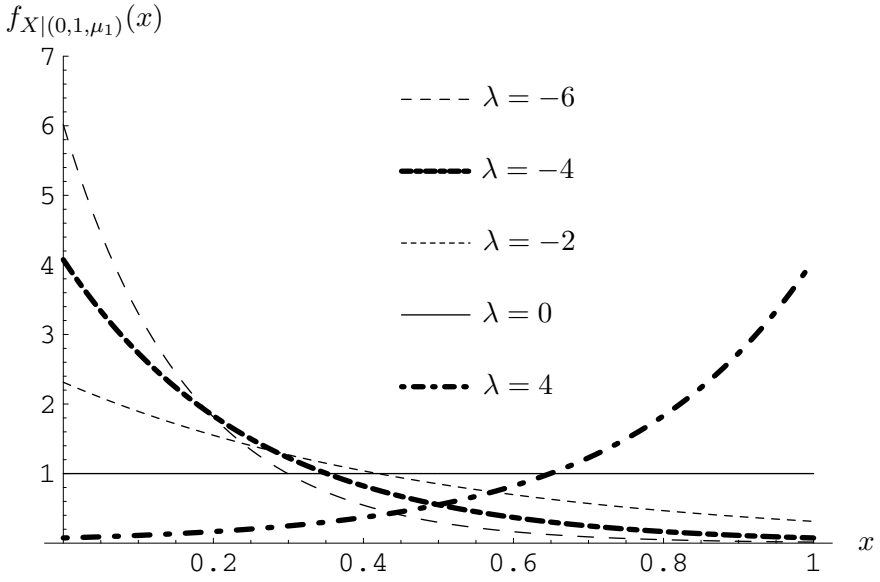


Figure 3.2.: Density function $f_{X|\{(0,1,\mu_1)\}}(x)$ for different values of $\lambda(0, 1, \mu_1)$

3.2. Properties

In this section properties of the distribution parameter λ, the corresponding moments and the n-fold convolution are listed. Some of them were already derived in [67].

3.2.1. Properties concerning λ

Even though Figure 3.1 indicates the uniqueness of $\lambda(a, b, \mu_1)$ for fixed a, b and μ_1, a small proof is necessary.

Proposition 3.3 (Uniqueness of λ) *Let $X = X|\{(a, b, \mu_1)\} \sim Mon(a, b, \mu_1)$. For fixed a, b and μ_1 with $-\infty < a < \mu_1 < b < +\infty$ the distribution parameter $\lambda = \lambda(a, b, \mu_1)$ is the unique solution of (3.6b) and (3.8a) respectively*

$$\mu_1 = \frac{\int_a^b x e^{\lambda x} \, \mathrm{d}x}{\int_a^b e^{\lambda x} \, \mathrm{d}x} = \frac{b e^{\lambda b} - a e^{\lambda a}}{e^{\lambda b} - e^{\lambda a}} - \frac{1}{\lambda} \,. \tag{3.10}$$

Proof of 3.3: First we will show, that for all μ_1, $a < \mu_1 < b$ there is a $\lambda \in \mathbb{R}$ which fulfils (3.10). Since μ_1 is continuous in λ it is sufficient to calculate the limits of μ_1 for λ to $\pm\infty$:

- $\lambda \to +\infty$:

$$\lim_{\lambda \to +\infty} \mu_1 = \lim_{\lambda \to +\infty} \left(\frac{b e^{\lambda b} - a e^{\lambda a}}{e^{\lambda b} - e^{\lambda a}} - \frac{1}{\lambda} \right) = \tag{3.11}$$

$$= \lim_{\lambda \to +\infty} \left(\frac{b - a e^{-\lambda(b-a)}}{1 - e^{-\lambda(b-a)}} \right) - 0 = \tag{3.12}$$

$$= \frac{b - a \cdot 0}{1 - 0} = b, \tag{3.13}$$

since $b > a$ and the limit of $e^{-\lambda(b-a)}$ is 0.

- $\lambda \to -\infty$:

$$\lim_{\lambda\to-\infty} \mu_1 = \lim_{\lambda\to-\infty} \left(\frac{be^{\lambda b} - ae^{\lambda a}}{e^{\lambda b} - e^{\lambda a}} - \frac{1}{\lambda} \right) = \tag{3.14}$$

$$= \lim_{\lambda\to-\infty} \left(\frac{be^{\lambda(b-a)} - a}{e^{\lambda(b-a)} - 1} \right) - 0 = \tag{3.15}$$

$$= \frac{b \cdot 0 - a}{0 - 1} = a, \tag{3.16}$$

since $b > a$ and the limit of $e^{\lambda(b-a)}$ is 0.

Thus we can conclude, that for every μ_1, $a < \mu_1 < b$, there is a $\lambda \in \mathbb{R}$ which fulfils (3.10). For the uniqueness it is now sufficient that the right hand side of (3.6b) is strictly increasing in λ. To proof this we differentiate (3.6b) with respect to λ[1]:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \left(\frac{\int_a^b xe^{\lambda x}\,\mathrm{d}x}{\int_a^b e^{\lambda x}\,\mathrm{d}x} \right) = \frac{\left(\int_a^b e^{\lambda x}\,\mathrm{d}x\right) \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\int_a^b xe^{\lambda x}\,\mathrm{d}x\right) - \left(\int_a^b xe^{\lambda x}\,\mathrm{d}x\right) \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\int_a^b e^{\lambda x}\,\mathrm{d}x\right)}{\left(\int_a^b e^{\lambda x}\,\mathrm{d}x\right)^2} \tag{3.17}$$

with $\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\int_a^b xe^{\lambda x}\,\mathrm{d}x\right) = \int_a^b \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(xe^{\lambda x}\right)\,\mathrm{d}x$ and $\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\int_a^b e^{\lambda x}\,\mathrm{d}x\right) = \int_a^b \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(e^{\lambda x}\right)\,\mathrm{d}x$ we obtain:

$$= \frac{\left(\int_a^b e^{\lambda x}\,\mathrm{d}x\right)\left(\int_a^b x^2e^{\lambda x}\,\mathrm{d}x\right) - \left(\int_a^b xe^{\lambda x}\,\mathrm{d}x\right)\left(\int_a^b xe^{\lambda x}\,\mathrm{d}x\right)}{\left(\int_a^b e^{\lambda x}\,\mathrm{d}x\right)^2} \tag{3.18}$$

$$= \frac{\int_a^b x^2e^{\lambda x}\,\mathrm{d}x}{\int_a^b e^{\lambda x}\,\mathrm{d}x} - \left(\frac{\int_a^b xe^{\lambda x}\,\mathrm{d}x}{\int_a^b e^{\lambda x}\,\mathrm{d}x} \right)^2 \tag{3.19}$$

$$= \mu_2 - (\mu_1)^2, \tag{3.20}$$

since the first term equals the second moment μ_2 of X and the second term equals μ_1.

[1] Obviously $e^{\lambda x}$ and $xe^{\lambda x}$ are continuous and their derivatives with respect to λ exist and are continuous as well.

The derived expression is nothing else than the variance σ^2 of X. We know, that $\sigma^2 > 0$ for all random variables, which have no one-point-distribution, and thus, we conclude that μ_1 is unique. Moreover μ_1 is strictly increasing in λ.

•

Proposition 3.4 (λ transformations) *Let* $X = X|\{(a,b,\mu_1)\} \sim Mon(a,b,\mu_1)$. *For the distribution parameter* $\lambda(a,b,\mu_1)$ *the following equations holds:*

(a) $\lambda(a,b,\mu_1) = \lambda(0,b-a,\mu_1-a)$.

(b) $\lambda(0,b,\mu_1) = \lambda(0,1,\frac{\mu_1}{b}) \cdot \frac{1}{b}$.

(c) $\lambda(a,b,\mu_1) = \lambda(0,1,\frac{\mu_1-a}{b-a}) \cdot \frac{1}{b-a}$.

Proof of 3.4: Proposition 3.3 ensures for fixed a and b the uniqueness of λ for given μ_1 (and vice versa). The proofs of (a) to (b) are straight forward by some simple transformations.

(a) Let $\lambda = \lambda(a,b,\mu_1)$ and $\tilde{\mu}_1 = \mu_1 - a$.

$$\tilde{\mu}_1 = \mu_1 - a = \frac{be^{\lambda b} - ae^{\lambda a}}{e^{\lambda b} - e^{\lambda a}} - \frac{1}{\lambda} - a \tag{3.21}$$

$$= \frac{be^{\lambda b} - ae^{\lambda a}}{e^{\lambda b} - e^{\lambda a}} - \frac{a(e^{\lambda b} - e^{\lambda a})}{e^{\lambda b} - e^{\lambda a}} - \frac{1}{\lambda} \tag{3.22}$$

$$= \frac{(b-a)e^{\lambda b}}{e^{\lambda b} - e^{\lambda a}} - \frac{1}{\lambda} \tag{3.23}$$

$$= \frac{(b-a)e^{\lambda(b-a)}}{e^{\lambda(b-a)} - 1} - \frac{1}{\lambda} \tag{3.24}$$

And this shows $\lambda(a,b,\mu_1) = \lambda(0,b-a,\mu_1-a)$.

(b) Let $\check{\mu}_1 = \frac{\mu_1}{b}$ and $\check{\lambda} = \lambda(0,1,\check{\mu}_1)$.

$$\check{\mu}_1 = \frac{e^{\check{\lambda}}}{e^{\check{\lambda}} - 1} - \frac{1}{\check{\lambda}} \tag{3.25}$$

$$= \frac{\frac{b}{b}e^{\check{\lambda}b/b}}{e^{\check{\lambda}b/b} - 1} - \frac{1}{\check{\lambda}\frac{b}{b}} \tag{3.26}$$

$$= \left(\frac{be^{\frac{\check{\lambda}}{b}b}}{e^{\frac{\check{\lambda}}{b}b} - 1} - \frac{1}{\frac{\check{\lambda}}{b}}\right) \cdot \frac{1}{b} \tag{3.27}$$

$$= \frac{\mu_1}{b} \tag{3.28}$$

And this shows $\lambda(0,b,\mu_1) = \lambda(0,1,\frac{\mu_1}{b}) \cdot \frac{1}{b}$.

(c) Follows directly from (a) and (b).

•

Notations

The above propositions show, how the general case $\mathcal{X} = \{x \,|\, a \leq x \leq b\}$ may be traced back to the special cases $\mathcal{X} = \{x \,|\, 0 \leq x \leq b\}$ and $\mathcal{X} = \{x \,|\, 0 \leq x \leq 1\}$ respectively. To reduce some effort in these cases, we will introduce some new notations.
For the range of variability we write

$$\mathcal{X}_b := \{x \,|\, 0 \leq x \leq b\}, \text{ in particular } \mathcal{X}_1 = \{x \,|\, 0 \leq x \leq 1\}. \tag{3.29}$$

For λ as function of $\mathcal{X}$ and μ_1 we write

$$\lambda(b, \mu_1) := \lambda(0, b, \mu_1) \tag{3.30}$$

$$\lambda(\mu_1) := \lambda(1, \mu_1) = \lambda(0, 1, \mu_1) \tag{3.31}$$

Because λ is a function of μ_1 and $\mathcal{X}$, we may conversely consider μ_1 as a function of λ and $\mathcal{X}$:

$$\mu_1(a, b, \lambda) := \begin{cases} \frac{be^{\lambda b} - ae^{\lambda a}}{e^{\lambda b} - e^{\lambda a}} - \frac{1}{\lambda} & \text{for } \lambda \neq 0 \\ \frac{a+b}{2} & \text{for } \lambda = 0 \end{cases}. \tag{3.32}$$

In analogy to the shorter notations (3.30) and (3.31), we introduce for μ_1

$$\mu_1(b, \lambda) := \mu_1(0, b, \lambda) \tag{3.33}$$

$$\mu_1(\lambda) := \mu_1(1, \lambda) = \mu_1(0, 1, \lambda) \tag{3.34}$$

To indicate, that a random variable X has a monotonic probability distribution in these special cases we define

$$X|\{(b, \mu_1)\} \sim Mon(b, \mu_1) \;:\Leftrightarrow\; X|\{(0, b, \mu_1)\} \sim Mon(0, b, \mu_1) \tag{3.35}$$

$$X|\{\mu_1\} \sim Mon(\mu_1) \;:\Leftrightarrow\; X|\{(0, 1, \mu_1)\} \sim Mon(0, 1, \mu_1) \tag{3.36}$$

Proposition 3.5 (lower bound of λ) *Let $X = X|\{\mu_1\} \sim Mon(\mu_1)$. Then the distribution parameter $\lambda(\mu_1)$ meets the following inequality:*

$$\lambda(\mu_1) > -\frac{1}{\mu_1} \quad \textit{for all } \mu_1, 0 < \mu_1 < 0.5 \tag{3.37}$$

Proof of 3.5: Since $\mu_1 < 0.5$ we know from proposition 3.2, that $\lambda(\mu_1) < 0$. Then $0 < 1 - e^{\lambda(\mu_1)} < 1$ and $\frac{e^{\lambda(\mu_1)}}{1-e^{\lambda(\mu_1)}} > 0$ implying

$$-\frac{1}{\lambda(\mu_1)} > -\frac{1}{\lambda(\mu_1)} - \frac{e^{\lambda(\mu_1)}}{1 - e^{\lambda(\mu_1)}} = \mu_1 \tag{3.38}$$

or rearranged (3.37)

$$\lambda(\mu_1) > -\frac{1}{\mu_1}\,. \tag{3.39}$$

•

Proposition 3.6 (mirror inverted λ) *Let $X = X|\{(a, b, \mu_1)\} \sim Mon(a, b, \mu_1)$. For the distribution parameter $\lambda(a, b, \mu_1)$ the following equation holds:*

$$\lambda(a, b, \mu_1) = -\lambda(a, b, (b + a) - \mu_1) \tag{3.40}$$

Proof of 3.6: The given statement becomes quite obvious by some considerations about the symmetry of the density functions of $X \sim Mon(a, b, \mu_1)$ and $Y \sim Mon(a, b, (b + a) - \mu_1)$, which are just mirror-inverted to the midpoint $\frac{a+b}{2}$ of $\mathcal{X}$. Nevertheless, the equation may be proved by a simple calculation: Let $\lambda = \lambda(a, b, \mu_1)$ and $\tilde{\mu}_1 = (b + a) - \mu_1$ then

$$\tilde{\mu}_1 = (b + a) - \mu_1 = (b + a) - \frac{be^{\lambda b} - ae^{\lambda a}}{e^{\lambda b} - e^{\lambda a}} + \frac{1}{\lambda} \tag{3.41}$$

$$= \frac{e^{-\lambda a-\lambda b}}{e^{-\lambda a-\lambda b}} \cdot \frac{ae^{\lambda b} - be^{\lambda a}}{e^{\lambda b} - e^{\lambda a}} - \frac{1}{-\lambda} \tag{3.42}$$

$$= \frac{ae^{-\lambda a} - be^{-\lambda b}}{e^{-\lambda a} - e^{-\lambda b}} - \frac{1}{-\lambda} \tag{3.43}$$

$$= \frac{be^{-\lambda b} - ae^{-\lambda a}}{e^{-\lambda b} - e^{-\lambda a}} - \frac{1}{-\lambda} \tag{3.44}$$

Again, with the uniqueness of λ the statement's correctness results immediatly.

•

Proposition 3.7 (limits of λ) *Let* $X = X|\{(a,b,\mu_1)\} \sim Mon(a,b,\mu_1)$. *For* μ_1 *and* $\lambda(a,b,\mu_1)$ *the following holds:*

(a) $\lim_{\mu_1 \to a} \lambda(a,b,\mu_1) \cdot (\mu_1 - a) = -1.$

(b) $\lim_{\mu_1 \to b} \lambda(a,b,\mu_1) \cdot (b - \mu_1) = 1.$

Proof of 3.7: With propositions 3.4 and 3.6 we are able to reduce the above two cases to the single case

$$\lim_{\mu_1 \to 0} \lambda(\mu_1) \cdot \mu_1 = -1\,, \tag{3.45}$$

which is shown by a short calculation:

$$\lim_{\mu_1 \to 0} \lambda(\mu_1) \cdot \mu_1 = \lim_{\mu_1 \to 0} \lambda(\mu_1) \cdot \left(\frac{e^{\lambda(\mu_1)}}{e^{\lambda(\mu_1)} - 1} - \frac{1}{\lambda(\mu_1)} \right) =$$

$$= \lim_{\mu_1 \to 0} \Bigg(\underbrace{\frac{\lambda(\mu_1) e^{\lambda(\mu_1)}}{e^{\lambda(\mu_1)} - 1}}_{\substack{\to 0, \text{ since} \\ \lim_{\mu_1 \to 0} \lambda(\mu_1) = -\infty}} - 1 \Bigg) = -1\,. \tag{3.46}$$

•

In proposition 3.5 we have given a lower bound for $\lambda(\mu_1)$. Now with 3.7 we are able to extend (3.37) with an upper bound.

Proposition 3.8 (lower and upper bounds of λ) *Let* $X = X|\{\mu_1\} \sim Mon(\mu_1)$. *Then the distribution parameter* $\lambda(\mu_1)$ *meets the following inequalities*

$$-\frac{1}{\mu_1} + 2 > \lambda(\mu_1) > -\frac{1}{\mu_1} \quad \textit{for all } 0 < \mu_1 < 0.5 \tag{3.47}$$

Proof of 3.8: First, $\mu_1 \to \frac{1}{2}$ implies $\lambda(\mu_1) \to 0 = -\frac{1}{\frac{1}{2}} + 2$. To proof the left hand side, we differentiate $\lambda(\mu_1)$ with respect to μ_1 and have to show that the derivative is larger than the derivative of $-\frac{1}{\mu_1} + 2$ for all $\mu_1 \in (0, 0.5)$. As aforementioned, μ_1 can be considered as a function of λ, and therefore $\mu_1(\lambda)$ may be differentiated with respect to λ. This was already exercised in the proof of proposition 3.3 where we found[2]

$$\frac{\mathrm{d}\mu_1(\lambda)}{\mathrm{d}\lambda} = \mu_2(\lambda) - (\mu_1(\lambda))^2\,. \tag{3.48}$$

[2]For μ_2 see (3.108).

Since $\lambda(\mu_1)$ is the inverse function of $\mu_1(\lambda)$ we get

$$\frac{\mathrm{d}\lambda(\mu_1)}{\mathrm{d}\mu_1} = \frac{1}{\mu_2 - \mu_1^2} \quad \text{with} \quad \mu_2 = \mu_2(\lambda(\mu_1))\,. \tag{3.49}$$

Together with $\frac{\mathrm{d}}{\mathrm{d}\mu_1}(-\frac{1}{\mu_1} + 2) = \frac{1}{\mu_1^2}$, we note that the following equivalence holds:

$$\frac{1}{\mu_1^2} < \frac{1}{\mu_2 - \mu_1^2} \iff 2\mu_1^2 - \mu_2 > 0\,. \tag{3.50}$$

In fact, we would need to analyse $2\mu_1^2 - \mu_2$ with respect to μ_1, but since $\lambda(\mu_1)$ is strictly monotone increasing in μ_1, we analyse $2\mu_1^2(\lambda) - \mu_2(\lambda)$ with respect to λ:

$$\begin{aligned} 2\mu_1^2 - \mu_2 &= 2\left(\frac{e^\lambda - \lambda e^\lambda - 1}{\lambda(1 - e^\lambda)}\right)^2 - \frac{-2e^\lambda + 2\lambda e^\lambda - \lambda^2 e^\lambda + 2}{\lambda^2(1 - e^\lambda)} = \\ &= \frac{-2\lambda e^{2\lambda} + 2\lambda e^\lambda + \lambda^2 e^{2\lambda} + \lambda^2 e^\lambda}{\lambda^2(1 - e^\lambda)^2} = \frac{-\lambda e^\lambda(2(e^\lambda - 1) - \lambda(e^\lambda + 1))}{\lambda^2(1 - e^\lambda)^2} \end{aligned} \tag{3.51}$$

For $0 < \mu_1 < 0.5$ we have $-\infty < \lambda(\mu_1) < 0$ implying $-\lambda e^\lambda > 0$ and we continue with analysing $h(\lambda) := 2(e^\lambda - 1) - \lambda(e^\lambda + 1)$. The limits of $h(\lambda)$ for $\lambda \to 0$ and $\lambda \to -\infty$, respectively, are

$$\lim_{\lambda \to 0} h(\lambda) = 0\,, \tag{3.52}$$

$$\lim_{\lambda \to -\infty} h(\lambda) = +\infty\,. \tag{3.53}$$

The first two derivatives of $h(\lambda)$ are:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} h(\lambda) = h'(\lambda) = e^\lambda - 1 - \lambda e^\lambda \tag{3.54}$$

$$\lim_{\lambda \to 0} h'(\lambda) = 0 \tag{3.55}$$

$$\lim_{\lambda \to -\infty} h'(\lambda) = -1 \tag{3.56}$$

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} h'(\lambda) = -\lambda e^\lambda > 0 \quad \text{for all } \lambda < 0 \tag{3.57}$$

Thus, we conclude, that $-1 < h'(\lambda) < 0$ and $h'(\lambda)$ is strictly monotone increasing in λ. With this it follows, that $0 < h(\lambda) < +\infty$ is strictly monotone decreasing in λ. Finally, $2\mu_1^2 - \mu_2 > 0$, i.e., $0 < \frac{1}{\mu_1^2} < \frac{1}{\mu_2 - \mu_1^2}$, and thus, $-\frac{1}{\mu_1} + 2 > \lambda(\mu_1)$.

•

Proposition 3.4 implies that $\lambda(b, \mu_1) = \lambda(\frac{\mu_1}{b})\frac{1}{b}$, which we can utilise to extend the above proposition 3.8 to the case $X|\{(b, \mu_1)\} \sim Mon(b, \mu_1)$ where $\mu_1 \in \left(0, \frac{b}{2}\right)$.

Corollary 3.9 (lower and upper bounds of λ) *Let $X = X|\{(b, \mu_1)\} \sim Mon(b, \mu_1)$. Then the distribution parameter $\lambda(b, \mu_1)$ meets the following inequaltities*

$$-\frac{1}{\mu_1} + \frac{2}{b} > \lambda(b, \mu_1) > -\frac{1}{\mu_1} \quad \text{for all } b \in (0, +\infty), \mu_1 \in \left(0, \frac{b}{2}\right) \tag{3.58}$$

Proposition 3.10 (limits of λ) *Let $X = X|\{(a, b, \mu_1)\} \sim Mon(a, b, \mu_1)$ with fixed μ_1. For $\lambda(a, b, \mu_1)$ the following holds:*

(a) $\lim_{b \to +\infty} \lambda(a, b, \mu_1) \cdot (b - a) = -\infty$.

(b) $\lim_{b \to +\infty} \lambda(a, b, \mu_1) = -\frac{1}{\mu_1 - a}$.

Proof of 3.10: Again, utilising propositions 3.4 and 3.6 we will show (b) at first:

$$\begin{aligned} \lim_{b \to +\infty} \lambda(a, b, \mu_1) &= \lim_{b \to +\infty} \lambda(0, b - a, \mu_1 - a) = \\ &= \lim_{b \to +\infty} \lambda(0, 1, \frac{\mu_1 - a}{b - a}) \frac{\mu_1 - a}{b - a} \cdot \frac{1}{\mu_1 - a} = -1 \cdot \frac{1}{\mu_1 - a} = -\frac{1}{\mu_1 - a} \,. \end{aligned} \tag{3.59}$$

Since $\mu_1 \in (a; b)$ is fixed (a) follows immediatly.

•

3.2.2. Properties concerning the distribution

Proposition 3.11 (distribution of standardisation) *Let $X \sim Mon(a, b, \mu_1)$, then for the random variable $Y := \frac{X-a}{b-a}$ it holds $Y \sim Mon(0, 1, \frac{\mu_1 - a}{b-a})$.*

Proof of 3.11: Let f_X and f_Y be the density functions of X and Y respectively, then it is known, that under the proposed linear transformation it holds $f_Y(y) = (b - a) f_X((b - a)y + a)$, which is

$$f_Y(y) = (b - a) f_X((b - a)y + a) \tag{3.60}$$

$$= (b - a) \frac{\lambda(a, b, \mu_1) e^{\lambda(a,b,\mu_1)((b-a)y+a)}}{e^{\lambda(a,b,\mu_1)b} - e^{\lambda(a,b,\mu_1)a}} \cdot \mathbb{1}_{[a,b]}((b - a)y + a) \tag{3.61}$$

$$= \frac{\lambda(a, b, \mu_1)(b - a) e^{\lambda(a,b,\mu_1)(b-a)y}}{e^{\lambda(a,b,\mu_1)(b-a)} - 1} \cdot \mathbb{1}_{[0,1]}(y) \tag{3.62}$$

$$= \frac{\lambda(\frac{\mu_1 - a}{b-a}) e^{\lambda(\frac{\mu_1 - a}{b-a})y}}{e^{\lambda(\frac{\mu_1 - a}{b-a})} - 1} \cdot \mathbb{1}_{[0,1]}(y) \tag{3.63}$$

with the last transformation obtained by using proposition 3.4. This already shows, that $Y \sim Mon(0, 1, \frac{\mu_1 - a}{b-a})$.

•

A kind of inversion of the above proposition is the following corollary.

Corollary 3.12 (distribution of affine transformation) *Let $X \sim Mon(a, b, \mu_1)$, then for the random variable $Z := cX + d$ with $c \neq 0$, $d \in \mathbb{R}$, it holds*

(a) for $c > 0$: $Z \sim Mon(ca + d, cb + d, c\mu_1 + d)$,

(b) for $c < 0$: $Z \sim Mon(cb + d, ca + d, c\mu_1 + d)$.

According to corollary 3.9 and proposition 3.10, $\lambda(b, \mu_1)$ approaches $-\frac{1}{\mu_1}$. Thus, in the case of $X \sim Mon(b, \mu_1)$ with large b and μ_1 very small compared to b, $\lambda(b, \mu_1)$ almost equals $-\frac{1}{\mu_1}$, and the similarity to the exponential distribution becomes evident. A random variable Y is said to be exponentially distributed with parameters $a \in \mathbb{R}$ and $\vartheta > a$, $Y \sim EXP(a, \vartheta)$, if its density function is

$$f_Y(y) = \begin{cases} 0 & \text{for } y < a, \\ \vartheta e^{-\vartheta(y-a)} & \text{for } y \geq a. \end{cases} \tag{3.64}$$

Then the distribution function F_Y of Y is

$$F_Y(y) = \begin{cases} 0 & \text{for } y < a, \\ 1 - e^{-\vartheta(y-a)} & \text{for } y \geq a, \end{cases} \tag{3.65}$$

and the first moment of Y is

$$\mathrm{E}[Y] = a + \frac{1}{\vartheta}\,. \tag{3.66}$$

Proposition 3.13 (convergence to the exponential distribution) *Let $a < b_0 \leq b$, $a < \mu_1 < \frac{a+b_0}{2}$, $X = X|\{(a, b, \mu_1)\} \sim Mon(a, b, \mu_1)$ and $Y \sim EXP(a, \vartheta)$ with $\vartheta = \frac{1}{\mu_1 - a}$. For constant a and μ_1, X converges in distribution to Y for $b \to +\infty$.*

Proof of 3.13: Because of proposition 3.4 we may set $a = 0$ without loss of generality. Let F_X and F_Y be the distribution functions of X and Y, respectively. We have to show that for every $\varepsilon > 0$ and all $x \in \mathbb{R}$ there exists a $B > b_0$ so that $|F_X(x) - F_Y(x)| < \varepsilon$ holds for all $b > B$.
For $x < 0$ the distribution functions of both, the monotonic and the exponential distribution are 0, i.e., they are identical in this case.
For $x \geq 0$ and since $b \to +\infty$ we only have to deal with the case $x \leq b$. We get

$$|F_X(x) - F_Y(x)| = \left| \frac{e^{\lambda(b,\mu_1)x} - 1}{e^{\lambda(b,\mu_1)b} - 1} - (1 - e^{-\vartheta x}) \right| \tag{3.67}$$

$$= \left| \frac{1 - e^{\lambda(b,\mu_1)x} - (1 - e^{-\vartheta x})(1 - e^{\lambda(b,\mu_1)b})}{1 - e^{\lambda(b,\mu_1)b}} \right| \tag{3.68}$$

$$= \frac{|e^{-\vartheta x} + e^{\lambda(b,\mu_1)b} - e^{\lambda(b,\mu_1)x} - e^{-\vartheta x + \lambda(b,\mu_1)b}|}{|1 - e^{\lambda(b,\mu_1)b}|} \tag{3.69}$$

$$\leq \frac{|e^{-\vartheta x} - e^{\lambda(b,\mu_1)x}| + |e^{\lambda(b,\mu_1)b}(1 - e^{-\vartheta x})|}{|1 - e^{\lambda(b,\mu_1)b}|} \tag{3.70}$$

Let $\varepsilon > 0$ be arbitrary. From 3.10(a) we know $\lambda(b, \mu_1)b \to -\infty$ for $b \to +\infty$, i.e., there exists a b_1, so that $e^{\lambda(b,\mu_1)b} < \frac{\varepsilon}{2+\varepsilon}$ and $1 - e^{\lambda(b,\mu_1)b} > 1 - \frac{\varepsilon}{2+\varepsilon}$ for all $b > b_1$. According to 3.10(b) $\lambda(b, \mu_1)$ converges to $-\frac{1}{\mu_1} = -\vartheta$, thus there is a b_2, so that $|e^{-\vartheta x} - e^{\lambda(b,\mu_1)x}| < \frac{\varepsilon}{2+\varepsilon}$ for all $b > b_2$. Together with $(1 - e^{-\vartheta x}) < 1$ we get for all $b > \max\{b_1, b_2\}$

$$|F_X(x) - F_Y(x)| \leq \frac{|e^{-\vartheta x} - e^{\lambda(b,\mu_1)x}| + |e^{\lambda(b,\mu_1)b}(1 - e^{-\vartheta x})|}{|1 - e^{\lambda(b,\mu_1)b}|} < \frac{\frac{\varepsilon}{2+\varepsilon} + \frac{\varepsilon}{2+\varepsilon}}{1 - \frac{\varepsilon}{2+\varepsilon}} = \varepsilon \,. \tag{3.71}$$

And this shows the convergence in distribution of X to Y.

•

Characteristic Function

Another important feature of a probability distribution is its characteristic function, which is defined as follows (see e.g. [29], pp. 44).

Definition 3.14 (characteristic function) *Let X be a continuous random variable with distribution function F_X and density function f_X. Then the* ***charac-***

teristic function $\varphi_{\mathbf{X}}$ *of* X *is defined for all* $t \in \mathbb{R}$ *by*

$$\varphi_X(t) = \mathrm{E}[e^{itX}] = \int_{-\infty}^{+\infty} e^{itx} \, \mathrm{d}F_X(x) = \int_{-\infty}^{+\infty} e^{itx} f_X(x) \, \mathrm{d}x \, . \tag{3.72}$$

Proposition 3.15 (characteristic function) *Let* $X \sim Mon(a, b, \mu_1)$ *and* $\lambda = \lambda(a, b, \mu_1)$*. Then the characteristic function* φ_X *is*

$$\varphi_X(t) = e^{ita} \frac{\lambda}{\left(e^{\lambda \cdot (b-a)} - 1\right)} \frac{1}{it + \lambda} \left(e^{(it+\lambda)(b-a)} - 1\right) . \tag{3.73}$$

Proof of 3.15: We will use a known property of the characteristic function under linear transformation of the random variable:

> Let Y and Z be random variables with $Z = \alpha + \beta Y$. Then for their characteristic functions it holds
>
> $$\varphi_Z(t) = e^{i\alpha t} \varphi_Y(\beta t) \, . \tag{3.74}$$

With this and since the general case $X \sim Mon(a, b, \mu_1)$ may be traced back to $Y \sim Mon(0, 1, \frac{\mu_1 - a}{b-a})$ by proposition 3.11 via the transformation $Y = \frac{X-a}{b-a}$, we at first derive $\varphi_Y(t)$ for $Y \sim Mon(0, 1, \mu_1)$. With $\lambda_Y = \lambda(0, 1, \mu_1)$ we calculate

$$\varphi_Y(t) = \int_{-\infty}^{+\infty} e^{itx} f_X(x) \, \mathrm{d}x \; = \; \int_0^1 e^{itx} \frac{\lambda_Y e^{\lambda_Y x}}{e^{\lambda_Y} - 1} \, \mathrm{d}x \tag{3.75}$$

$$= \frac{\lambda_Y}{e^{\lambda_Y} - 1} \int_0^1 e^{(it+\lambda_Y)x} \, \mathrm{d}x \tag{3.76}$$

$$= \frac{\lambda_Y}{e^{\lambda_Y} - 1} \frac{1}{it + \lambda_Y} \left[e^{(it+\lambda_Y)x} \right]_0^1 \tag{3.77}$$

$$= \frac{\lambda_Y}{e^{\lambda_Y} - 1} \frac{1}{it + \lambda_Y} \left(e^{it+\lambda_Y} - 1 \right) \tag{3.78}$$

For the general case we remind that $\lambda_X = \lambda(a, b, \mu_1) = \lambda(\frac{\mu_1 - a}{b-a})\frac{1}{b-a} = \lambda_Y \frac{1}{b-a}$ (see 3.4). Then we finally get $\varphi_X(t)$ for $X \sim Mon(a, b, \mu_1)$

$$\varphi_X(t) = e^{iat}\varphi_Y((b-a)t) \tag{3.79}$$

$$= e^{iat}\frac{\lambda_Y}{e^{\lambda_Y} - 1}\frac{1}{i(b-a)t + \lambda_Y}\left(e^{i(b-a)t+\lambda_Y} - 1\right) \tag{3.80}$$

$$= e^{iat}\frac{\lambda_X(b-a)}{e^{\lambda_X(b-a)} - 1}\frac{1}{i(b-a)t + \lambda_X(b-a)}\left(e^{i(b-a)t+\lambda_X(b-a)} - 1\right) \tag{3.81}$$

$$= e^{iat}\frac{\lambda_X}{e^{\lambda_X(b-a)} - 1}\frac{1}{it + \lambda_X}\left(e^{(it+\lambda_X)(b-a)} - 1\right). \tag{3.82}$$

•

From the properties of the characteristic function it is known, that the convergence in distribution is equivalent to the convergence of the corresponding characteristic functions. According to (3.7b), for $[a, b] = [0, 1]$ and $\lambda \to 0$ or equivalent $\mu_1 \to \frac{1}{2}$ the monotonic distribution $Mon(0, 1, \mu_1)$ converges to the constant distribution $\mathrm{U}(0, 1)$, the same holds for the characteristic function. With $X \sim Mon(0, 1, \mu_1)$ we get

$$\lim_{\lambda\to 0}\varphi_X(t) = \lim_{\lambda\to 0}\frac{\lambda}{e^{\lambda} - 1}\frac{1}{it + \lambda}\left(e^{it+\lambda} - 1\right) \tag{3.83}$$

$$= \frac{1}{it}\left(e^{it} - 1\right), \tag{3.84}$$

since $\lim_{\lambda\to 0}\frac{\lambda}{e^{\lambda}-1} = 1$. And this is just the characteristic function φ_Y of a constant distributed random variable Y on the closed unit interval $[0, 1]$, i.e., $Y \sim \mathrm{U}(0, 1)$. In 3.13 we have shown the convergence in distribution of a monotonic distributed random variable $X \sim Mon(a, b, \mu_1)$ for μ_1 constant to an exponentially distributed random variable $Y \sim EXP(a, \vartheta)$ as the upper bound of $\mathcal{X}$ converges to $+\infty$. Transferred to the corresponding characteristic functions φ_X and $\varphi_Y(t) = e^{iat}\frac{\vartheta}{\vartheta - it}$ that means that

$$|\varphi_X(t) - \varphi_Y(t)| \xrightarrow{b\to+\infty} 0 \tag{3.85}$$

for every $t \in \mathbb{R}$. Since $\varphi_X(t)$ and $\varphi_Y(t)$ respectively are complex functions, the direct proof of this convergence needs some laborious calculations. As in 3.13 we

set $a = 0$ without loss of generality, then we get

$$|\varphi_X(t) - \varphi_Y(t)| = \frac{t^2(\vartheta + \lambda(\frac{\mu_1}{b}))^2 + A + B + C + D}{\left(e^{b\lambda(\frac{\mu_1}{b})} - 1\right)^2 (t^2 + \vartheta^2)\left(t^2 + \lambda^2(\frac{\mu_1}{b})\right)} \tag{3.86}$$

with

$$A = -2e^{b\lambda(\frac{\mu_1}{b})}t^2(\vartheta^2 + \vartheta\lambda(\frac{\mu_1}{b}))\,, \tag{3.87}$$

$$B = \quad e^{2b\lambda(\frac{\mu_1}{b})}(t^2\vartheta^2 + t^2\lambda^2(\frac{\mu_1}{b}) + 2\vartheta^2\lambda^2(\frac{\mu_1}{b}))\,, \tag{3.88}$$

$$C = -2e^{b\lambda(\frac{\mu_1}{b})}\lambda(\frac{\mu_1}{b})\left(\left(\vartheta - e^{b\lambda(\frac{\mu_1}{b})}\vartheta + \lambda(\frac{\mu_1}{b})\right)t^2 + e^{b\lambda(\frac{\mu_1}{b})}\vartheta^2\lambda(\frac{\mu_1}{b})\right)\cos(bt)\,, \tag{3.89}$$

$$D = -2e^{b\lambda(\frac{\mu_1}{b})}t\vartheta\lambda(\frac{\mu_1}{b})\left(e^{b\lambda(\frac{\mu_1}{b})} - 1\right)(\vartheta + \lambda(\frac{\mu_1}{b}))\sin(bt)\,. \tag{3.90}$$

Since $b\lambda(\frac{\mu_1}{b}) \to -\infty$ for $b \to +\infty$ in all four terms A, B, C and D the factor $e^{b\lambda(\frac{\mu_1}{b})}$ and $e^{2b\lambda(\frac{\mu_1}{b})}$, respectively, converges to 0. Simultaneous, the other factors remain bounded. Because $\vartheta = \frac{1}{\mu_1}$ and $\lambda(\frac{\mu_1}{b}) \to -\frac{1}{\mu_1}$ for $b \to +\infty$ we get $(\vartheta + \lambda(\frac{\mu_1}{b})) \to 0$ as well. Finally, since the denominator obviously does not converge to 0, convergence in distribution of X to Y follows again.

Quantile Functions

First we give a very general definition of *quantiles* and *quantile functions*.

Definition 3.16 (upper and lower quantile function) *Let X be a univariate random variable with distribution function $F_X\colon \mathbb{R} \to \mathbb{R}$. For $\gamma \in [0,1]$ any $x \in \mathbb{R}$ solving simultaneously the two inequalities*

$$\begin{gathered} \mathrm{P}(X \le x) \ge \gamma\,, \quad \mathrm{P}(X \ge x) \ge 1 - \gamma \\ \Longleftrightarrow \\ \lim_{t \nearrow x} F_X(t) =: F(x^-) \le \gamma \le F_X(x) \end{gathered} \tag{3.91}$$

is called γ-quantile.
The lower quantile function $z_X^{(L)} : (0,1) \to \mathbb{R}$ is defined by

$$z_X^{(L)}(\gamma) \;=\; \inf\{x \,|\, F(x^-) \le \gamma \le F(x)\}\,. \tag{3.92}$$

The upper quantile function $z_X^{(U)} : (0,1) \to \mathbb{R}$ *is defined by*

$$z_X^{(U)}(\gamma) \;=\; \sup\{x \,|\, F(x^-) \le \gamma \le F(x)\}\,. \tag{3.93}$$

It is known, that for a continuous and strictly increasing distribution function F_X we have $z_X^{(L)}(\gamma) = z_X^{(U)}(\gamma)$ for all $\gamma \in (0,1)$, i.e., the γ-quantile $z_X(\gamma)$ is unique and defined by $F_X(z_X(\gamma)) = \gamma$.
Since the distribution function of $X|\{(a,b,\mu_1)\} \sim Mon(a,b,\mu_1)$ is continuous and strictly increasing on the compact interval $[a,b]$ with $F_{X|\{(a,b,\mu_1)\}}(a) = 0$ and $F_{X|\{(a,b,\mu_1)\}}(b) = 1$ we get the quantile function $z_{X|\{(a,b,\mu_1)\}}(\gamma)$ for all $\gamma \in [0,1]$ as follows

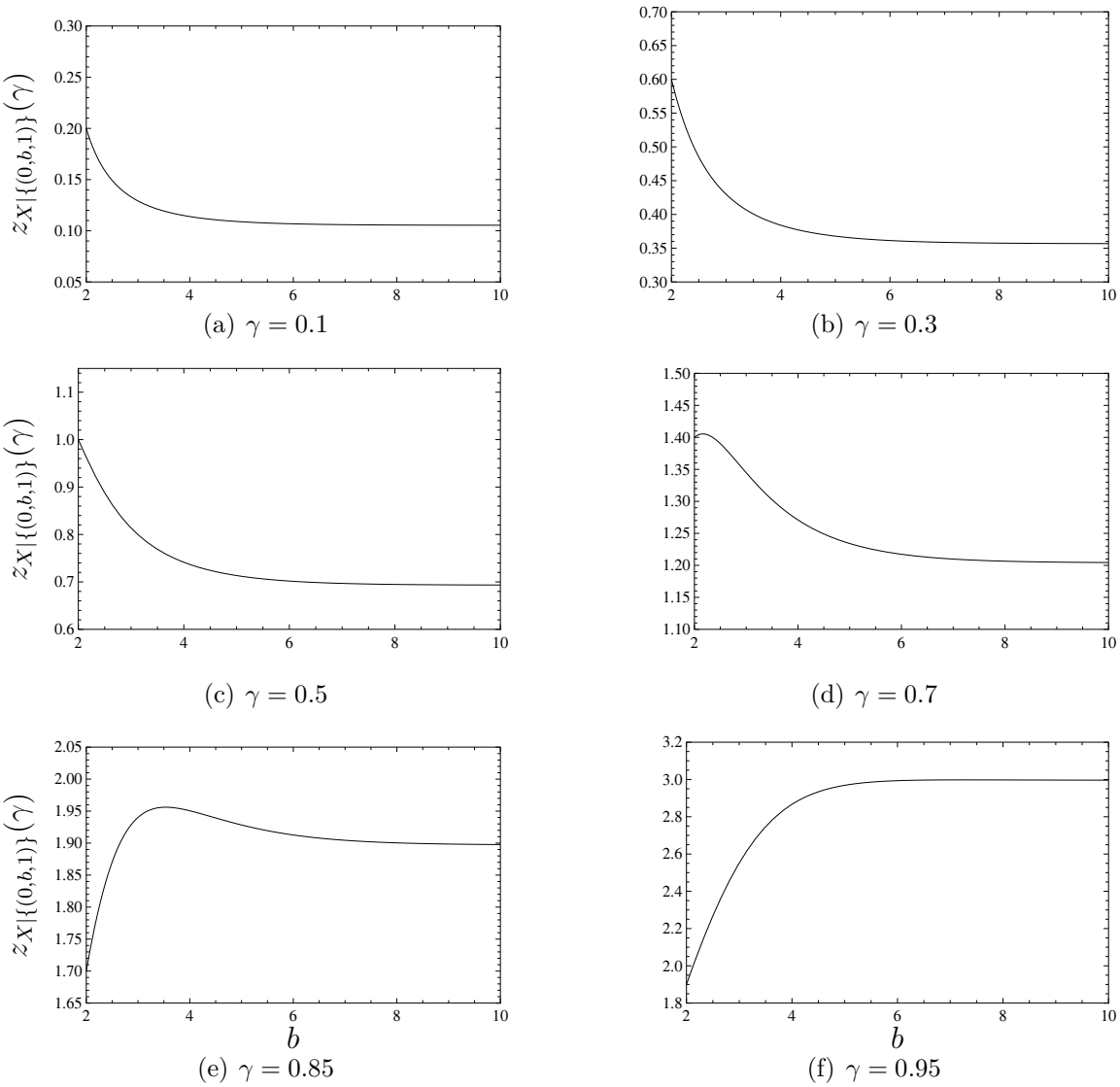
$$\begin{aligned} z_{X|\{(a,b,\mu_1)\}}(\gamma) \;&=\; F^{-1}_{X|\{(a,b,\mu_1)\}}(\gamma) = \\ &= \frac{1}{\lambda(a,b,\mu_1)} \ln\Big(\gamma\big(e^{\lambda(a,b,\mu_1)b} - e^{\lambda(a,b,\mu_1)a}\big) + e^{\lambda(a,b,\mu_1)a}\Big)\,. \end{aligned} \tag{3.94}$$

To get an impression of the dependency of the quantile function on b let $a = 0$ and $\mu_1 = 1$ be fixed while $b > 2\mu_1 = 2$, i.e., we have a monotonously decreasing density function for $X|\{(0,b,\mu_1)\}$. Figure 3.3 displays its courses for $\gamma = 0.1, 0.3, 0.5, 0.7, 0.85, 0.95$. The limiting values to the left, i.e., for $b \searrow 2\mu_1$, and to the right, i.e., for $b \to +\infty$, are explained by the convergence of the montone distribution. For $b \searrow 2\mu_1$ the monotonic distribution converges to the uniform distribution $\mathrm{U}(0, 2\mu_1)$ and, thus, the quantile $z_{X|\{(0,b,\mu_1)\}}(\gamma)$ converges to $\gamma \cdot 2\mu_1$. For $b \to +\infty$ the monotonic distribution converges to the exponential distribution $EXP(1/\mu_1)$ and the quantile $z_{X|\{(0,b,\mu_1)\}}(\gamma)$ converges to the quantile of the exponential distribution $-\mu_1 \ln(1-\gamma)$. Not only, that the relation between the limiting values changes[3], but also the course of $z_{X|\{(0,b,\mu_1)\}}(\gamma)$ is neither monotoneously decreasing nor increasing in b for some values γ. It seems, that values $\gamma \ge 0.95$ lead to a monotonously increasing quantile function in b with limiting value $-\mu_1 \ln(1-\gamma)$ for $b \to +\infty$.

Exponential Family

If a family of distributions is an exponential family with respect to a distribution parameter $\vartheta \in \Theta \subseteq \mathbb{R}^d$, some useful properties result for statistical inference.

[3] $\gamma \cdot 2\mu_1$ is larger than $-\mu_1 \ln(1-\gamma)$ for approximatly $\gamma < 0.796812$.



(a) $\gamma = 0.1$ (b) $\gamma = 0.3$ (c) $\gamma = 0.5$ (d) $\gamma = 0.7$ (e) $\gamma = 0.85$ (f) $\gamma = 0.95$

Figure 3.3.: The quantile function $z_{X|\{(0,b,1)\}}(\gamma)$ in dependency of b for different values γ.

A family of distributions is said to be a k-parameter exponential family if the density functions can be written as

$$f_{X|\{\vartheta\}}(x) = h(x)c(\vartheta)\exp\Big(\sum_{i=1}^{k} w_i(\vartheta)t_i(x)\Big)\,, \tag{3.95}$$

with functions

$$\begin{aligned} &h : \mathcal{X} \to \mathbb{R},\ h(x) \geq 0 \text{ for all } x \in \mathcal{X}\,,\\ &t_i : \mathcal{X} \to \mathbb{R} \text{ for all } i = 1,\dots,k\,,\\ &c : \Theta \to \mathbb{R},\ c(\vartheta) \geq 0 \text{ for all } \vartheta \in \Theta\,,\\ &w_i : \Theta \to \mathbb{R} \text{ for all } i = 1,\dots,k\,. \end{aligned} \tag{3.96}$$

We should note, that h and all t_i do not depend on ϑ, and c and all w_i do not depend on x.

The density functions $f_{X|\{(a,b,\mu_1)\}}$ of the monotonic distribution $Mon(a,b,\mu)$ generate an exponential family in $\mu_1 \in (a,b)$ and $\lambda \in \mathbb{R}$, since for all $\mu_1 \in (a,b)$ and $\lambda \in \mathbb{R}$, respectively, the density functions can be written as (3.95) with

$$\begin{aligned} h(x) &= \mathbb{1}_{[a,b]}(x)\,,\\ t(x) &= x\,,\\ c(\lambda(a,b,\mu_1)) &= \frac{1}{\int_a^b e^{\lambda(a,b,\mu_1)t}\,\mathrm{d}t}\,,\\ w(\lambda(a,b,\mu_1)) &= \lambda(a,b,\mu_1)\,. \end{aligned} \tag{3.97}$$

Function $c(\lambda(a,b,\mu_1))$ is written with the integral in the denominator, since in this way we do not have to treat the case $\mu_1 = \frac{a+b}{2}$ ($\Leftrightarrow \lambda = 0$) separately.

It is worth mentioning, that the monotonic distribution $Mon(a,b,\mu)$ does not generate an exponential family neither in a, nor in b, nor in $\mathcal{X} = [a,b]$. This results from function $h(x) = \mathbb{1}_{[a,b]}(x)$, which cannot be separated into a function depending on x alone and functions depending alone on a, b and $[a,b]$, respectively.

3.2.3. Moments of $X|(a, b, \mu_1)$

In this section we want to analyse some properties of the moments of $X \sim Mon(a, b, \mu_1)$.

All moments $\mathrm{E}[X^k]$ of any random variable X – in case of existence – may be calculated from the first k moments $\mathrm{E}[Y^i]$, $1 \leq i \leq k$, of a linear transformation $Y = \frac{X-a}{b-a}$ of X by the well known formulas:

$$\mathrm{E}[Y^k] = \mathrm{E}\left[\left(\frac{X-a}{b-a}\right)^k\right] = \frac{1}{(b-a)^k}\sum_{i=0}^{k}\binom{k}{i}a^i\mathrm{E}[X^{k-i}] \tag{3.98a}$$

$$\mathrm{E}[X^k] = \mathrm{E}\left[((b-a)Y+a)^k\right] = \sum_{i=0}^{k}\binom{k}{i}a^i(b-a)^{k-i}\mathrm{E}[Y^{k-i}] \tag{3.98b}$$

Therefore, we firstly examine the case $X \sim Mon(0, 1, \mu_1)$ and write $\lambda = \lambda(\mu_1)$.

The k-th moment of X is then calculated in the usual way by the integral

$$\mathrm{E}[X^k] \;=\; \frac{\int_0^1 x^k e^{\lambda x}\,\mathrm{d}x}{\int_0^1 e^{\lambda x}\,\mathrm{d}x} = \int_0^1 \frac{\lambda x^k}{e^\lambda - 1}e^{\lambda x}\,\mathrm{d}x\,. \tag{3.99}$$

By successive integration by parts the above integral can be explicitly solved and we immediately get the following proposition 3.17:

Proposition 3.17 (moments) *Let $X \sim Mon(0, 1, \mu_1)$. Then for $k \in \mathbb{N}$ the k-th moment of X is*

$$\mathrm{E}[X^k] \;= \mu_k \;=\; \left(1 - \frac{\sum_{\ell=1}^{k}\frac{(-\lambda)^\ell}{\ell!}}{e^{-\lambda}-1}\right)\frac{k!}{(-\lambda)^k}\,. \tag{3.100}$$

With the above equation (3.100) the moments are at hand. But for the sake of completeness we also give the moment generating function $M_X(t) := \mathrm{E}[e^{tX}]$ of X. The calculation of M_X would be straight forward and utilises the analogue to the property of characteristic function under linear transformation of the random variable:

Let Y and Z be random variables with $Z = \alpha + \beta Y$. Then for their moment generating functions it holds

$$M_Z(t) = e^{\alpha t} M_Y(\beta t)\,. \tag{3.101}$$

Proposition 3.18 (moment generating function) *Let $X \sim Mon(a, b, \mu_1)$. The moment generating function M_X of X is*

$$M_X(t) \;=\; \frac{\lambda}{e^{\lambda b} - e^{\lambda a}} \frac{e^{(t+\lambda)b} - e^{(t+\lambda)a}}{t+\lambda}\,. \tag{3.102}$$

Remark In [13] (p. 137) a formula of the moment generating function in case of an exponential family of distributions is referenced:

> According to Brown (1986, Section 1.1)[4], to define an exponential family of distributions we start with a nonnegative function $\nu(x)$ and define the set $\mathcal{N}$ by
>
> $$\mathcal{N} = \left\{\theta : \int_{\mathcal{X}} e^{\theta x}\nu(x)\,\mathrm{d}x < \infty\right\}. \tag{3.103}$$
>
> If we let $\lambda(\theta) = \int_{\mathcal{X}} e^{\theta x}\nu(x)\,\mathrm{d}x$, the set of probability densities defined by
>
> $$f(x|\theta) = \frac{e^{\theta x}\nu(x)}{\lambda(\theta)},\ x \in \mathcal{X},\ \theta \in \mathcal{N}, \tag{3.104}$$
>
> is an *exponential family*. The moment-generating function of $f(x|\theta)$ is
>
> $$M_X(t) = \int_{\mathcal{X}} e^{tx} f(x|\theta)\,\mathrm{d}x = \frac{\lambda(t+\theta)}{\lambda(\theta)} \tag{3.105}$$
>
> and hence exists by construction.

Adapting the notation to ours and with function $c(\lambda) = \frac{\lambda}{e^{\lambda b} - e^{\lambda a}}$ from (3.97), we may rewrite (3.102) as

$$M_X(t) = \frac{c(\lambda)}{c(t+\lambda)}\,. \tag{3.106}$$

[4] Brown, L.D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory.* Institute of Mathematical Statistics Lecture Notes – Monograph Series. IMS, Hayward, CA.

In addition we provide the first four moments of $Y \sim Mon(\mu_1)$ with $\lambda = \lambda(\mu_1)$:

$$\mathrm{E}[Y] = \mu_1 = \left(1 - \frac{-\lambda}{e^{-\lambda} - 1}\right) \frac{1}{-\lambda} = \frac{1 - \lambda - e^{-\lambda}}{\lambda(e^{-\lambda} - 1)} \tag{3.107}$$

$$\mathrm{E}[Y^2] = \mu_2 = \left(1 - \frac{-\lambda + \frac{\lambda^2}{2}}{e^{-\lambda} - 1}\right) \frac{2}{\lambda^2} = \frac{-2 + 2\lambda - \lambda^2 + 2e^{-\lambda}}{\lambda^2(e^{-\lambda} - 1)} \tag{3.108}$$

$$\mathrm{E}[Y^3] = \mu_3 = \left(1 - \frac{-\lambda + \frac{\lambda^2}{2} - \frac{\lambda^3}{6}}{e^{-\lambda} - 1}\right) \frac{6}{-\lambda^3} = \frac{6 - 6\lambda + 3\lambda^2 - \lambda^3 - 6e^{-\lambda}}{\lambda^3(e^{-\lambda} - 1)} \tag{3.109}$$

$$\begin{aligned} \mathrm{E}[Y^4] = \mu_4 &= \left(1 - \frac{-\lambda + \frac{\lambda^2}{2} - \frac{\lambda^3}{6} + \frac{\lambda^4}{24}}{e^{-\lambda} - 1}\right) \frac{24}{\lambda^4} = \\ &= \frac{-24 + 24\lambda - 12\lambda^2 + 4\lambda^3 - \lambda^4 + 24e^{-\lambda}}{\lambda^4(e^{-\lambda} - 1)} \end{aligned} \tag{3.110}$$

The obvious pattern in the sequence of moments for $Y \sim Mon(\mu_1)$ can be easily deduced by successive partial integration:

$$\mathrm{E}[Y^k] = (-1)^k \frac{k!}{\lambda^k} \left(\frac{e^{-\lambda} - 1 - \sum_{l=1}^{k} (-1)^l \frac{\lambda^l}{l!}}{e^{-\lambda} - 1}\right) \tag{3.111}$$

And with (3.98b) we get for the corresponding first four moments of $X \sim Mon(a, b, \mu_1)$ with $\lambda = \lambda(a, b, \mu_1)$:

$$\mathrm{E}[X] = \mu_1 = \frac{(-1 + \lambda b)e^{\lambda b} - (-1 + \lambda a)e^{\lambda a}}{\lambda(e^{\lambda b} - e^{\lambda a})} \tag{3.112}$$

$$\mathrm{E}[X^2] = \mu_2 = \frac{(2 - 2\lambda b + \lambda^2 b^2)e^{\lambda b} - (2 - 2\lambda a + \lambda^2 a^2)e^{\lambda a}}{\lambda^2(e^{\lambda b} - e^{\lambda a})} \tag{3.113}$$

$$\mathrm{E}[X^3] = \mu_3 = \frac{(-6 + 6\lambda b - 3\lambda^2 b^2 + \lambda^3 b^3)e^{\lambda b} - (-6 + 6\lambda a - 3\lambda^2 a^2 + \lambda^3 a^3)e^{\lambda a}}{\lambda^3(e^{\lambda b} - e^{\lambda a})} \tag{3.114}$$

$$\begin{aligned} \mathrm{E}[X^4] = \mu_4 = \frac{1}{\lambda^4(e^{\lambda b} - e^{\lambda a})} \Big(&(24 - 24\lambda b + 12\lambda^2 b^2 - 4\lambda^3 b^3 + \lambda^4 b^4)e^{\lambda b} - \\ &- (24 - 24\lambda a + 12\lambda^2 a^2 - 4\lambda^3 a^3 + \lambda^4 a^4)e^{\lambda a} \Big) \end{aligned} \tag{3.115}$$

The central moments $\tilde{\mu}_k := \mathrm{E}[(X - \mu_1)^k]$ are related to the moments $\mu_k = \mathrm{E}[X^k]$ by the well known transformation

$$\mathrm{E}[(X - \mu_1)^k] \; = \tilde{\mu}_k \; = \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-i} \mu_i \mu_1^{k-i} \tag{3.116}$$

With the above stated formulas it is quite easy to derive the expression for the variance $\mathrm{Var}[X] = \tilde{\mu}_2 = \sigma^2$ of $X \sim Mon(\mu_1)$:

$$\mathrm{Var}[X] = \tilde{\mu}_2 = \sigma^2 \;=\; \frac{1 - e^{-\lambda}(2 + \lambda^2) + e^{-2\lambda}}{\lambda^2 (e^{-\lambda} - 1)^2} \,. \tag{3.117}$$

Higher central moments, together with the skewness $\gamma_1 = \frac{\mathrm{E}[(X-\mu_1)^3]}{\sigma^3} = \frac{\tilde{\mu}_3}{\tilde{\mu}_2^{3/2}}$ and kurtosis $\delta = \frac{\mathrm{E}[(X-\mu_1)^4]}{\sigma^4} = \frac{\tilde{\mu}_4}{\tilde{\mu}_2^2}$ and kurtosis excess $\gamma_2 = \delta_X - 3$ respectively, lack of a intuitionally comprehensive (clearly arranged) form and are omitted here – but their dependency on μ_1 is illustrated in figure 3.4 for $X \sim Mon(\mu_1)$ and $\mu_1 \in (0, 1)$. We may note, that the kurtosis excess converges to 6 for $\mu_1 \to 0$ and $\mu_1 \to 1$ respectively. The former expresses the convergence in distribution to the exponential distribution whose kurtosis excess equals exactly 6.

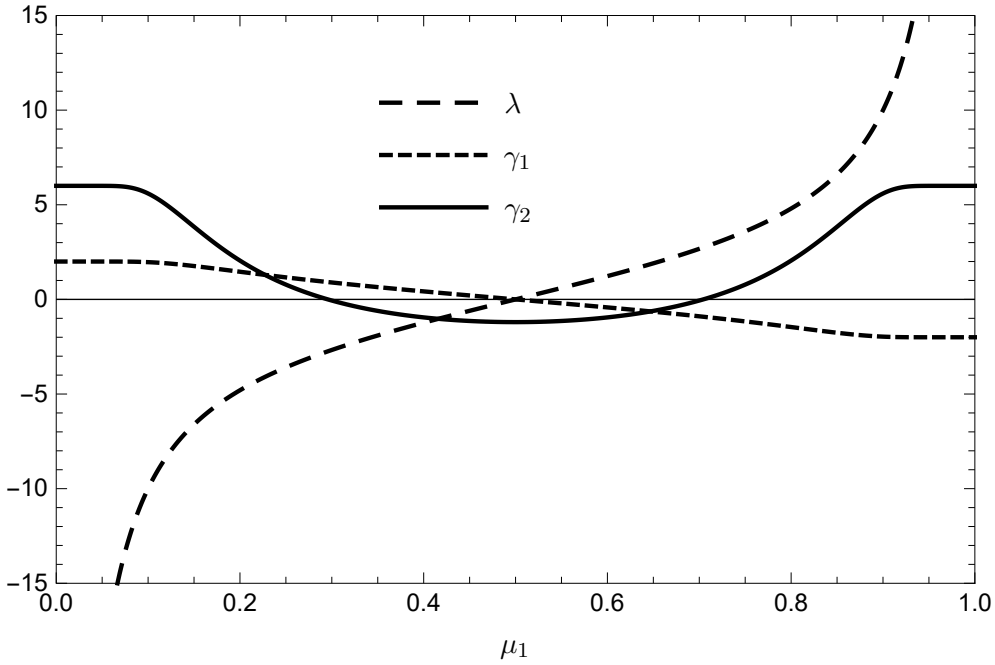


Figure 3.4.: Plot of distribution parameter λ, skewness γ_1 and kurtosis excess γ_2 as functions of μ_1 for $X \sim Mon(\mu_1)$ and $\mu_1 \in (0, 1)$.

3.3. Sample Distributions

If nothing else is stated, $X_1, X_2, \ldots, X_n$ should be an i.i.d. sample for $X \sim Mon(a, b, \mu_1)$, abbreviated with $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$. Troughout this section we want to derive the distributions of functions $T : \mathbb{R}^n \to \mathbb{R}^q$, $q \in \mathbb{N}$, of $\boldsymbol{X}$, i.e., statistics of $\boldsymbol{X}$.

3.3.1. Joint Distribution of $X_1, \ldots, X_n$

Since $X_1, \ldots, X_n$ are independent their joint probability density function equals the product of the individual densities:

$$\begin{aligned} f_{\boldsymbol{X}|\{(a,b,\mu_1)\}}(\boldsymbol{x}) &= \prod_{i=1}^{n} f_{X_i|\{(a,b,\mu_1)\}}(x_i) \\ &= \frac{e^{\lambda(a,b,\mu_1)\sum_{i=1}^{n} x_i}}{\left(\int_a^b e^{\lambda(a,b,\mu_1)t}\,\mathrm{d}t\right)^n} \mathbb{1}_{[a,b]^n}(\boldsymbol{x}) \end{aligned} \tag{3.118}$$

We note, that for all $\boldsymbol{x} \in \{(x_1, \ldots, x_n) \,|\, a \leq x_i \leq b\}$ the value of the joint density function depends only on the sum $\sum_{i=1}^{n} x_i$.

3.3.2. Sum of $X_1, \ldots, X_n$

Let $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i =: K_n$. Since the general case $\mathcal{X} = [a, b]$ again may be traced back to the special case $\mathcal{X} = [0, 1]$ we assume $X_i \sim Mon(\mu_1)$.
Firstly let $n = 2$. To derive the density function of K_2 we calculate the convolution of $f_{X_1|\{\mu_1\}}$ and $f_{X_2|\{\mu_1\}}$:

$$\begin{aligned} f_{K_2|\{\mu_1\}}(x) &= f_{(X_1+X_2)|\{\mu_1\}}(x) = \int f_{X_1|\{\mu_1\}}(y) f_{X_2|\{\mu_1\}}(x-y)\,\mathrm{d}y \\ &= \int \frac{\lambda}{e^\lambda - 1} e^{\lambda y} \mathbb{1}_{[0,1]}(y) \cdot \frac{\lambda}{e^\lambda - 1} e^{\lambda(x-y)} \mathbb{1}_{[0,1]}(x-y)\,\mathrm{d}y \\ &= \left(\frac{\lambda}{e^\lambda - 1}\right)^2 e^{\lambda x} \underbrace{\int \mathbb{1}_{[0,1]}(y) \mathbb{1}_{[0,1]}(x-y)\,\mathrm{d}y}_{(*)} \end{aligned} \tag{3.119}$$

Where $(*)$ is the convolution operator of two independent random variables uniformly distributed on $[0,1]$. The resulting density function may be found in many textbooks, e.g. [66, 67]:

Let $U_1, U_2, \ldots, U_n$ be independent random variables uniformly distributed on $[0,1]$, i.e.,

$$U_i \sim \mathrm{U}(0,1)\,,\ i = 1, \ldots, n$$

Then the density function of $\sum_{i=1}^{n} U_i$ is:

$$f_{\sum_{i=1}^{n} U_i}(x) = \frac{1}{(n-1)!} \sum_{i=0}^{\lfloor x \rfloor} (-1)^i \binom{n}{i} (x-i)^{n-1} \mathbb{1}_{[0,n]}(x) \tag{3.120}$$

$$\text{with} \qquad \lfloor x \rfloor = \max\{q \in \mathbb{Z} | q \leq x\}$$

With this on hand we are able to derive the density function of K_n, provided by the subsequent proposition 3.19.

Proposition 3.19 (sum of i.i.d. monotonic distributed random variables) *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with $X_i \sim Mon(\mu_1)$, $U_1, U_2, \ldots, U_n$ be i.i.d. random variables with $U_i \sim \mathrm{U}(0,1)$, and $f_{\sum_{i=1}^{n} U_i}(x)$ the density function of $\sum_{i=1}^{n} U_i$. Then the following holds for the density function of $K_n = \sum_{i=1}^{n} X_i$*

$$f_{K_n|\{\mu_1\}}(x) = \left(\frac{\lambda}{e^{\lambda} - 1}\right)^n e^{\lambda x} f_{\sum_{i=1}^{n} U_i}(x) \tag{3.121}$$

Proof of 3.19: The proof will be performed by induction with respect to n. For $n = 2$ the proposition holds as shown by (3.119). Next assume that (3.121) is true for any $m \leq n$. The density function $f_{K_{n+1}|\{\mu_1\}}$ of K_{n+1} is obtained by the

convolution of $f_{K_n|\{\mu_1\}}$ and $f_{X_{n+1}|\{\mu_1\}}$ yielding:

$$\begin{aligned}
f_{K_{n+1}|\{\mu_1\}}(x) &= \int f_{K_n|\{\mu_1\}}(y) f_{X_{n+1}|\{\mu_1\}}(x-y)\,\mathrm{d}y \\
&= \int \left(\frac{\lambda}{e^\lambda - 1}\right)^n e^{\lambda y} f_{\sum_{i=1}^n U_i}(y) \left(\frac{\lambda}{e^\lambda - 1}\right) e^{\lambda(x-y)} f_{U_{n+1}}(x-y)\,\mathrm{d}y \\
&= \left(\frac{\lambda}{e^\lambda - 1}\right)^{n+1} e^{\lambda x} \int f_{\sum_{i=1}^n U_i}(y) f_{U_{n+1}}(x-y)\,\mathrm{d}y \\
&= \left(\frac{\lambda}{e^\lambda - 1}\right)^{n+1} e^{\lambda x} f_{\sum_{i=1}^n U_i + U_{n+1}}(x) \\
&= \left(\frac{\lambda}{e^\lambda - 1}\right)^{n+1} e^{\lambda x} f_{\sum_{i=1}^{n+1} U_i}(x)
\end{aligned}$$

•

To summarise, the probability density function of $K_n = \sum_{i=1}^n X_i$ with $X_i \sim Mon(\mu_1)$ is

$$f_{K_n|\{\mu_1\}}(x) = \left(\frac{\lambda}{e^\lambda - 1}\right)^n e^{\lambda x} \frac{1}{(n-1)!} \sum_{i=0}^{\lfloor x \rfloor} (-1)^i \binom{n}{i} (x-i)^{n-1} \mathbb{1}_{[0,n]}(x) \tag{3.122}$$

With (3.122) the density function of the mean $\overline{X} := \frac{1}{n} K_n$ is obtained:

$$f_{\overline{X}|\{\mu_1\}}(x) = n f_{K_n|\{\mu_1\}}(nx) = n \left(\frac{\lambda}{e^\lambda - 1}\right)^n e^{\lambda n x} f_{\sum_{i=1}^n U_i}(nx) \tag{3.123}$$

Furthermore the distribution function $F_{K_n|\{\mu_1\}}$ of K_n for $x \in [0, n]$ is given by:

$$\begin{aligned}
F_{K_n|\{\mu_1\}}(x) = \left(\frac{1}{e^\lambda - 1}\right)^n &\cdot \\
\cdot \Bigg\{ e^{\lambda x} \left[\sum_{i=0}^{\lfloor x \rfloor} (-1)^i \binom{n}{i} \left(\sum_{k=0}^{n-2} (-1)^k (x-i)^{n-1-k} \frac{\lambda^{n-1-k}}{(n-1-k)!} \right) \right] + & \\
+ \left[\sum_{i=0}^{\lfloor x \rfloor} (-1)^{n+1+i} \binom{n}{i} (e^{\lambda x} - e^{\lambda i}) \right] \Bigg\} &
\end{aligned} \tag{3.124}$$

If $X_i \sim Mon(a, b, \mu_1)$, then according to proposition 3.11 $Y_i := \frac{X_i - a}{b-a} \sim Mon(0, 1, \frac{\mu_1 - a}{b-a})$. Let $P_n := \sum_{i=1}^n Y_i$ and $K_n := \sum_{i=1}^n X_i$, then

$$P_n = \sum_{i=1}^n Y_i = \sum_{i=1}^n \frac{X_i - a}{b-a} = \frac{(\sum_{i=1}^n X_i) - na}{b-a} = \frac{K_n - na}{b-a}\,. \tag{3.125}$$

Writing $\tilde{\mu}_1 = \frac{\mu_1 - a}{b-a}$, the density function $f_{K_n|\{(a,b,\mu_1)\}}$ of K_n again results from (3.122) using proposition 3.4

$$\begin{aligned} f_{K_n|\{(a,b,\mu_1)\}}(x) &= \frac{1}{b-a} f_{P_n|\{\tilde{\mu}_1\}}\left(\frac{x-na}{b-a}\right) \\ &= \frac{1}{b-a}\left(\frac{\lambda(\tilde{\mu}_1)}{e^{\lambda(\tilde{\mu}_1)}-1}\right)^n e^{\lambda(\tilde{\mu}_1)\frac{x-na}{b-a}} \frac{1}{(n-1)!} \cdot \\ &\quad \cdot \sum_{i=0}^{\lfloor \frac{x-na}{b-a} \rfloor} (-1)^i \binom{n}{i} \left(\tfrac{x-na}{b-a} - i\right)^{n-1} \mathbb{1}_{[0,n]}\left(\tfrac{x-na}{b-a}\right) \qquad (3.126) \\ &= \left(\frac{(b-a)\lambda(a,b,\mu_1)}{e^{(b-a)\lambda(a,b,\mu_1)}-1}\right)^n e^{\lambda(a,b,\mu_1)(x-na)} \frac{1}{(n-1)!(b-a)^n} \cdot \\ &\quad \cdot \sum_{i=0}^{\lfloor \frac{x-na}{b-a} \rfloor} (-1)^i \binom{n}{i} (x - na - i(b-a))^{n-1} \mathbb{1}_{[na,nb]}(x) \\ &= \left(\frac{\lambda(a,b,\mu_1)}{e^{\lambda(a,b,\mu_1)b} - e^{\lambda(a,b,\mu_1)a}}\right)^n e^{\lambda(a,b,\mu_1)x} \frac{1}{(n-1)!} \cdot \\ &\quad \cdot \sum_{i=0}^{\lfloor \frac{x-na}{b-a} \rfloor} (-1)^i \binom{n}{i} (x - na - i(b-a))^{n-1} \mathbb{1}_{[na,nb]}(x) \qquad (3.127) \end{aligned}$$

and the distribution function $f_{K_n|\{(a,b,\mu_1)\}}$ of K_n for $x \in [na, nb]$ is

$$\begin{aligned} F_{K_n|\{(a,b,\mu_1)\}}(x) &= \left(\frac{1}{e^{(b-a)\lambda(a,b,\mu_1)}-1}\right)^n \left\{ e^{\lambda(a,b,\mu_1)(x-na)} \cdot \right. \\ &\cdot \left[\sum_{i=0}^{\lfloor \frac{x-na}{b-a} \rfloor} (-1)^i \binom{n}{i} \left(\sum_{k=0}^{n-2} (-1)^k (x - na - i(b-a))^{n-1-k} \frac{(\lambda(a,b,\mu_1))^{n-1-k}}{(n-1-k)!} \right) \right] + \\ &\left. + \left[\sum_{i=0}^{\lfloor \frac{x-na}{b-a} \rfloor} (-1)^{n+1+i} \binom{n}{i} \left(e^{\lambda(a,b,\mu_1)(x-na)} - e^{(b-a)\lambda(a,b,\mu_1)i}\right) \right] \right\} \qquad (3.128) \end{aligned}$$

No simplification arises for the respective terms of the density and distribution function of the mean $\overline{X} = \frac{1}{n} K_n$, thus we only give the generally known formulas

$$\begin{aligned} F_{\overline{X}|\{(a,b,\mu_1)\}}(x) &= F_{K_n|\{(a,b,\mu_1)\}}(nx)\,, \qquad (3.129) \\ f_{\overline{X}|\{(a,b,\mu_1)\}}(x) &= n f_{K_n|\{(a,b,\mu_1)\}}(nx)\,. \qquad (3.130) \end{aligned}$$

3.3.3. Order Statistics

Arranging the random variables $X_1, \ldots, X_n$ in ascending order yields the *order statistics* $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, i.e., $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. Therefore, for example, $X_{(1)} = \min\{X_1, \ldots, X_n\}$ is the sample minimum and $X_{(n)} = \max\{X_1, \ldots, X_n\}$ is the sample maximum, respectively. Since the general formulas of the (joint) distributions of order statistics are known, e.g. see [39], some of them are explicitly derived here.

3.3.3.1. Distributions concerning $X_{(1)}$ and $X_{(n)}$

In general, the distribution function $F_{X_{(n)}}$ of $X_{(n)}$ given an i.i.d. sample of a random variable X is

$$F_{X_{(n)}}(x) = (F_X(x))^n \tag{3.131}$$

with distribution function F_X of X. The density function $f_{X_{(n)}}$ of $X_{(n)}$ is

$$f_{X_{(n)}}(x) = n(F_X(x))^{n-1} f_X(x) \tag{3.132}$$

with density function f_X of X.
Consequently, for $X \sim Mon(a, b, \mu_1)$ we get for $x \in \mathcal{X}$

$$F_{X_{(n)}|\{(a,b,\mu_1)\}}(x) = \left(\frac{e^{\lambda x} - e^{\lambda a}}{e^{\lambda b} - e^{\lambda a}}\right)^n , \tag{3.133}$$

and

$$f_{X_{(n)}|\{(a,b,\mu_1)\}}(x) = n\left(\frac{e^{\lambda x} - e^{\lambda a}}{e^{\lambda b} - e^{\lambda a}}\right)^{n-1} \frac{\lambda e^{\lambda x}}{e^{\lambda b} - e^{\lambda a}} . \tag{3.134}$$

Similar, the distribution function $F_{X_{(1)}}$ of $X_{(1)}$ is

$$F_{X_{(1)}}(x) = 1 - (1 - F_X(x))^n \tag{3.135}$$

and the density function $f_{X_{(1)}}$ of $X_{(1)}$ is

$$f_{X_{(1)}}(x) = n(1 - F_X(x))^{n-1} f_X(x) . \tag{3.136}$$

This yields for $X \sim Mon(a, b, \mu_1)$ for $x \in \mathcal{X}$

$$F_{X_{(1)}|\{(a,b,\mu_1)\}}(x) = 1 - \left(\frac{e^{\lambda b} - e^{\lambda x}}{e^{\lambda b} - e^{\lambda a}}\right)^n , \tag{3.137}$$

and

$$f_{X_{(1)}|\{(a,b,\mu_1)\}}(x) = n\left(\frac{e^{\lambda b} - e^{\lambda x}}{e^{\lambda b} - e^{\lambda a}}\right)^{n-1} \frac{\lambda e^{\lambda x}}{e^{\lambda b} - e^{\lambda a}}\,. \tag{3.138}$$

The joint distribution function of $X_{(1)}$ and $X_{(n)}$ for $x_1 < x_n$ is

$$F_{X_{(1)},X_{(n)}}(x_1, x_n) = (F_X(x_n))^n - (F_X(x_n) - F_X(x_1))^n \tag{3.139}$$

and the joint density function for $x_1 < x_n$ is

$$f_{X_{(1)},X_{(n)}}(x_1, x_n) = n(n-1) f_X(x_1) f_X(x_n)(F_X(x_n) - F_X(x_1))^{n-2}\,. \tag{3.140}$$

For $X \sim Mon(a, b, \mu_1)$ and $x_1, x_n \in \mathcal{X}$, $x_1 < x_n$, we get

$$F_{(X_{(1)},X_{(n)})|\{(a,b,\mu_1)\}}(x_1, x_n) = \frac{(e^{\lambda x_n} - e^{\lambda a})^n - (e^{\lambda x_n} - e^{\lambda x_1})^n}{(e^{\lambda b} - e^{\lambda a})^n} \tag{3.141}$$

and

$$f_{(X_{(1)},X_{(n)})|\{(a,b,\mu_1)\}}(x_1, x_n) = n(n-1)\frac{\lambda^2 e^{\lambda(x_1+x_n)}(e^{\lambda x_n} - e^{\lambda x_1})^{n-2}}{(e^{\lambda b} - e^{\lambda a})^n}\,. \tag{3.142}$$

For the random variable $R = X_{(n)} - X_{(1)}$, i.e., the range of the sample, the densitiy function f_R is

$$f_R(r) = \int_{-\infty}^{+\infty} n(n-1) f_X(t+r) f_X(t)(F_X(t+r) - F_X(t))^{n-2}\,\mathrm{d}t \tag{3.143}$$

and the distribution function F_R is

$$F_R(r) = \int_{-\infty}^{+\infty} n f_X(t)(F_X(t+r) - F_X(t))^{n-1}\,\mathrm{d}t\,. \tag{3.144}$$

The density function $f_{R|\{(a,b,\mu_1)\}}$ of the range R of the sample is obtained as

$$f_{R|\{(a,b,\mu_1)\}}(r) = (n-1)\frac{\lambda e^{\lambda r}(e^{\lambda r} - 1)^{n-2}(e^{n\lambda(b-r)} - e^{n\lambda a})}{(e^{\lambda b} - e^{\lambda a})^n}\mathbb{1}_{[0,b-a]}(r)\,. \tag{3.145}$$

Consequently, the distribution function $F_{R|\{(a,b,\mu_1)\}}$ for $r \in [0; b-a]$ is given by

$$F_{R|\{(a,b,\mu_1)\}}(r) = \frac{(e^{\lambda r} - 1)^{n-1}(e^{\lambda r} e^{n\lambda(b-r)} - e^{n\lambda a})}{(e^{\lambda b} - e^{\lambda a})^n}\,. \tag{3.146}$$

3.3.3.2. Distributions concerning $X_{(n)} = \max X_i$ and $K_n = \sum X_i$

Here we assume $X_i \sim Mon(b, \mu_1)$ and first derive the joint distribution of

$$\begin{aligned}(X_{(n)}, K_n)|\{(b,\mu_1)\} &= (X_{(n)}|\{(b,\mu_1)\}, K_n|\{(b,\mu_1)\}) \\ &= (\max_{1\le i\le n} X_i|\{(b,\mu_1)\}, \sum_{i=0}^{n} X_i|\{(b,\mu_1)\})\,. \end{aligned} \tag{3.147}$$

In [62] we find a derivation of the joint density for the case of n i.i.d. exponential random variables. There, the following notations are used: $E_1, \ldots, E_n$ i.i.d. with $E_i \sim EXP(\beta)$, $X = \sum E_i$, $Y = \max E_i$. The authors call the joint distribution of X and Y *bivariate distribution with gamma and generalized exponential marginals (BGGE distribution)*. This leads to the following proposition 3.20.

> **Proposition 3.20** ***(joint distribution of sum and maximum of i.i.d. exponential random variables)*** *For $n \ge 2$, let f_n be the PDF of $(X, Y) \sim BGGE(\beta, n)$. Then for $k = 1, 2, \ldots, n-1$ we have*
>
> $$f_n(x,y) = \beta^n e^{-\beta x} \sum_{s=1}^{k} \frac{n(n-1)}{(s-1)!(n-s)!}(x-sy)^{n-2}(-1)^{s+1}\,,\ (x,y) \in S_k\,,$$
>
> *where S_k is the kth sector defined in (2.20)* [which are similar to those in figure 3.5].

Incorporating the boundedness of the range of variability $\mathcal{X}$ of $X|\{(b,\mu_1)\} \sim Mon(b,\mu_1)$, and adapting proposition 3.20 to our notations leads to the desired joint density of $(X_{(n)}, K_n)|\{(b,\mu_1)\}$.

Corollary 3.21 (joint distribution of of sample maximum and sample sum) *Let $X_1|\{(b,\mu_1)\}, X_2|\{(b,\mu_1)\}, \ldots, X_n|\{(b,\mu_1)\}$ be i.i.d. with $X_i|\{(b,\mu_1)\} \sim Mon(b,\mu_1)$, $b > 0$, $0 < \mu_1 < b$. Then for $n \ge 2$ the joint density $f_{(X_{(n)},K_n)|\{(b,\mu_1)\}}(t,s)$ of the maximum $X_{(n)}|\{(b,\mu_1)\} := \max_{1\le i\le n} X_i|\{(b,\mu_1)\}$ and the sum $K_n|\{(b,\mu_1)\} :=$*

$\sum_{i=1}^{n} X_i|\{(b,\mu_1)\}$ *is given by*

$$f_{(X_{(n)},K_n)|\{(b,\mu_1)\}}(t,s) =$$

$$\begin{cases} \left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)b}-1}\right)^n e^{\lambda(b,\mu_1)s} \sum_{m=1}^{k} \frac{n(n-1)}{(m-1)!(n-m)!}(s-mt)^{n-2}(-1)^{m+1} \\ \qquad\qquad \frac{1}{1+k}s \le t \le \frac{1}{k}s, \\ \qquad\qquad \text{for } 0 \le t \le b, \qquad k = 1,\dots,n-1, \\ \qquad\qquad 0 \le s \le nb, \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{else.} \end{cases} \tag{3.148}$$

Figure 3.5 displays the regions $\frac{1}{1+k}s \le t \le \frac{1}{k}s$ which define the value of k in (3.148).

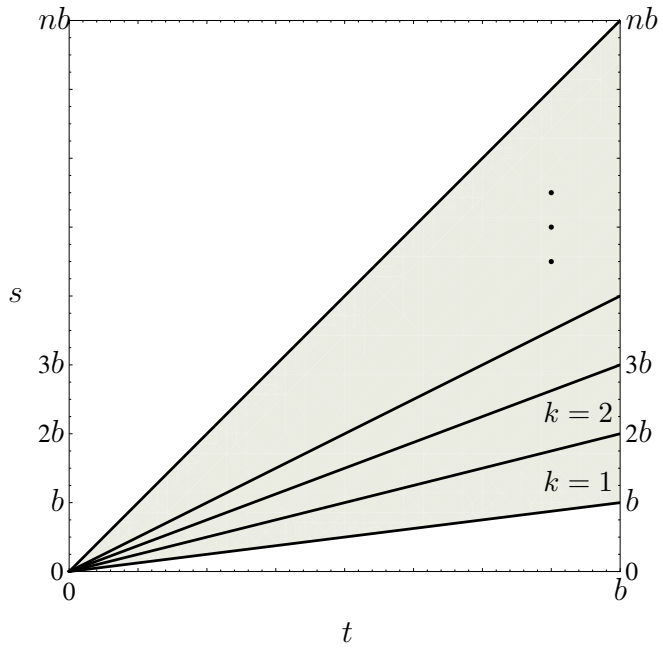


Figure 3.5.: The different regions $\frac{1}{1+k}s \le t \le \frac{1}{k}s$ defining k in the definition (3.148) of the joint density function $f_{(X_{(n)},K_n)|\{(b,\mu_1)\}}(t,s)$. Note the different scaling of the axes.

The marginal densities, i.e., $f_{X_{(n)}|\{(b,\mu_1)\}}$ and $f_{K_n|\{(b,\mu_1)\}}$, may be obtained by integration over the respective other variable – of course, we would get the already derived expressions (3.134) for $f_{X_{(n)}|\{(b,\mu_1)\}}$ and (3.127) for $f_{K_n|\{(b,\mu_1)\}}$, respectively:

$$f_{X_{(n)}|\{(b,\mu_1)\}}(t) = n\frac{\lambda(b,\mu_1)e^{\lambda(b,\mu_1)t}}{e^{\lambda(b,\mu_1)b}-1}\left(\frac{e^{\lambda(b,\mu_1)t}-1}{e^{\lambda(b,\mu_1)b}-1}\right)^n \tag{3.149}$$

$$\begin{aligned} f_{K_n|\{(b,\mu_1)\}}(s) &= \left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)b}-1}\right)^n e^{\lambda(b,\mu_1)s}\frac{1}{(n-1)!}\cdot \\ &\quad \cdot\sum_{i=0}^{\lfloor\frac{s}{b}\rfloor}(-1)^i\binom{n}{i}(s-ib)^{n-1}\mathbb{1}_{[0,nb]}(s) \end{aligned} \tag{3.150}$$

In figures 3.6 and 3.7 we display the joint density $f_{(X_{(n)},K_n)|\{(b,\mu_1)\}}$ together with the marginal densities $f_{K_n|\{(b,\mu_1)\}}$ and $f_{X_{(n)}|\{(b,\mu_1)\}}$ for $b=2$, $n=8$ and $\mu_1=\frac{9}{10}$, $\frac{4}{10}$.

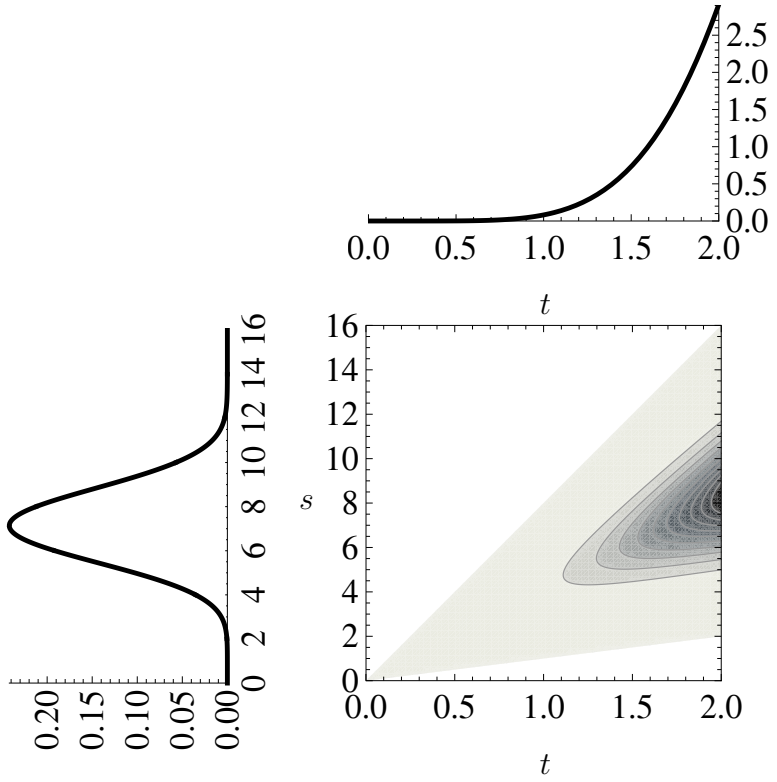


Figure 3.6.: The joint density $f_{(X_{(8)},K_8)|\{(2,\frac{9}{10})\}}$ displayed at the lower right in form of a contour plot (darker regions indicate higher values of the density). The marginal densities are to the left ($f_{K_8|\{(2,\frac{9}{10})\}}$) and on top ($f_{X_{(8)}|\{(2,\frac{9}{10})\}}$).

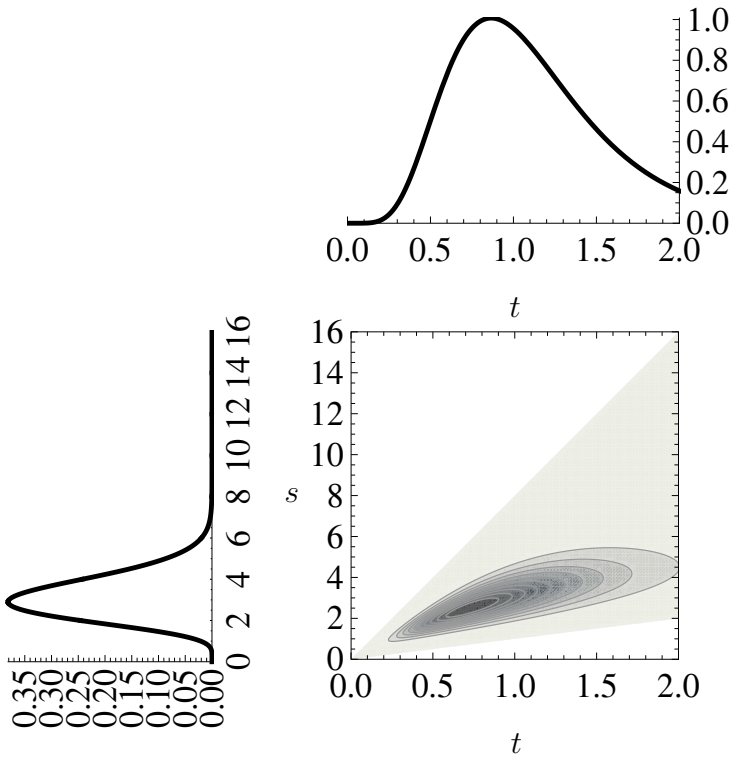


Figure 3.7.: The joint density $f_{(X_{(8)},K_8)|\{(2,\frac{2}{10})\}}$ displayed at the lower right in form of a contour plot (darker regions indicate higher values of the density). The marginal densities are to the left ($f_{K_8|\{(2,\frac{2}{10})\}}$) and on top ($f_{X_{(8)}|\{(2,\frac{2}{10})\}}$).

Next we will derive the distribution of the ratio of $X_{(n)}$ to K_n, the so-called *peak to sum ratio*. Defining the transformed random variable $(U, V) := (X_{(n)}/K_n, K_n)$, i.e., the mapping $g(t, s) = (\frac{t}{s}, s)$, the joint density of (U, V) is known to be

$$f_{(U,V)|\{(b,\mu_1)\}}(u, v) \;=\; f_{(X_{(n)},K_n)|\{(b,\mu_1)\}}(v, \frac{v}{u})\frac{|v|}{u^2}\,. \tag{3.151}$$

That is, for $0 \le v \le b$, $\frac{1}{k+1} \le u \le \frac{1}{k}$, $k = 1, 2, \ldots, n-1$ we get with $c_n(b, \mu_1) := \left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)b}-1}\right)^n$

$$f_{(U,V)|\{(b,\mu_1)\}}(u, v) = \tag{3.152}$$

$$= c_n(b, \mu_1)e^{\lambda(b,\mu_1)v/u}\left(\sum_{m=1}^{k}\frac{n(n-1)}{(m-1)!(n-m)!}(\frac{v}{u} - mv)^{n-2}(-1)^{m+1}\right)\frac{v}{u^2} \tag{3.153}$$

$$= c_n(b, \mu_1)e^{\lambda(b,\mu_1)v/u}\frac{v^{n-1}}{u^n}\left(\sum_{m=1}^{k}\frac{n(n-1)}{(m-1)!(n-m)!}(1 - mu)^{n-2}(-1)^{m+1}\right) \tag{3.154}$$

Integrating with respect to v with the substitution $z = -\frac{\lambda(b,\mu_1)v}{u}$ yields the density $f_{U|\{(b,\mu_1)\}}(u)$ for $\frac{1}{k+1} \le u \le \frac{1}{k}$, $k = 1, 2, \ldots, n-1$:

$$f_{U|\{(b,\mu_1)\}}(u) = \int_0^b f_{(U,V)|\{(b,\mu_1)\}}(u, v)\,\mathrm{d}v \tag{3.155}$$

$$= c_n(b, \mu_1)\left(\sum_{m=1}^{k}\frac{n(n-1)}{(m-1)!(n-m)!}(1 - mu)^{n-2}(-1)^{m+1}\right)\frac{1}{u^n}\int_0^b v^{n-1}e^{\lambda(b,\mu_1)v/u}\,\mathrm{d}v \tag{3.156}$$

$$= c_n(b, \mu_1)\left(\sum_{m=1}^{k}\frac{n(n-1)}{(m-1)!(n-m)!}(1 - mu)^{n-2}(-1)^{m+1}\right)\cdot$$

$$\cdot\,\frac{1}{u^n}\left(-\frac{u}{\lambda(b, \mu_1)}\right)^n\int_0^{-\frac{\lambda(b,\mu_1)b}{u}} z^{n-1}e^{-z}\,\mathrm{d}z \tag{3.157}$$

$$= c_n(b, \mu_1)\left(\sum_{m=1}^{k}\frac{n(n-1)}{(m-1)!(n-m)!}(1 - mu)^{n-2}(-1)^{m+1}\right)\frac{\gamma(n, -\frac{\lambda(b,\mu_1)b}{u})}{(-\lambda(b, \mu_1))^n} \tag{3.158}$$

where $\gamma(n, -\frac{\lambda(b,\mu_1)b}{u})$ is the lower incomplete gamma function. Replacing $c_n(b, \mu_1)$ again and noting, that $\lambda(b, \mu_1)b = \lambda(\frac{\mu_1}{b})$, finally yields $\frac{1}{k+1} \leq u \leq \frac{1}{k}$, $k = 1, 2, \ldots, n-1$:

$$f_{U|\{(b,\mu_1)\}}(u) = \frac{\gamma(n, -\frac{\lambda(\frac{\mu_1}{b})}{u})}{(1 - e^{\lambda(\frac{\mu_1}{b})})^n} \left(\sum_{m=1}^{k} \frac{n(n-1)}{(m-1)!(n-m)!} (1 - mu)^{n-2} (-1)^{m+1} \right), \tag{3.159}$$

and $f_{U|\{(b,\mu_1)\}}(u) = 0$ for $u < \frac{1}{n}$ and $u > 1$, respectively. Figure 3.8 shows $f_{U|\{(b,\mu_1)\}}(u)$ for $b = 2$, $n = 8$ and $\mu_1 = 0.9, 0.7, 0.55, 0.4, 0.2$.

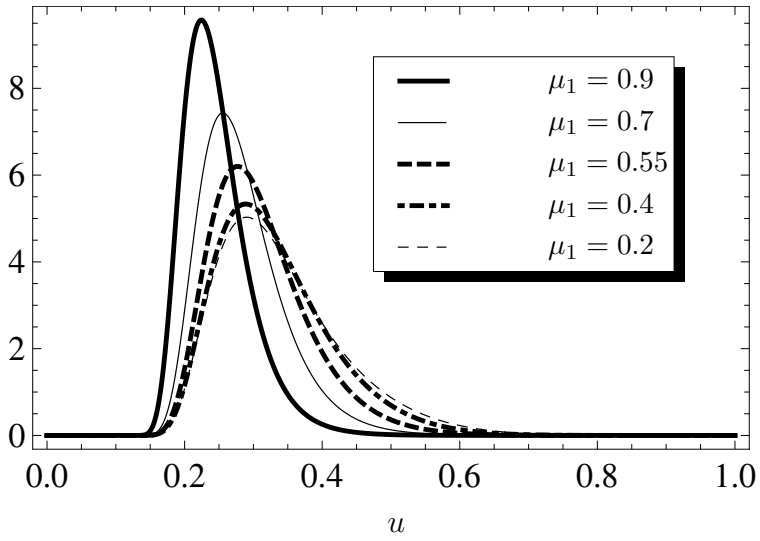


Figure 3.8.: Density $f_{U|\{(b,\mu_1)\}}$ of $U = X_{(n)}/K_n$ for $b = 2$, $n = 8$ and $\mu_1 = 0.9, 0.7, 0.55, 0.4, 0.2$.

Conclusion

In this chapter we gave an extensive characterisation of the monotonic probability distribution $Mon(a, b, \mu_1)$ and derived several of its properties. Some of them will help us in the following chapters to derive estimation procedures for one of its parameters. Further analysis with respect to the dependency of μ_1, λ and b from each other is presented in Appendix A.

4. General Concepts of Inference

4.1. Metrology – The Science of Measurement

Statistical *estimation* deals with the determination of unknown, but fixed values of deterministic variables involved in a random phenomenon. Therefore, the methods applied in estimation may be understood as measurement procedures and, consequently, should meet the same requirements as technical measurement devices. The science of measurement is called *metrology*. Short introductions to metrology were given for example by the *Bureau International des Poids et Mesures BIPM (International Bureau of Weights and Measures)* [8] or by the *European Association of National Metrology Institutes EURAMET* [26]. The last one starts with a historical note:

> "The death penalty faced those who forgot or neglected their duty to calibrate the standard unit of length at each full moon. Such was the peril courted by the royal site architects responsible for building the temples and pyramids of the Pharaohs in ancient Egypt, 3000 years BC. The first royal cubit was defined as the length of the forearm from elbow to tip of the extended middle finger of the ruling Pharaoh, plus the width of his hand. The original measurement was transferred to and carved in black granite. The workers at the building sites were given copies in granite or wood and it was the responsibility of the architects to maintain them." ([26], p. 8)

Even though the example stems from long ago, the main tasks in nowadays' metrology remained almost unchanged (cf. [26], p. 9): defining standard units (*royal cubit*) which are internationally accepted (by the pharaoh's authority), realising these standard units by scientific methods (measuring the length of the pharaoh's forearm and the width of his hand), and finally, establishing "of traceability chains by determining and documenting the value and accuracy of a measurement" (transferring to and carving in black granite and making copies in

granite or wood). Today, the international standard unit of length is the *metre* (settled by the *Convention du Métre* on 20 May 1875), whose unit of measurement is realised by the use of lasers (since 1983 in fact by the length of the path traveled by light in vacuum during the time interval of 1/299 792 458 of a second). The traceability chain is realised by documented relationships between manufactures of products of which one characteristic is length and some accredited laboratories for optical length metrology.

One branch of metrology is *industrial metrology*, which "has to ensure the adequate functioning of measurement instruments used in industry, in production and testing processes, for ensuring quality of life for citizens and for academic research" ([26], p. 10). The functioning of measurement instruments (or in general any measurement device) is ensured by calibration. In [26], p. 17, we find four main aims of calibration:

1. To establish and demonstrate traceability. [By providing a calibration certificate.]
2. To ensure readings from the instrument are consistent with other measurements.
3. To determine the accuracy of the instrument readings.
4. To establish the reliability of the instrument i.e. that it can be trusted.

The above list of aims contains two terms, which have great importance for the characterisation of measurement procedures: *accuracy* and *reliability*. Accuracy means that the measurement result, which in metrology is always given by one single value, is in a certain sense close to the unknown value of the quantity to be measured (measurand). Reliability refers to repeated measurements and means repeatability or consistency of measurements, i.e., a measurement procedure is called reliable, if it yields when repeated similar results. Closely related to accuracy and reliability are two other terms: *precision* and *uncertainty*. A measurement procedure is called precise, if repeated measurements of the same quantity under the same conditions yield results which are close to each other. Uncertainty is widely understood as a quality criterion of a measurement result, in the sense that it enables the comparison of different results (cf. [26], p. 21). Uncertainty is a "non-negative parameter" (cf [12], 2.26), which together with

the measurement result determines a set of values that should contain the true but unknown one. In connection with accuracy and reliability, uncertainty may be regarded as a combination of both: If the accuracy and/or reliability are low, measurement uncertainty is large. Conversely, if both accuracy and reliability are high, measurement uncertainty is small. Therefore, stating a measurement result by a single value alone is useless, but it should always be given together with its uncertainty. The *Guide to the expression of uncertainty in measurement (GUM)* [9] states, that "In general, the result of a measurement (...) is complete only when accompanied by a statement of the uncertainty of that estimate [of the value of the measurand]." ([9], 3.1.2) while the GUM defines "uncertainty (of measurement) [as a] parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand" ([9], 2.2.3). As the *GUM* may be considered as the *de facto* standard for expressing uncertainty in all fields of measurement we want to review shortly its underlying concepts, definitions and recommended guidelines.

4.1.1. Review of the *Guide to the expression of uncertainty in measurement (GUM)*

The GUM [9] was published for the first time 1995 by the member organisations of the Joint Committee for Guides in Metrology (JCGM)[1]. Even though the most recent edition of the GUM was published 2008, its subtitle 'GUM 1995 with minor corrections' indicates that it is almost identical to the first edition. A good overview of the GUM's evolution and related documents like the supplements 1 and 2 [10, 11] gives [4]. With respect to the "fundamental concepts and principles" ([4], p. S162) it summarises ([4], p. S162-S163):

- Measurement model relating functionally one or more output

[1] JCGM member organisations are: International Bureau of Weights and Measures (BIPM), International Electrotechnical Commission (IEC), International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), International Laboratory Accreditation Cooperation (ILAC), International Organization for Standardization (ISO), International Union of Pure and Applied Chemistry (IUPAC), International Union of Pure and Applied Physics (IUPAP), International Organization of Legal Metrology (OIML).

quantities, about which information is required, to input quantities, about which information is available.

- Modelling of measurement knowledge about a quantity in terms of a probability distribution.
- Expectation (estimate) and standard deviation (standard uncertainty) of a quantity characterized by a probability distribution.
- Use of new information to update an input probability density function: Bayes' theorem.
- Assignment of a probability density function to a quantity using the Principle of Maximum (Information) Entropy.
- Determination of the distribution for an output quantity (or the joint distribution for more than one output quantity) using the propagation of distributions.

The measurement equation

An output quantity Y is a measurand which is subject to measurement (since the GUM "treats the measurand as a scalar" ([9], 3.1.7), we adopt this for simplification of notation). The "measurand Y is (...) determined from N other quantities $X_1, X_2, \ldots, X_N$ through a functional relationship f: $Y = f(X_1, X_2, \ldots, X_N)$" ([9], 4.1.1). The function f is often denoted as the *measurement function* (e.g. in [41]). The input quantities themselves may be measurands and, therefore, depend on further quantities, thus, the measurement function may be at least very complicated or even cannot be stated explicitly ([9], 4.1.2). A result of the measurement, i.e., an estimate y of the measurand Y, is obtained by substitution of the input quantities $X_1, X_, \ldots, X_N$ with their input estimates $x_1, x_2, \ldots, x_N$ in f: $y = f(x_1, x_2, \ldots, x_N)$ ([9], 4.1.4).

Example 4.1 *This example stems from Annex H of the GUM and treats the calibration of an end gauge ([9], H.1). The task is to determine the "length of a nominally 50 mm end gauge [...] by comparing it with a known standard of the same nominal length." The set-up of the measurement function yields a linear*

approximation ([9], (H.3)):

$$l = f(l_S, d, \alpha_S, \theta, \delta\alpha, \delta\theta) = l_S + d - l_S(\delta\alpha \cdot \theta + \alpha_S \cdot \delta\theta) \tag{4.1}$$

where (compare [9], H.1.1 and H.1.2)

l	*is the measurand, that is, the length at 20 °C of the end gauge being calibrated;*
l_S	*is the length of the standard at 20 °C as given in its calibration certificate;*
d	*is the difference in their lengths;*
α, α_S *and* $\delta\alpha$	*are the coefficients of thermal expansion, respectively, of the gauge being calibrated and the standard, and* $\delta\alpha = \alpha - \alpha_S$*;*
θ, θ_S *and* $\delta\theta$	*are the deviations in temperature from the 20 °C reference temperature, respectively, of the gauge and the standard, and* $\delta\theta = \theta - \theta_S$*.*

"The differences $\delta\theta$ and $\delta\alpha$ [...] are estimated to be zero; [...] It thus follows [...] that the estimate of the value of the measurand l may be obtained from the expression $l_S + \overline{d}$, where l_S is the length of the standard at 20 °C as given in the calibration certificate and d is estimated by $\overline{d}$, the arithmetic mean of [...] independent repeated observations." ([9], H.1.2)

Law of propagation of uncertainty

The uncertainty of the result of the measurement then is determined by the uncertainties of the input estimates. The GUM defines *standard uncertainty* with the "uncertainty of the result of a measurement expressed as a standard deviation", noted as $u(y)$. Since each input estimate x_i carries some uncertainty, expressed as a standard uncertainty $u(x_i)$, which contributes to the uncertainty of the result of the measurement y, the standard uncertainty of y is calculated as a *combined standard uncertainty*, denoted by $u_c(y)$, "equal to the positive square root of a sum of terms, the terms being the variances or covariances of these other quantities weighted according to how the measurement result varies with changes in these quantities" ([9], 2.3.4). This is a verbal translation of "a first-order Taylor

series approximation of $Y = f(X_1, X_2, \ldots, X_N)$" given by

$$u_c(y) = \left(\sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} u(x_i, x_j) \right)^{1/2} , \tag{4.2}$$

where "$u(x_i, x_j)$ is the estimated covariance associated with x_i and x_j" ([9], 5.1.2 and 5.2.2) and "the partial derivatives $\partial f / \partial x_i = \partial f / \partial X_i$ evaluated at $X_i = x_i$" ([9], 5.1.3). If the input quantities are all uncorrelated, the second sum in (4.2) equals 0. Through this linear approximation (4.2) the uncertainities $u(x_i)$ of the input quantities X_i are propagated and determine the uncertainty of the result of the measurement. The *GUM* terms this as the *law of propagation of uncertainty* ([9], 5.1.2).

Example 4.2 *(Continuation of example 4.1) The involved quantities in the measurement equation* (4.1) *are assumed to be uncorrelated ([9], H.1.2), thus, applying equation* (4.2) *to the measurement function* (4.1) *yields ([9], equation (H.4)):*

$$u_c^2(l) = c_S^2 u^2(l_S) + c_d^2 u^2(d) + c_{\alpha_S}^2 u^2(\alpha_S) + c_\theta^2 u^2(\theta) + c_{\delta\alpha}^2 u^2(\delta\alpha) + c_{\delta\theta}^2 u^2(\delta\theta) \tag{4.3}$$

with coefficients c_i equal the partial derivatives of the measurement function (and taking into account that $\delta\theta$ and $\delta\alpha$ are estimated to be zero) (compare [9], H.1.3):

$$\begin{aligned}
c_S &= \frac{\partial f}{\partial l_S} = 1 - (\delta\alpha \cdot \theta + \alpha_S \cdot \delta\theta) = 1 \\
c_d &= \frac{\partial f}{\partial d} = 1 \\
c_{\alpha_S} &= \frac{\partial f}{\partial \alpha_S} = -l_S \delta\theta = 0 \\
c_\theta &= \frac{\partial f}{\partial \theta} = -l_S \delta\alpha = 0 \\
c_{\delta\alpha} &= \frac{\partial f}{\partial \delta\alpha} = -l_S \theta \\
c_{\delta\theta} &= \frac{\partial f}{\partial \delta\theta} = -l_S \alpha_S
\end{aligned}$$

It thus follows the combined standard uncertainty

$$u_c^2(l) = u^2(l_S) + u^2(d) + l_S^2 \theta^2 u^2(\delta\alpha) + l_S^2 \alpha_S^2 u^2(\delta\theta) \,. \tag{4.4}$$

The *expanded uncertainty* U is the "quantity defining an interval about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand" ([9], 2.3.1 and 2.3.5). This expanded uncertainty is in general a certain multiple k of the combined standard uncertainty, i.e., $U = ku_c(y)$, where k is called *coverage factor*. Together with "the best estimate of the value attributable to the measurand Y is y, and that $y - U$ to $y + U$ is an interval (...) that encompasses a large fraction p of the probability distribution characterized by the result and its combined standard uncertainty, and p is the *coverage probability* or *level of confidence*[2] of the interval" ([9], 6.2.1 and 6.2.2).

Determining the appropriate coverage factor for some desired coverage probability is a difficult task. Section 6 and Annex G of the *GUM* present the interpretation of the expanded uncertainty, and give an answer to the question of how the coverage factor k should be determined (we come to this point later). These recommendations are closely related to the *GUM*'s concept and evaluation of uncertainty.

Types of evaluation of uncertainty

The *GUM* indicates two types of evaluation of uncertainty. If uncertainties of input quantities are calculated according to a "method of evaluation (...) by the statistical analysis of series of observations", this evaluation is called *Type A* evaluation, whereas a *Type B* evaluation of uncertainties is a "method of evaluation (...) by means other than the statistical analysis of series of observations" ([9], 2.3.2 and 2.3.3). More specific is 4.1.6 of [9]:

> Each input estimate x_i and its associated standard uncertainty $u(x_i)$ are obtained from a distribution of possible values of the input quantity X_i. This probability distribution may be frequency based, that is, based on a series of observations $X_{i,k}$ of X_i, or it may be an *a priori* distribution. Type A evaluations of standard uncertainty components are founded on frequency distributions while Type B evaluations are

[2] Section 6.2.2 of [9] stresses that *level of confidence* must not be mistaken as *confidence level*, and, consequently, the interval defined by y and U not as a confidence interval.

> founded on *a priori* distributions. It must be recognized that in both cases the distributions are models that are used to represent the state of our knowledge.

Section 4.3 of the *GUM* gives some description of Type B evaluations and in which situations they are applied ([9], 4.3.1):

> For an estimate x_i of an input quantity X_i that has not been obtained from repeated observations, the associated estimated variance $u^2(x_i)$ or the standard uncertainty $u(x_i)$ [i.e., the standard deviation] is evaluated by scientific judgement based on all of the available information on the possible variability of X_i. The pool of information may include
>
> - previous measurement data;
> - experience with or general knowledge of the behaviour and properties of relevant materials and instruments;
> - manufacturer‘s specifications;
> - data provided in calibration and other certificates;
> - uncertainties assigned to reference data taken from handbooks.

The subsequent paragraphs to section 4.3 of [9] discuss cases of available information and the resulting evaluation of standard uncertainties. The stated cases and examples describe the translation of the available information about the input quantity X_i into some probability distribution of X_i. For example, in 4.3.5 we find:

> Consider the case where, based on the available information, one can state that “there is a fifty-fifty chance that the value of the input quantity X_i lies in the interval a_- to a_+” (in other words, the probability that X_i lies within this interval is 0,5 or 50 percent). If it can be assumed that the distribution of possible values of X_i is approximately normal, then the best estimate x_i of X_i can be taken to be the midpoint of the interval. Further, if the half-width of the interval is denoted by $a = (a_+ - a_-)/2$, one can take $u(x_i) = 1,48a$, because for a normal distribution with expectation μ and standard deviation

> σ the interval $\mu \pm \sigma/1,48$ encompasses approximately 50 percent of the distribution.

In addition to section 4 annex F of [9] provides a "practical guidance on evaluating uncertainty components".
Section F.1 treats Type A evaluations: In section F.1.1 with respect to the two questions "To what extent are the repeated observations completely independent repetitions of the measurement procedure?" (F.1.1.2) and "it must be asked whether all of the influences that are assumed to be random really are random" (F.1.1.3), and section F.1.2 discusses cases where the correlation of two input quantities has to be taken into account.
Type B evaluations are treated in section F.2 which starts with F.2.1 "The need for Type B evaluations":

> If a measurement laboratory had limitless time and resources, it could conduct an exhaustive statistical investigation of every conceivable cause of uncertainty, for example, by using many different makes and kinds of instruments, different methods of measurement, different applications of the method, and different approximations in its theoretical models of the measurement. The uncertainties associated with all of these causes could then be evaluated by the statistical analysis of series of observations and the uncertainty of each cause would be characterized by a statistically evaluated standard deviation. In other words, all of the uncertainty components would be obtained from Type A evaluations. Since such an investigation is not an economic practicality, many uncertainty components must be evaluated by whatever other means is practical.

The subsequent subsections to F.2.1 discuss several cases where a Type B evaluation of uncertainties becomes necessary.

Example 4.3 *(Continuation of example 4.2) Representative for the different components of the combined standard uncertainty* $u_c^2(l)$ (4.4) *of the length of the end gauge, we illustrate the determination of the "uncertainty of the measured difference in lengths,* $u(d)$*" ([9], H.1.3.2), where both Type A and Type B evaluations are utilised.*

Several sources of uncertainty contribute to the uncertainty $u(d)$: the uncertainty of the arithmetic mean of the readings of difference of lengths $u(\overline{d})$ and the uncertainty of the length carried by the comparator (i.e., the test bed where the standard end gauge and the to be calibrated end gauge are compared), which itself consists of two components, that are the uncertainty "due to random errors" yielding the standard uncertainty $u(d_1)$, and the uncertainty "due to systematic errors" yielding the standard uncertainty $u(d_2)$. Therefore,

$$u^2(d) = u^2(\overline{d}) + u^2(d_1) + u^2(d_2).$$

The uncertainty component $u(\overline{d})$ is determined by repeated measurements of the difference d of lengths l and l_S. Since a pooled experimental standard deviation (see [9], 4.2.4) s_p is available which is 13 nm = 13 $\times$ 10^{-9} m, the standard uncertainty of the arithmetic mean of the readings of difference of lengths based on 5 observations is

$$u(\overline{d}) = \frac{s_p}{\sqrt{5}} \approx 5.8nm.$$

The calibration certificate of the comparator states, that the uncertainty "due to random errors" is based on 6 replicate measurements and is reported to be ± 0.01 μm at a level of confidence of 95 %, that is, with the t-factor $t_{0.95}(5) = 2.57$ for 5 = 6-1 degrees of freedom (see [9], Annex G, Table G.2) we get (see [9], 4.3, in particular 4.3.4)

$$u(d_1) = \frac{0.01\mu m}{2.57} \approx 3.9nm.$$

The uncertainty "due to systematic errors" is also given by the comparator's calibration certificate as 0.02 μm at the "three sigma level", which means that the standard uncertainty of this component results to be

$$u(d_2) = \frac{0.02\mu m}{3} \approx 6.7nm.$$

All three components together yield the uncertainty $u(d)$:

$$u(d) = \sqrt{u^2(\overline{d}) + u^2(d_1) + u^2(d_2)} \approx \sqrt{93nm^2} \approx 9.7nm.$$

We may categorise the evaluations of the three uncertainty components into Type A and Type B the following way. The uncertainty component $u(\overline{d})$ is determined

by repeated measurements and is, therefore, based on a Type A evaluation. The components $u(d_1)$ and $u(d_2)$ which are contributed by the calibration certificate (no matter whether they are originally based on repeated measurements) are considered to be based on Type B evaluations.

In fact, the differentiation of Type A and Type B evaluations is not necessary, since the GUM treats them essentially in the same way (see [9], Annex E) with respect to the calculation of the combined standard uncertainty $u_c(y)$. Almost the only important impact appears, when the expanded uncertainty $U = ku_c(y)$ and the coverage factor k in particular is calculated. The latter is connected to the degrees of freedom ν_i for each component $u_i(y) = c_i u(x_i)$ and the effective degrees of freedom ν_{eff} for $u_c(y)$. Their derivations are discussed in section G and "a summary of the preferred method" is given in G.6.4 of [9].

Reporting uncertainty

When the combined standard uncertainty $u_c(y)$ is calculated the GUM gives some concrete suggestions in section 7 how a measurement result should be reported ([9], 7.2.1):

a) give a full description of how the measurand Y is defined;

b) give the estimate y of the measurand Y and its combined standard uncertainty $u_c(y)$; the units of y and $u_c(y)$ should always be given;

c) include the relative combined standard uncertainty $u_c(y)/y$, $y \neq 0$, when appropriate;

d) give the information outlined in 7.2.7 or refer to a published document that contains it.

Additionally, it is stated, that it may be useful, to include (ibid.)

– the estimated effective degrees of freedom ν_{eff} (see G.4);

– the Type A and Type B combined standard uncertainties $u_{cA}(y)$ and $u_{cB}(y)$ and their estimated effective degrees of freedom ν_{effA} and ν_{effB} (see G.4.1, Note 3).

Similar, the GUM gives guidances in other cases, i.e., when the expanded uncertainty $U = ku_c(y)$ should be the measure of uncertainty (7.2.3), or more than one measurand is determined simultaneously (7.2.5).

4.1.2. Metrological Requirements for Statistical Estimation

Statistical estimation deals with the task to determine the unknown value of a deterministic variable D involved in a random phenomenon. Similar as in the case of technical measurements, the deterministic variable D cannot be observed directly, but only through its impact on a random variable representing the considered random phenomenon. In statistics two methods of estimation are usually distinguished: the *point estimation* and the *set estimation*. We will see, that there are sometimes connections between these two methods.

Usually, the deterministic variable D represents a distribution parameter with unknown value ϑ of the probability distribution $P_{X|\{\vartheta\}}$ of the random variable $X|\{\vartheta\}$. Then, based on a random experiment, the value $(x_1, \ldots, x_n)$ of a random sample $(X_1|\{\vartheta\}, \ldots, X_n|\{\vartheta\})$ is observed. Note, that the random sample in estimation is equivalent to the measurement process for technical measurements. But there is a second difference between statistical estimation and technical measurement: For estimation the sample size n is in general larger than one, while for measurement it is in general equal to one.

The point estimation method (point estimator) maps the observed value $(x_1, \ldots, x_n)$ onto one value $\widehat{\vartheta}$, the *estimate* of D. This mapping may be regarded as a measurement procedure, i.e., a point estimator is a function W from the set of possible values of the sample (the range of variability) into a set which (at least) contains the true value ϑ of D. In which sense does a point estimator meet the measurement requirements of metrology? Accuracy for a point estimator W is interpreted as its so-called *bias*, i.e. the difference between the true value ϑ and the first moment $E[W]$ of the estimator. Maximum accuracy means that the bias is zero, in statistics this property of a point estimator is called *unbiasedness*. To call a point estimator reliable makes not much sense since it is almost in all situation impossible that an estimate $\widehat{\vartheta}$ equals the true value ϑ of D. Therefore, the term

reliability is not used in the context of point estimators. If the reliability would be interpreted as the probability, that the point estimation procedure yields the true value of the parameter then one would get in general zero[3]. Even though there are numerous methods to rate the quality of point estimators, this fact disqualifies the concept of point estimators. Their use becomes only justified by the fact that some set estimation methods are based on point estimators, again.

Analogous to point estimators, set estimation methods (set estimators) are based on a random sample. A set estimator maps the observed value $(x_1, \ldots, x_n)$ of the sample not onto one single value, but onto a set $C(\{x_1, \ldots, x_n\})$, i.e., a set estimator is a function C from the range of variability of the sample into a system of subsets of the set of values ϑ of D which cannot be excluded. Again, the question is in which manner a set estimator meets the measurement requirements of metrology.

Pretty much obvious is the interpretation of accuracy: The sizes of the resulting sets obtained by a set estimator may be regarded as a measure of its accuracy. Since the size might depend on the observed value of the random sample, i.e., the sets might not have some constant size, some more detailed definition would be necessary. The reliability of a set estimator should indicate how often we might expect to get a correct result. A result is regarded as correct if the true value ϑ of D is covered by the set estimate $C(\{x_1, \ldots, x_n\})$. The expectation of getting a correct result when applying a set estimator is appropriately quantified by the corresponding probability. Thus, the metrologistic terms accuracy and reliability have meaningful correspondents in the field of set estimators. The analogue of uncertainty would be the value of accuracy multiplied by 0.5. We will see, that as long as the sample size is not increased, accuracy and reliability are connected to each other in some reciprocal way, i.e., if the accuracy should be increased we have to accept a decrease in reliability and vice versa.

In the following sections we present point and set estimation methods in a general way as we may find them in many standard textbooks on statistical inference,

[3] A Situation may be constructed where the parameter space Θ consists of a finite number k of discrete values ϑ_i, $i = 1, \ldots, k$, and the variability spaces $\mathcal{X}_i$ of $X|\{\vartheta_i\}$ are depending on the respective ϑ_i in a way, that $(\mathcal{X}_i)_{i=1,\ldots,k}$ is a disjoint family. Then, realisations may match to exactly one value ϑ_i.

e.g. [13] and [6]. A nice basic introduction is also provided by [75].

Examples will be given to illustrate the manageability and potential shortcomings of the presented methods. The probability distributions used for illustration are the exponential distributions $EXP(a, \vartheta)$ and $EXP(\vartheta)$ when $a = 0$, respectively, the uniform distribution $\mathrm{U}(a, b)$ and the normal distribution $N(\mu, \sigma^2)$. The monotonic distribution $Mon(a, b, \mu_1)$ will not be used for illustration since this is exclusively discussed in the following chapter. The distributional parameter for all considered probability distributions is denoted D. However, to prevent misunderstandings the values of D are denoted as shown above differently.

4.2. Point Estimation

In [13] we find a very short and utmost general definition of point estimators:

Definition 4.4 (point estimator) *A point estimator is any function $W(X_1, \ldots, X_n)$ of a sample; that is, any statistic[4] is a point estimator.*

Following the definition, it is outlined, that it "makes no mention of any correspondence between the estimator and the parameter it is to estimate"([13], p. 311). Assume D is the parameter to be estimated with parameter space Θ, i.e., the set of possible values ϑ of D, the range of variability of the sample $\boldsymbol{X}|\{\vartheta\} = (X_1|\{\vartheta\}, \ldots, X_n|\{\vartheta\})$ is denoted $\mathcal{X}^n = \{(x_1, \ldots, x_n)|x_i \in \mathcal{X}_i\}$, and the range of variability of $W(\boldsymbol{X}|\{\vartheta\})$ is denoted $\mathcal{W} = \{W(x_1, \ldots, x_n)|x_i \in \mathcal{X}_i\}$. We note, that the equality $\mathcal{W} = \Theta$ is not required.
As a function of random variables a point estimator itself is a random variable. After the value $\boldsymbol{x} = (x_1, \ldots, x_n)$, i.e., the realisation of the sample $\boldsymbol{X}|\{\vartheta\}$, is available, the value $W(\boldsymbol{x})$ is called *(point) estimate*.

4.2.1. Method of Moments

The simplest and most intuitional approach to derive point estimators of parameters of probability distributions is the so-called *method of moments*.

[4] A "statistic" is also just any function of a sample.

Definition 4.5 (moment estimator) *Let $X|\{\vartheta\}$ be a random variable with values ϑ of the distribution parameter D and parameter space $\Theta \subset \mathbb{R}^d$. Let g be a function of moments $\mu_{t_1}, \mu_{t_2}, \ldots, \mu_{t_s}$, $1 \leq t_1 < t_2 < \cdots < t_s$, of $X|\{\vartheta\}$, so that*

$$g(\mu_{t_1}, \mu_{t_2}, \ldots, \mu_{t_s}) = \vartheta\,. \tag{4.5}$$

Let $M_{n,j}$ be the empirical moment of order j for an i.i.d. sample of $X|\{\vartheta\}$, i.e.,

$$M_{n,j} := \frac{1}{n}\sum_{i=1}^{n} X_i^j\,. \tag{4.6}$$

Then $g(M_{n,t_1}, M_{n,t_2}, \ldots, M_{n,t_s})$ is called a moment estimator for D, and for realisations $m_{n,j}$ of $M_{n,j}$ the value $g(m_{n,t_1}, m_{n,t_2}, \ldots, m_{n,t_s})$ is called the estimate of the unknown value ϑ of D.

Setting $s = 1$, $m_1 = m$ and $g = id$ yields $g(\mu_m) = \mu_m$ and

$$g(M_{n,m}) = M_{n,m} = \frac{1}{n}\sum_{i=1}^{n} X_i^m\,. \tag{4.7}$$

This means, that the empirical moment $M_{n,m}$ is a moment estimator of μ_m. Especially $\overline{X} = M_{n,1}$ is a moment estimator of μ_1.
Since there may be another function $\tilde{g}$ of a different number and sort of moments $\mu_{\tilde{m}_1}, \mu_{\tilde{m}_2}, \ldots, \mu_{\tilde{m}_t}$ whith $\tilde{g}(\mu_{\tilde{m}_1}, \mu_{\tilde{m}_2}, \ldots, \mu_{\tilde{m}_t}) = \vartheta$, we note that in general the moment estimator for D is not unique. The question which of the competing moment estimators should be choosen is subject of the next section, where quality characteristics of estimators are presented.

Example 4.6 *Let $\boldsymbol{X}|\{\vartheta\}$ be an i.i.d. sample for $X|\{\vartheta\} \sim EXP(\vartheta)$. Since $\mathrm{E}[X|\{\vartheta\}] = \frac{1}{\vartheta} = \mu_1$, we have $\vartheta = g(\mu_1) = \frac{1}{\mu_1}$. Thus, $g(\overline{X}) = \frac{1}{\overline{X}}$ is a moment estimator of ϑ. But since $\mathrm{E}[X^2|\{\vartheta\}] = \frac{2}{\vartheta^2} = \mu_2$, and hence $\vartheta = \tilde{g}(\mu_2) = \sqrt{\frac{2}{\mu_2}}$, we get with $\tilde{g}(M_{n,2}) = \sqrt{\frac{2}{M_{n,2}}}$ another moment estimator of ϑ.*

Example 4.7 *Let $\boldsymbol{X}|\{(\mu, \sigma^2)\}$ be an i.i.d. sample for $X|\{(\mu, \sigma^2)\} \sim N(\mu, \sigma^2)$ and $\vartheta = (\mu, \sigma^2)$. Since $\mathrm{E}[X|\{(\mu, \sigma^2)\}] = \mu = \mu_1$ and $\mathrm{E}[X^2|\{(\mu, \sigma^2)\}] = \sigma^2 + \mu^2 = \mu_2$, we have $(\mu, \sigma^2) = g(\mu_1, \mu_2) = (\mu_1, \mu_2 - \mu_1^2)$. Thus, $g(\overline{X}, M_{n,2}) = (\overline{X}, M_{n,2} - \overline{X}^2)$ is a moment estimator of (μ, σ^2).*

Especially in cases of a restricted parameter space Θ, the moment estimate may not be element of Θ, i.e., the moment estimator may yield a value which is inconsistent with the parameter space. Furthermore, the estimate may even contradict the observations.

Example 4.8 *Let $\boldsymbol{X}||\{b\}$ be an i.i.d. sample for $X|\{b\} \sim \mathrm{U}(0, b)$ and $\vartheta = b$. Since $\mathrm{E}[X|\{b\}] = \frac{b}{2} = \mu_1$, we have $b = g(\mu_1) = 2\mu_1$. Thus, $g(\overline{X}) = 2\overline{X}$ is a moment estimator of b. Now, suppose that the values*

$$(x_1, x_2, x_3, x_4, x_5) = (4.3, 1.7, 2.6, 0.9, 1.1)$$

are the realisation of the i.i.d. sample $X_1, \ldots, X_5$. Then $2\overline{x} = 4.24$ is the estimate of b. But the maximum of the sample's realisation is 4.3. Thus, the unknown value b has to be at least 4.3, otherwise 4.3 could not have been observed.

4.2.2. Maximum Likelihood Principle

Probably the most often applied method to derive point estimators is the so-called *Maximum Likelihood Principle* and the derived point estimators are consequently called the *Maximum Likelihood Estimators*. We note, that for the following definition we assume that the random variable X has a absolute continuous distribution.

Definition 4.9 (maximum likelihood estimator) *Let $f_{\boldsymbol{X}|\{\vartheta\}}$ be the density function of the random variable $\boldsymbol{X}|\{\vartheta\} = (X_1|\{\vartheta\}, \ldots, X_n|\{\vartheta\})$, $\vartheta \in \Theta \subset \mathbb{R}^d$. For each realisation $\boldsymbol{x} = (x_1, \ldots, x_n)$ of $\boldsymbol{X}|\{\vartheta\}$ define the function*

$$L(\vartheta; \boldsymbol{x}) := f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{x}) \tag{4.8}$$

for $\vartheta \in \Theta$ and given $\boldsymbol{x} \in \mathcal{X}$. Then $L(\vartheta; \boldsymbol{x}) : \Theta \to \mathbb{R}$ is called Likelihood Function *w.r.t. the realisation $\boldsymbol{x}$. Each value $\vartheta^* \in \Theta$ where $L(\vartheta; \boldsymbol{x})$ attains its maximum is called* Maximum Likelihood Estimate (MLE) *of ϑ w.r.t. $\boldsymbol{x}$, i.e.,*

$$\vartheta^* \in \Theta \text{ is MLE w.r.t. } \boldsymbol{x} \iff L(\vartheta^*; \boldsymbol{x}) = \sup_{\vartheta \in \Theta} L(\vartheta; \boldsymbol{x}) \,. \tag{4.9}$$

A function $\widehat{\vartheta}(\boldsymbol{X})$ *is called* Maximum Likelihood Estimator (also MLE) *for* D *if* $\widehat{\vartheta}(\boldsymbol{x})$ *is a maximum likelihood estimate of* ϑ *for all* $\boldsymbol{x} \in \mathcal{X} \setminus N$*, with*

$$\mathrm{P}_{\boldsymbol{X}|\{\vartheta\}}(\{\boldsymbol{x}|\boldsymbol{x} \in N\}) = \int_N f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = 0 \quad \textit{for all} \quad \vartheta \in \Theta\,. \tag{4.10}$$

The random variable $\boldsymbol{X}|\{\vartheta\} = (X_1|\{\vartheta\}, \ldots, X_n|\{\vartheta\})$ in the definition is not assumed to be a sample for $X|\{\vartheta\}$, but only to be a n-dimensional random variable. In the case when $\boldsymbol{X}|\{\vartheta\}$ is an i.i.d. sample for a random variable $X|\{\vartheta\}$ with density function $f_{X|\{\vartheta\}}$, the likelihood function $L(\vartheta; \boldsymbol{x})$ simplifies to:

$$L(\vartheta; \boldsymbol{x}) = \prod_{i=1}^{n} f_{X|\{\vartheta\}}(x_i)\,. \tag{4.11}$$

Remarks: In contrast to the moment estimator the MLE only yields values which are elements of Θ by construction. Finding the global maximum of $L(\vartheta; \boldsymbol{x})$ is more or less a problem of differential calculus, i.e., finding the zeros of the (partial) derivative(s) of $L(\vartheta; \boldsymbol{x})$ (solving the likelihood equations) and checking whether they are really global maxima. Difficulties arise when no explicit solutions may be given and numerical algorithms have to be utilised. Besides finding maxima in the inner of Θ, one has also to check the existence of boundary maxima. The latter case often appears when the parameter space is restricted.

Since the logarithm function is monotone the *log-likelihood function*

$$l(\vartheta; \boldsymbol{x}) := \log L(\vartheta; \boldsymbol{x}) \tag{4.12}$$

is analysed instead of $L(\vartheta; \boldsymbol{x})$. Maxima of $l(\vartheta; \boldsymbol{x})$ are maxima of $L(\vartheta; \boldsymbol{x})$ and vice versa, but the derivatives of $l(\vartheta; \boldsymbol{x})$ are often simpler.

Example 4.10 *Let* $\boldsymbol{X}|\{\vartheta\}$ *be an i.i.d. sample for* $X|\{\vartheta\} \sim EXP(\vartheta)$ *with* $\vartheta \in \Theta = (0, +\infty)$*. Then the likelihood function is*

$$L(\vartheta; \boldsymbol{x}) = \prod_{i=1}^{n} f_{X|\{\vartheta\}}(x_i) = \vartheta^n e^{-\vartheta \sum_{i=1}^{n} x_i} \mathbb{1}_{[0,+\infty)^n}(\boldsymbol{x}) \tag{4.13}$$

with log-likelihood function

$$l(\vartheta;\boldsymbol{x}) = n\log\vartheta - \vartheta\sum_{i=1}^{n} x_i\,. \tag{4.14}$$

We first note, that $\mathrm{P}_{\boldsymbol{X}|\{\vartheta\}}(\{\boldsymbol{x}\,|\,x_i \le 0, i=1,\ldots,n\}) = 0$ *for all* $\vartheta > 0$ *and hence we do not have to take the case* $\sum x_i = 0$ *into consideration. Differentiating* $l(\vartheta;\boldsymbol{x})$ *with respect to* ϑ *yields*

$$\frac{\partial}{\partial\vartheta} l(\vartheta;\boldsymbol{x}) = \frac{n}{\vartheta} - \sum_{i=1}^{n} x_i\,. \tag{4.15}$$

Equation $\frac{\partial}{\partial\vartheta} l(\vartheta;\boldsymbol{x}) = 0$ *has obviously exactly one solution which is a global maximum and hence the maximum likelihood estimate of* ϑ *is*

$$\vartheta^* = \frac{n}{\sum\limits_{i=1}^{n} x_i} = \frac{1}{\overline{x}}\,, \tag{4.16}$$

and the maximum likelihood estimator $\widehat{\vartheta} : (0,+\infty)^n \to \Theta$ *for* D *is*

$$\widehat{\vartheta}(\boldsymbol{X}) = \frac{1}{\overline{X}}\mathbb{1}_{(0,+\infty)^n}(\boldsymbol{X})\,. \tag{4.17}$$

Now suppose, that $\vartheta \ge c > 0$*, then* $\Theta = [c,+\infty)$*. A realisation of the sample with* $\overline{x} > \frac{1}{c}$ *implies* $\widehat{\vartheta}(\boldsymbol{x}) = \frac{1}{\overline{x}} < c$*, i.e., the estimate of* ϑ *is not an element of* Θ*. For* $\vartheta > \frac{1}{\overline{x}}$ *we have* $\frac{\partial}{\partial\vartheta} l(\vartheta;\boldsymbol{x}) < 0$*, i.e.,* $l(\vartheta;\boldsymbol{x})$ *is decreasing, therefore* $l(\vartheta;\boldsymbol{x})$ *attains its maximum at* $\widehat{\vartheta} = c$*. Consequently the maximum likelihood estimator for* D *is given by:*

$$\widehat{\vartheta}(\boldsymbol{X}) = \frac{1}{\overline{X}}\mathbb{1}_{(0,1/c]}(\overline{X}) + c\mathbb{1}_{(1/c,+\infty)}(\overline{X})\,. \tag{4.18}$$

Example 4.11 *Let* $\boldsymbol{X}|\{(\mu,\sigma^2)\}$ *be an i.i.d. sample for* $X|\{(\mu,\sigma^2)\} \sim N(\mu,\sigma^2)$ *and* $\vartheta = (\mu,\sigma^2)$ *with* $\Theta = \mathbb{R}\times(0,+\infty)$*. Then the likelihood function for any* $\boldsymbol{x} \in \mathbb{R}^n$ *is*

$$L(\mu,\sigma^2;\boldsymbol{x}) = \prod_{i=1}^{n} f_{X|\{(\mu,\sigma^2)\}}(x_i) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)(\sum\limits_{i=1}^{n}(x_i-\mu)^2)} \tag{4.19}$$

and the log-likelihood function is

$$l(\mu,\sigma^2;\boldsymbol{x}) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\,. \tag{4.20}$$

The system of likelihood equations is

$$\begin{aligned}\frac{\partial}{\partial\mu}l(\mu,\sigma^2;\boldsymbol{x}) &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 0\\ \frac{\partial}{\partial\sigma^2}l(\mu,\sigma^2;\boldsymbol{x}) &= -\frac{n}{2}\frac{1}{\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0\end{aligned} \tag{4.21}$$

which has the unique solution

$$\vartheta^* = \left(\mu^*,(\sigma^2)^*\right) = \left(\overline{x},\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2\right) \tag{4.22}$$

as long as $(\sigma^2)^* > 0 \Leftrightarrow \boldsymbol{x} \notin N := \{\boldsymbol{x}\,|\,x_1 = x_2 = \ldots = x_n\}$. *But since*

$$\mathrm{P}_{\boldsymbol{X}|\{(\mu,\sigma^2)\}}(\{\boldsymbol{x}\,|\,\boldsymbol{x}\in N\}) = 0 \;\; \textit{for all} \;\; (\mu,\sigma^2)\in\mathbb{R}\times(0,+\infty)\,, \tag{4.23}$$

this case needs not to be considered. With this, the maximum likelihood estimator $\widehat{\vartheta}:\mathbb{R}^n\to\Theta$ *for* D *is given by:*

$$\widehat{\vartheta}(\boldsymbol{X}) = \left(\widehat{\mu}(\boldsymbol{X}),\widehat{\sigma}^2(\boldsymbol{X})\right) = \left(\overline{X},\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right). \tag{4.24}$$

Example 4.12 *Let* $\boldsymbol{X}|\{b\}$ *be an i.i.d. sample for* $X|\{b\}\sim\mathrm{U}(0,b)$ *with* $b\in\Theta = (0,+\infty)$. *Then the likelihood function is*

$$\begin{aligned}L(b;\boldsymbol{x}) &= \prod_{i=1}^{n}\frac{1}{b}\mathbb{1}_{[0,b]}(x_i)\\ &= \frac{1}{b^n}\mathbb{1}_{[\max\{x_1,\ldots,x_n\},+\infty)}(b)\,.\end{aligned} \tag{4.25}$$

Obviously, $L(b;\boldsymbol{x})$ *attains its maximum in* $b^* = \max\{x_1,\ldots,x_n\}$ *and, therefore,* $\widehat{b}(\boldsymbol{X}) = \max\{X_1,\ldots,X_n\}$ *is the maximum likelihood estimator of* b.

An important property of the MLEs is the so-called *invariance property*:

Proposition 4.13 (invariance of MLE) *Let* $\widehat{\vartheta}$ *be the maximum likelihood estimator of* ϑ *and* $h:\Theta\to\mathbb{R}^k$ *an abritrary function. Then the maximum likelihood estimator of* $h(\vartheta)$ *is* $h(\widehat{\vartheta})$.

A proof may be found in many standard textbooks like [13, 2].

Example 4.14 *Let $\boldsymbol{X}|\{\vartheta\}$ be an i.i.d. sample for $X|\{\vartheta\} \sim EXP(\vartheta)$. If not the value ϑ of D but $\mu_1 = \mathrm{E}[X|\{\vartheta\}]$ shall be estimated, then from example 4.10 and with proposition 4.13 we readily conclude, that the maximum likelihood estimator of μ_1 is*

$$\widehat{\mu}_1(\boldsymbol{X}) = \overline{X}\,. \tag{4.26}$$

Example 4.15 *Let $\boldsymbol{X}|\{\sigma\}$ be an i.i.d. sample for $X|\{\sigma\} \sim N(0, \sigma^2)$ and $\sigma \in \Theta = (0, +\infty)$. From example 4.11 we have that the maximum likelihood estimator of σ^2 is $n^{-1}\sum_{i=1}^n X_i^2$. With proposition 4.13 it follows, that the maximum likelihood estimator of $\sigma = \sqrt{\sigma^2}$ is*

$$\widehat{\sigma}(\boldsymbol{X}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}\,. \tag{4.27}$$

Maximum Likelihood in Exponential Families

In section 3.2.2 we have shortly introduced the concept of an exponential family. The density functions of distributions generating an exponential family are of form (3.95):

$$f_{X|\{\vartheta\}}(x) = h(x)c(\vartheta)\exp\Big(\sum_{j=1}^{k} w_j(\vartheta)t_j(x)\Big) \tag{4.28}$$

with functions h, c, w_i, t_i defined in (3.96). Often the reparameterisation $\eta_j = w_j(\vartheta)$, $j = 1, \ldots, k$, is performed[5] and (4.28) written in the form

$$f_{X|\{\eta\}}(x) = h(x)\exp\Big(\sum_{j=1}^{k} \eta_j t_j(x) - A(\eta)\Big) \tag{4.29}$$

which is called *natural parameterisation* or *canonical* form. In fact, the set of possible values $\eta = (\eta_1, \ldots, \eta_k)$ need not arise from the reparameterisation, but is defined by ([13], p.114)

$$\mathcal{H} = \{\eta = (\eta_1, \ldots, \eta_k) \,|\, \int_{\mathbb{R}} h(x)\exp\Big(\sum_{j=1}^{k} \eta_j t_j(x)\Big)\,\mathrm{d}x < \infty\} \tag{4.30}$$

[5] In some cases $w_j(\vartheta) = \vartheta_j$ for all $j = 1, \ldots, k$ with the dimension d of ϑ equals the dimension k of η, i.e., $k = d$.

which is called the *natural parameter space*. Additionally, the term

$$A(\eta) = \ln \int_{\mathbb{R}} h(x) \exp\Big(\sum_{j=1}^{k} \eta_j t_j(x)\Big) \, \mathrm{d}x \tag{4.31}$$

plays not only the role of a normalisation factor, but has also some interesting connection to the moment-generating function $M_{(t_1(X),\ldots,t_k(X))}$ (see [6], p.59).

Example 4.16 *Let* $X|\{(\mu,\sigma^2)\} \sim N(\mu,\sigma^2)$. *The density function of* $X|\{(\mu,\sigma^2)\}$ *may be written in the form*

$$f_{X|\{(\mu,\sigma^2)\}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}\right) . \tag{4.32}$$

Defining

$$h(x) = 1 \quad \textit{for all } x \in \mathbb{R}, \tag{4.33}$$

$$c(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \quad \textit{for } \mu \in \mathbb{R}, \sigma > 0, \tag{4.34}$$

$$w_1(\mu,\sigma^2) = -\frac{1}{2\sigma^2} \quad \textit{for } \sigma^2 > 0, \qquad w_2(\mu,\sigma^2) = \frac{\mu}{\sigma^2} \quad \textit{for } \sigma^2 > 0, \tag{4.35}$$

$$t_1(x) = x^2 \quad \textit{and} \quad t_2(x) = x \, , \tag{4.36}$$

yields

$$f_{X|\{(\mu,\sigma^2)\}}(x) = h(x)c(\mu,\sigma^2) \exp\big(w_1(\mu,\sigma^2)t_1(x) + w_2(\mu,\sigma^2)t_2(x)\big) \, , \tag{4.37}$$

implying that the density functions of the normal distribution form an exponential family with $k = 2$. *Identifying* η_j *with* $w_j(\mu,\sigma^2)$, $j = 1, 2$, *we easily get the natural parameterisation form*

$$f_{X|\{(\eta_1,\eta_2)\}}(x) = \frac{1}{\sqrt{\pi}} \exp\left(\eta_1 x^2 + \eta_2 x - \left(-\frac{\eta_2^2}{4\eta_1} + \frac{1}{2}\ln(-\eta_1)\right)\right) \tag{4.38}$$

with $\eta_1 < 0$ *and* $\eta_2 \in \mathbb{R}$.

Example 4.17 *Let* $X|\{(a,\vartheta)\} \sim EXP(a,\vartheta)$. *The density function of* $X|\{(a,\vartheta)\}$ *is*

$$f_{X|\{(a,\vartheta)\}}(x) = \vartheta \exp\big(-\vartheta(x-a)\big)\mathbb{1}_{[a,+\infty)}(x) = \\ = \mathbb{1}_{[a,+\infty)}(x) \exp\big(-\vartheta x - (-\vartheta a - \ln\vartheta)\big) \, . \tag{4.39}$$

Obviously, the density functions generate an exponential family in ϑ, but not in a, since $\mathbb{1}_{[a,+\infty)}(x)$ cannot be seperated into a function solely of x and a function solely of a. With $\eta = -\vartheta < 0$, we get the natural parameterisation form

$$f^{(a)}_{X|\{\eta\}}(x) = \mathbb{1}_{[a,+\infty)}(x) \exp\big(\eta x - (\eta a - \ln(-\eta))\big), \tag{4.40}$$

i.e., $h(x) = \mathbb{1}_{[a,+\infty)}(x)$, $t(x) = x$ and $A(\eta) = \eta a - \ln(-\eta)$.

Let $\boldsymbol{X}|\{\eta\}$, $\eta = (\eta_1, \ldots, \eta_k)$, be an i.i.d. sample for $X|\{\eta\}$, then the joint density function in the natural parameterisation form and the likelihood function, respectively, is

$$\prod_{i=1}^{n} f_{X|\{\eta\}}(x_i) = \Big(\prod_{i=1}^{n} h(x_i)\Big) \exp\Big(\sum_{j=1}^{k} \eta_j T_{nj}(\boldsymbol{x}) - nA(\eta)\Big) \tag{4.41}$$

with

$$T_{nj}(\boldsymbol{x}) = \sum_{i=1}^{n} t_j(x_i)\,. \tag{4.42}$$

First, we note, that the joint densities (4.41) form again an exponential family. Calculating the log-likelihood function $l(\eta; \boldsymbol{x})$ and partially differentiating it with respect to η_s yields

$$\frac{\partial l}{\partial \eta_s}(\eta; \boldsymbol{x}) = T_{ns}(\boldsymbol{x}) - n\frac{\partial A(\eta)}{\partial \eta_s}\,, s = 1, \ldots, k. \tag{4.43}$$

Then, from the theory of exponential families it is known, that

$$\mathrm{E}[t_s(X)|\{\eta\}] = \frac{\partial A(\eta)}{\partial \eta_s}\,, s = 1, \ldots, k, \tag{4.44}$$

and, thus, $\frac{\partial l}{\partial \eta_s}(\eta; \boldsymbol{x}) = 0$ yields the system of equations

$$\frac{1}{n} T_{ns}(\boldsymbol{x}) = \mathrm{E}[t_s(X)|\{\eta\}]\,, s = 1, \ldots, k. \tag{4.45}$$

Then, under some often met conditions (e.g. see [6], pp.121), the maximum likelihood estimator $\widehat{\eta}$ exists, is unique and is a solution of (4.45).

Example 4.18 *(Continuation of example 4.16) Let $n > 1$ and $\boldsymbol{X}|\{(\eta_1, \eta_2)\}$ be an i.i.d. sample for $X|\{(\eta_1, \eta_2)\} \sim N(-\frac{\eta_2}{2\eta_1}, -\frac{1}{2\eta_1})$. The system of equations*

(4.45) *yields*

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{\eta_2^2}{4\eta_1^2} - \frac{1}{2\eta_1}\,, \tag{4.46}$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x} = -\frac{\eta_2}{2\eta_1}\,, \tag{4.47}$$

which has as unique solution the maximum likelihood estimate η^* *and, since* $\mathrm{P}_{\boldsymbol{X}|\{(\eta_1,\eta_2)\}}(\{\boldsymbol{x}\,|\,x_1 = x_2 = \cdots = x_n\}) = 0$*, i.e., for* $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

$$\mathrm{P}_{S^2|\{(\eta_1,\eta_2)\}}(\{0\}) = 0 \tag{4.48}$$

holds, the MLE $\widehat{\eta}(\boldsymbol{X})$*:*

$$\widehat{\eta}(\boldsymbol{X}) = \left(-\frac{1}{2S^2}, \frac{\overline{X}}{S^2}\right)\,. \tag{4.49}$$

Not surprisingly, $\widehat{\eta}$ *would also be obtained by applying the one-to-one-mapping between the original parameter values* (μ, σ^2) *of the normal distribution and* η *onto the MLE* $(\widehat{\mu}, \widehat{\sigma}^2)$*, see example 4.11.*

Example 4.19 *(Continuation of example 4.17) Let* $n > 1$ *and* $\boldsymbol{X}|\{a, \vartheta\}$ *be an i.i.d. sample for* $X|\{\vartheta\} \sim EXP(a, \vartheta)$ *with known value* a*. Without loss of generality we may set* $a = 0$*, then the parameter value* η *is one-dimensional, and* (4.45) *yields only one single equation:*

$$\frac{1}{n}\sum_{i=1}^{n} x_i = -\frac{1}{\eta}\,. \tag{4.50}$$

Since $\mathrm{P}_{X||\{\eta\}}(\{x|x > 0\}) = 1$ *this yields the MLE of* η *to be*

$$\widehat{\eta}(\boldsymbol{X}|\{\eta\}) \;=\; -\frac{1}{\overline{X}}\,. \tag{4.51}$$

Again, $\widehat{\eta}$ *is also obtained from the mapping* $\vartheta \mapsto \eta = -\vartheta$*.*

4.2.3. Bayes Estimator

The previous two methods require that the probability distribution $\mathrm{P}_{\boldsymbol{X}}$ of the i.i.d. sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a member of a family of probability distributions parameterised by the values $\vartheta \in \Theta$ of the distribution parameter D, i.e.,

$\mathrm{P}_{\boldsymbol{X}} = \mathrm{P}_{\boldsymbol{X}|\{\vartheta\}}$. Because the true value ϑ of D is unknown but fixed it is subject to estimation. The Bayesian approach differs in this point by assuming D itself as a random variable. To stress this crucial point, we will change our notation slightly and denote the random parameter with the capital letter $\boldsymbol{Y}$ (also in bold, because $\boldsymbol{Y}$ may be multi-dimensional), whose values $\boldsymbol{y}$ are the values of the distribution parameter of $\boldsymbol{X}$.

The probability distribution of $\boldsymbol{Y}$ is called *a-priori-distribution* or just *prior distribution*, and we will denote it by $\pi_{\boldsymbol{Y}}(\boldsymbol{y})$[6]. Then, given the realisation $\boldsymbol{x}$ of $\boldsymbol{X}$, the *a-posteriori-distribution* or simply *posterior distribution* is

$$\pi_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{x}) = \frac{f_{\boldsymbol{X}|\{\boldsymbol{Y}=\boldsymbol{y}\}}(\boldsymbol{x})\pi_{\boldsymbol{Y}}(\boldsymbol{y})}{\int f_{\boldsymbol{X}|\{\boldsymbol{Y}=\boldsymbol{y}\}}(\boldsymbol{x})\pi_{\boldsymbol{Y}}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}}\,. \tag{4.52}$$

We note, that the denominator in (4.52) is the marginal distribution of $\boldsymbol{X}$, often denoted by $m(\boldsymbol{x})$

$$m(\boldsymbol{x}) = \int f_{\boldsymbol{X}|\{\boldsymbol{Y}=\boldsymbol{y}\}}(\boldsymbol{x})\pi_{\boldsymbol{Y}}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \tag{4.53}$$

Point estimates $\widehat{\boldsymbol{Y}}$ for $\boldsymbol{Y}$ may now be derived from the posterior distribution, e.g. the mean of $\pi_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{x})$, or if $\boldsymbol{Y} = Y$ is one-dimensional also the mode and median may be considered as point estimates.

Example 4.20 *Let $X_i|\{Y = y\} \sim EXP(y)$ and the prior distribution of Y is the gamma distribution, i.e.,*

$$\pi_Y(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-y\beta} \tag{4.54}$$

for $y \geq 0$ and with known parameters $\alpha, \beta > 0$. Then, for all $\boldsymbol{x} \in [0, +\infty)^n$ the posterior distribution results to be

$$\pi_Y(y|\boldsymbol{x}) = \frac{(n\overline{x} + \beta)^{\alpha+n}}{\Gamma(\alpha+n)} y^{(\alpha+n)-1} e^{-y(n\overline{x}+\beta)}\,. \tag{4.55}$$

We see, that $\pi_Y(y|\boldsymbol{x})$ is again a gamma distribution with parameters $\alpha + n$ and $n\overline{x} + \beta$. Assume, we have $\alpha = 1$, $\beta = 0.5$, $n = 10$ and $\overline{x} = 0.3$. Figure 4.1 displays both $\pi_Y(y)$ (dashed line) and $\pi_Y(y|\boldsymbol{x})$ (solid line). Now, point estimates

[6] In a more general way, the prior distribution $\pi_{\boldsymbol{Y}|\{\eta\}}(\boldsymbol{y})$ is also parameterised by a so-called *hyperparameter* η. We will only introduce the Bayesian method in the case where η is known, thus, we may omit it in the notation.

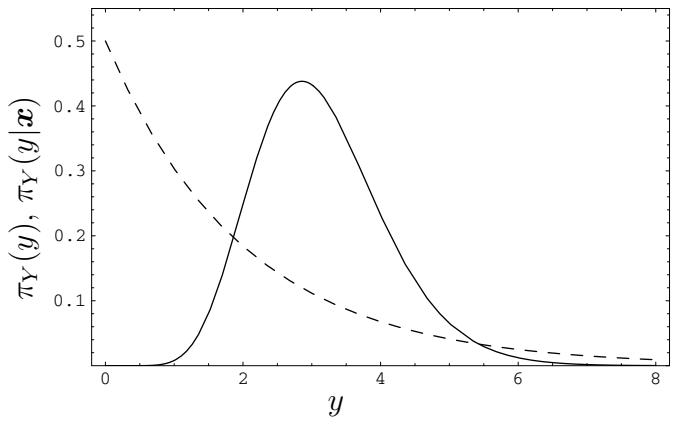


Figure 4.1.: Prior distribution $\pi_Y(y)$ (dashed line) and resulting posterior distribution $\pi_Y(y|\boldsymbol{x})$ (solid line) for $\alpha = 1$, $\beta = 0.5$, $n = 10$ and $\overline{x} = 0.3$.

$\widehat{Y}$ for Y may be the mean of $\pi_Y(y|\boldsymbol{x})$, which is

$$\mathrm{E}[Y|\boldsymbol{x}] = \frac{\alpha + n}{n\overline{x} + \beta}\,, \tag{4.56}$$

or the mode M of $\pi_Y(y|\boldsymbol{x})$, which is

$$M = \frac{\alpha + n - 1}{n\overline{x} + \beta}\,. \tag{4.57}$$

4.3. Evaluating Point Estimators

The methods in the previous section describe ways of deriving point estimators. Sometimes, with different methods the same estimator of the value ϑ of a parameter D is obtained, but many times this is not the case. In the latter, the question arises which of the competing point estimators is more favourable. To answer this we need some methods of evaluation – the basic ones we want to present here following again [13].

4.3.1. Mean Squared Error MSE, Bias

Definition 4.21 (mean squared error) *Let $W|\{\vartheta\} = W(\boldsymbol{X}|\{\vartheta\})$ be a point estimator of the value ϑ of a parameter D. Then the mean squared error MSE*

as a function of $\vartheta \in \Theta$ *is defined by*

$$\mathrm{MSE}[W|\{\vartheta\}] = \mathrm{E}[(W|\{\vartheta\} - \vartheta)^2]\,. \tag{4.58}$$

Equation (4.58) may be easily rewritten into

$$\mathrm{MSE}[W|\{\vartheta\}] = \mathrm{Var}[W|\{\vartheta\}] + (\mathrm{E}[W|\{\vartheta\}] - \vartheta)^2\,, \tag{4.59}$$

where the second term is called the bias of the estimator $W|\{\vartheta\}$:

Definition 4.22 (bias) *Let* $W|\{\vartheta\}$ *be a point estimator of the value* ϑ *of a parameter* D*. Then the bias of* $W|\{\vartheta\}$ *is defined by*

$$\mathrm{bias}[W|\{\vartheta\}] = \mathrm{E}[W|\{\vartheta\}] - \vartheta\,. \tag{4.60}$$

If the bias of an estimator $W|\{\vartheta\}$ *equals* 0 *for all* ϑ*, i.e.,* $\mathrm{E}[W|\{\vartheta\}] = \vartheta$ *for all* $\vartheta \in \Theta$*, then the estimator is called unbiased.*

If an estimator $W|\{\vartheta\}$ is unbiased, we easily see, that its MSE equals the variance, i.e., $\mathrm{MSE}[W|\{\vartheta\}] = \mathrm{Var}[W|\{\vartheta\}]$.
Although, unbiased estimators are most commonly preferred, their MSE may be larger than that of a biased estimator:

Example 4.23 *Let* $\boldsymbol{X}|\{(\mu,\sigma^2)\}$ *be an i.i.d. sample for* $X|\{(\mu,\sigma^2)\} \sim N(\mu,\sigma^2)$*. The method of moments and the ML method yield both the same estimator* $\widehat{\sigma}^2(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ *of* σ^2 *(see examples 4.7 and 4.11). Since*

$$\mathrm{E}[\widehat{\sigma}^2(\boldsymbol{X})|\{(\mu,\sigma^2)\}] = \frac{n-1}{n}\sigma^2\,, \tag{4.61}$$

$\widehat{\sigma}^2(\boldsymbol{X})$ *is a biased estimator of* σ^2*. From this, it is equally easy to derive the well-known unbiased estimator* $S^2(\boldsymbol{X}) = \frac{n}{n-1}\widehat{\sigma}^2(\boldsymbol{X})$ *of* σ^2*. Then, the MSEs of the two estimators are given as follows:*

$$\begin{aligned}\mathrm{MSE}[\widehat{\sigma}^2(\boldsymbol{X})|\{(\mu,\sigma^2)\}] &= \mathrm{Var}[\widehat{\sigma}^2(\boldsymbol{X})|\{(\mu,\sigma^2)\}] + \mathrm{bias}[\widehat{\sigma}^2(\boldsymbol{X})|\{(\mu,\sigma^2)\}] \\ &= \frac{2n-1}{n^2}\sigma^4\,, \end{aligned} \tag{4.62}$$

$$\mathrm{MSE}[S^2(\boldsymbol{X})|\{(\mu,\sigma^2)\}] = \mathrm{Var}[S^2(\boldsymbol{X})|\{(\mu,\sigma^2)\}] = \frac{2}{n-1}\sigma^4\,, \tag{4.63}$$

implying $MSE[\widehat{\sigma}^2(\boldsymbol{X})|\{(\mu,\sigma^2)\}] < MSE[S^2(\boldsymbol{X})|\{(\mu,\sigma^2)\}]$.

For some further discussion of the usefullness of the MSE as a criterion for choosing between competing estimators see [13], p. 332.

Example 4.24 *Let* $\boldsymbol{X}|\{b\}$ *be an i.i.d sample for* $X|\{b\} \sim \mathrm{U}(0,b)$. *Example 4.8 yields the moment estimator* $2\overline{X}$ *of* b *and example 4.12 the ML estimator* $\widehat{b}_{MLE}(\boldsymbol{X}) = \max\{X_1,\ldots,X_n\}$. *Since*

$$\mathrm{E}[2\overline{X}|\{b\}] = b\,, \tag{4.64}$$

$$\mathrm{E}[\widehat{b}_{MLE}(\boldsymbol{X})|\{b\}] = \frac{n}{n+1}b\,, \tag{4.65}$$

we see that the moment estimator is unbiased, while the ML estimator is biased. Multiplying the latter with $\frac{n+1}{n}$ *yields an unbiased estimator* $\frac{n+1}{n}\widehat{b}_{MLE}(\boldsymbol{X})$ *of* b. *For the MSEs we have*

$$\mathrm{MSE}[2\overline{X}|\{b\}] = \mathrm{Var}[2\overline{X}|\{b\}] = \frac{b^2}{3n}\,, \tag{4.66}$$

$$\mathrm{MSE}[\frac{n+1}{n}\widehat{b}_{MLE}(\boldsymbol{X})|\{b\}] = \mathrm{Var}[\frac{n+1}{n}\hat{b}_{MLE}(\boldsymbol{X})|\{b\}] = \frac{b^2}{n(n+2)} \tag{4.67}$$

$$\mathrm{MSE}[\widehat{b}_{MLE}(\boldsymbol{X})|\{b\}] = \frac{2b^2}{(n+1)(n+2)}\,. \tag{4.68}$$

For $n \geq 2$ *the MSE of* $2\overline{X}$ *is always at least as large as that of* $\widehat{b}_{MLE}(\boldsymbol{X})$, *which itself is always larger than that of* $\frac{n+1}{n}\widehat{b}_{MLE}(\boldsymbol{X})$. *Thus, basing our choice on the MSE we would select* $\frac{n+1}{n}\max\{X_1|\{b\},\ldots,X_n|\{b\}\}$ *as an estimator of* b.

Example 4.25 *Let* $\boldsymbol{X}|\{\vartheta\}$ *be an i.i.d. sample for* $X|\{\vartheta\} \sim EXP(\vartheta)$. *From example 4.10 we have the MLE* $\widehat{\vartheta}(\boldsymbol{X}) = \frac{1}{\overline{X}}$ *of* ϑ, *which is also a moment estimator. It is then known, that* $n\overline{X} = \sum_{i=1}^{n} X_i$ *is distributed acccording to the gamma distribution* $GAM(\vartheta, n)$ *with density function given by*

$$f_{\sum_{i=1}^{n} X_i|\{\vartheta\}}(t) = \frac{\vartheta^n t^{n-1} e^{-\vartheta t}}{\Gamma(n)} \mathbb{1}_{[0,+\infty)}(t)\,, \tag{4.69}$$

and thus $\overline{X}|\{(n,\vartheta)\} \sim GAM(n\vartheta, n)$. *With this, we obtain for* $n > 1$ *the expected value of* $\frac{1}{\overline{X}}$ *as*

$$\mathrm{E}\left[\frac{1}{\overline{X}}|\{(n,\vartheta)\}\right] = \frac{n}{n-1}\vartheta \tag{4.70}$$

which shows that $\frac{1}{\overline{X}}$ *is a biased estimator of* ϑ*. Multiplying it with* $\frac{n-1}{n}$ *yields the unbiased estimator* $\frac{n-1}{n}\frac{1}{\overline{X}}$ *of* ϑ*. For the variance we get for* $n > 2$

$$\begin{aligned}&\mathrm{Var}\left[\frac{n-1}{n}\frac{1}{\overline{X}}|\{(n,\vartheta)\}\right] = \\ &= \left(\frac{n-1}{n}\right)^2 \mathrm{Var}\left[\frac{1}{\overline{X}}|\{(n,\vartheta)\}\right] = \left(\frac{n-1}{n}\right)^2 \frac{n^2}{(n-1)^2(n-2)}\vartheta^2 = \frac{1}{n-2}\vartheta^2\,.\end{aligned} \tag{4.71}$$

Comparing the MSE of $\frac{1}{\overline{X}}$ *with that of* $\frac{n-1}{n}\frac{1}{\overline{X}}$ *shows that we should favour* $\frac{n-1}{n}\frac{1}{\overline{X}}$ *as an estimator of* ϑ*.*

4.3.2. Uniformly Minimum Variance Unbiased Estimator UMVUE

In traditional statistics, almost only unbiased estimators are further analysed. This fact arise from the interpretation of unbiasedness: If we could draw infinitely often a sample under the same conditions, i.e., the samples themselves are i.i.d. too, the mean of the resulting point estimates would converge to the true, but unknown value ϑ of the distribution parameter D. From this perception, choosing the variance as selection criterion among competing unbiased estimators seems to be reasonable, because an unbiased estimator with small variance would yield more often estimates which are close to the unknown value of the parameter than an unbiased estimator with larger variance.

Definition 4.26 (UMVUE) *Let* $W_1|\{\vartheta\}$ *and* $W_2|\{\vartheta\}$ *be two unbiased estimators of a function* $\gamma(\vartheta)$ *of the values* $\vartheta \in \Theta$ *of a distribution parameter* D*. Then* $W_1|\{\vartheta\}$ *is called uniformly better than* $W_2|\{\vartheta\}$ *if*

$$\mathrm{Var}[W_1|\{\vartheta\}] \leq \mathrm{Var}[W_2|\{\vartheta\}] \quad \textit{for all } \vartheta \in \Theta\,. \tag{4.72}$$

If inequality (4.72) *holds for any other unbiased estimator* $W_2|\{\vartheta\}$ *of* $\gamma(\vartheta)$*, than* $W_1|\{\vartheta\}$ *is called uniformly minimum variance unbiased estimator (UMVUE) of* $\gamma(\vartheta)$*.*

Example 4.27 *In example 4.24 we have already seen that the variance of the unbiased estimator* $2\overline{X}$ *of* b *is larger than that of the likewise unbiased estimator* $\frac{n+1}{n}\max\{X_1,\ldots,X_n\}$. *Hence,* $\frac{n+1}{n}\max\{X_1,\ldots,X_n\}$ *is uniformly better than* $2\overline{X}$.

The well-known *Cramér-Roa inequality* establishes a lower bound of the variance of unbiased estimators (see [13], pp. 335).

Theorem 4.28 (Cramér-Roa inequality) *Let* $\boldsymbol{X}|\{\vartheta\} = (X_1|\{\vartheta\},\ldots,X_n|\{\vartheta\})$ *be a sample with joint density function* $f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{x})$, *and let* $W|\{\vartheta\} = W(\boldsymbol{X}|\{\vartheta\})$ *be an estimator satisfying*

$$\frac{\mathrm{d}}{\mathrm{d}\vartheta}\mathrm{E}[W|\{\vartheta\}] = \int_{\mathcal{X}^n} \frac{\partial}{\partial\vartheta}\big(W(\boldsymbol{x})f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{x})\big)\,\mathrm{d}\boldsymbol{x} \quad \textit{and} \quad \mathrm{Var}[W|\{\vartheta\}] < +\infty\,. \tag{4.73}$$

Then

$$\mathrm{Var}[W|\{\vartheta\}] \geq \frac{\frac{\mathrm{d}}{\mathrm{d}\vartheta}\mathrm{E}[W|\{\vartheta\}]}{\mathrm{E}\Big[\big(\frac{\partial}{\partial\vartheta}\log f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{X})\big)^2|\{\vartheta\}\Big]}\,. \tag{4.74}$$

If $\boldsymbol{X}|\{\vartheta\}$ is an i.i.d. sample for $X|\{\vartheta\}$ with density $f_{X|\{\vartheta\}}(x)$, then the inequality (4.74) simplifies to

$$\mathrm{Var}[W|\{\vartheta\}] \geq \frac{\frac{\mathrm{d}}{\mathrm{d}\vartheta}\mathrm{E}[W|\{\vartheta\}]}{n\mathrm{E}\Big[\big(\frac{\partial}{\partial\vartheta}\log f_{X|\{\vartheta\}}(X)\big)^2|\{\vartheta\}\Big]}\,. \tag{4.75}$$

The denominator in (4.74) and (4.75), respectively, is called *Fisher information* $I(\vartheta)$:

$$I(\vartheta) := \mathrm{E}\Big[\big(\frac{\partial}{\partial\vartheta}\log f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{X})\big)^2|\{\vartheta\}\Big] \tag{4.76}$$

Furthermore it is known, that if the density function $f_{X|\{\vartheta\}}(x)$ satisfies

$$\frac{\mathrm{d}}{\mathrm{d}\vartheta}\mathrm{E}\Big[\frac{\partial}{\partial\vartheta}\log f_{X|\{\vartheta\}}(X)|\{\vartheta\}\Big] = \int \frac{\partial}{\partial\vartheta}\left[\frac{\partial}{\partial\vartheta}\log f_{X|\{\vartheta\}}(x)\right] f_{X|\{\vartheta\}}(x)\,\mathrm{d}x \tag{4.77}$$

then the Fisher information meets the following relation:

$$\mathrm{E}\Big[\big(\frac{\partial}{\partial\vartheta}\log f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{X})\big)^2|\{\vartheta\}\Big] = -\mathrm{E}\Big[\frac{\partial^2}{\partial\vartheta^2}\log f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{X})|\{\vartheta\}\Big]\,. \tag{4.78}$$

The condition is especially met for exponential families[7].

The following proposition may represent a constructive method to find an UMVUE.

[7] See [13], p. 338.

Proposition 4.29 (Cramér-Rao lower bound) *Let $\boldsymbol{X}|\{\vartheta\}$ be an i.i.d. sample for $X|\{\vartheta\}$ with density $f_{X|\{\vartheta\}}(x)$, where $f_{X|\{\vartheta\}}(x)$ satisfies condition (4.73). Let $L(\vartheta;\boldsymbol{x}) = \prod_{i=1}^{n} f_{X|\{\vartheta\}}(x_i)$ denote the likelihood function of $\boldsymbol{x} = (x_1,\ldots,x_n)$. If $W|\{\vartheta\}$ is any unbiased estimator of $\gamma(\vartheta)$, then $W|\{\vartheta\}$ attains the Cramér-Rao Lower Bound if and only if*

$$a(\vartheta)\big[W(\boldsymbol{x}) - \gamma(\vartheta)\big] = \frac{\partial}{\partial\vartheta}\log L(\vartheta;\boldsymbol{x}) \tag{4.79}$$

for some function $a(\vartheta)$.

4.3.2.1. Sufficient, minimal sufficient and complete statistics

In the search for uniformly minimum variance unbiased estimators, sufficiency und completeness of a statistic $T(\boldsymbol{X}|\{\vartheta\})$, i.e., any function of a sample $\boldsymbol{X}|\{\vartheta\}$, play some important role. That is, if we are able to find a sufficient and complete statistic $T(\boldsymbol{X}|\{\vartheta\})$, the following propositions show a way, to find a uniformly minimum variance unbiased estimator. We confine ourself by only presenting the definition and some basic propositions. We illustrate the statements by the uniform distribution $\mathrm{U}(0,b)$, $b>0$. As usual let $\mathcal{X}^n$ be the range of variability of the sample $\boldsymbol{X}|\{\vartheta\}$ and $\mathcal{T}$ be the codomain of $T(\boldsymbol{X}|\{\vartheta\})$.

Definition 4.30 (sufficient statistic) *Let $T : \mathcal{X}^n \to \mathcal{T}$ be a statistic, then $T(\boldsymbol{X}|\{\vartheta\})$ is called a sufficient statistic for ϑ if the conditional distribution of $\boldsymbol{X}|\{\vartheta\}$ for given value of $T(\boldsymbol{X}|\{\vartheta\})$ does not depend on ϑ.*

Proposition 4.31 (sufficient statistic) *Let $T : \mathcal{X}^n \to \mathcal{T}$ be a statistic, $f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{x})$ be the density function of $\boldsymbol{X}|\{\vartheta\}$ and $g_{T(\boldsymbol{X}|\{\vartheta\})}(t)$ the density function of $T(\boldsymbol{X}|\{\vartheta\})$. Then $T(\boldsymbol{X}|\{\vartheta\})$ is a sufficient statistic for ϑ if and only if*

$$\frac{f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{x})}{g_{T(\boldsymbol{X}|\{\vartheta\})}(T(\boldsymbol{x}))} = c(\boldsymbol{x}) \quad \text{for every} \quad \boldsymbol{x}\in\mathcal{X}^n\,, \tag{4.80}$$

i.e., is only a function of $\boldsymbol{x}$ but not of ϑ.

Proposition 4.32 (sufficient statistic, factorisation) *Let $f_{\boldsymbol{X}|\{\vartheta\}}$ be the density function of $\boldsymbol{X}|\{\vartheta\}$ and $T : \mathcal{X}^n \to \mathcal{T}$ be a statistic. Then $T(\boldsymbol{X}|\{\vartheta\})$ is a*

sufficient statistic for ϑ if and only if there exist functions $q_\vartheta(t)$ and $h(\boldsymbol{x})$ such that

$$f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{x}) = q_\vartheta(T(\boldsymbol{x}))h(\boldsymbol{x}) \tag{4.81}$$

for all $\boldsymbol{x} \in \mathcal{X}$ and all $\vartheta \in \Theta$.

Example 4.33 *Let $\boldsymbol{X}|\{b\}$ be an i.i.d. sample for $X|\{b\} \sim \mathrm{U}(0,b)$ and $T(\boldsymbol{X}|\{b\}) = \max_{1\leq i\leq n} X_i|\{b\}$. With*

$$q_b(t) = \frac{1}{\vartheta^n}\mathbb{1}_{[0,b]}(t)\ ,\ h(\boldsymbol{x}) = \mathbb{1}_{[0,+\infty)^n}(\boldsymbol{x}) \tag{4.82}$$

we may factorise $f_{\boldsymbol{X}|\{b\}}(\boldsymbol{x})$ according to proposition 4.32 into

$$f_{\boldsymbol{X}|\{b\}}(\boldsymbol{x}) = q_b(T(\boldsymbol{x}))h(\boldsymbol{x}) = \frac{1}{b^n}\mathbb{1}_{[0,b]}(\max_{1\leq i\leq n} x_i)\mathbb{1}_{[0,+\infty)^n}(\boldsymbol{x})\,. \tag{4.83}$$

This shows the sufficiency of $T(\boldsymbol{X}|\{b\})$.

Definition 4.34 (minimal sufficient statistic) *Let $T(\boldsymbol{X}|\{\vartheta\})$ be a sufficient statistic for ϑ. Then $T(\boldsymbol{X}|\{\vartheta\})$ is called a minimal sufficient statistic if $T(\boldsymbol{X}|\{\vartheta\})$ is a function of any other sufficient statistic $U(\boldsymbol{X}|\{\vartheta\})$ for ϑ.*

Proposition 4.35 (minimal sufficient statistic) *Let $f_{\boldsymbol{X}|\{\vartheta\}}$ be the density function of $\boldsymbol{X}|\{\vartheta\}$ and $T(\boldsymbol{X}|\{\vartheta\})$ be a statistic. Then $T(\boldsymbol{X}|\{\vartheta\})$ is a minimal sufficient statistic for ϑ if*

$$\frac{f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{x})}{f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{y})} = c(\boldsymbol{x},\boldsymbol{y}) \ \Leftrightarrow\ T(\boldsymbol{x}) = T(\boldsymbol{y}) \ \text{ for all } \ \boldsymbol{x},\boldsymbol{y} \in \mathcal{X}^n\,. \tag{4.84}$$

Example 4.36 *In continuation of example 4.33 we now show, that $T(\boldsymbol{X}|\{b\}) = \max_{1\leq i\leq n} X_i|\{b\}$ is also a minimal sufficient statistic for b. For $\boldsymbol{x},\boldsymbol{y} \in \mathcal{X}^n$ the ratio in proposition 4.35 is*

$$\frac{f_{\boldsymbol{X}|\{b\}}(\boldsymbol{x})}{f_{\boldsymbol{X}|\{b\}}(\boldsymbol{y})} = \frac{\frac{1}{b^n}\mathbb{1}_{[0,b]}(\max_{1\leq i\leq n} x_i)\mathbb{1}_{[0,+\infty)^n}(\boldsymbol{x})}{\frac{1}{b^n}\mathbb{1}_{[0,b]}(\max_{1\leq i\leq n} y_i)\mathbb{1}_{[0,+\infty)^n}(\boldsymbol{y})}\,. \tag{4.85}$$

If $\max_{1\leq i\leq n} x_i = \max_{1\leq i\leq n} y_i$ the equality of nominator and denominator is obvious, thus, the ratio equals 1, that is, it is constant. Conversely, if the ratio is not a function of b, then $\mathbb{1}_{[0,b]}(\max_{1\leq i\leq n} x_i) = \mathbb{1}_{[0,b]}(\max_{1\leq i\leq n} y_i)$ needs to hold. Since, $\mathbb{1}_{[0,b]}(z) = \mathbb{1}_{[z,+\infty)}(b)$, the equality $\mathbb{1}_{[\max_{1\leq i\leq n} x_i,+\infty)}(b) = \mathbb{1}_{[\max_{1\leq i\leq n} y_i,+\infty)}(b)$ only holds, if $\max_{1\leq i\leq n} x_i = \max_{1\leq i\leq n} y_i$. Together, it follows that $T(\boldsymbol{X}|\{b\}) = \max_{1\leq i\leq n} X_i|\{b\}$ is minimal sufficient.

The Rao-Blackwell theorem states how to find an uniformly better unbiased estimator when an unbiased estimator and a sufficient statistic are already known:

Proposition 4.37 (Rao-Blackwell) *Let $T(\boldsymbol{X}|\{\vartheta\})$ be a sufficient statistic for ϑ, $V(\boldsymbol{X}|\{\vartheta\})$ be an unbiased estimator of $\eta = \eta(\vartheta) \in \mathbb{R}$. Then $\mathrm{E}[V|T]$ is an unbiased estimator of η and is uniformly better than V.*

Definition 4.38 (complete statistic) *Let $T(\boldsymbol{X}|\{\vartheta\})$ be a statistic. Then $T(\boldsymbol{X}|\{\vartheta\})$ is called a complete statistic for ϑ if for any function r and all $\vartheta \in \Theta$*

$$\mathrm{E}[r(T(\boldsymbol{X}|\{\vartheta\}))] = 0 \;\Rightarrow\; \mathrm{P}_{r(T(\boldsymbol{X}|\{\vartheta\}))}(\{0\}) = 1\,. \tag{4.86}$$

Example 4.39 *In continuation of example 4.36 we will show, that $T(\boldsymbol{X}|\{b\}) = \max_{1\le i\le n} X_i|\{b\}$ is also a complete statistic for b. For this, let r be a function with $\mathrm{E}[r(T(\boldsymbol{X}|\{b\}))] = 0$ for all $b > 0$. Since the density of $T(\boldsymbol{X}|\{b\}) = \max_{1\le i\le n} X_i|\{b\}$ is $\frac{n}{b^n}t^{n-1}\mathbb{1}_{[0,b]}(t)$ we have*

$$\mathrm{E}[r(T(\boldsymbol{X}|\{b\}))] = \frac{n}{b^n}\int_0^b r(t)t^{n-1}\,\mathrm{d}t = 0 \quad \text{for all} \quad b > 0\,. \tag{4.87}$$

Defining $R(b) = \int_0^b r(t)t^{n-1}\,\mathrm{d}t$ and differentiating $R(b)$ w.r.t. b we get almost everywhere

$$\frac{\mathrm{d}}{\mathrm{d}b}R(b) = r(b)b^{n-1}\,. \tag{4.88}$$

Since $R(b) = 0$ for all $b > 0$, we conclude $r(b)b^{n-1} = 0$ almost everywhere. Therefore, $r(t) = 0$ almost everywhere, that is, $\mathrm{P}_{r(T(\boldsymbol{X}|\{b\}))}(\{0\}) = 1$ for all $b > 0$. And this shows the completeness of $T(\boldsymbol{X}|\{b\}) = \max_{1\le i\le n} X_i|\{b\}$ for b.

Finally, the Lehmann-Scheffé theorem states how an UMVUE may be derived from a sufficient and complete statistic:

Proposition 4.40 (Lehmann-Scheffé) *Let $T(\boldsymbol{X}|\{\vartheta\})$ be a sufficient and complete statistic for ϑ, $V(\boldsymbol{X}|\{\vartheta\}) = r(T(\boldsymbol{X}|\{\vartheta\}))$ be an unbiased estimator of $\eta = \eta(\vartheta) \in \mathbb{R}$. Then $V(\boldsymbol{X}|\{\vartheta\})$ is a uniformly minimum variance unbiased estimators (UMVUE) of η.*

Example 4.41 *Continuation of example 4.39.* $T(\boldsymbol{X}|\{b\}) = \max_{1\leq i\leq n} X_i|\{b\}$ *is a sufficient and complete statistic for* b*. In example 4.24 it is shown, that*

$$\mathrm{E}[\max_{1\leq i\leq n} X_i|\{b\}] = \frac{n}{n+1}b \tag{4.89}$$

Therefore,

$$V(\boldsymbol{X}|\{b\}) = r(T(\boldsymbol{X}|\{b\})) = \frac{n+1}{n}\max_{1\leq i\leq n} X_i \tag{4.90}$$

is an unbiased estimator of b *and from proposition 4.40 it follows, that it is the UMVUE.*

4.4. Set Estimation

A very general definition for *set estimators* may be given similar to that of point estimators 4.4[8]:

Definition 4.42 (set estimator, set estimate) *A set estimator of the value* $\vartheta \in \mathbb{R}^q$ *of a distribution parameter* D *is a function* $C(\boldsymbol{X}|\{\vartheta\})$ *of a sample* $\boldsymbol{X}|\{\vartheta\}$ *with* $C(\boldsymbol{X}|\{\vartheta\}) \in \mathcal{B}$*, where* $\mathcal{B}$ *is the Borel* σ*-algebra over* $\mathbb{R}^q$*. The random set* $C(\boldsymbol{X}|\{\vartheta\})$ *is called set estimator of* ϑ*. If* $\boldsymbol{x}$ *is observed, the set* $C(\boldsymbol{x})$ *is called set estimate of* ϑ*.*

Note, that no constraints are made to the set estimators for example with respect to their geometric design. Actually, set estimators are often defined as rectangular sets and, in the case of $\vartheta \in \mathbb{R}$, as intervals, respectively. For the latter we find a definition in [13], p. 417:

Definition 4.43 (interval estimator, interval estimate) *An interval estimate of* $\vartheta \in \mathbb{R}$ *is any pair of functions,* $L(\boldsymbol{x})$ *and* $U(\boldsymbol{x})$*, of a sample that satisfy* $L(\boldsymbol{x}) \leq U(\boldsymbol{x})$ *for all* $\boldsymbol{x} \in \mathcal{X}^n$*. If* $\boldsymbol{x}$ *is observed, than the interval*

$$C(\boldsymbol{x}) = \{\vartheta \,|\, L(\boldsymbol{x}) \leq \vartheta \leq U(\boldsymbol{x})\} = [L(\boldsymbol{x}), U(\boldsymbol{x})] \tag{4.91}$$

[8] We follow the definition of interval estimators given by [13]

is called interval estimate of ϑ. The random interval

$$C(\boldsymbol{X}|\{\vartheta\}) = [L(\boldsymbol{X}|\{\vartheta\}), U(\boldsymbol{X}|\{\vartheta\})] \tag{4.92}$$

is called interval estimator of ϑ.

The terms *coverage probability*, *confidence coefficient*, *confidence level* or *reliability* are used as quality criterions for set estimators. The coverage probability is defined in [13], p. 418, as follows:

Definition 4.44 (coverage probability) *For a set estimator $C(\boldsymbol{X}|\{\vartheta\})$ of ϑ, the coverage probability of $C(\boldsymbol{X}|\{\vartheta\})$ is defined by*

$$\mathrm{P}_{\boldsymbol{X}|\{\vartheta\}}(\vartheta \in C(\boldsymbol{X}|\{\vartheta\}))\,, \tag{4.93}$$

that is the probability that the random set estimator $C(\boldsymbol{X}|\{\vartheta\})$ covers the true value ϑ of the distribution parameter D.

We note, that the coverage probability is a function of ϑ, i.e., it varies with ϑ.

Definition 4.45 (confidence level) *For a set estimator $C(\boldsymbol{X}|\{\vartheta\})$ of ϑ, the infimum of the coverage probability over all $\vartheta \in \Theta$ defines the confidence level β:*

$$\beta = \inf_{\vartheta \in \Theta} \mathrm{P}_{\boldsymbol{X}|\{\vartheta\}}(\vartheta \in C(\boldsymbol{X}|\{\vartheta\})) \tag{4.94}$$

A set estimator with a specified confidence level β is then called *confidence set estimator* or in the one-dimensional case *confidence interval estimator*.

Example 4.46 *Let $\boldsymbol{X}|\{b\}$ be an i.i.d. sample for $X|\{b\} \sim \mathrm{U}(0, b)$. The MLE of b is $\widehat{b}(\boldsymbol{X}) = \max_{1 \le i \le n} X_i|\{b\}$. It is reasonable, that a confidence interval estimator $C(\boldsymbol{X}|\{b\})$ of b contains the point estimator $\widehat{b}(\boldsymbol{X})$ and, since the true value b is at least as large as $\widehat{b}(\boldsymbol{X})$, it should mark the left bound of $C(\boldsymbol{X}|\{b\})$,*

e.g. the interval $\left[\widehat{b}(\boldsymbol{X}), a\widehat{b}(\boldsymbol{X})\right]$ *with* $a > 1$ *some constant represents a confidence interval estimator. The coverage probability then is calculated to be*

$$\begin{aligned}
\mathrm{P}_{\boldsymbol{X}|\{b\}}\left(b \in C(\boldsymbol{X}|\{b\})\right) &= \mathrm{P}_{\boldsymbol{X}|\{b\}}\left(b \in \left[\widehat{b}(\boldsymbol{X}), a\widehat{b}(\boldsymbol{X})\right]\right) \\
&= \mathrm{P}_{\boldsymbol{X}|\{b\}}\left(\widehat{b}(\boldsymbol{X}) \leq b \leq a\widehat{b}(\boldsymbol{X})\right) \\
&= \mathrm{P}_{\boldsymbol{X}|\{b\}}\left(\frac{1}{a}b \leq \widehat{b}(\boldsymbol{X}) \leq b\right) \\
&= 1 - \mathrm{P}_{\boldsymbol{X}|\{b\}}\left(\widehat{b}(\boldsymbol{X}) \leq \frac{1}{a}b\right) \\
&= 1 - \left(\frac{\frac{1}{a}b}{b}\right)^n = 1 - \left(\frac{1}{a}\right)^n
\end{aligned} \tag{4.95}$$

Remarkably is that $\mathrm{P}_{\boldsymbol{X}|\{b\}}(b \in C(\boldsymbol{X}|\{b\}))$ *does not depend on* b*, but is constant for any fixed sample size* n*, i.e., the coverage probability already equals the confidence level. Given a certain value* $\beta \in (0,1)$ *as confidence level, it is then possible to solve* (4.95) *for* a*:*

$$a_\beta = \left(\sqrt[n]{1-\beta}\right)^{-1} \tag{4.96}$$

4.4.1. Approximative Confidence Interval Estimator

If the sample size n is sufficiently large, the central limit theorem provides a way to derive an approximative confidence interval estimator. The following proposition states in a very general way the asymptotic of the distribution of the empirical moments $M_{n,j}$ as defined in 4.5.

Proposition 4.47 (asymptotic of empirical moments) *Let* $\boldsymbol{X}$ *be an i.i.d. sample for* X *with* $\mu_j = \mathrm{E}[X^j]$*,* $j = 1, 2, \ldots$*, and* $t_1, t_2, \ldots, t_s \in \mathbb{N}$*. Let*

$$\boldsymbol{M}_n = (M_{n,t_1}, M_{n,t_2}, \ldots, M_{n,t_s})^\top \quad \text{and} \quad \boldsymbol{\mu} = (\mu_{t_1}, \mu_{t_2}, \ldots, \mu_{t_s})^\top . \tag{4.97}$$

Then $\sqrt{n}(\boldsymbol{M}_n - \boldsymbol{\mu})$ *converges in distribution to a random variable with the* s*-dimensional normal distribution* $N(\mathbf{0}, \boldsymbol{\Xi})$ *with covariance matrix* $\boldsymbol{\Xi}$*, for which holds*

$$\boldsymbol{\Xi} = \left(\mu_{t_i+t_j} - \mu_{t_i}\mu_{t_j}\right)_{1 \leq i,j \leq s} = n\mathrm{Cov}[\boldsymbol{M}_n] . \tag{4.98}$$

Applying the delta method gives the asymptotic distribution of moment estimators. We give this method for the case of estimating an one-dimensional distribution parameter:

Proposition 4.48 (asymptotic of moment estimator) *Let $\boldsymbol{M}_n$ and $\boldsymbol{\mu}$ as defined in* (4.97)*, let $g(\boldsymbol{M}_n)$ be a moment estimator as defined in 4.5 of the value $\vartheta = g(\boldsymbol{\mu})$ of a distribution parameter D with $g : \mathbb{R}^s \to \mathbb{R}$ differentiable in $\boldsymbol{\mu}$. Then $\sqrt{n}(g(\boldsymbol{M}_n) - g(\boldsymbol{\mu}))$ converges in distribution to a random variable with normal distribution $N(0, \mathrm{grad}g(\boldsymbol{\mu})\boldsymbol{\Xi}(\mathrm{grad}g(\boldsymbol{\mu}))^\top)$, where* grad$g$ *is the gradient of g.*

Now, for a large sample size n, the distribution of $\sqrt{n}(g(\boldsymbol{M}_n) - g(\boldsymbol{\mu}))$ is approximately the normal distribution $N(0, \mathrm{grad}g(\boldsymbol{\mu})\boldsymbol{\Xi}(\mathrm{grad}g(\boldsymbol{\mu}))^\top)$. Transformation to the standard normal distribution and utilising the empirical moments as moment estimators of all involved moments μ_i, finally leads to a construction of an approximative confidence interval estimator. We illustrate this with two examples:

Example 4.49 *Let $s = 1$ and $t = t_1 = 1$, then $\sqrt{n}(M_{n,1} - \mu_1)$ converges in distribution to a random variable Y with $Y \sim N(0, \mu_2 - \mu_1^2)$. Since $M_{n,1} = \overline{X}$ and $\mu_2 - \mu_1^2 = \sigma^2$, this is just a variation of the central limit theoreme as it may be found in many introductiory books on statistics. We are interested in an approximative confidence interval estimator of μ_1. For n large we have approximately*

$$\frac{\sqrt{n}(\overline{X} - \mu_1)}{\sigma} \sim N(0, 1)\,. \tag{4.99}$$

With $z(p)$ the p-quantile of the standard normal distribution we have

$$z(\alpha_1) \leq \frac{\sqrt{n}(\overline{X} - \mu_1)}{\sigma} \leq z(1 - \alpha_2)\,, \tag{4.100}$$

where the coverage probability is approximated by $1 - (\alpha_1 + \alpha_2) = \beta$. Solving the inequality for μ_1 yields the approximative confidence interval estimator of μ_1

$$C_{\alpha_1,\alpha_2}(\boldsymbol{X}) = \left\{ \mu_1 \;\middle|\; \overline{X} - z(1 - \alpha_2)\frac{\sigma}{\sqrt{n}} \leq \mu_1 \leq \overline{X} - z(\alpha_1)\frac{\sigma}{\sqrt{n}} \right\}. \tag{4.101}$$

If σ^2 is not known, it is common usage to substitute σ by

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}\,. \tag{4.102}$$

In the case that the sample elements are all normal distributed, it is commonly known that the ratio

$$\frac{\sqrt{n}(\overline{X}-\mu_1)}{S} \tag{4.103}$$

has Student's t-distribution with $n-1$ degrees of freedom. Then the p-quantiles $t_{n-1}(p)$ of the t-distribution have to be used instead of the normal quantiles $z(p)$. The resulting confidence interval estimator then is not an approximation, but is exact with confidence level $1-(\alpha_1+\alpha_2)$. Applying this also in the approximative case, an approximative confidence interval estimator then is

$$C_{\alpha_1,\alpha_2}(\boldsymbol{X}) = \left\{ \mu_1 \;\middle|\; \overline{X} - t_{n-1}(1-\alpha_2)\frac{S}{\sqrt{n}} \le \mu_1 \le \overline{X} - t_{n-1}(\alpha_1)\frac{S}{\sqrt{n}} \right\} . \tag{4.104}$$

Since both the standard normal and the t-distribution are symmetric with respect to 0, choosing $\alpha_1 = \alpha_2 = \frac{1-\beta}{2}$ yields shortest (approximative) confidence interval estimates.

Example 4.50 *Let $s = 1$ and $t = t_1 = 2$, then $\sqrt{n}(M_{n,2} - \mu_2)$ converges in distribution to a random variable Y with $Y \sim N(0, \mu_4 - \mu_2^2)$. Now, let $\vartheta = \mu_2$, i.e., $g = id$, then for n large we have approximately*

$$\frac{\sqrt{n}(M_{n,2}-\mu_2)}{\sqrt{\mu_4-\mu_2^2}} \sim N(0,1) . \tag{4.105}$$

With $z(p)$ the p-quantile of the standard normal distribution we first have

$$z(\alpha_1) \le \frac{\sqrt{n}(M_{n,2}-\mu_2)}{\sqrt{\mu_4-\mu_2^2}} \le z(1-\alpha_2) , \tag{4.106}$$

which has a probability approximately $1-(\alpha_1+\alpha_2) = \beta$, or equivalent

$$M_{n,2} - z(1-\alpha_2)\frac{\sqrt{\mu_4-\mu_2^2}}{\sqrt{n}} \le \mu_2 \le M_{n,2} - z(\alpha_1)\frac{\sqrt{\mu_4-\mu_2^2}}{\sqrt{n}} . \tag{4.107}$$

Now, substituting μ_4 and μ_2 (of course, only on the left and right side of the inequality) by $M_{n,4}$ and $M_{n,2}$ yields

$$\begin{aligned} &C_{\alpha_1,\alpha_2}(\boldsymbol{X}) = \\ &= \left\{ \mu_2 \;\middle|\; M_{n,2} - z(1-\alpha_2)\frac{\sqrt{M_{n,4}-M_{n,2}^2}}{\sqrt{n}} \le \mu_2 \le M_{n,2} - z(\alpha_1)\frac{\sqrt{M_{n,4}-M_{n,2}^2}}{\sqrt{n}} \right\} , \end{aligned} \tag{4.108}$$

which is an approximative confidence interval estimator of μ_2. *Symmetry and, thus, minimal length of the interval with respect to the moment estimator* $M_{n,2}$ *may be achieved by choosing* $\alpha_1 = \alpha_2 = \frac{1-\beta}{2}$.

Very analogue arguments are used in the case where a maximum likelihood estimator $\widehat{\vartheta}(\boldsymbol{X})$ is utilised to estimate the value ϑ of D. Then under some regularity conditions ([13], p. 516) we know, that $\sqrt{nI(\vartheta)}(\widehat{\vartheta}(\boldsymbol{X}) - \vartheta)$ converges for $n \to \infty$ to a random variable Y which is standard normal distributed. $I(\vartheta)$ is the Fisher information defined by (4.76). That is, we can procede as in the case of the moment estimators. For n large, $\sqrt{nI(\vartheta)}(\widehat{\vartheta}(\boldsymbol{X}) - \vartheta)$ is approximately standard normal distributed, then the probability for

$$z(\alpha_1) \leq \sqrt{nI(\vartheta)}(\widehat{\vartheta}(\boldsymbol{X}) - \vartheta) \leq z(1-\alpha_2) \tag{4.109}$$

is approximately $1-(\alpha_1+\alpha_2) = \beta$. Solving the inequality for ϑ and substituting ϑ in the Fisher information $I(\vartheta)$ by $\widehat{\vartheta}(\boldsymbol{X})$ yields a confidence interval estimator of ϑ

$$C(\boldsymbol{X}|\{\vartheta\}) = \left[\widehat{\vartheta}(\boldsymbol{X}) - \frac{z(1-\alpha_2)}{\sqrt{nI(\widehat{\vartheta}(\boldsymbol{X}))}}, \widehat{\vartheta}(\boldsymbol{X}) - \frac{z(\alpha_1)}{\sqrt{nI(\widehat{\vartheta}(\boldsymbol{X}))}}\right] \tag{4.110}$$

with an approximate confidence level $1-(\alpha_1+\alpha_2) = \beta$. Choosing $\alpha_1 = \alpha_2 = \frac{1-\beta}{2}$ results into confidence interval estimates with minimal lengths.

Example 4.51 *Let* $\boldsymbol{X}|\{\vartheta\}$ *be an i.i.d. sample for* $X|\{\vartheta\} \sim EXP(\vartheta)$. *From example 4.10 we know the MLE of* ϑ *to be* $\widehat{\vartheta}(\boldsymbol{X}) = \frac{1}{\overline{X}}$. *The Fisher information then is*

$$I(\vartheta) = \frac{n}{\vartheta^2} \tag{4.111}$$

and, thus, the confidence interval estimator of ϑ *with approximate confidence level* β *is given by:*

$$C(\boldsymbol{X}|\{\vartheta\}) = \left[\frac{1}{\overline{X}} - \frac{z((1+\beta)/2)}{n\overline{X}}, \frac{1}{\overline{X}} + \frac{z((1+\beta)/2)}{n\overline{X}}\right] \tag{4.112}$$

4.4.2. Pivotal Variables

Example 4.46 already illustrates, that for some probability distributions it might be not too difficult to construct confidence set estimators, if we are able to find a random variable with a distribution independent of ϑ. In the example the random variable independent of $\vartheta = b$ is $\frac{\widehat{b}(\boldsymbol{X})}{b}$. These variables are called *pivotal variables*:

Definition 4.52 (pivotal variable) *Let Q be a function of the range of variability $\mathcal{X}^n$ of a sample $\boldsymbol{X}|\{\vartheta\}$ and the parameter space Θ into a set $\mathcal{Q}$. Then the random variable $Q(\boldsymbol{X}|\{\vartheta\}, \vartheta)$ is called pivotal variable or pivot if*

$$\mathrm{P}_{\boldsymbol{X}|\{\vartheta\}}\left(Q(\boldsymbol{X}|\{\vartheta\}, \vartheta) \in \boldsymbol{B}\right) = \mathrm{P}(\boldsymbol{B}) \quad \textit{for each} \quad \boldsymbol{B} \subseteq \mathcal{Q}\,, \tag{4.113}$$

that is, the probability of $\boldsymbol{B}$ does not depend on ϑ.

Given a certain confidence level β, then the task is to find a set $\boldsymbol{B}$ with $\mathrm{P}(\boldsymbol{B}) = \beta$, from which the confidence set estimator $C(\boldsymbol{X}|\{\vartheta\})$ of ϑ with confidence level β easily arises through

$$C(\boldsymbol{X}|\{\vartheta\}) = \{\vartheta | Q(\boldsymbol{X}|\{\vartheta\}, \vartheta) \in \boldsymbol{B}\}\,. \tag{4.114}$$

If the pivot $Q(\boldsymbol{X}|\{\vartheta\}, \vartheta)$ is real-valued and has a continuous distribution, say F_Q, a reasonable set B would be an interval. That is, we have to determine two quantiles $c^{(\ell)}$ and $c^{(u)}$ of F_Q, with $c^{(\ell)} < c^{(u)}$ for which $F_Q(c^{(u)}) - F_Q(c^{(\ell)}) = \beta$ holds. Defining the quantile function $q(\varrho)$ for $\varrho \in (0, 1)$ we get a confidence interval estimator of ϑ with confidence level β for each $t \in (0, 1)$:

$$C_\beta(\boldsymbol{X}|\{\vartheta\}) = \{\vartheta \,|\, q\left((1-t)(1-\beta)\right) \leq Q(\boldsymbol{X}|\{\vartheta\}, \vartheta) \leq q\left(1 - t(1-\beta)\right)\} \tag{4.115}$$

The question, in which cases it is possible to find a pivotal variable is answered for example in [13] where it is shown, that it is always possible to determine pivots for location-scale families of distributions. On page 427 of [13] we find table 4.1 with the different cases of the corresponding pivotal quantities.

The prime example for the location-scale case is the normal distribution $N(\mu, \sigma^2)$, where μ is called the location parameter and σ is called the scale parameter. The most interesting case is, when both μ and $\sigma = \sqrt{\sigma^2}$ should be estimated:

Form of pdf	Type of pdf	Pivotal quantity
$f(x-\mu)$	Location	$\overline{X}-\mu$
$\frac{1}{\sigma}f(\frac{x}{\sigma})$	Scale	$\frac{\overline{X}}{\sigma}$
$\frac{1}{\sigma}f(\frac{x-\mu}{\sigma})$	Location-scale	$\frac{\overline{X}-\mu}{\sigma}$

Table 4.1.: Location-scale pivots

Example 4.53 *Let $\boldsymbol{X}|\{(\mu,\sigma^2)\}$ be an i.i.d. sample for $X|\{(\mu,\sigma^2)\} \sim N(\mu,\sigma^2)$ with μ and σ^2 unknown. Each of the following pivotal variables alone lead already to confidence set estimators of (μ,σ):*

- *The random variable*

$$V = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \tag{4.116}$$

has chi-square distribution with $n-1$ degrees of freedom, $V \sim \chi^2_{n-1}$. Then a confidence set estimator $C_{V,\beta}(\boldsymbol{X})$ of (μ,σ^2) with confidence level $1-(\alpha_1+\alpha_2)=\beta$ is derived as follows:

$$C_{V,\beta}(\boldsymbol{X}) = \left\{(\mu,\sigma^2) \;\middle|\; \frac{(n-1)S^2}{\chi^2_{n-1}(1-\alpha_2)} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1}(\alpha_1)}\right\} \tag{4.117}$$

where $\chi^2_{n-1}(p)$ is the p-quantile of the χ^2_{n-1} distribution. We note, that this confidence set estimator $C_{V,\beta}(\boldsymbol{X})$ actually does not yield any information about μ, but only delivers a confidence interval of σ^2. Taking the square root on both sides gives a confidence set estimator $\widetilde{C}_{V,\beta}(\boldsymbol{X})$ of (μ,σ):

$$\widetilde{C}_{V,\beta}(\boldsymbol{X}) = \left\{(\mu,\sigma) \;\middle|\; \sqrt{\frac{(n-1)S^2}{\chi^2_{n-1}(1-\alpha_2)}} \le \sigma \le \sqrt{\frac{(n-1)S^2}{\chi^2_{n-1}(\alpha_1)}}\right\} \tag{4.118}$$

- *The random variable*

$$Z = \frac{\overline{X}-\mu}{\sigma/n} \tag{4.119}$$

has standard normal distribution, $Z \sim N(0,1)$. A confidence set estimator $C_{Z,\beta}(\boldsymbol{X})$ of (μ,σ^2) with confidence level $1-(\alpha_1+\alpha_2)=\beta$ then is

$$C_{Z,\beta}(\boldsymbol{X}) = \left\{(\mu,\sigma^2) \;\middle|\; \overline{X} - z(1-\alpha_2)\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} - z(\alpha_1)\frac{\sigma}{\sqrt{n}}\right\} \tag{4.120}$$

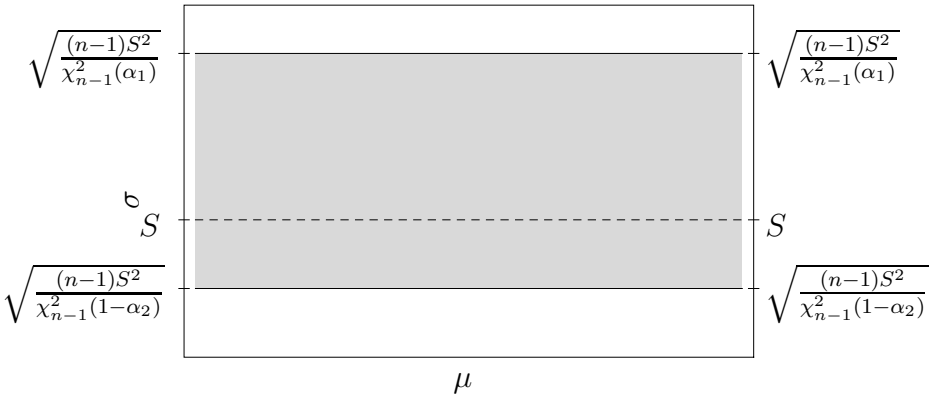


Figure 4.2.: Confidence set estimator $\widetilde{C}_{V,\beta}(\boldsymbol{X})$ of (μ, σ) based on the pivotal variable $V = \frac{(n-1)S^2}{\sigma^2}$.

where $z(p)$ is the p-quantile of the $N(0,1)$-distribution. Here we note, that $C_{Z,\beta}(\boldsymbol{X})$ does yield some information about both parameters.

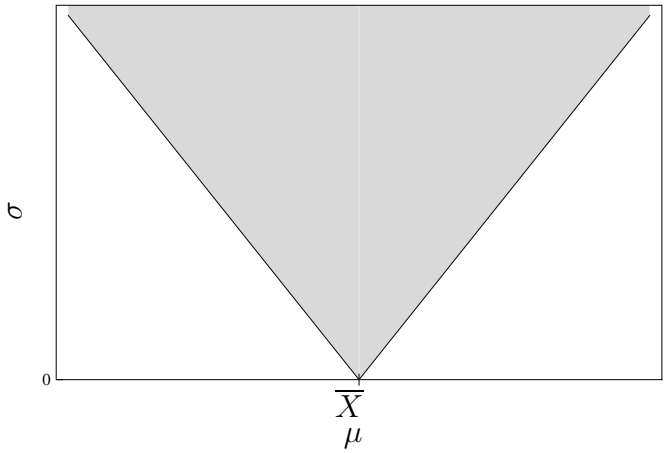


Figure 4.3.: Confidence set estimator $C_{Z,\beta}(\boldsymbol{X})$ of (μ, σ) based on the pivotal variable $Z = \frac{\overline{X}-\mu}{\sigma/n}$.

- *The random variable*

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \tag{4.121}$$

has Student's t-distribution with $n-1$ degrees of freedom, $T \sim t_{n-1}$. This yields as a confidence set estimator $C_{T,\beta}(\boldsymbol{X})$ of (μ, σ) with confidence level

$$1-(\alpha_1+\alpha_2)=\beta$$

$$C_{T,\beta}(\boldsymbol{X})=\left\{(\mu,\sigma^2)\;\middle|\;\overline{X}-t_{n-1}(1-\alpha_2)\frac{S}{\sqrt{n}}\leq\mu\leq\overline{X}-t_{n-1}(\alpha_1)\frac{S}{\sqrt{n}}\right\} \tag{4.122}$$

where $t_{n-1}(p)$ is the p-quantile of the t_{n-1} distribution. Now, $C_{T,\beta}(\boldsymbol{X})$ does not give any information about σ, but only gives a confidence interval estimator of μ.

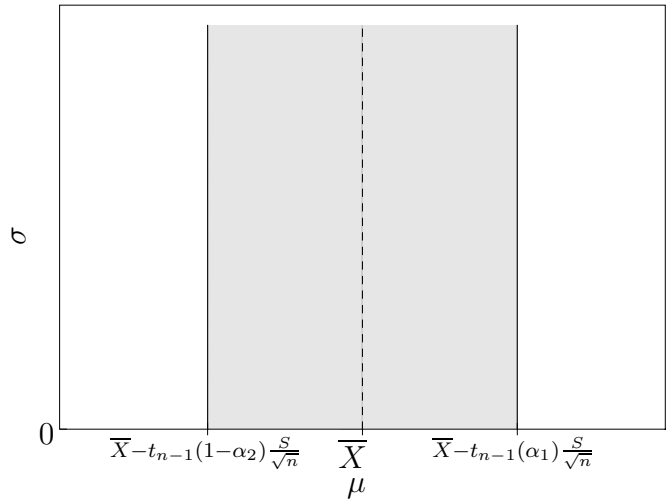


Figure 4.4.: Confidence set estimator $C_{T,\beta}(\boldsymbol{X})$ of (μ,σ) based on the pivotal variable $T=\frac{\overline{X}-\mu}{S/\sqrt{n}}$.

Intersection of $C_{Z,\beta_1}(\boldsymbol{X})$ and $\widetilde{C}_{V,\beta_2}(\boldsymbol{X})$ yields a confidence set estimator $C_{(Z,V),\beta}(\boldsymbol{X})$ with trapeziodal shape. Since $\overline{X}$ and S^2 are independent for normal i.i.d. samples, we get a confidence level β of $C_{(Z,V),\beta}(\boldsymbol{X})$ with $\beta=\beta_1\cdot\beta_2$.
Intersection of $C_{T,\beta_1}(\boldsymbol{X})$ and $\widetilde{C}_{V,\beta_2}(\boldsymbol{X})$ yields a confidence set estimator $C_{(T,V),\beta}(\boldsymbol{X})$ with rectangular shape. Now, using Bonferroni's rule we get for the confidence level β of $C_{(T,V),\beta}(\boldsymbol{X})$ only a lower bound $\beta\geq 1-(1-\beta_1)(1-\beta_2)$.

The exponential distribution yields another prominent example.

Example 4.54 *Let $\boldsymbol{X}|\{\vartheta\}$ be an i.i.d. sample for $X|\{\vartheta\}\sim EXP(\vartheta)$. Then with the sufficient statistic $T(\boldsymbol{X}|\{\vartheta\})=\sum_{i=1}^n X_i$ we define the statistic*

$$Q(\boldsymbol{X}|\{\vartheta\},\vartheta)=2T(\boldsymbol{X}|\{\vartheta\})\vartheta\,. \tag{4.123}$$

$Q(\boldsymbol{X}|\{\vartheta\},\vartheta)$ has chi-square distribution with $2n$ degrees of freedom, $Q(\boldsymbol{X}|\{\vartheta\},\vartheta) \sim \chi^2_{2n}$, and, thus, is a pivotal variable. Now, given α_1 and α_2 with $1-(\alpha_1+\alpha_2)=\beta$ and $\chi^2_{2n}(p)$ the p-quantile of the chi-square distribution as in example 4.53 we have

$$\mathrm{P}\left(\chi^2_{2n}(\alpha_1) \leq Q(\boldsymbol{X}|\{\vartheta\},\vartheta) \leq \chi^2_{2n}(1-\alpha_2)\right) = \beta\,. \tag{4.124}$$

Solving the inequality in Q for ϑ yields a confidence interval estimator of ϑ with confidence level $1-(\alpha_1+\alpha_2)=\beta$:

$$C_{T,\beta}(\boldsymbol{X}|\{\vartheta\}) = \left\{ \vartheta \;\middle|\; \frac{\chi^2_{2n}(\alpha_1)}{2T(\boldsymbol{X}|\{\vartheta\})} \leq \vartheta \leq \frac{\chi^2_{2n}(1-\alpha_2)}{2T(\boldsymbol{X}|\{\vartheta\})} \right\} \tag{4.125}$$

In both examples 4.53 and 4.54 the resulting confidence set estimators with confidence level β are not unique, since the involved quantiles, or better, the quantiles' orders α_1 and α_2 are almost arbitrary, as long as $1-(\alpha_1+\alpha_2)=\beta$ holds. Since the volume of a confidence set estimate may be understood as the accuracy of the estimation, α_1 and α_2 should be choosen in such way, that the accuracy becomes maximal, i.e., the volume minimal. For confidence interval estimates that means that we are seeking minimal lengths. If the distribution of the pivotal variable has a symmetric density with respect to 0, the solution $\alpha_1 = \alpha_2 = \frac{1-\beta}{2}$ follows immediately. We also note, that in this case the symmetry leads to equation $q(\varrho) = -q(1-\varrho)$ for the quantiles. Therefore, in example 4.53 we get confidence interval estimators yielding estimates with minimal lengths as follows:

$$C_{Z,\beta}(\boldsymbol{X}) = \left\{ (\mu,\sigma) \;\middle|\; \overline{X} - z(\frac{1+\beta}{2})\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z(\frac{1+\beta}{2})\frac{\sigma}{\sqrt{n}} \right\}, \tag{4.126}$$

$$C_{T,\beta}(\boldsymbol{X}) = \left\{ (\mu,\sigma) \;\middle|\; \overline{X} - t_{n-1}(\frac{1+\beta}{2})\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{n-1}(\frac{1+\beta}{2})\frac{S}{\sqrt{n}} \right\}. \tag{4.127}$$

Since the chi-square distribution is not symmetric, the task is to find optimal α_1 and α_2 for $C_{V,\beta}$ in example 4.53 and for $C_{T,\beta}$ in example 4.54:

- For minimising the length of the confidence interval estimator $C_{V,\beta}$ in example 4.53, we have to minimise

$$\frac{1}{\chi^2_{n-1}(\alpha_1)} - \frac{1}{\chi^2_{n-1}(1-\alpha_2)}\,. \tag{4.128}$$

E.g. for $n = 21$ and $\beta = 0.9$ we would get minimal length for $\alpha_1 = 0.085874$ and $\alpha_2 = 0.014126$.

- For minimising the length of the confidence interval estimator $C_{T,\beta}$ in example 4.54, we have to minimise

$$\chi^2_{2n}(1-\alpha_2) - \chi^2_{2n}(\alpha_1)\,. \tag{4.129}$$

E.g. for $n = 10$ and $\beta = 0.9$ we would get minimal length for $\alpha_1 = 0.028110$ and $\alpha_2 = 0.071890$.

4.4.3. Inverting Acceptance Regions

The following method to construct confidence set estimators was originally developed J. Neyman [57]. It is based on so-called acceptance regions A, i.e., subsets of the variability space $\mathcal{X}^n$. For every possible value ϑ of the distribution parameter D an acceptance region $A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)$ is constructed, which has at least a probability of β, i.e., $\mathrm{P}_{\boldsymbol{X}|\{\vartheta\}}\left(\boldsymbol{X}|\{\vartheta\} \in A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)\right) \geq \beta$. In fact, these acceptance regions may also be called predictions. The following theorem states the relation between acceptance regions and confidence set estimates.

Theorem 4.55 (duality of acceptance regions and set estimates) *For every $\vartheta \in \Theta$, let $A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)$ be an acceptance region with probability of at least β. For every realisation $\boldsymbol{x}$ of $\boldsymbol{X}|\{\vartheta\}$ define the set $C(\boldsymbol{x})$ by*

$$C(\boldsymbol{x}) = \{\vartheta \in \Theta \,|\, \boldsymbol{x} \in A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)\}\,. \tag{4.130}$$

Then $C(\boldsymbol{X}|\{\vartheta\})$ is a confidence set estimator of ϑ with confidence level β.

Since $\beta \leq \mathrm{P}_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{X}|\{\vartheta\} \in A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)) = \mathrm{P}_{\boldsymbol{X}|\{\vartheta\}}(\vartheta \in C(\boldsymbol{X}|\{\vartheta\}))$ the proof is already achieved.

Although, the idea appears to be quite simple, the major problems arise in the construction of the acceptance regions. We present two ways of construction: based on highest probability regions and based on the likelihood ratio test statistic.

4.4.3.1. Highest Probability Region

As we have discussed subsequent to examples 4.53 and 4.54, the volume of a confidence set estimate may be interpreted as its accuracy. In this sense, we might get the idea to construct acceptance regions with lowest volume. For discrete distributions that would be achieved for each $\vartheta \in \Theta$ by selecting those elements $\boldsymbol{x}$ of the range of variability $\mathcal{X}^n$ of $\boldsymbol{X}|\{\vartheta\}$ with highest probability till some given probability is reached. For continuous distributions we would select those elements $\boldsymbol{x} \in \mathcal{X}^n$ with highest value of the density function $f_{\boldsymbol{X}|\{\vartheta\}}$. Thus, in the *continuous case* we define the acceptance regions by the **largest** r so that

$$A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta) = \{\boldsymbol{x} \,|\, f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{x}) \geq r\} \tag{4.131}$$

with

$$\int_{A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)} f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \geq \beta \,. \tag{4.132}$$

Therefore, those regions are named *highest probability regions* (HPR).
Acceptance regions constructed for the original random sample vector $\boldsymbol{X}|\{\vartheta\} \in \mathbb{R}^n$ are n-dimensional as well. It is comprehensible, that is not very manageable in many cases. That is, one may wish to construct acceptance regions for a real-valued (one-dimensional) function $T|\{\vartheta\} = T(\boldsymbol{X}|\{\vartheta\})$ of $\boldsymbol{X}|\{\vartheta\}$, i.e., $T\colon \mathcal{X}^n \to \mathcal{T} \subseteq \mathbb{R}$. Then equations (4.131) and (4.132) become

$$A_{T|\{\vartheta\}}(\vartheta) = \{t \,|\, f_{T|\{\vartheta\}}(t) \geq r\} \tag{4.133}$$

with

$$\int_{A_{T|\{\vartheta\}}(\vartheta)} f_{T|\{\vartheta\}}(t) \, \mathrm{d}t \geq \beta \,. \tag{4.134}$$

Obviously, this is often much more easier to treat, but we have to note that the resulting acceptance regions may not be equivalent to the original ones.

Remark: In general, the inequalities (4.132) and (4.134) cannot be refined to equalities. For a discrete distribution this is quite obviuous: by the selection

criterion (4.131), the probability of the acceptance region is regularly larger than the demanded β. But even for continuous distributions it may happen, that the density function is constant for a whole set of values $\boldsymbol{x}$ which has non-zero probability. If those values have to be included in the acceptance region, i.e., the whole set is a subset of the acceptance region, the demanded probability may be exceeded. The next example illustrates this case.

Example 4.56 *Let $\boldsymbol{X}|\{b\}$ be an i.i.d. sample for $X|\{b\} \sim \mathrm{U}(0,b)$ with $b > 0$. The joint density of $\boldsymbol{X}|\{b\}$ is*

$$f_{\boldsymbol{X}|\{b\}}(\boldsymbol{x}) = \frac{1}{b^n} \mathbb{1}_{[0,b]^n}(\boldsymbol{x})\,. \tag{4.135}$$

That is, for all $\boldsymbol{x} \in \mathcal{X}^n = [0,b]^n$ the value of the density function is constant with $\frac{1}{b^n}$. Thus, the selection criterion in equation (4.131) yields for the largest r the value $\frac{1}{b^n}$ and for the acceptance region the whole range of variability, i.e., $A_{\boldsymbol{X}|\{b\}}(b) = \mathcal{X}^n = [0,b]^n$ for each b.

This defect may be avoided by requiring some minimality property of the acceptance regions. Thus, we add to equations (4.131) and (4.132) an additional requirement:

$$A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta) = \{\boldsymbol{x} \,|\, f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{x}) \geq r\} \tag{4.136}$$

with

$$\int_{A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)} f_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \geq \beta \tag{4.137}$$

and

$$A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta) \;\text{ minimal size} \tag{4.138}$$

If the density function $f_{T|\{\vartheta\}}$ of a real-valued random variable $T|\{\vartheta\}$ is unimodal, we find in [13] (p. 441), that the resulting acceptance region is in fact an interval, which has minimal length:

Proposition 4.57 (unimodal PDF, minimal interval) *Let f be an unimodal density function. If the interval $[\ell, u]$ satisfies*

(i) $\int_\ell^u f(x)\,\mathrm{d}x = \beta$,

(ii) $f(\ell) = f(u) > 0$, *and*

(iii) $\ell \le x^* \le u$, *where* x^* *is a mode of* $f(x)$,

then $[\ell, u]$ *is the shortest among all intervals that satisfy (i).*

Example 4.58 *Let* $\boldsymbol{X}|\{\vartheta\} = (X_1|\{\vartheta\}, X_2|\{\vartheta\})$ *be an i.i.d. sample for* $X|\{\vartheta\} \sim EXP(\vartheta)$ *with* $\vartheta \in \Theta = (0, +\infty)$. *The joint density function of* $\boldsymbol{X}|\{\vartheta\}$ *is*

$$\begin{aligned} f_{\boldsymbol{X}|\{\vartheta\}}(x_1, x_2) &= f_{X_1|\{\vartheta\}}(x_1) f_{X_2|\{\vartheta\}}(x_2) \\ &= \vartheta^2 e^{-\vartheta(x_1+x_2)} \mathbb{1}_{[0,+\infty)^2}(x_1, x_2)\,. \end{aligned} \tag{4.139}$$

Firstly, the value of $f_{\boldsymbol{X}|\{\vartheta\}}$ *only depends on the sum* $x_1 + x_2$, *i.e., given two pairs* $(x_1, x_2), (y_1, y_2) \in [0, +\infty)^2$ *with* $x_1 + x_2 = y_1 + y_2$, *then* $f_{\boldsymbol{X}|\{\vartheta\}}(x_1, x_2) = f_{\boldsymbol{X}|\{\vartheta\}}(y_1, y_2)$. *Secondly, the smaller* $x_1 + x_2$ *the larger* $f_{\boldsymbol{X}|\{\vartheta\}}(x_1, x_2)$, *with absolute maximum in* $(x_1, x_2) = (0, 0)$. *Thus, the acceptance regions* $A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)$ *are filled with pairs* (x_1, x_2) *with* $x_1 + x_2 = c$ *starting with* $c = 0$, *i.e.,* $(x_1, x_2) = (0, 0)$, *till some* $c^* = c^*(\vartheta)$ *for which holds*

$$\begin{aligned} &\mathrm{P}_{\boldsymbol{X}|\{\vartheta\}}(\boldsymbol{X}|\{\vartheta\} \in A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)) = \\ &\quad = \int_0^{c^*} \int_0^{c^*-t_2} \vartheta^2 e^{-\vartheta(t_1+t_2)}\,\mathrm{d}t_1\,\mathrm{d}t_2 = 1 - e^{-c^*\vartheta}(1 + c^*\vartheta) = \beta\,. \end{aligned} \tag{4.140}$$

Summarised, the acceptance region for ϑ *for some* $\beta \in (0, 1)$ *may be represented as follows*

$$\begin{aligned} &A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta) = \\ &\quad = \left\{(x_1, x_2) \;\middle|\; x_1 \ge 0, x_2 \ge 0, x_1 + x_2 \le c^* \text{ with } 1 - e^{-c^*\vartheta}(1 + c^*\vartheta) = \beta \right\}. \end{aligned} \tag{4.141}$$

We note, that the determination of $c^* = c^*(\vartheta)$ *need not be done for every value* ϑ, *since we may easily derive the equation* $c^*(\vartheta) = \frac{c^*(1)}{\vartheta}$.

Figure 4.5 illustrates the acceptance regions $A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)$ *for* $\vartheta = 1, 2, 3, 4$ *and* $\beta = 0.9$. *Each has the form of an isosceles triangle with an area of*

$$|A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)| = \frac{1}{2}\frac{(c^*(1))^2}{\vartheta^2}\,. \tag{4.142}$$

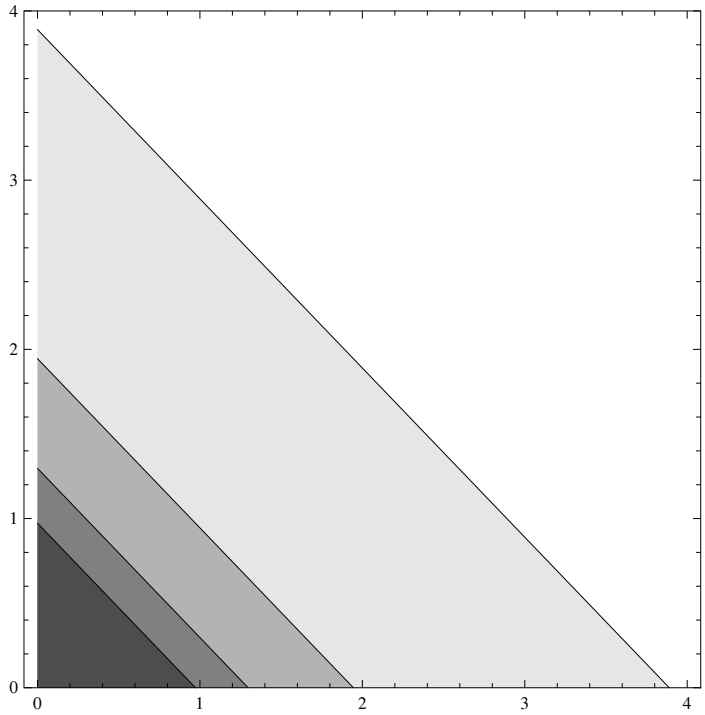


Figure 4.5.: Acceptance regions $A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)$ for $\vartheta = 1, 2, 3, 4$ and $\beta = 0.9$. The largest one corresponds to $\vartheta = 1$, the second largest to $\vartheta = 2$ etc.

We realise, that for increasing ϑ the acceptance regions become smaller and are included in all acceptance regions for smaller ϑ. That is, given one pair (x_1, x_2) which meets $x_1 + x_2 = c^*(\vartheta)$ for some ϑ, it is included in $A_{\boldsymbol{X}|\{\vartheta'\}}(\vartheta')$ for all $\vartheta' \leq \vartheta$, but not in $A_{\boldsymbol{X}|\{\vartheta''\}}(\vartheta'')$ for all $\vartheta'' > \vartheta$.
By inversion of the acceptance regions we get confidence interval estimates as follows

$$C(x_1, x_2) =$$
$$= \left\{ \vartheta \;\middle|\; 0 \leq \vartheta \leq \vartheta_0 \;\; \textit{with} \;\; 1 - e^{-(x_1+x_2)\vartheta_0}(1 + (x_1 + x_2)\vartheta_0) = \beta \right\} . \quad (4.143)$$

In fact, these confidence interval estimates yield only information about the upper bound for the values ϑ of D since the lower bound 0 is the natural lower bound for ϑ. This is not due to the sample size $n = 2$. For general n, the selection criterion tells us to fill the acceptance regions with all values $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in$

$[0,+\infty)^n$ *with* $x_1+x_2+\cdots+x_n \leq c^*(\vartheta)$*, where* $c^*(\vartheta)$ *is determined by*

$$\int_{x_1+x_2+\cdots+x_n\leq c^*} \vartheta^n e^{-\vartheta(x_1+x_2+\cdots+x_n)}\,\mathrm{d}\boldsymbol{x} = \beta \tag{4.144}$$

$$\Leftrightarrow 1-e^{-c^*\vartheta}\left(1+\frac{(c^*\vartheta)^1}{1!}+\frac{(c^*\vartheta)^2}{2!}+\cdots+\frac{(c^*\vartheta)^{n-1}}{(n-1)!}\right)=\beta\,. \tag{4.145}$$

That is, the filling of the acceptance regions always starts at $\boldsymbol{x}=(0,0,\ldots,0)\in\mathbb{R}^n$*. We realise, that the integral in* (4.144) *represents the distribution function* $G_{\vartheta,n}$ *of the gamma distribution* $GAM(\vartheta,n)$ *in* c^**. It is nearby to use this fact to solve for* $c^*(\vartheta)$ *numerically. Finally, the confidence interval estimates are always as follows*

$$C(\boldsymbol{x})=\left\{\vartheta\,\middle|\,0\leq\vartheta\leq\vartheta_0 \;\; \textit{with} \right.$$
$$\left. 1-e^{-\vartheta_0\sum x_i}\left(1+\frac{(\vartheta_0\sum x_i)^1}{1!}+\frac{(\vartheta_0\sum x_i)^2}{2!}+\cdots+\frac{(\vartheta_0\sum x_i)^{n-1}}{(n-1)!}\right)=\beta\right\}. \tag{4.146}$$

And again, the connection to the gamma distribution function, that is, $G_{\vartheta_0,n}(\sum x_i)=\beta$*, should be used to solve numerically for* ϑ_0*.*

According to equation (4.139) the joint density $f_{\boldsymbol{X}|\{\vartheta\}}(x_1,x_2)$ is constant for all (x_1,x_2) with $x_1+x_2=t$. Therefore, one might have the idea to move from the random variable $\boldsymbol{X}|\{\vartheta\}=(X_1|\{\vartheta\},X_2|\{\vartheta\})$ to the random variable $X_1|\{\vartheta\}+X_2|\{\vartheta\}$. We will discuss this in the next example.

Example 4.59 *Equation* (4.139) *in the previous example 4.58 suggests that it might be better to construct the acceptance regions for* ϑ *not based on the joint density* $f_{\boldsymbol{X}|\{\vartheta\}}$ *of* $\boldsymbol{X}|\{\vartheta\}$ *but based on the density* $f_{T|\{\vartheta\}}$ *of the sum* $T|\{\vartheta\}=(X_1+X_2)|\{\vartheta\}=X_1|\{\vartheta\}+X_2|\{\vartheta\}$*. The random variable* $T|\{\vartheta\}$ *is distributed according to the gamma distribution* $GAM(\vartheta,2)$ *with density function given by*

$$f_{T|\{\vartheta\}}(t)=\vartheta^2 t e^{-\vartheta t}\mathbb{1}_{[0,+\infty)}(t)\,. \tag{4.147}$$

It is already known from example 4.54, that by the transformation

$$Q(T|\{\vartheta\},\vartheta)=2T|\{\vartheta\}\vartheta \tag{4.148}$$

we get a pivotal random variable Q which has chi-square distribution with 4 degrees of freedom. Therefore, we have to find only one acceptance interval A_Q for Q for some $\beta \in (0,1)$ which then yields by retransformation the acceptance interval $A_{T|\{\vartheta\}}(\vartheta)$ for each value ϑ of D:

$$A_Q = \left\{q \,|\, \chi^2_4(\alpha_1) \le q \le \chi^2_4(1-\alpha_2)\right\} \quad \textit{with} \quad 1-(\alpha_1+\alpha_2) = \beta \tag{4.149}$$

$$\Longleftrightarrow$$

$$A_{T|\{\vartheta\}}(\vartheta) = \left\{ t \;\middle|\; \frac{\chi^2_4(\alpha_1)}{2\vartheta} \le t \le \frac{\chi^2_4(1-\alpha_2)}{2\vartheta} \right\} \tag{4.150}$$

By inversion we get confidence interval estimates of ϑ with confidence level $\beta = 1-(\alpha_1+\alpha_2)$:

$$C(t) = \left\{ \vartheta \;\middle|\; \frac{\chi^2_4(1-\alpha_2)}{2t} \le \vartheta \le \frac{\chi^2_4(\alpha_1)}{2t} \right\} \tag{4.151}$$

This is the same result as in example 4.54 for $n = 2$. We note, that minimising the lengths of the acceptance intervals, i.e., minimising

$$\chi^2_4(1-\alpha_2) - \chi^2_4(\alpha_1)\,, \tag{4.152}$$

yields also confidence interval estimates of minimal lengths. For example for $\beta = 0.9$ we get $\alpha_1 = 0.003322$ and $\alpha_2 = 0.096678$.

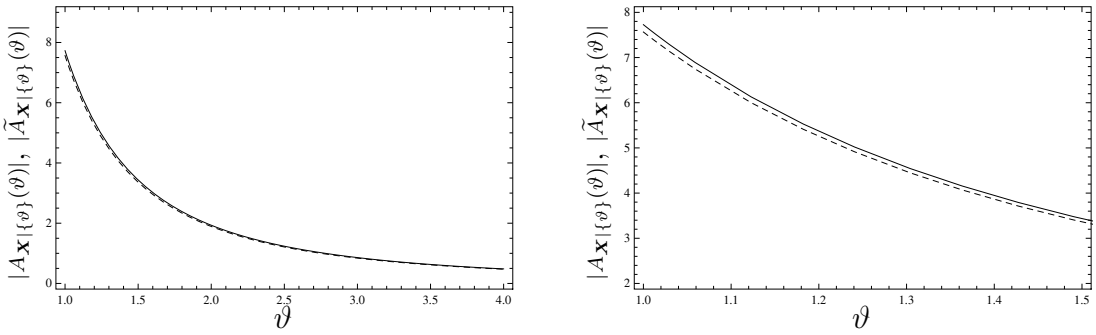
By this change from $\boldsymbol{X}|\{\vartheta\}$ to $T|\{\vartheta\}$ we expect also a change of the area of the acceptance regions in example 4.59 compared to those in example 4.58. The acceptance intervals $A_{T|\{\vartheta\}}(\vartheta)$ from example 4.59 may be transformed to acceptance regions $\widetilde{A}_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)$ similar to those in example 4.58:

$$A_{T|\{\vartheta\}}(\vartheta) = \left\{ t \;\middle|\; \frac{\chi^2_4(\alpha_1)}{2\vartheta} \le t \le \frac{\chi^2_4(1-\alpha_2)}{2\vartheta} \right\} \tag{4.153}$$

$$\Rightarrow\ \widetilde{A}_{\boldsymbol{X}|\{\vartheta\}}(\vartheta) = \left\{ (x_1,x_2) \;\middle|\; x_1 \ge 0, x_2 \ge 0, \frac{\chi^2_4(\alpha_1)}{2\vartheta} \le x_1 + x_2 \le \frac{\chi^2_4(1-\alpha_2)}{2\vartheta} \right\} \tag{4.154}$$

The acceptance regions $\widetilde{A}_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)$ now have the form of an isosceles trapezium with an area of

$$|\widetilde{A}_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)| = \frac{1}{8\vartheta^2}\left((\chi^2_4(1-\alpha_2))^2 - (\chi^2_4(\alpha_1))^2\right) \tag{4.155}$$



(a) $\vartheta \in [1, 4]$ (b) $\vartheta \in [1, 1.5]$

Figure 4.6.: Both areas $|A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)|$ (dashed line) and $|\widetilde{A}_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)|$ (solid line) for $\vartheta \in [1, 4]$. Only by reducing the plotting range to $\vartheta \in [1, 1.5]$ the difference becomes visible.

Figure 4.6 displays both areas $|A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)|$ and $|\widetilde{A}_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)|$ for $\vartheta \in [1, 4]$ and $\beta = 0.9$. We see, that the area of $A_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)$ is always smaller than that of $\widetilde{A}_{\boldsymbol{X}|\{\vartheta\}}(\vartheta)$, but the difference is rather small.

Once the sample value $\boldsymbol{x}$, and thus, $t = T(\boldsymbol{x})$ is observed, it is a bit easier to get the confidence interval estimate $C(t)$ given by equation (4.151), then the confidence interval estimate $C(\boldsymbol{x})$ given by equation (4.146).

The next example illustrates that the minimisation of the lengths of the acceptance intervals does not always result into confidence intervals with minimal lengths.

Example 4.60 *Let $\boldsymbol{X}|\{(\mu, \sigma^2)\}$ be an i.i.d. sample for $X|\{(\mu, \sigma^2)\} \sim N(\mu, \sigma^2)$, where the value μ of the first moment of $X|\{(\mu, \sigma^2)\}$ is known. Thus, w.l.o.g. we may set $\mu = 0$ and write $X|\{\sigma^2\} = X|\{(0, \sigma^2)\}$. The value $\sigma^2 > 0$ of the variance of $X|\{\sigma^2\}$ should be estimated. Without considering the joint distribution of $\boldsymbol{X}|\{\sigma^2\}$, we directly move to the random variable V defined by (cf. example 4.53)*

$$V = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \tag{4.156}$$

which has chi-square distribution with $n - 1$ degrees of freedom. Then the acceptance interval A_V for V for some $\beta \in (0, 1)$ and by retransformation the

acceptance intervals $A_{S^2|\{\sigma^2\}}(\sigma^2)$ *for each value* $\sigma^2 > 0$ *are as follows*

$$A_V = \left\{ v \;\middle|\; \chi^2_{n-1}(\alpha_1) \leq v \leq \chi^2_{n-1}(1-\alpha_2) \right\} \quad \textit{with} \quad 1-(\alpha_1+\alpha_2) = \beta \tag{4.157}$$

$$\Longleftrightarrow$$

$$A_{S^2|\{\sigma^2\}}(\sigma^2) = \left\{ s^2 \;\middle|\; \frac{\chi^2_{n-1}(\alpha_1)}{n-1}\sigma^2 \leq s^2 \leq \frac{\chi^2_{n-1}(1-\alpha_2)}{n-1}\sigma^2 \right\} \tag{4.158}$$

By inversion we get confidence interval estimates of σ^2 *with confidence level* $\beta = 1-(\alpha_1+\alpha_2)$*:*

$$C(s^2) = \left\{ \sigma^2 \;\middle|\; \frac{n-1}{\chi^2_{n-1}(1-\alpha_2)}s^2 \leq \sigma^2 \leq \frac{n-1}{\chi^2_{n-1}(\alpha_1)}s^2 \right\} \tag{4.159}$$

which is the same result as in example 4.53, equation (4.118)*. In contrast to the previous example 4.59 we note, that minimisation of the length of the acceptance intervals does not yield confidence interval estimates of* σ^2 *of minimal lengths: For example for* $n = 21$ *and* $\beta = 0.9$ *minimising the lengths of* $A_{S^2|\{\sigma^2\}}(\sigma^2)$ *yields* $\alpha_1 = 0.028110$ *and* $\alpha_2 = 0.071890$ *and, thus, lengths of* $1.004499\sigma^2$*. Taking these values for the confidence interval estimates* $C(s^2)$ *would yield lengths of* $1.374320s^2$*. But minimising the lengths of the confidence interval estimates* $C(s^2)$ *yields* $\alpha_1 = 0.085874$ *and* $\alpha_2 = 0.014126$ *and, thus, lengths of* $1.108131s^2$*, which are vastly shorter.*

4.4.3.2. Likelihood Ratio Test Statistic

From the theory of hypothesis tests, a very common way is to utilise the likelihood ratio test statistic

$$\frac{L(\vartheta_0;\boldsymbol{x})}{\sup_{\vartheta\in\Theta} L(\vartheta;\boldsymbol{x})} \tag{4.160}$$

where $L(\vartheta;\boldsymbol{x})$ is the usual likelihood function. The acceptance regions are then defined by the *largest* r so that

$$A(\vartheta_0) = \left\{ \boldsymbol{x} \;\middle|\; \frac{L(\vartheta_0;\boldsymbol{x})}{\sup_{\vartheta\in\Theta} L(\vartheta;\boldsymbol{x})} \geq r \right\} \tag{4.161}$$

with

$$\mathrm{P}_{\boldsymbol{X}|\{\vartheta_0\}}(\boldsymbol{X}|\{\vartheta_0\} \in A(\vartheta_0)) \geq \beta \tag{4.162}$$

In general, the side condition $\mathrm{P}_{\boldsymbol{X}|\{\vartheta_0\}}(\boldsymbol{X}|\{\vartheta_0\} \in A(\vartheta_0)) \geq \beta$ cannot be refined to $\mathrm{P}_{\boldsymbol{X}|\{\vartheta_0\}}(\boldsymbol{X}|\{\vartheta_0\} \in A(\vartheta_0)) = \beta$, especially not in the case of discrete distributions. We also see the close relation to the maximum likelihood estimator: suppose $\widehat{\vartheta}$ is maximising $L(\vartheta; \boldsymbol{x})$ for given $\boldsymbol{x}$, then $\sup_{\vartheta \in \Theta} L(\vartheta; \boldsymbol{x}) = L(\widehat{\vartheta}; \boldsymbol{x})$.
At last, we want to illustrate this method by an example concerning the uniform distribution $\mathrm{U}(0, b)$.

Example 4.61 *Let $\boldsymbol{X}|\{b\}$ be an i.i.d. sample for $X|\{b\} \sim \mathrm{U}(0, b)$ and the value $b > 0$ of the distribution parameter D should be estimated. Since the MLE of b is $\max\limits_{1 \leq i \leq n} X_i$, the likelihood ratio test statistic becomes*

$$\frac{\frac{1}{b_0^n} \mathbb{1}_{[0,b_0]^n}(\boldsymbol{x})}{\sup\limits_{b>0} L(b; \boldsymbol{x})} = \frac{\left(\max\limits_{1 \leq i \leq n} x_i\right)^n}{b_0^n} \mathbb{1}_{[0,b_0]}\left(\max_{1 \leq i \leq n} x_i\right) . \tag{4.163}$$

Then the acceptance region $A(b_0)$ is

$$A(b_0) = \left\{ \boldsymbol{x} \in [0, b_0]^n \;\middle|\; \frac{\left(\max\limits_{1 \leq i \leq n} x_i\right)^n}{b_0^n} \geq r^* \right\} , \tag{4.164}$$

with a certain value r^* for which $\mathrm{P}_{\boldsymbol{X}|\{b_0\}}(\boldsymbol{X}|\{b_0\} \in A(b_0)) = \beta$ is satified. The inequaltiy defining $A(b_0)$ is only fixing the maximum $\max\limits_{1 \leq i \leq n} x_i$. Since $\max\limits_{1 \leq i \leq n} x_i \leq b_0$, we conclude that $A(b_0)$ may be expressed as an interval for $\max\limits_{1 \leq i \leq n} x_i$

$$A(b_0) = \{ \max_{1 \leq i \leq n} x_i | \sqrt[n]{r^*} b_0 \leq \max_{1 \leq i \leq n} x_i \leq b_0 \} \tag{4.165}$$

Incorporating the side condition we get

$$\begin{aligned} \mathrm{P}_{\boldsymbol{X}|\{b_0\}}(\boldsymbol{X}|\{b_0\} \in A(b_0)) &= \mathrm{P}_{\boldsymbol{X}|\{b_0\}}\left(\sqrt[n]{r^*} b_0 \leq \max_{1 \leq i \leq n} X_i \leq b_0 \right) = \\ &= 1 - \left(F_{X|\{b_0\}}(\sqrt[n]{r^*} b_0) \right)^n = 1 - \frac{r^* b_0^n}{b_0^n} = 1 - r^* = \beta , \end{aligned} \tag{4.166}$$

which yields $r^* = 1 - \beta$ and the acceptance regions with probability β are

$$A(b_0) = \left\{ \max_{1 \leq i \leq n} x_i \;\middle|\; \sqrt[n]{1 - \beta} b_0 \leq \max_{1 \leq i \leq n} x_i \leq b_0 \right\} . \tag{4.167}$$

Then by inversion of the acceptance region we get a confidence set estimator $C(\max_{1\le i\le n} X_i)$ *of* b *with confidence level* β

$$C(\max_{1\le i\le n} X_i) = \left\{ b \;\middle|\; \max_{1\le i\le n} X_i \le b \le \frac{\max\limits_{1\le i\le n} X_i}{\sqrt[n]{1-\beta}} \right\} . \tag{4.168}$$

We note, that this is the same result as in example 4.46.

4.4.4. Bayesian Set Estimation

Assume the posterior distribution $\pi_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{x})$ of $\boldsymbol{Y}$ for a given realisation $\boldsymbol{x}$ of $\boldsymbol{X}$. For any subset $\boldsymbol{B}$ of the parameter space Θ, the *credible probability* is defined as follows

$$\mathrm{P}_{\boldsymbol{Y}}(\boldsymbol{Y} \in \boldsymbol{B}|\boldsymbol{x}) = \int_{\boldsymbol{B}} \pi_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{x})\, \mathrm{d}\boldsymbol{y} . \tag{4.169}$$

The set $\boldsymbol{B}$ is called *credible set* for $\boldsymbol{Y}$.

Example 4.62 *Let* $\boldsymbol{X}|\{y\}$ *be an i.i.d. sample for* $X|\{y\} \sim EXP(y)$ *where the prior distribution of* Y *is the gamma distribution* $GAM(\alpha,\beta)$ *with known* $\alpha,\beta > 0$. *In example 4.20 we have shown, that the density* $\pi_Y(y|\boldsymbol{x})$ *of the posterior distribution then is the gamma distribution* $GAM(\alpha+n, n\overline{x}+\beta)$. *We exploit the unimodality of* $\pi_Y(y|\boldsymbol{x})$ *with maximum in* $\frac{\alpha+n-1}{n\overline{x}+\beta}$, *then the shortest credible interval* B *with credible probability* ϱ *is defined by (cf. proposition 4.57)*

$$B(\boldsymbol{x}) = \{y \,|\, \ell \le y \le u \;\; \textit{with} \;\; \int_{\ell}^{u} \pi_Y(y|\boldsymbol{x})\, \mathrm{d}y = \varrho \;\; \textit{and} \;\; \pi_Y(\ell|\boldsymbol{x}) = \pi_Y(u|\boldsymbol{x})\} . \tag{4.170}$$

With the values $\alpha = 1$, $\beta = 0.5$, $n = 10$ *and an observed* $\boldsymbol{x}$ *which yields* $\overline{x} = 0.3$ *we get for a credible probability* $\varrho = 0.9$ *the credible interval*

$$B(\overline{x} = 0.3) = \{y \,|\, 1.60837 \le y \le 4.62826\} . \tag{4.171}$$

Figure 4.7 illustrates the concrete derivation of (4.170).

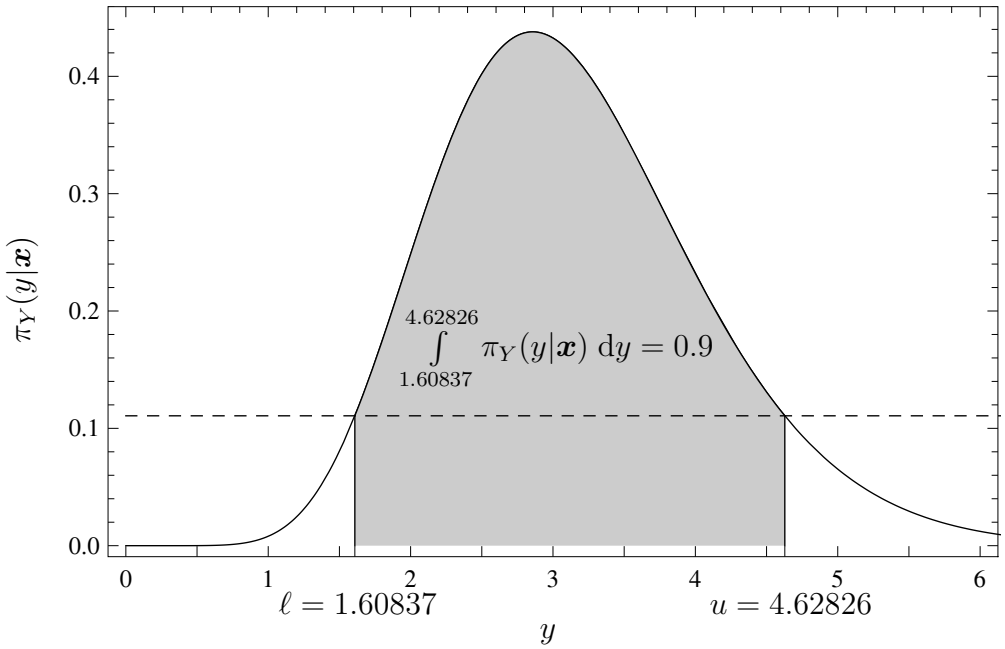


Figure 4.7.: Posterior distribution $\pi_Y(y|\boldsymbol{x})$ for $\alpha = 1$, $\beta = 0.5$, $n = 10$ and $\overline{x} = 0.3$. The shadowed area marks the credible probability of 0.9 which is covered by the credible interval $B(\overline{x} = 0.3) = \{y|1.60837 \leq y \leq 4.62826\}$.

4.4.5. Confidence Intervals vs. Credible Intervals

In this section we want to illustrate the difference between confidence intervals and credible intervals. We will do this by continuing example 4.62. First we derive the confidence interval, therefore, let $\boldsymbol{X}|\{\vartheta\}$ be an i.i.d. sample for $X|\{\vartheta\} \sim EXP(\vartheta)$ with sample size $n = 10$. According to example 4.54 we get the confidence interval estimator of ϑ for a confidence level of 0.9 with shortest length as follows

$$C(\overline{X}|\{\vartheta\}) = \left\{ \vartheta \;\middle|\; \frac{\chi^2_{20}(0.028110)}{20\overline{X}} \leq \vartheta \leq \frac{\chi^2_{20}(0.928110)}{20\overline{X}} \right\} \tag{4.172}$$

$$= \left\{ \vartheta \;\middle|\; \frac{0.489295}{\overline{X}} \leq \vartheta \leq \frac{1.493794}{\overline{X}} \right\}. \tag{4.173}$$

Figure 4.8 displays the confidence interval estimates (4.173) and the credibel interval defined by (4.170) in dependency of the observed value $\overline{x}$. We note, that

the confidence interval and the credible interval vastly differ for small values of $\overline{x}$, while in the displayed region from $\overline{x} = 0.5$ to $\overline{x} = 0.7$ they are almost identical. Now, we want to consider the confidence interval (4.173) as a credible set, i.e.,

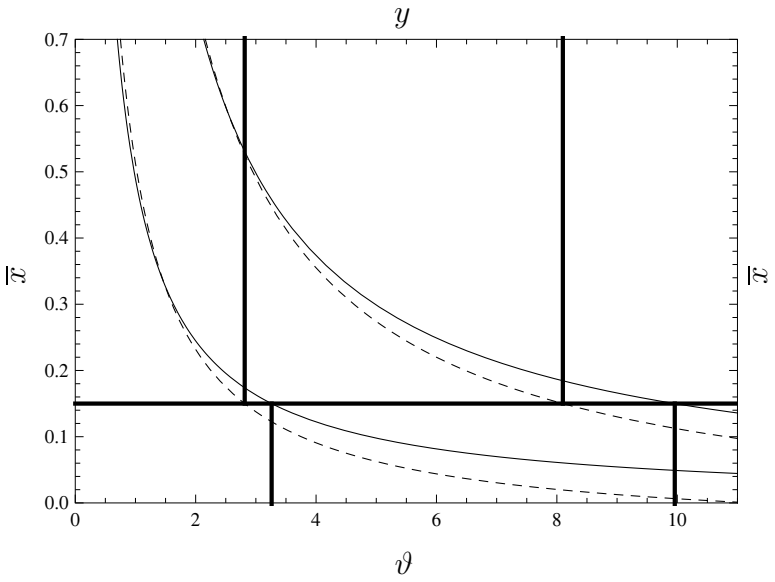


Figure 4.8.: The area between the dashed lines indicates the set of the credible intervals, the area between the solid lines indicates the set of the confidence intervals. Both depend on the observed value $\overline{x}$. The confidence level and the credible probability, respectively, is 0.9. The intersection of the regions with the horizontal line at $\overline{x} = 0.15$ marks the resulting confidence and credible interval estimate, respectively.

with the adjusted notation and for the observed value $\overline{x}$ of $\overline{X}|\{\vartheta\}$ as

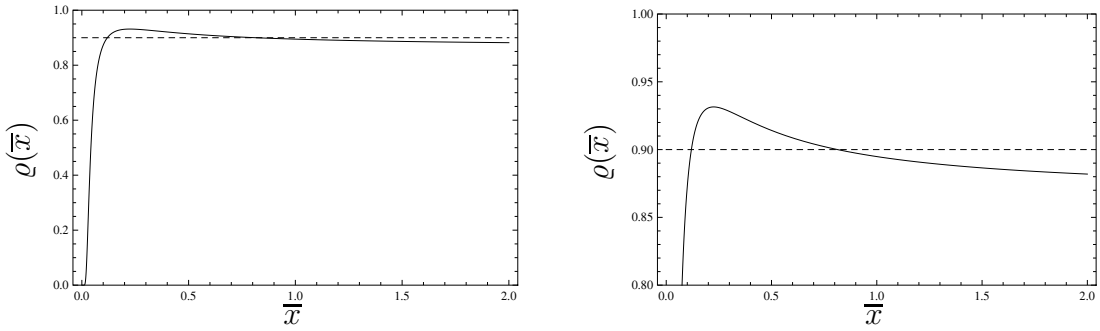
$$\left\{ y \;\middle|\; \frac{0.489295}{\overline{x}} \leq y \leq \frac{1.493794}{\overline{x}} \right\} \tag{4.174}$$

where y are the values of Y which has gamma distribution $GAM(\alpha + n, n\overline{x} + \beta) = GAM(11, 11\overline{x} + 0.5)$ (cf. example 4.62). For the credible probability in

dependency of $\overline{x}$ we get:

$$\varrho(\overline{x}) := \mathrm{P}_{Y|\overline{x}}\left(\left\{y \;\middle|\; \frac{0.489295}{\overline{x}} \leq y \leq \frac{1.493794}{\overline{x}}\right\}\right)$$
$$= \int_{\frac{0.489295}{\overline{x}}}^{\frac{1.493794}{\overline{x}}} \pi_Y(y|\overline{x})\,\mathrm{d}y = \int_{\frac{0.489295}{\overline{x}}}^{\frac{1.493794}{\overline{x}}} \frac{(10\overline{x}+0.5)^{11}}{\Gamma(11)} y^{10} e^{-y(10\overline{x}+0.5)}\,\mathrm{d}y \quad (4.175)$$

For $\overline{x} \to 0$ both bounds tend to $+\infty$ while the integrand tends to $\frac{0.5^{11}}{\Gamma(11)} y^{10} e^{-0.5y}$ and, thus, the integral to 0. For $\overline{x} \to +\infty$ the analysis is more difficult and we confine ourselves by illustrating the credibel probability $\varrho(\overline{x})$ in dependency of $\overline{x}$ with figure 4.9. Only for a small region of values of $\overline{x}$ the credible probability



(a) $\overline{x} \in (0, 2]$ (b) $\overline{x} \in (0, 2]$

Figure 4.9.: The credible probability $\varrho(\overline{x}) = \mathrm{P}_{Y|\overline{x}}(\{y|\frac{0.489295}{\overline{x}} \leq y \leq \frac{1.493794}{\overline{x}}\})$ in dependency of $\overline{x}$ of the confidence interval estimate $\{\vartheta|\frac{0.489295}{\overline{x}} \leq \vartheta \leq \frac{1.493794}{\overline{x}}\}$ considered as a credible interval.

surpasses the 0.9, while for the rest it falls below 0.9.

On the other hand, considering the credible intervals as confidence intervals we may calculate their coverage probability. That is, the probability of (4.170) in dependency of ϑ. The lower and the upper bound, $\ell(\overline{X})$ and $u(\overline{X})$, of the credible interval are now considered as random variables which depend on $\overline{X}|\{\vartheta\}$, thus, we get

$$\mathrm{P}_{\overline{X}|\{\vartheta\}}\left(\ell(\overline{X}) \leq \vartheta \leq u(\overline{X})\right) = 1 - \left(\mathrm{P}_{\overline{X}|\{\vartheta\}}\left(\ell(\overline{X}) > \vartheta\right) + \mathrm{P}_{\overline{X}|\{\vartheta\}}\left(u(\overline{X}) < \vartheta\right)\right) \quad (4.176)$$

Since, both bounds $\ell(\overline{x})$ and $u(\overline{x})$ are decreasing with $\overline{x}$, we need to determine the values $\overline{x}_\ell$ and $\overline{x}_u$ for which

$$\ell(\overline{x}_\ell) = \vartheta \text{ and } u(\overline{x}_u) = \vartheta \tag{4.177}$$

hold. Then the coverage probability becomes

$$\mathrm{P}_{\overline{X}|\{\vartheta\}}(\ell(\overline{X}) \leq \vartheta \leq u(\overline{X})) = 1 - \left(\mathrm{P}_{\overline{X}|\{\vartheta\}}(\overline{X} < \overline{x}_\ell) + \mathrm{P}_{\overline{X}|\{\vartheta\}}(\overline{X} > \overline{x}_u)\right) . \tag{4.178}$$

Figure 4.10 illustrates the procedure.

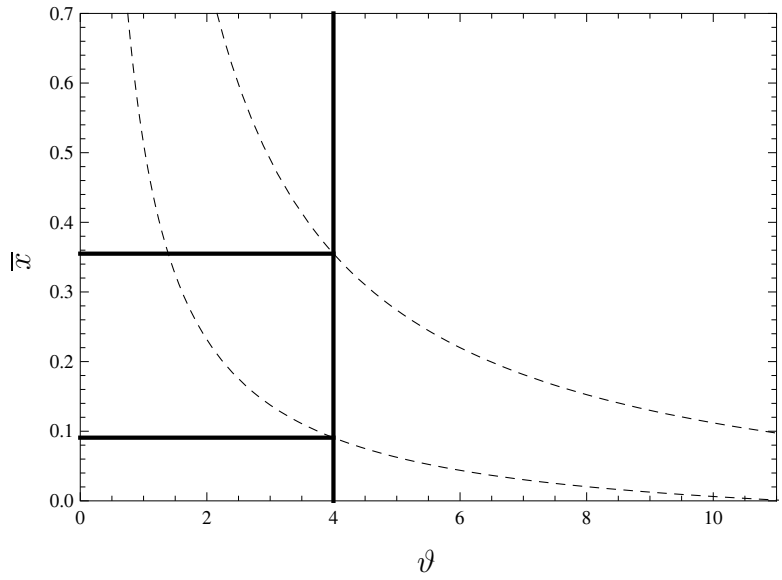


Figure 4.10.: The area between the dashed lines indicates the set of credible intervals. Considered as a confidence interval, the coverage probability of a credibel interval for a certain value ϑ_0 is the probability for observing values $\overline{x}$ of $\overline{X}|\{\vartheta\}$ outside the intersection of the area with the vertical line at ϑ_0.

Since, $\overline{X}|\{\vartheta\}$ is distributed according to the gamma distribution $GAM(n\vartheta, n)$ (cf. example 4.14), the only problem lies in the necessary numerical determination of $\overline{x}_\ell$ and $\overline{x}_u$ for each ϑ. Figure 4.11 displays the coverage probability of the credible set as function of ϑ. Till some certain value of ϑ the coverage probability is at least the credible probability of 0.9, but from there on it falls below. An explanation lies in the prior distribution of Y, which was the gamma distribution

$GAM(\alpha, \beta)$ with $\alpha = 1$ and $\beta = 0.5$. As illustrated in figure 4.1, the prior distribution is monotonously decreasing, i.e., it reflects the (subjective) belief, that small values y (which are ϑ in the sense of confidence intervals) of Y are more probable than large ones. This leads to the effect, that the credible intervals are somehow towed to smaller values of y and ϑ, respectively.

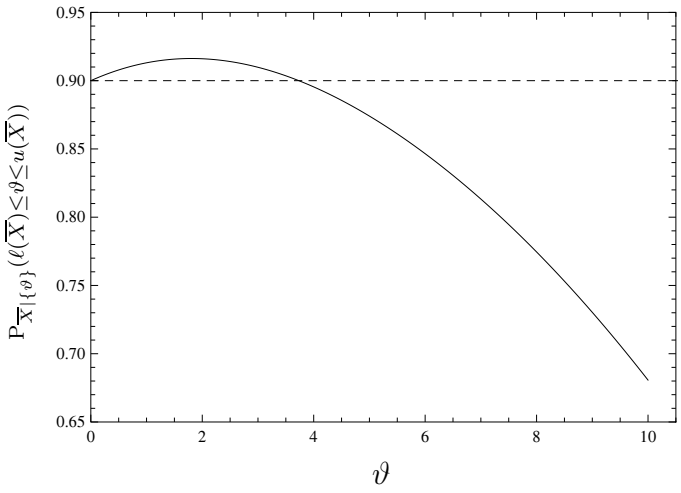


Figure 4.11.: The coverage probability (4.178) as function of ϑ. It never falls below the credible probability of 0.9 for small values of ϑ, but from some value on it underruns 0.9 and steadily decreases.

We end this consideration by realising, that the Bayesian concept of estimation by credible intervals is totally different to that of confidence intervals and it is hardly possible to compare them. In particular, the choice of the prior distribution plays a crucial role and significantly governs the posterior distribution and the credible interval, respectively. Due to similarities of the proposed Neyman measurement procedure in Chapter 6 with the Bayesian approach we will draw some comparisons between them in Chapter 9.

4.4.6. Evaluating confidence intervals

In the above examples, we already derived confidence interval estimators with a given confidence level optimal in the sense of yielding confidence interval estimates of minimal lengths. In the introduction to this chapter we have drawn parallels of technical measurement procedures to statistical estimation. The accuracy and the reliability of a measurement procedure are two of its important features. Accuracy in the sense of statistical estimation might be interpreted as the lengths of the resulting interval estimates. Therefore, the length of the confidence interval estimates (or the volumes of a confidence set estimates for a multidimensional parameter) is for sure an important criterion to judge competitive confidence interval estimators. If the length itself depends somehow on the random sample $\boldsymbol{X}$ and, thus, is also a random variable, it may be better to take the expected length, i.e., the mean length into account. In metrology, a measurement procedure is called reliable, if it yields similar results when repeated. The corresponding term in statistical estimation is the confidence level, thus, we recall what the confidence level means: If the random sample could be realised infinitely often, the relative frequency a confidence interval estimator with a confidence level of $\beta \in (0,1)$ yields interval estimates which cover the true value ϑ of the parameter D would equal β. In (traditional) statistical estimation, set estimators are not evaluated by their (mean) volume in first place, but via their false coverage probability. For the sake of completeness, we give here the relevant definitions. Since we introduce in Chapter 6 an alternative way to evaluate set estimators, we do not further work out details nor give examples.

Definition 4.63 (false coverage) *For a set estimator $C(\boldsymbol{X}|\{\vartheta\})$ of ϑ with a confidence level of $\beta \in (0,1)$, the probability of false coverage is definded by*

$$\mathrm{P}_{\boldsymbol{X}|\{\vartheta\}}(\vartheta' \in C(\boldsymbol{X}|\{\vartheta\})) \quad \textit{for all } \vartheta' \in \Theta \setminus \{\vartheta\}\,. \tag{4.179}$$

Definition 4.64 (uniformly accurate) *Let $C_1(\boldsymbol{X}|\{\vartheta\})$ and $C_2(\boldsymbol{X}|\{\vartheta\})$ be two set estimators of ϑ both with a confidence level of $\beta \in (0,1)$. Then $C_1(\boldsymbol{X}|\{\vartheta\})$ is called uniformly more accurate than $C_2(\boldsymbol{X}|\{\vartheta\})$ if*

$$\mathrm{P}_{\boldsymbol{X}|\{\vartheta\}}(\vartheta' \in C_1(\boldsymbol{X}|\{\vartheta\})) \leq \mathrm{P}_{\boldsymbol{X}|\{\vartheta\}}(\vartheta' \in C_2(\boldsymbol{X}|\{\vartheta\})) \quad \textit{for all } \vartheta \in \Theta \setminus \{\vartheta\}\,. \tag{4.180}$$

If inequality (4.180) *holds for any other set estimator* $C_2(\boldsymbol{X}|\{\vartheta\})$ *of* ϑ *with confidence level* $\beta \in (0,1)$*, then* $C_1(\boldsymbol{X}|\{\vartheta\})$ *is called uniformly most accurate confidence set estimator.*

Definition 4.65 (unbiased set estimator) *A set estimator* $C(\boldsymbol{X}|\{\vartheta\})$ *of* ϑ *with a confidence level of* $\beta \in (0,1)$ *is called unbiased if*

$$\mathrm{P}_{\boldsymbol{X}|\{\vartheta\}}(\vartheta' \in C(\boldsymbol{X}|\{\vartheta\})) \leq \beta \quad \text{for all } \vartheta' \in \Theta \setminus \{\vartheta\}\,. \tag{4.181}$$

4.5. Concluding Thoughts on Classical Point and Set Estimation

At the beginning the two relevant features for evaluating the methods used in metrology have been specified, namely reliability and accuracy. Reliability of a method refers to how often correct results are obtained when applying the method, while accuracy refers to the usability of correct results. In this paper two statistical approaches for solving the related problems in quality control are described and briefly analyzed namely *point estimators* and *set estimators*.

As already mentioned, neither reliability nor accuracy is clearly introduced in the theory of statistical point estimators. Any point estimator yields one value called *estimate*. The only founded statement about any obtained estimate is that it is almost certainly not equal with the true value. Consequently, point estimators have a zero reliability and relying on achieved results may lead to wrong decisions and provisions. The situation is even worse with regard to accuracy. In classical statistics unbiasedness of the point estimator is generally required, although it may be counterproductive and lead to estimates with larger mean squared error particular in the more realistic cases of a restricted parameter space. By the way, the same arguments produced against point estimators are also valid in view of technical measurement procedures in metrology and metrology can therefore not be used in order to improve statistical point estimators.

The situation with respect to set estimators is better. The reliability of a set estimator is appropriately given by the coverage probability, i.e., the probability

that an application of the method will yield an estimate which covers the true but unknown value. The confidence level, on the other hand, represents a reliability requirement or specification for the method to be applied. The inaccuracy of an obtained set estimate is given in a natural way by its size (by the length in case of an interval estimate). The smaller the size, the more accurate and thus the more usable is the set estimate. Maximum accuracy is obtained by set estimates being singletons. Note, however, that the larger the confidence level and hence the reliability requirement, the larger is the size and hence the inaccuracy of the resulting estimates. Therefore it is reasonable to define an optimal set estimator by the solution of an optimization problem with respect to accuracy with a side condition concerning the required reliability. The problem then is to determine a set estimator with minimum size of the estimates meeting the side condition given by the confidence level. However, there are also some drawbacks with respect to set estimators:

- Classical statistics does not provide a measure of accuracy for the estimator, but only for estimates.
- The uncontrolled use of asymptotic models may lead to misleading set estimators and thus to wrong decisions.
- Set estimators in classical statistics are often based on unbiased estimators representing an unnecessary restriction.

Finally, there are Bayesian set estimators which actually become more and more popular, because the resulting Bayesian set estimates exhibit a larger accuracy then the competing non-Bayesian set estimators. However, in the Bayesian context the concept of probability is not used as an objective property of a future event, but as a measure of a subjective opinion or belief. Consequently, there is no objective statement possible about the actually achieved reliability.

5. Inference Applied

The monotonic probability distribution $Mon(a, b, \mu_1)$ is fully determined by the range of variability $\mathcal{X} = \{x \,|\, a \leq x \leq b\}$ and the value of the first moment μ_1. Consequently, the distribution parameter consists of the three components a, b and μ_1. Each and all combinations of them could be subject of inference.

The main scope of this chapter is to derive a measurement procedure for the expectation $\mathrm{E}[X]$, i.e., the first moment of $X \sim Mon(a, b, \mu_1)$. Thus, the afore presented methods should be mainly applied with respect to $\mathrm{E}[X]$, but we will also shortly discuss the measurement of the upper bound b.

5.1. Point Estimation

The essential properties leading to the monotonic probability distribution include the knowledge about the monotonic behaviour of the density function, i.e., whether it is monotonic increasing or decreasing. These essential properties result into quantitave properties in the way, that only certain values for the components of the distribution parameter (a, b, μ_1) are possible. Besides the trivial restriction $-\infty < a < b < +\infty$, i.e., the range of variability is neither unbounded nor a degenerated interval, the monotonicity results into a restriction of the possible value of μ_1. If the density function of X is monotone decreasing it follows, that $\mu_1 \in \{m \,|\, a < m < \frac{a+b}{2}\}$, and conversely, if the density function of X is monotone increasing we have $\mu_1 \in \{m \,|\, \frac{a+b}{2} < m < b\}$. These conditions together only determine the maximal possible parameter space. Other essential properties and quantitative knowledge respectively may lead to further restrictions, e.g. the random variable may only adopt non-negative values, which results into $0 \leq a < b < +\infty$. Similar to proposition 3.2 we have:

Corollary 5.1 (maximum parameter space) *The parameter space Θ for a monotonic distributed random variable $X \sim Mon(a, b, \mu_1)$ is atmost either*

(a) $\Theta = \{(a, b, \mu_1) \in \mathbb{R}^3 \mid -\infty < a < \mu_1 < \frac{a+b}{2} < b < +\infty\}$, *if the densitiy function of* X *is strictly decreasing, or*

(b) $\Theta = \{(a, b, \mu_1) \in \mathbb{R}^3 \mid -\infty < a < \frac{a+b}{2} < \mu_1 < b < +\infty\}$, *if the densitiy function of* X *is strictly increasing.*

With the alternative parameterisation (a, b, λ) we have:

Corollary 5.2 (maximum parameter space) *The parameter space* Θ *for a monotonic distributed random variable* $X \sim Mon(a, b, \lambda)$ *is atmost either*

(a) $\Theta = \{(a, b, \lambda) \in \mathbb{R}^3 \mid -\infty < a < b < +\infty, -\infty < \lambda < 0\}$, *if the densitiy function of* X *is strictly decreasing, or*

(b) $\Theta = \{(a, b, \lambda) \in \mathbb{R}^3 \mid -\infty < a < b < +\infty, 0 < \lambda < +\infty\}$, *if the densitiy function of* X *is strictly increasing.*

5.1.1. Method of Moments

Moment estimators for up to three parameters could be of interest. Thus, for estimating all three parameters we need a function $g = (g_1, g_2, g_3, \ldots)$ with at least three component functions g_i, $i = 1, 2, \ldots$, which describe the relations between (a, b, μ_1) (or (a, b, λ)) and the moments $\mu_1, \mu_2, \ldots$. (3.112), (3.113) and (3.114) yield the system of equations

$$\mathrm{E}[X] = \mu_1 = \frac{(-1 + \lambda b)e^{\lambda b} - (-1 + \lambda a)e^{\lambda a}}{\lambda(e^{\lambda b} - e^{\lambda a})}$$

$$\mathrm{E}[X^2] = \mu_2 = \frac{(2 - 2\lambda b + \lambda^2 b^2)e^{\lambda b} - (2 - 2\lambda a + \lambda^2 a^2)e^{\lambda a}}{\lambda^2(e^{\lambda b} - e^{\lambda a})}$$

$$\mathrm{E}[X^3] = \mu_3 = \frac{(-6 + 6\lambda b - 3\lambda^2 b^2 + \lambda^3 b^3)e^{\lambda b} - (-6 + 6\lambda a - 3\lambda^2 a^2 + \lambda^3 a^3)e^{\lambda a}}{\lambda^3(e^{\lambda b} - e^{\lambda a})}$$

According to definition 4.5 we actually need a function g, which maps (μ_1, μ_2, μ_3) on (a, b, μ_1). Since the equation for μ_1 cannot be explicity solved for one of the other parameters a, b, numerical calculations become necessary.

In the following we derive moment estimators for μ_1 and b under different assumptions of existing knowledge about the parameters.

5.1.1.1. Estimation of μ_1

Assuming that $\mathcal{X}$ is known, we may set $\mathcal{X} = [0,1]$ without loss of generality. Then the first empirical moment $M_{n,1} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is already the desired moment estimator of μ_1. Although $\overline{X}$ is an unbiased estimator, it does not always yield reasonable results. Since it is assumed that the monotonicity of the density function $f_{X|\{\mu_1\}}$ of X is known, the estimate $\overline{x}$ may nevertheless be not element of Θ: Albeit $f_{X|\{\mu_1\}}$ being monotone decreasing, i.e., the upper bound of Θ is $\frac{1}{2}$, $\overline{x}$ may be larger than $\frac{1}{2}$.

5.1.1.2. Estimation of the upper bound b of $\mathcal{X}$

Assuming that μ_1 and the lower bound a of $\mathcal{X}$ are known, we may set $a = 0$ without loss of generality. The upper bound b of $\mathcal{X}$ should be unknown.

Since b is the only parameter to be estimated, the first idea is, that similar to the estimation of μ_1, only one empirical moment is necessary. The first moment μ_1 of X is assumed to be known and one may consider the first empirical moment not suitable and necessary to find the moment estimator of b. Thus, we consider the second moment μ_2 of X estimated by the second empirical moment:

$$\mathrm{E}[X^2] = \mu_2(b,\lambda) = \frac{(2 - 2\lambda b + \lambda^2 b^2)e^{\lambda b} - 2}{\lambda^2(e^{\lambda b} - 1)} \quad \text{with} \quad \lambda = \lambda(b, \mu_1) \tag{5.1}$$

$$M_{n,2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \tag{5.2}$$

We note, that μ_2 depends on b and on λ, where the latter itself is a function of b and μ_1. Equating the realisation $m_{n,2}$ of $M_{n,2}$ with $\mu_2(b,\lambda)$ and solving for b is not possible, since we may only find a set of possible values of (b,λ), which all solve

$$m_{n,2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{(2 - 2\lambda b + \lambda^2 b^2)e^{\lambda b} - 2}{\lambda^2(e^{\lambda b} - 1)}\,. \tag{5.3}$$

Of course, this holds for $\mu_2(b,\lambda)$ as well - different pairs of (b,λ) yield the same value of μ_2:

Example 5.3 *For $(b,\lambda) = (4,-1)$ we have $\mu_2(4,-1) = 1.55222$, but the same value results for $(b,\lambda) = (5,-1.082987)$.*

Analogue to μ_2, μ_1 is considered as a function of b and λ, and thus, a concrete value of μ_1 determines a set of possible values of (b, λ):

Example 5.4 *For* $(b, \lambda) = (4, -1)$ *we have* $\mu_1(4, -1) = 0.92537$*, which is the same value of* μ_1 *as for* $(b, \lambda) = (5, -1.050820)$*.*

Therefore, the known value of μ_1 determines a set $S^{(1)}_{\mu_1} := \{(b, \lambda) \,|\, \mu_1(b, \lambda) = \mu_1\}$. Then, the task is, to find that element of $S^{(1)}_{\mu_1}$ for which equation (5.3) holds.

Example 5.5 *Let* $\mu_1 = 0.9$ *and we assume, that the density function is strictly monotone decreasing, which yields* $b \geq 2\mu_1 = 1.8$ *and* $\lambda < 0$*. It follows, that* $S^{(1)}_{\mu_1=0.9} = \{(b, \lambda) \,|\, \mu_1(b, \lambda) = 0.9\} \subset \{(b, \lambda) \,|\, b > 1.8, \lambda < 0\}$*. Assume, the sample results into* $m_{n,2} = 1.53411$*, then a small numerical calculation yields* $\lambda = -1.07606$ *and as estimate of* b *the value* 4.73003*.*

Two questions arise at this point. First, whether it is always possible to find a solution, i.e., an estimate for b for all combinations of values of μ_1 and $m_{n,2}$, and, second, whether a feasible solution is also unique.

We realise, that the derivation of point estimates for b by means of the (elementary) method of moments becomes a laborious task. Therefore, it is hardly applicable and we will not pursue this method further.

5.1.2. Maximum Likelihood

5.1.2.1. MLE of μ_1 and λ, respectively

With (3.97) we have already seen, that the density functions $f_{X|\{(a,b,\mu_1\}}$ of $X \sim Mon(a, b, \mu_1)$ generate an exponential family in $\mu_1 \in (a, b)$ and $\lambda \in \mathbb{R}$, respectively. Although, μ_1 is only implictly expressed via λ, we utilise the one-to-one relation between them and may apply the invariance property 4.13 on the MLE for λ, which we want to derive for the case of a monotone decreasing density of X, i.e., $\lambda \leq 0$. Without loss of generality we set $\mathcal{X} = [0, 1]$.

The density function written in the natural parameterisation form with $\eta = \lambda$ then is

$$f_{X|\{\lambda\}}(x) = \frac{\lambda}{e^\lambda - 1}\mathbb{1}_{[0,1]}(x)e^{\lambda x} = \mathbb{1}_{[0,1]}(x)\exp(\lambda x - (\ln(1-e^\lambda) - \ln(-\lambda)))\,, \quad (5.4)$$

yielding $A(\lambda) = \ln(1-e^\lambda) - \ln(-\lambda)$. Therefore, the equation to be solved for the MLE $\hat{\lambda}$ is given by

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x} = \frac{\lambda e^\lambda - e^\lambda + 1}{\lambda(e^\lambda - 1)}\,. \quad (5.5)$$

Equation (5.5) is nothing else, then equation (3.8a) which determines λ given μ_1, but now with $\overline{x}$ instead of μ_1. Since we presume $\lambda \leq 0$, values of $\overline{x} > 0.5$ would lead to $\lambda(\overline{x}) > 0$ which would be incompatible to the parameter space. Thus, we examine the log-likelihood function $l(\lambda; \boldsymbol{x})$ for $\overline{x} > 0.5$ with respect to λ:

$$l(\lambda; \boldsymbol{x}) = \lambda n\overline{x} + n\ln\frac{\lambda}{e^\lambda - 1} \quad (5.6)$$

and the derivative with respect to λ yields

$$\frac{\partial l(\lambda; \boldsymbol{x})}{\partial\lambda} = n\overline{x} - n\frac{\lambda e^\lambda - e^\lambda + 1}{\lambda(e^\lambda - 1)}\,. \quad (5.7)$$

For $\lambda \nearrow 0$ we have $\frac{\lambda e^\lambda - e^\lambda + 1}{\lambda(e^\lambda - 1)} \nearrow 0.5$, i.e., $\frac{\partial l(\lambda;\boldsymbol{x})}{\partial\lambda} > 0$ for $\overline{x} > 0.5$. That means, that $l(\lambda; \boldsymbol{x})$ for $\overline{x} > 0.5$ attains its maximum at $\lambda = 0$. Together we get the MLE for λ:

$$\hat{\lambda}(\boldsymbol{X}) = \begin{cases} \lambda \text{ with } \frac{\lambda e^\lambda - e^\lambda + 1}{\lambda(e^\lambda - 1)} = \overline{x} & \text{for } \overline{x} < 0.5, \\ 0 & \text{for } \overline{x} \geq 0.5. \end{cases} \quad (5.8)$$

Now, from the invariance property of the MLE 4.13, we easily get the MLE for μ_1:

$$\hat{\mu}_1(\boldsymbol{X}) = \begin{cases} \overline{x} & \text{for } \overline{x} < 0.5, \\ 0.5 & \text{for } \overline{x} \geq 0.5. \end{cases} \quad (5.9)$$

If we restrict the parameter space even further, e.g. $\lambda \in [\lambda^\ell, \lambda^u] \subset (-\infty, 0] \Leftrightarrow \mu_1 \in [\mu_1(\lambda^\ell), \mu_1(\lambda^u)] =: [\mu_1^\ell, \mu_1^u] \subset (0, 0.5]$, we note that $\frac{\partial l(\lambda;\boldsymbol{x})}{\partial\lambda} < 0$ for $\overline{x} < \mu_1^\ell$,

and $\frac{\partial l(\lambda;\boldsymbol{x})}{\partial \lambda} > 0$ for $\overline{x} > \mu_1^u$. Therefore, the MLE for λ results to be

$$\hat{\lambda}(\boldsymbol{X}) = \begin{cases} \lambda^\ell & \text{for } \overline{x} \leq \mu_1^\ell, \\ \lambda \text{ with } \frac{\lambda e^\lambda - e^\lambda + 1}{\lambda(e^\lambda - 1)} = \overline{x} & \text{for } \mu_1^\ell < \overline{x} < \mu_1^u, \\ \lambda^u & \text{for } \overline{x} \geq \mu_1^u, \end{cases} \tag{5.10}$$

and so, the MLE for μ_1 is

$$\hat{\mu}_1(\boldsymbol{X}) = \begin{cases} \mu_1^\ell & \text{for } \overline{x} \leq \mu_1^\ell, \\ \overline{x} & \text{for } \mu_1^\ell < \overline{x} < \mu_1^u, \\ \mu_1^u & \text{for } \overline{x} \geq \mu_1^u. \end{cases} \tag{5.11}$$

Evaluating the MLE of μ_1

The least knowledge about the distribution is about its monotonic behaviour. That is, we first want to deal with the in this sense unrestricted parameter space $\mu_1 \in (0, 0.5]$ where we get (5.9) for the MLE $\hat{\mu}_1(\boldsymbol{X})$ of μ_1. The expected value of $\hat{\mu}_1(\boldsymbol{X})$ may be assessed in the following way:

$$\begin{aligned} \mathrm{E}_{\mu_1}[\hat{\mu}_1(\boldsymbol{X})] &= \mathrm{E}[\overline{X}|\overline{X} < 0.5] + \mathrm{E}[0.5|\overline{X} \geq 0.5] && (5.12) \\ &= \int_0^{0.5} \overline{X} \, \mathrm{dP} + 0.5 \int_{0.5}^1 \mathrm{dP} && (5.13) \\ &= \mu_1 - \left(\int_{0.5}^1 \overline{X} \, \mathrm{dP} - 0.5 \int_{0.5}^1 \mathrm{dP} \right) && (5.14) \\ &= \mu_1 - \left(\int_{0.5}^1 (\overline{X} - 0.5) \, \mathrm{dP} \right) && (5.15) \end{aligned}$$

Since $\overline{X} \geq 0.5$ under the integral, the integrand and so the integral are positive and, thus, $\mathrm{E}_{\mu_1}[\hat{\mu}_1(\boldsymbol{X})] < \mu_1$. That is, the MLE of μ_1 is not unbiased. Therefore, the discussion of all other evaluation criterions for unbiased estimators becomes redundant.

5.1.2.2. MLE of b

We want to derive the MLE of b, assuming that the first moment μ_1 of $X \sim Mon(b, \mu_1)$ is known and that the density function is monotone decreasing. From

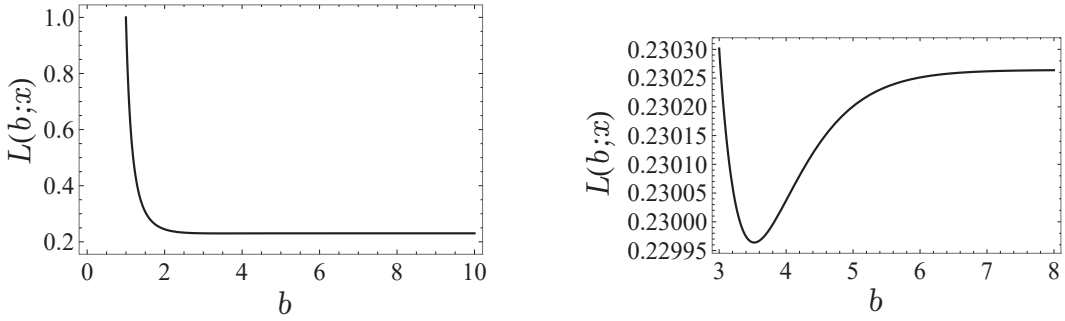
the latter, we already get a restriction of possible values of b: $\mu_1 \leq \frac{b}{2}$, i. e., $b \geq 2\mu_1$. Again, without loss of generality we may assume $a = 0$. The likelihood function given a realisation $\boldsymbol{x} = (x_1, \ldots, x_n)$ for $b \geq 2\mu_1$ is

$$L(b; \boldsymbol{x}) = \left(\frac{\lambda(b, \mu_1)}{e^{\lambda(b,\mu_1)} - 1} \right)^n e^{\lambda(b,\mu_1) \sum_{i=1}^n x_i} \mathbb{1}_{[0,b]^n}(\boldsymbol{x}) \tag{5.16}$$

$$= \left(\frac{\lambda(b, \mu_1)}{e^{\lambda(b,\mu_1)} - 1} \right)^n e^{\lambda(b,\mu_1) \sum_{i=1}^n x_i} \mathbb{1}_{[\max\{2\mu_1, \max x_i\}, +\infty)}(b) \,. \tag{5.17}$$

Note, that the term $\mathbb{1}_{[\max\{2\mu_1, \max x_i\}, +\infty)}(b)$ contains the condition $b \geq 2\mu_1$ which is incorporated besides a possible further restriction of potential values of b maximising the likelihood function: Since all $X_i \sim Mon(b, \mu_1)$, we have $x_i \leq b$ for all $i = 1, \ldots, n$, i. e., $b \geq \max x_i$. Before analysing the likelihood function with respect to b, we want to give some numerical examples, which already indicate possible occurring situations. Throughout the examples, we let $\mu_1 = 0.5$, $n = 10$ and give no further restrictions with respect to b, and vary the values of $\sum_{i=1}^{10} x_i$ and $\max x_i$.

Example 5.6 *Let* $\sum_{i=1}^{10} x_i = 4.2$ *and* $\max x_i = 0.9$. *Then the values of* b *are only restricted by* $\mu_1 = 0.5$, *i.e.,* $b \geq 1$. *Figure 5.1(a) shows, that the likelihood*



(a) $b \in [1, 10]$ (b) $b \in [3, 8]$

Figure 5.1.: Likelihood function $L(b; \boldsymbol{x})$ for $b \in [1, 10]$ and $b \in [3, 8]$, respectively, with $\mu_1 = 0.5$, $\sum_{i=1}^{10} x_i = 4.2$ and $\max x_i = 0.9$.

function attains its maximum in $b = 1$, *i.e.,* $\hat{b} = 1$ *is the MLE for* b. *The maximal value then is* $L(1; \boldsymbol{x}) = 1$. *Taking a closer look onto the course of the function for* $b \in [3, 8]$ *in figure 5.1(b), a minimum around* 3.5 *becomes visible. Right from this minimum* $L(b; \boldsymbol{x})$ *appears to be monotone increasing against an upper limit.*

Example 5.7 *Let* $\sum_{i=1}^{10} x_i = 3.4$ *and* $\max x_i = 0.9$. *Then the values of* b *are only restricted by* $\mu_1 = 0.5$, *i.e.,* $b \geq 1$. *Figure 5.2 shows, that* $L(b; \boldsymbol{x})$ *attains again a*

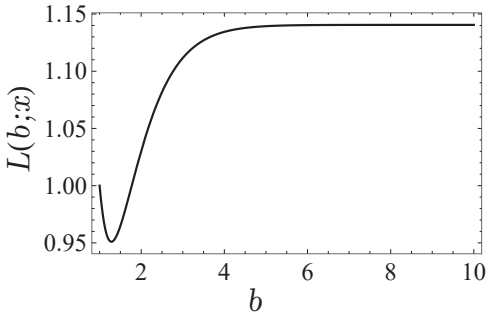


Figure 5.2.: Likelihood function $L(b; \boldsymbol{x})$ for $b \in [1, 10]$ with $\mu_1 = 0.5$, $\sum_{i=1}^{10} x_i = 3.4$ and $\max x_i = 0.9$.

maximum in $b = 1$ *with* $L(1; \boldsymbol{x}) = 1$. *But now it is only a local maximum, since* $L(b; \boldsymbol{x})$ *appears to be approaching an upper limit for* $b \to +\infty$ *which is larger than* 1. *This leads to the MLE* $\hat{b} = +\infty$ *for* b.

Example 5.8 *Let* $\sum_{i=1}^{10} x_i = 3.2$ *and* $\max x_i = 0.9$. *Then the values of* b *are only restricted by* $\mu_1 = 0.5$, *i.e.,* $b \geq 1$. *In figure 5.3* $L(b; \boldsymbol{x})$ *seems to have no*

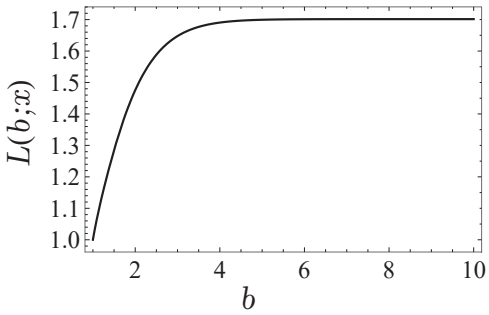


Figure 5.3.: Likelihood function $L(b; \boldsymbol{x})$ for $b \in [1, 10]$ with $\mu_1 = 0.5$, $\sum_{i=1}^{10} x_i = 3.2$ and $\max x_i = 0.9$.

local maximum in $b = 1$ *anymore, but approaches an upper limit for* $b \to +\infty$. *Again, this leads to the MLE* $\hat{b} = +\infty$ *for* b.

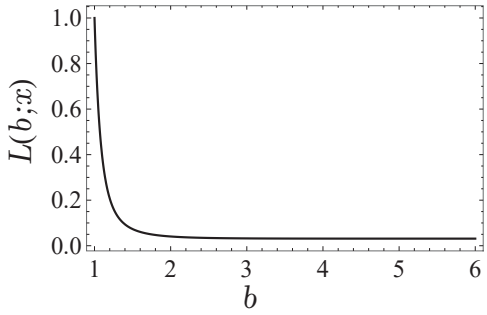
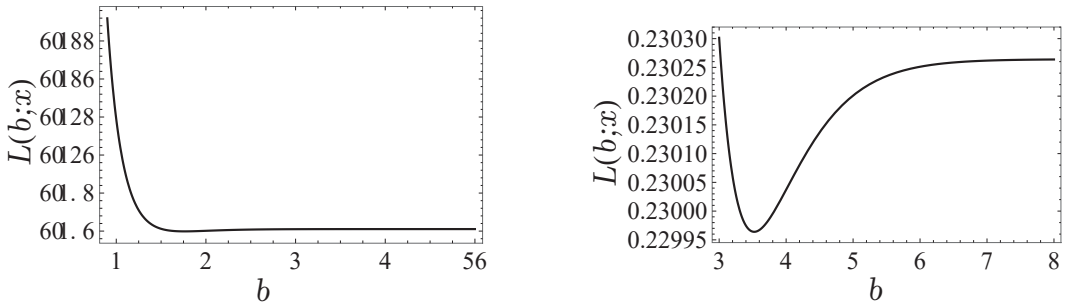


Figure 5.4.: Likelihood function $L(b;\boldsymbol{x})$ for $b \in [1,6]$ with $\mu_1 = 0.5$, $\sum_{i=1}^{10} x_i = 5.2$ and $\max x_i = 0.9$.

Example 5.9 *Let* $\sum_{i=1}^{10} x_i = 5.2$ *and* $\max x_i = 0.9$. *Then the values of* b *are only restricted by* $\mu_1 = 0.5$, *i.e.,* $b \geq 1$. *In figure 5.4* $L(b;\boldsymbol{x})$ *appears to have its global maximum in* $b = 1$, *again. Thus, the MLE for* b *is* $\hat{b} = 1$. *The course of the function now seems to be decreasing against some lower bound for* $b \to +\infty$ *– at least a local minimum like in example 5.6 cannot be detected.*

Example 5.10 *Let* $\sum_{i=1}^{10} x_i = 4.2$ *and* $\max x_i = 1.8$. *Then the latter restricts the values of* b *to* $b \geq 1.8$. *Similar to example 5.6 figure 5.5(a) indicates, that*



(a) $b \in [1.8, 10]$ (b) $b \in [3, 8]$

Figure 5.5.: Likelihood function $L(b;\boldsymbol{x})$ for $b \in [1.8, 10]$ and $b \in [3, 10]$, respectively, with $\mu_1 = 0.5$, $\sum_{i=1}^{10} x_i = 4.2$ and $\max x_i = 1.8$.

$L(b;\boldsymbol{x})$ *attains its global maximum in the smallest of the potential values of* b, *i.e., the MLE is* $\hat{b} = 1.8$. *There is a local minimum around* 3.5 *as well, from which on* $L(b;\boldsymbol{x})$ *appears to be increasing against an upper bound for* $b \to +\infty$.

Example 5.11 *Let* $\sum_{i=1}^{10} x_i = 6.7$ *and* $\max x_i = 1.8$*, and again, the latter restricts the values of* b *to* $b \geq 1.8$*. As in example 5.9, the course of* $L(b; \boldsymbol{x})$ *(see*

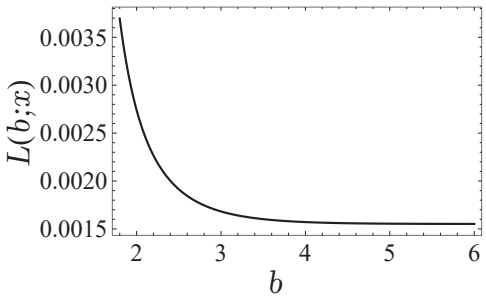


Figure 5.6.: Likelihood function $L(b; \boldsymbol{x})$ for $b \in [1.8, 6]$ with $\mu_1 = 0.5$, $\sum_{i=1}^{10} x_i = 6.7$ and $\max x_i = 1.8$.

figure 5.6) seems to attain its global maximum in the smallest of the potential values of b*, i.e., the MLE is* $\hat{b} = 1.8$*. Furthermore it appears to be monotone decreasing.*

The above examples indicate that there are only two possibilities for the MLE of b: either the likelihood function attains its maximum in the smallest of the potential values of b, determined by $2\mu_1$ or $\max x_i$, or in $+\infty$. We want to structure our analyse by the values of $\max x_i$ in comparison to $2\mu_1$. And again, we assume no upper bound for the potential values of b.

$\max x_i < 2\mu_1$

No further restriction on the potential values of b arise in this case. Then the limiting values of $L(b; \boldsymbol{x})$ for $b = 2\mu_1$ and $b \to +\infty$, respectively, are easily derived via the convergence behaviour of $Mon(b, \mu_1)$:

a. The distribution $Mon(b, \mu_1)$ coincides with the uniform distribution $\mathrm{U}(0, 2\mu_1)$ for $b = 2\mu_1$.

b. The distribution $Mon(b, \mu_1)$ converges to the exponential distribution $EXP(\mu_1)$ for $b \to +\infty$.

Thus, this yields for the likelihood function:

a. $L(b; \boldsymbol{x}) \overset{b=2\mu_1}{=} \left(\frac{1}{2\mu_1}\right)^n$

b. $L(b;\boldsymbol{x}) \overset{b\to+\infty}{\to} \left(\frac{1}{\mu_1}\right)^n e^{-\frac{1}{\mu_1}\sum_{i=1}^n x_i}$

Comparing both values, leads to

$$\left(\frac{1}{2\mu_1}\right)^n > \left(\frac{1}{\mu_1}\right)^n e^{-\frac{1}{\mu_1}\sum_{i=1}^n x_i} \Leftrightarrow \sum_{i=1}^n x_i > n\mu_1 \cdot \ln 2\,. \tag{5.18}$$

This comparison is sufficient, if we can show, that there is no global maximum between $2\mu_1$ and $+\infty$. Therefore, we analyse the monotone behaviour of $L(b;\boldsymbol{x})$ by differentiation with respect to b:

$$\begin{aligned}&\frac{\partial}{\partial b}L(b;\boldsymbol{x}) = \\ &= \left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)b}-1}\right)^n e^{\lambda(b,\mu_1)\sum x_i}\left[\frac{\partial\lambda(b,\mu_1)}{\partial b}(\sum x_i - n\mu_1) - n\frac{\lambda(b,\mu_1)e^{\lambda(b,\mu_1)b}}{e^{\lambda(b,\mu_1)b}-1}\right]\end{aligned} \tag{5.19}$$

Since

$$\begin{aligned}\frac{\lambda(b,\mu_1)e^{\lambda(b,\mu_1)b}}{e^{\lambda(b,\mu_1)b}-1} &= f_{X|\{(b,\mu_1)\}}(b) > 0\,,\\ \frac{\partial\lambda(b,\mu_1)}{\partial b} &< 0 \text{ (since } \lambda(b,\mu_1) = \lambda(\mu_1/b) \text{)},\\ \text{and } \left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)b}-1}\right)^n e^{\lambda(b,\mu_1)\sum x_i} &> 0\,,\end{aligned}$$

the sign of $\frac{\partial}{\partial b}L(b;\boldsymbol{x})$ for a certain value of b depends only on $\sum x_i$. Let $\sum x_i \geq n\mu_1$, then $\frac{\partial}{\partial b}L(b;\boldsymbol{x}) < 0$ for all $b \geq 2\mu_1$, i.e., $L(b;\boldsymbol{x})$ is monotone decreasing. Thus, the MLE is given by $\hat{b} = 2\mu_1$.
If only equation (5.18), i.e., $\sum x_i > n\mu_1 \cdot \ln 2$, holds, we have

$$\frac{\partial\lambda(b,\mu_1)}{\partial b} \overset{b=2\mu_1}{=} -\frac{3}{2}\frac{1}{\mu_1^2} \tag{5.20}$$

and with $f_{X|\{(2\mu_1,\mu_1)\}}(2\mu_1) = \frac{1}{2\mu_1}$ we get

$$\begin{aligned}&\left[\frac{\partial\lambda(b,\mu_1)}{\partial b}\left(\sum x_i - n\mu_1\right) - n\,\frac{\lambda(b,\mu_1)e^{\lambda(b,\mu_1)b}}{e^{\lambda(b,\mu_1)b}-1}\right] = \\ &\overset{b=2\mu_1}{=} -\frac{3}{2}\frac{\sum x_i}{\mu_1^2} + n\frac{1}{\mu_1} < -\frac{3}{2}\frac{n\mu_1\ln 2}{\mu_1^2} + n\frac{1}{\mu_1} = \frac{n}{\mu_1}(1-\ln\sqrt{8}) < 0\,,\end{aligned} \tag{5.21}$$

that is, $L(b;\boldsymbol{x})$ is monotone decreasing in $b = 2\mu_1$. Like in example 5.7, if $n\mu_1 > \sum x_i > n\mu_1 \cdot \ln 2$ holds, i.e.,

$$L(2\mu_1;\boldsymbol{x}) = \left(\frac{1}{2\mu_1}\right)^n < \left(\frac{1}{\mu_1}\right)^n e^{-\frac{1}{\mu_1}\sum x_i} = \lim_{b\to+\infty} L(b;\boldsymbol{x})\,, \tag{5.22}$$

then there is a local minimum. Examples 5.6 and 5.7 suggest, that this is also a global minimum and that there are no other local minima. This would lead to the conclusion, that in these cases, the comparison of the values of $L(b;\boldsymbol{x})$ in $b = 2\mu_1$ and for $b \to +\infty$ is sufficient for identifying the MLE $\hat{b}$: it is either $2\mu_1$ or $+\infty$.

On the other hand, $L(b;\boldsymbol{x})$ is monotone increasing for $b = 2\mu_1$, if

$$-\frac{3}{2}\frac{\sum x_i}{\mu_1^2} + n\frac{1}{\mu_1} > 0 \;\Leftrightarrow\; \sum x_i < \frac{2}{3}n\mu_1\,. \tag{5.23}$$

Then, $\sum x_i - n\mu_1 < \sum x_i - \frac{2}{3}n\mu_1 < 0$ and example 5.8 indicates, that $L(b;\boldsymbol{x})$ is monotone increasing for all $b \geq 2\mu_1$.
Finally, it is in fact only important, which of the two values $L(2\mu_1;\boldsymbol{x})$ and $\lim_{b\to+\infty} L(b;\boldsymbol{x})$ is the largest one, yielding the MLE either $\hat{b} = 2\mu_1$ or $\hat{b} = +\infty$.

$\max x_i \geq 2\mu_1$

Here, we have a further restriction on the potential values of b by $b \geq \max x_i$. But since, $\max x_i$ has no influence on the value of $L(b;\boldsymbol{x})$ if only $b \geq \max x_i$ holds, the analysis is almost identical to the preceding paragraph 5.1.2.2. We have to compare the values $L(\max x_i;\boldsymbol{x})$ and $\lim_{b\to+\infty} L(b;\boldsymbol{x})$, yielding for the MLE either $\hat{b} = \max x_i$ or $\hat{b} = +\infty$.

Further restrictions on the potential values of b

Next assume that there is more knowledge about the potential values of b given by $b \in [b^\ell, b^u] \subset [2\mu_1, +\infty)$. Then for the analysis of $L(b;\boldsymbol{x})$ the same holds again as in paragraph 5.1.2.2. The course of $L(b;\boldsymbol{x})$ is not affected by the further restriction if only $\max\{\max x_i, b^\ell\} \leq b \leq b^u$, i.e., only the values of $L(b;\boldsymbol{x})$ in the boundary points determine the MLE:

- Either $L(\max\{\max x_i, b^\ell\};\boldsymbol{x}) > L(b^u;\boldsymbol{x}) \Rightarrow \hat{b} = \max\{\max x_i, b^\ell\}$

- or $L(\max\{\max x_i, b^\ell\}; \boldsymbol{x}) < L(b^u; \boldsymbol{x}) \Rightarrow \hat{b} = b^u$.

We have seen, that the MLE for b behaves in a not very welcomed way. By some small variation of $\max x_i$, the MLE jumps from $\max x_i$ directly to $+\infty$. Thus, we conclude, that the Maximum Likelihood Method is not really suitable to derive a point estimator for the upper bound b of $\mathcal{X}$.

5.1.2.3. $\max X_i$ a Natural Point Estimator of b?

Although the MLE for b is not suitable, we may get the idea just to take $\max X_i$ as an estimator for b. Of course, this would be a very natural estimator for an upper bound of the range of variability.

Its expected value is

$$\mathrm{E}_b[\max X_i] = b - \int_0^b \left(\frac{e^{\lambda(0,b,\mu_1)x} - 1}{e^{\lambda(0,b,\mu_1)b} - 1} \right)^n \mathrm{d}x \tag{5.24}$$

$$= b \left(1 - \int_0^1 \left(\frac{e^{\lambda(0,1,\mu_1/b)y} - 1}{e^{\lambda(0,1,\mu_1/b)} - 1} \right)^n \mathrm{d}y \right) \qquad < b \tag{5.25}$$

Unsurprisingly, $\max X_i$ is not an unbiased estimator for b. As generally known, $\max X_i$ always underestimates the upper bound b of the range of variability (see example 4.24 in the case of an uniform distribution $\mathrm{U}(0, b)$), but the bias

$$\mathrm{bias}_b[\max X_i] = b \int_0^1 \left(\frac{e^{\lambda(0,1,\mu_1/b)y} - 1}{e^{\lambda(0,1,\mu_1/b)} - 1} \right)^n \mathrm{d}y \tag{5.26}$$

decreases with increasing sample size n. The relative bias of $\mathrm{E}_b[\max X_i]$ from b is given by

$$\int_0^1 \left(\frac{e^{\lambda(0,1,\mu_1/b)y} - 1}{e^{\lambda(0,1,\mu_1/b)} - 1} \right)^n \mathrm{d}y \tag{5.27}$$

which depends on the ratio $\frac{\mu_1}{b}$ and on n. In table 5.1 we illustrate this dependency by some numerical examples. For $b = 2\mu_1$ we have in fact the uniform distribution $\mathrm{U}(0, b)$, i.e., as known from example 4.24, the bias of $\max X_i$ is $\frac{n}{n+1}b$ and the relative bias is $\frac{1}{n+1}$, respectively. Although, the relative bias is decreasing with increasing n, the estimator $\max X_i$ will often yield bad estimates if b is large compared to μ_1. The explanation is that the probability

$$\mathrm{P}_{\max X_i|\{(0,b,\mu_1)\}}(\max X_i > y) = 1 - (F_{X|\{(0,b,\mu_1)\}}(y))^n \tag{5.28}$$

	$\frac{\mu_1}{b}$				
n	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{10}$	$\frac{1}{20}$	$\frac{1}{50}$
10	0.090909	0.339591	0.707338	0.853552	0.941421
20	0.047619	0.233904	0.640737	0.820113	0.928045
50	0.019608	0.128774	0.551358	0.77504	0.910016
100	0.009901	0.0760034	0.483682	0.740631	0.896252
500	0.001996	0.0184201	0.330119	0.66036	0.864144

Table 5.1.: Relative bias of $\mathrm{E}_b[\max X_i]$ from b for different values of n and $\frac{\mu_1}{b}$.

	y			
b	$0.8b$	$0.9b$	$0.95b$	$0.99b$
4	0.77903	0.459449	0.243526	0.0504449
10	0.0144514	0.00390817	0.00147746	0.000239696
20	$< 10^{-5}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-7}$
50	≈ 0	≈ 0	≈ 0	≈ 0

Table 5.2.: Probability that $\max X_i$ exceeds the value y, i.e., $\mathrm{P}_{\max X_i|\{(0,b,\mu_1)\}}(\max X_i > y)$, in dependency of b with $n = 50$ and $\mu_1 = 1$.

rapidly decreases with y, which is illustrated in table 5.2 with $n = 50$ and $\mu_1 = 1$. That is, even with considerable large sample size n it happens very seldom, that we observe a sample with maximal value somewhere near the unknown upper bound b.

5.2. Set Estimation

The traditional point estimators for μ_1 and the upper bound b, respectively, reveal some poor characteristics like biasedness or an abrupt change of its value. Therefore, the traditional set estimators which are based on point estimators would surely inherit this poorness, and we will not work out any further analysis.

Those set estimators based on the inversion of acceptance regions would be worth analysing, but this task should be postponed to Chapter 7, where they serve as comparisons to the next presented Neyman measurement procedure. A Bayesian set estimator will be discussed in Chapter 9.

6. Neyman Measurement Procedures

In the previous chapters we gave an overview over the most common approaches in classical statistics to estimate the unknown value ϑ of a distribution parameter D. We have seen, that for the monotonic distribution, which is based on qualitative knowledge about its monotonicity and quantitative knowledge about its range of variability and its first moment, these traditional approaches may yield useless results.

In this chapter we want to present the approach developed by von Collani and Dumitrescu (2001) [16] and von Collani et al. (2001) [17]. In contrast to the classical statistics, the authors embed their approach to estimation in the metrological framework and, therefore, denote them as *measurement procedures*. This embedding is derived in more details in [22]. Since we have already drawn some parallels between statistics and metrology, this approach seems to be reasonable.

6.1. The Bernoulli-Space

In [19] and [20] the fundamental ideas to this approach are developed in detail. Some of the following explanations may seem to be dispensable, since they may be considered to describe well-established ideas. However, a full comprehension can only be achieved, if we present them here in an appropriate coverage.

6.1.1. The Model

The variables involved in a random phenomenon are the *random variable* X and the *deterministic variable* D. The random variable X describes the quantity of interest, where X may be multi-dimensional, i.e., the quantity of interest may have several components. The values x the random variable X may adopt form

the *range of variability* $\mathcal{X}$. From a realistic point of view, $\mathcal{X}$ has to be a bounded set. Every random, i.e., future development is influenced by some past development, i.e., by fixed *initial conditions*. The determinstic variable D (one- or multi-dimensional) describes these aspects which have influence on the random variable X. The values d of D form the range of D denoted by $\mathcal{D}$. If we have complete knowledge about the past development, the actual value $d = d_0$ of D is known and the range $\mathcal{D}$ is a singleton, i.e., $\mathcal{D} = \{d_0\}$. If ignorance exists about the initial conditions, i.e., we have only partial knowledge about the past, then $\mathcal{D}$ consists of all values d of D which cannot be excluded with certainty. Typically, $\mathcal{D}$ contains a series or interval (hypercube) of values d. Since $\mathcal{D}$ specifies the amount of ignorance, it is called the *ignorance space*.

To indicate that the actual value d of the deterministic variable D influences upon the random variable X and to distinguish different initial conditions yielding different random variables we write $X|\{d\}$ with range of variability $\mathcal{X}(\{d\})$. For the sake of simplicity we consider $\mathcal{X}(\{d\})$ as a subset of $\mathbb{R}^k$, $k \geq 1$. Then, the *variability function* $\mathcal{X}$ describes the relation between the initial conditions and the future's variability:

$$\mathcal{X} \colon \mathbb{B}_{\mathcal{D}} \to \mathbb{B}_{\mathbb{R}^k} \tag{6.1}$$

where $\mathbb{B}_{\mathcal{D}}$ and $\mathbb{B}_{\mathbb{R}^k}$ are the Borel σ-fields over the respective set.

Not only the range of variability $\mathcal{X}(\{d\})$ of $X|\{d\}$ is influenced by the initial conditions, but also the structure of randomness, i.e., the probabilities of events $E \in \mathbb{B}_{\mathbb{R}^k}$. Denoting the set of all probability measures with $\mathbb{P}$, we may define the *random structure function* $\mathcal{P}$ which maps each value $d \in \mathcal{D}$ onto the probability measure of $X|\{d\}$:

$$\mathcal{P} \colon \mathcal{D} \to \mathbb{P} \tag{6.2}$$

with $\mathcal{P}(d) = \mathrm{P}_{X|\{d\}}$ and $\mathrm{P}_{X|\{d\}}$ the probability measure, i.e., the probability distribution of $X|\{d\}$. The set of all probability distribution of $X|\{d\}$ for $d \in \mathcal{D}$ is denoted $\mathcal{P}_{X,D}$:

$$\mathcal{P}_{X,D} = \{\mathrm{P}_{X|\{d\}} \mid d \in \mathcal{D}\} \tag{6.3}$$

With the ignorance space $\mathcal{D}$, the variability function $\mathcal{X}$ and the random structure function $\mathcal{P}$, the *Bernoulli-Space* $\mathbb{B}_{X,D}$ for the pair of variables (X, D) is defined

as follows:

$$\mathbb{B}_{X,D} = (\mathcal{D}, \mathcal{X}, \mathcal{P}) \tag{6.4}$$

Example 6.1 *Consider the pair of variables* (X, D) *where the determinsitic variable* D *consists of three components* $D = (A, B, \mathrm{E}[X])$ *with*

- A *specifying the lower bound of the range of variability,*
- B *specifying the upper bound of the range of variability and*
- $\mathrm{E}[X]$ *specifying the expectation of the random variable* X*.*

The value of D *is denoted* $d = (a, b, \mu_1)$*. The ignorance space* $\mathcal{D}$ *should be*

$$\mathcal{D} = \left\{ (a, b, \mu_1) \;\middle|\; 0 \leq a < b \leq u, a < \mu_1 < \frac{a+b}{2} \right\} \tag{6.5}$$

with some value $u > 0$*. The variability function* $\mathcal{X}$ *is specified by its images*

$$\mathcal{X}(\{(a, b, \mu_1)\}) = \{x \,|\, a \leq x \leq b\}\,. \tag{6.6}$$

The determinstic variable D *also specifies the probability distribution of* X *through the random structure function* $\mathcal{P}$ *and its images, respectively,*

$$\mathcal{P}(a, b, \mu_1) = Mon(a, b, \mu_1)\,. \tag{6.7}$$

That is, the random variable $X|\{(a, b, \mu_1)\}$ *is distributed according to the monotonic distribution* $Mon(a, b, \mu_1)$*. Since* $a < \mu_1 < \frac{a+b}{2}$*, the density functions* $f_{X|\{(a,b,\mu_1)\}}$ *are monotonously decreasing on* $\mathcal{X}(\{(a, b, \mu_1)\})$*.*

6.1.2. The Sources of Uncertainty

Humans are permanently concerned on the one hand with lack of complete knowledge, i.e., ignorance about already terminated developments, e. g. physical measures like the weight of some amount of corn. And on the other hand with the not deterministic, i.e., random behaviour of future developments, e. g. world's climate development or the next day's weather. That is, ignorance and randomness are the sources of *uncertainty about future developments*, and both are contained in the Bernoulli-Space: ignorance refers to D and is represented by $\mathcal{D}$, randomness refers to $X|\{d\}$ and is represented by $\mathcal{X}$ in connection with $\mathcal{P}$.

6.1.2.1. No Ignorance – Complete Knowledge

In the case of no ignorance or equivalent complete knowledge, the ignorance space $\mathcal{D}$ contains only the known actual value d_0 of D, i.e., $\mathcal{D} = \{d_0\}$. Then for the Bernoulli-Space for the pair of variables (X, D) we may omit its functional components and represent it by the triple

$$\left(\{d_0\}, \mathcal{X}(\{d_0\}), \mathrm{P}_{X|\{d_0\}}\right) \tag{6.8}$$

where

- $\mathcal{X}(\{d_0\})$ is the range of variability of $X|\{d_0\}$ and
- $\mathrm{P}_{X|\{d_0\}}$ is the probability distribution of $X|\{d_0\}$.

Although there is no ignorance, uncertainty stills exists through randomness described by $\mathcal{X}(\{d_0\})$ and $\mathrm{P}_{X|\{d_0\}}$.

6.1.2.2. No randomness

In the case of no randomness, we would have a deterministic cause-effect relation between the initial conditions and the future development. For each $d \in \mathcal{D}$ the range of variability $\mathcal{X}(\{d\})$ would contain only one element, say $x(\{d\})$, and the probability distribution $\mathrm{P}_{X|\{d\}}$ would degenerate to a one-point distribution, i.e., $\mathrm{P}_{X|\{d\}}(E) = \mathbb{1}_E(x(\{d\}))$ for every event E with respect to $X|\{d\}$. Uncertainty in this case only arises through ignorance about the value of D.

6.1.2.3. Ignorance and Randomness

Ignorance refers in the first instance to the deterministic variable D in the way that its actual value d_0 is not known, thus, the ignorance space $\mathcal{D}$ contains more than one element. Consequently, the image sets of the variability function $\mathcal{X}$, i.e., $\{\mathcal{X}(\{d\}) \,|\, d \in \mathcal{D}\}$, and of the random structure function $\mathcal{P}$, i.e., $\mathcal{P}_{X,D}$, may also contain more than one element. The size of $\mathcal{D}$, i.e., number of elements or volume, may be interpreted as the degree of ignorance. But it seems to be reasonable to assume that in all real world scenarios the ignorance is somehow

bounded, i.e., that some knowledge about the possible values of D exists – in this sense ignorance may be understood as incomplete knowledge.

But ignorance also refers to the variability function $\mathcal{X}$, which assigns to each value $d \in \mathcal{D}$ a range of variability $\mathcal{X}(\{d\})$ of the $X|\{d\}$. Only in very rare situations, the range of variability and the variability function $\mathcal{X}$, respectively, are known exactly. In the more frequent situations this is not the case, thus, the image $\mathcal{X}(\{d\})$ of the variability function $\mathcal{X}$ for each $d \in \mathcal{D}$ should at least cover the true range of variability.

Given the ignorance space $\mathcal{D}$, the overall range of variability $\mathcal{X}(\mathcal{D})$ for X is given by

$$\mathcal{X}(\mathcal{D}) = \bigcup_{d \in \mathcal{D}} \mathcal{X}(\{d\})\,. \tag{6.9}$$

Since every future event is a subset of $\mathcal{X}(\mathcal{D})$, it is called *prediction space*.

Both, the ranges of variability $\mathcal{X}(\{d\})$ for each $d \in \mathcal{D}$ and the ignorance space $\mathcal{D}$ represent the range of uncertainty and together build up the *uncertainty space* $\mathcal{U}_{X,D}$ as follows

$$\mathcal{U}_{X,D} = \bigcup_{d \in \mathcal{D}} \{d\} \times \mathcal{X}(\{d\}) = \{(d, x) \,|\, d \in \mathcal{D}, x \in \mathcal{X}(\{d\})\}\,. \tag{6.10}$$

Example 6.2 *Consider the following Bernoulli-Space $\mathbb{B}_{X,D} = (\mathcal{D}, \mathcal{X}, \mathcal{P})$ for the pair of variables (X, D) with*

- $\mathcal{D} = \{(a, b, \mu_1) \,|\, a = 0, b = 1, 0 < m_\ell \leq \mu_1 \leq m_u < 0.5\}$
- $\mathcal{X}(\{(0, 1, \mu_1)\}) = \{x \,|\, 0 \leq x \leq 1\} = [0, 1]$, *i.e., independent of the value μ_1 of* $\mathrm{E}[X]$
- $\mathcal{P}(0, 1, \mu_1) = Mon(0, 1, \mu_1)$

That is, we have ignorance only about the value μ_1 of $\mathrm{E}[X]$*. About the other components of D, i.e., the range of variability, we have complete knowledge. Therefore, we consider* $\mathrm{E}[X]$ *as the determinstic variable. Thus, the ignorance space $\mathcal{D}$ is the set of values of* $\mathrm{E}[X]$*: $\mathcal{D} = \{\mu_1 \,|\, 0 < m_\ell \leq \mu_1 \leq m_u < 0.5\}$. The uncertainty space then is*

$$\mathcal{U}_{X,\mathrm{E}[X]} = \bigcup_{\mu_1 \in \mathcal{D}} \{\mu_1\} \times \mathcal{X}(\{\mu_1\}) = [m_\ell, m_u] \times [0, 1]\,, \tag{6.11}$$

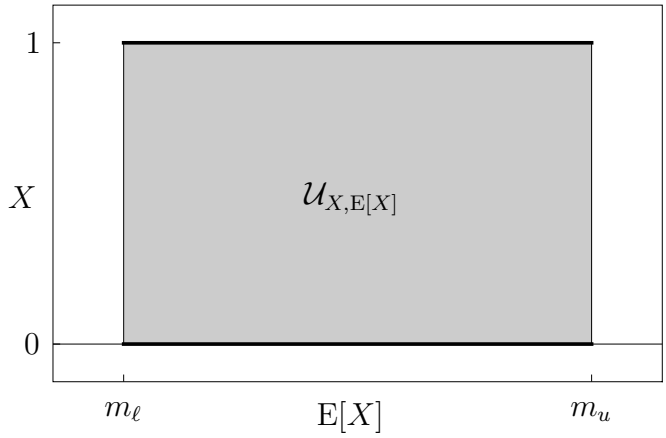


Figure 6.1.: Uncertainty space $\mathcal{U}_{X,\mathrm{E}[X]}$, where $X|\{\mu_1\} \sim Mon(0,1,\mu_1)$ and $\mu_1 \in \mathcal{D} = \{\mu_1 \,|\, 0 < m_\ell \leq \mu_1 \leq m_u < 0.5\}$. Because there is only ignorance about the expectation $\mathrm{E}[X]$ of X, we consider $\mathrm{E}[X]$ as the deterministic variable.

that is, a rectangle in the $\mathrm{E}[X]$*-*X*-plane displayed in figure 6.1.*

Example 6.3 *Next consider the pair of variables* (X, D) *and the corresponding Bernoulli-Space* $\mathbb{B}_{X,D} = (\mathcal{D}, \mathcal{X}, \mathcal{P})$ *for the case that there is ignorance only about the upper bound* B *of the range of variability and complete knowledge about the lower bound* A *with value* $a = 0$ *and the expectation* $\mathrm{E}[X]$ *with value* $\mu_1 = m$*. Thus, we consider* B *as the deterministic variable. We get*

- $\mathcal{D} = \{b \,|\, 0 < b_\ell \leq b \leq b_u\}$ *with* $b_\ell > 2m$
- $\mathcal{X}(\{(0, b, m)\}) = \{x \,|\, 0 \leq x \leq b\} = [0, b]$
- $\mathcal{P}(0, b, m) = Mon(0, b, m)$

The uncertainty space displayed in figure 6.2 then is

$$\mathcal{U}_{X,B} = \bigcup_{b \in \mathcal{D}} \{b\} \times \mathcal{X}(\{b\})\,. \tag{6.12}$$

Figure 6.3 displays the uncertainty space together with some density functions $f_{X|\{b\}}$*.*

In both examples 6.2 and 6.3 the random structure function $\mathcal{P}$ maps the values of the respective determinstic variable on a member of the family of monotonic

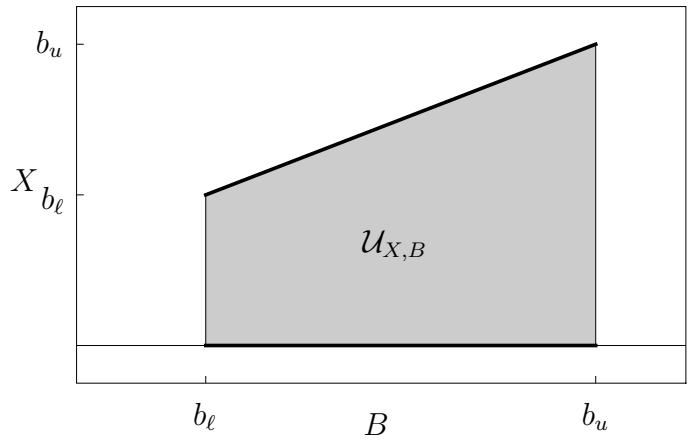


Figure 6.2.: Uncertainty space $\mathcal{U}_{X,B}$, where $X|\{b\} \sim Mon(0, b, m)$, $b \in \mathcal{D} = \{b \,|\, 0 < b_\ell \leq b \leq b_u\}$ with $b_\ell > 2m$. Because there is only ignorance about the upper bound B of the range of variability, we consider B as the deterministic variable.

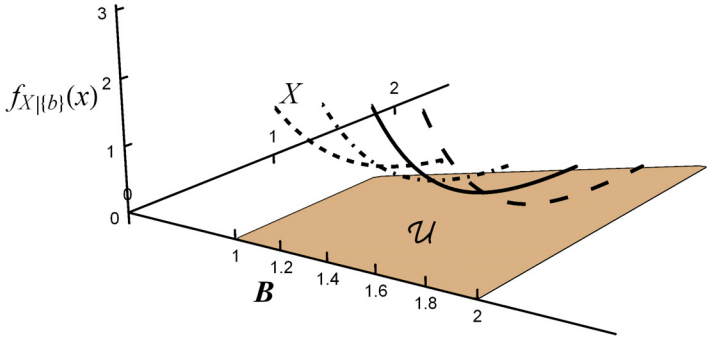


Figure 6.3.: Uncertainty space $\mathcal{U}_{X,B}$ from figure 6.2 with some density functions $f_{X|\{b\}}$ for different values b of B.

distributions $Mon(a, b, \mu_1)$. This is due to the fact, that ignorance has to be incorporated also into the structure of randomness. In Chapter 2 we have presented a general concept of the derivation of probability distributions based on available qualitative knowledge. This qualitative knowledge constitutes which quantitative knowledge about the range of variability and the number of moments is necessary to determine the probability distribution. That is, a generic deterministic variable consists of the range of variability (in fact, its bounds) and a certain number of moments. Having the range of variability and the number of moments at hand, the random structure function maps each value on a member of the corresponding family of minimum information representatives. In the case of the aboves examples 6.2 and 6.3, available knowledge refers to the monotone behaviour of the density function and, thus, that the random structure function maps each value of the deterministic variable on a monotonic distribution.

In contrast to models used in classical statistics, the choice of probability distribution and random structure function, respectively, is only based on the available knowledge. In particular, it is not governed by some implicit considerations about the (analytic) feasibility of the problem at hand. In cases where the choice of a statistical distribution is justified as a sufficiently good approximation, the deterministic variable would consist of the distribution parameters. But since for many statistical distributions, their parameters may be substituted by a certain number of moments, we also might keep the moments as components of a generic determinstic variable.

6.2. Reliable Stochastic Procedures

Stochastic procedures aim at reducing uncertainty given by the uncertainty space (6.10). Since uncertainty has two sources we distinguish between prediction procedures and measurement procedures:

- *Prediction procedures* aim at predicting the future outcome of X in form of a subset of $\mathcal{X}(\mathcal{D})$ based on the knowledge or ignorance given by $\mathcal{D}$.
- *Measurement procedures* aim at measuring the unknown actual value d_0 of D in form of a subset of $\mathcal{D}$ based on an observed value of X.

Formally, prediction and measurement procedures define mappings between $\mathbb{B}_{\mathcal{D}}$ and $\mathbb{B}_{\mathcal{X}(\mathcal{D})}$:

- Prediction procedure:

$$A_X : \mathbb{B}_{\mathcal{D}} \to \mathbb{B}_{\mathcal{X}(\mathcal{D})} \tag{6.13}$$

- Measurement procedure:

$$C_D : \mathbb{B}_{\mathcal{X}(\mathcal{D})} \to \mathbb{B}_{\mathcal{D}} \tag{6.14}$$

The main requirement a stochastic procedure should meet is the predefined *reliability*. As introduced at the beginning of Chapter 4, the reliability of a procedure stands for its ability to yield correct results. For a prediction procedure a prediction, i.e., a subset $A_X(\mathcal{D}_0)$ of $\mathcal{X}(\mathcal{D})$ should be considered as correct if the outcome x of X is element of $A_X(\mathcal{D}_0)$, where $\mathcal{D}_0$ is a subset of $\mathcal{D}$. The measurement $C_D(\mathcal{X}_0) \subset \mathcal{D}$ of a measurement procedure should be considered as correct if the actual value d_0 of D is element of $C_D(\mathcal{X}_0)$, where $\mathcal{X}_0$ is a subset of $\mathcal{X}(\mathcal{D})$, i.e., an event. The reliability of a procedure is then defined as the probability that it will yield a correct result when applied. The lower bound of this probability is called *reliability level* and denoted by β[1]. To derive a stochastic procedure with a certain realibility level β it is necessary to take the random structure, i.e., the probability distributions $\mathcal{P}_{X,D}$ into account.

6.2.1. β-Prediction and β-Measurement

The derivation of β-measurement procedures introduced in [20] has been originally inspired by J. Neyman [57], since the idea is very similar to that of the inversion of acceptance regions in section 4.4.3. The start of this derivation is to define β-predictions. Since the final aim is to develop β-measurement procedures, we concentrate on introducing β-prediction procedures only for the subsystem of sets $\big\{\{d\} \,|\, d \in \mathcal{D}\big\}$ of $\mathbb{B}_{\mathcal{D}}$.

[1] In classical statistics this would be the confidence level of an estimation procedure.

6.2.1.1. β-Prediction Procedure

Definition 6.4 (β-prediction procedure, β-uncertainty space of X) *Let* $\mathcal{I} = \{\{d\} \mid d \in \mathcal{D}\}$. *A mapping*

$$A_X^{(\beta)} : \mathcal{I} \to \mathbb{B}_{\mathcal{X}(\mathcal{D})} \tag{6.15}$$
$$\text{with } A_X^{(\beta)}(\{d\}) \subset \mathcal{X}(\{d\}) \text{ for } d \in \mathcal{D}$$

is called β*-prediction procedure for* X *if*

$$\mathrm{P}_{X|\{d\}}(A_X^{(\beta)}(\{d\}) \geq \beta \text{ for } d \in \mathcal{D}\,. \tag{6.16}$$

The set

$$\mathcal{U}_X^{(\beta)} = \{(d, x) \mid d \in \mathcal{D}, x \in A_X^{(\beta)}(\{d\})\} \tag{6.17}$$

is called β*-uncertainty space of* X.

Obviously, for all β-uncertainty spaces we have $\mathcal{U}_X^{(\beta)} \subset \mathcal{U}_{X,D}$.

We note, that for β-prediction procedures (6.15) the reliability requirement is met by condition (6.16). But besides this we also note, that the definition does not uniquely define a β-prediction procedure, since the selection of the predictions $A_X^{(\beta)}(\{d\})$ for each $d \in \mathcal{D}$ is not further specified.

Example 6.5 *Consider a Bernoulli-Space for the pair of variables* (X, D). *Again,* X *is distributed according to the monotonic distribution* $Mon(a, b, \mu_1)$. *Assuming* $a = 0$ *and* $b = 1$, *i.e., there is no ignorance about the range of variability, and only ignorance about* $\mathrm{E}[X]$ *of* X, *say* $0.05 \leq \mu_1 \leq 0.45$. *Thus, the ignorance space is given by* $\mathcal{D} = \{\mu_1 \mid 0.05 \leq \mu_1 \leq 0.45\}$. *Then, the set of probability distributions* $\mathcal{P}_{X,\mathrm{E}[X]}$ *contains only monotonic distributions* $Mon(\mu_1)$ *with decreasing density functions* $f_{X|\{\mu_1\}}(x)$ *for* $x \in \mathcal{X}(\{mu_1\}) = [0, 1]$. *The uncertainty space* $\mathcal{U}_{X,\mathrm{E}[X]}$ *is a rectangle in the* $\mathrm{E}[X]$-X*-plane:*

$$\mathcal{U}_{X,\mathrm{E}[X]} = [0.05, 0.45] \times [0, 1] \tag{6.18}$$

To illustrate that the predictions $A_X^{(\beta)}(\{\mu_1\})$ *for* $\mu_1 \in \mathcal{D}$ *are not unique, we display three different possibilities. Each yielding a different* β*-uncertainty space* $\mathcal{U}_X^{(\beta)}$ *for*

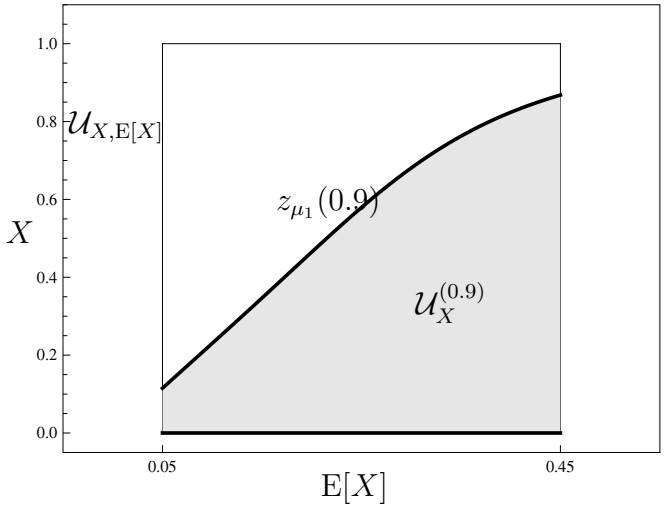


Figure 6.4.: Uncertainty space $\mathcal{U}_{X,\mathrm{E}[X]}$ as in example 6.2 with $\mu_1 \in \mathcal{D} = [0.05, 0.45]$. For $\beta = 0.9$ and for each value μ_1 of $\mathrm{E}[X]$ the predictions are defined by $A_X^{(0.9)}(\{\mu_1\}) = [0, z_{\mu_1}(0.9)]$.

X. *The upper quantile function* $z_{\mu_1}(\gamma)$ *for* $X|\{\mu_1\} \sim Mon(0, 1, \mu_1)$ *is defined as usual:*

$$z_{\mu_1}(\gamma) = F^{-1}_{X|\{\mu_1\}}(\gamma) \text{ for } \gamma \in (0, 1) \tag{6.19}$$

$$z_{\mu_1}(0) = 0 \tag{6.20}$$

$$z_{\mu_1}(1) = 1 \tag{6.21}$$

Note that we have to consider $z_{\mu_1}(\gamma)$ *also as a function of* μ_1. *Figures 6.4, 6.5 and 6.6 display three reasonable* β*-uncertainty spaces* $\mathcal{U}_X^{(\beta)}$.

6.2.1.2. β-Measurement Procedure

We start with the general definition of a β-measurement procedure. It is similiar to definition 4.42 of confidence set estimators but with the newly introduced terminology and notation.

Definition 6.6 (β-measurement procedure, β-uncertainty space of $\boldsymbol{D}$) *Let*

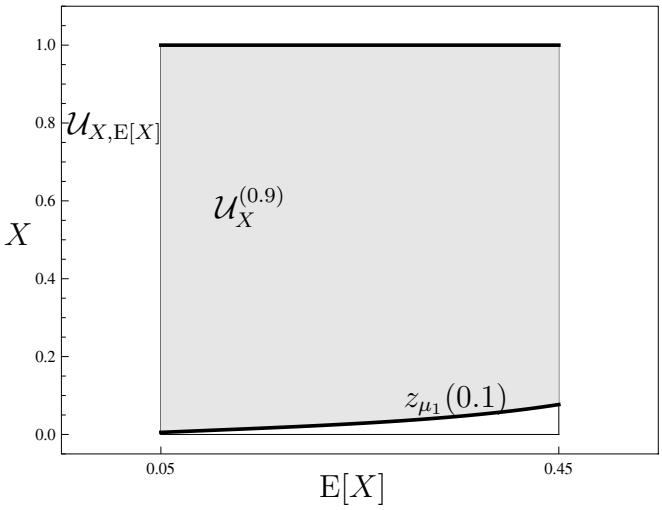


Figure 6.5.: Uncertainty space $\mathcal{U}_{X,\mathrm{E}[X]}$ as in example 6.2 with $\mu_1 \in \mathcal{D} = [0.05, 0.45]$. For $\beta = 0.9$ and for each value μ_1 of $\mathrm{E}[X]$ the predictions are defined by $A_X^{(0.9)}(\{\mu_1\}) = [z_{\mu_1}(0.1), 1]$.

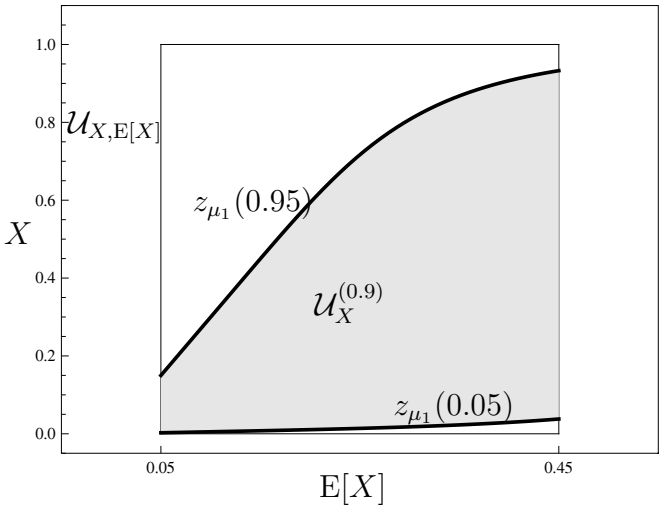


Figure 6.6.: Uncertainty space $\mathcal{U}_{X,D}$ as in example 6.2 with $\mu_1 \in \mathcal{D} = [0.05, 0.45]$. For $\beta = 0.9$ and for each value μ_1 of $\mathrm{E}[X]$ the predictions are defined by $A_X^{(0.9)}(\{\mu_1\}) = [z_{\mu_1}(0.05), z_{\mu_1}(0.95)]$.

$\mathcal{E} = \{\{x\} \,|\, x \in \mathcal{X}(\mathcal{D})\}$. *A mapping*

$$C_D^{(\beta)} : \mathcal{E} \to \mathbb{B}_{\mathcal{D}} \quad \text{with } C_D^{(\beta)}(\{x\}) \subset \mathcal{D} \text{ for } x \in \mathcal{X}(\mathcal{D}) \tag{6.22}$$

is called β-measurement procedure for D if

$$\mathrm{P}_{X|\{d\}}(\{x \,|\, d \in C_D^{(\beta)}(\{x\})\}) \geq \beta \text{ for } d \in \mathcal{D}\,. \tag{6.23}$$

The set

$$\mathcal{U}_D^{(\beta)} = \{(d, x) \,|\, x \in \mathcal{X}(\mathcal{D}), d \in C_D^{(\beta)}(\{x\})\} \tag{6.24}$$

is called β-uncertainty space of D.

As in section 4.4.3 a β-measurement procedure based on a β-prediction procedure is derived as follows:

Theorem 6.7 (duality of β-prediction procedure and β-measurement procedure) *Let $A_X^{(\beta)}$ be a β-prediction procedure according to definition 6.4 and*

$$\{A_X^{(\beta)}(\{d\}) \,|\, d \in \mathcal{D}\} \tag{6.25}$$

the resulting set of β-predictions. If

$$\bigcup_{d \in \mathcal{D}} A_X^{(\beta)}(\{d\}) = \mathcal{X}(\mathcal{D}) \tag{6.26}$$

then for each $x \in \mathcal{X}(\mathcal{D})$ the sets

$$\{d \in \mathcal{D} \,|\, x \in A_X^{(\beta)}(\{d\})\} \tag{6.27}$$

are non-empty subsets of $\mathcal{D}$ and are the images of a β-measurement procedure $C_D^{(\beta)}$, i.e.,

$$C_D^{(\beta)}(\{x\}) = \{d \in \mathcal{D} \,|\, x \in A_X^{(\beta)}(\{d\})\}\,. \tag{6.28}$$

Non-emptiness of $C_D^{(\beta)}(\{x\})$ for all $x \in \mathcal{X}(\mathcal{D})$ arises directly through condition (6.26). The reliability requirement is also met through

$$\beta \leq \mathrm{P}_{X|\{d\}}(A_{X|\{d\}}^{(\beta)}(\{d\})) = \mathrm{P}_{X|\{d\}}(\{x \,|\, d \in C^{(\beta)}(\{x\})\})\,. \tag{6.29}$$

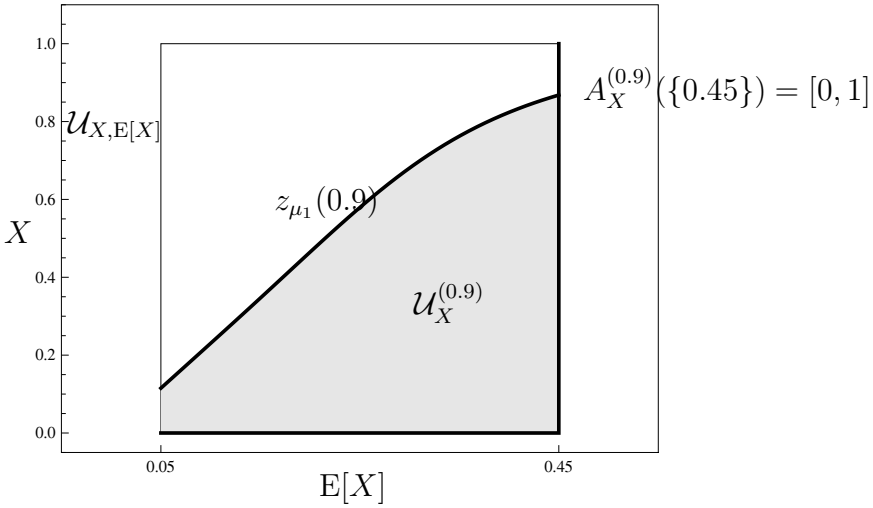


Figure 6.7.: The adjusted β-uncertainty space $\mathcal{U}_X^{(\beta)}$ for X yields a β-measurement procedure $C_{\mathrm{E}[X]}^{(\beta)}$ which assigns to all $x \in \mathcal{X}(\mathcal{D}) = [0,1]$ a non-empty subset $C_{\mathrm{E}[X]}^{(\beta)}(\{x\})$ of $\mathcal{D}$.

Example 6.8 *None of the β-uncertainty spaces $\mathcal{U}_X^{(\beta)}$ for X in example 6.5 yields a β-measurement procedure which assures non-emptiness of $C_{\mathrm{E}[X]}^{(\beta)}(\{x\})$ for all $x \in \mathcal{X}(\mathcal{D})$. A nearby idea is to assign for all x where $C_{\mathrm{E}[X]}^{(\beta)}(\{x\}) = \emptyset$ a certain fixed value $m_0 \in \mathcal{D}$ (see the remarks to theorem 6.11). In the case of $\mathcal{U}_X^{(\beta)}$ in figure 6.4 this occurs for values $x > z_{0.45}(0.9)$. It makes sense to assign to all such values x the value 0.45 of $\mathrm{E}[X]$ as the measurement result: $C_{\mathrm{E}[X]}^{(\beta)}(\{x\}) = \{0.45\}$ for all $x > z_{0.45}(0.9)$. With regard to the prediction procedure, this can be achieved by adding the set $\{0.45\} \times [0,1]$ to the former β-uncertainty space which is displayed in figure 6.7. The result is a β-measurement procedure $C_{\mathrm{E}[X]}^{(\beta)}$ which yields non-empty subsets of $\mathcal{D}$ for all $x \in \mathcal{X}(\mathcal{D})$. For the β-uncertainty spaces displayed in figures 6.5 and 6.6 a completion may be done very similar.*
Figure 6.8 displays a β-uncertainty space $\mathcal{U}_X^{(\beta)}$ based on predictions $A_X^{(\beta)}(\{\mu_1\})$ whose union equals $\mathcal{X}(\mathcal{D})$. Thus, no further adjustments are necessary to obtain a β-measurement procedure $C_{\mathrm{E}[X]}^{(\beta)}$ which yields non-empty meaurement sets $C_{\mathrm{E}[X]}^{(\beta)}(\{x\})$ for all $x \in \mathcal{X}(\mathcal{D})$. The predictions are derived as follows: for $\beta \in (0,1)$ define a function $\gamma(\mu_1)$ depending on $\mu_1 \in \mathcal{D} = [m_\ell, m_u]$ by

$$\gamma(\mu_1) = \frac{1-\beta}{m_u - m_\ell}(\mu_1 - m_\ell) + \beta\,. \tag{6.30}$$

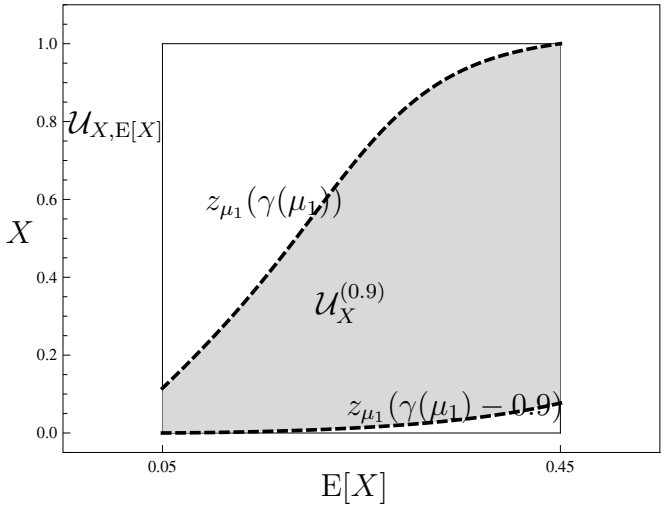


Figure 6.8.: The β-uncertainty space $\mathcal{U}_X^{(\beta)}$ for X needs no further adjustment to yield a β-measurement procedure, since $\bigcup_{\mu_1 \in \mathcal{D}} A_X^{(\beta)}(\{\mu_1\}) = \mathcal{X}(\mathcal{D})$.

That is, $\gamma(\mu_1)$ is a linear increasing function in μ_1 with $\gamma(m_\ell) = \beta$ and $\gamma(m_u) = 1$. With the upper quantile function $z_{\mu_1}(\gamma)$ from example 6.5 we get predictions $A_X^{(0.9)}(\{\mu_1\}) = [z_{\mu_1}(\gamma(\mu_1) - 0.9), z_{\mu_1}(\gamma(\mu_1))]$ for $\mu_1 \in \mathcal{D}$ whose union equals $\mathcal{X}(\mathcal{D}) = [0, 1]$.

Point β-Measurement: As outlined in the introduction to Chapter 4 point measurement procedures, i.e., point estimators, have in general a reliability of zero. In [20] this fact is also stressed and it is concluded that "point measurement should be used only in conjunction with a β-measurement procedure and the value $T_D(x)$ should necessarily be an element of the corresponding β-measurement $C_D^{(\beta)}(\{x\})$. If these two conditions are met, the measurement is called a *point β-measurement*, and is represented as $T_D^{(\beta)}(x)$." That is, a point measurement is in general a pointwise mapping

$$T_D \colon \mathcal{X}(\mathcal{D}) \to \mathcal{D} \tag{6.31}$$

while a point β-measurement additionaly meets the condition that $T_D^{(\beta)}(x) \in C_D^{(\beta)}(\{x\})$ for each $x \in \mathcal{X}(\mathcal{D})$.

Since the use of point measurement is at least disputable, we will not follow up further analysis.

6.3. Neyman Measurement Procedure

Up to now, besides condition (6.26) in theorem 6.7, we have no further guidelines how to construct a suitable β-prediction procedure which leads to a meaningful β-measurement procedure. In [17] a construction is developed by taking the procedure's *precision* into account. That is, by demanding that a procedure should yield minimal measurement sets in the sense of their volume while meeting a given reliability level.

6.3.1. Measures on $\mathcal{U}_{X,D}$

Since β-measurement sets $C_D^{(\beta)}(\{x\})$ and β-prediction sets $A_X^{(\beta)}(\{d\})$ are subsets in $\mathcal{U}_{X,D}$, it is necessary to derive suitable measures on the measurable space

$$(\mathcal{U}_{X,D}, \mathbb{B}_{\mathcal{U}_{X,D}})\,. \tag{6.32}$$

The start might be to find a suitable measure on $(\mathcal{D}, \mathbb{B}_{\mathcal{D}})$. Two arguments yield that it should be based on the Lebesgue measure or the counting measure:

- A measurement procedure should yield reliable and precise measurement sets. Precision of a set is defined by its volume, which is measured by the Lebesgue measure in case of continuous sets and by the counting measure in case of discrete sets.
- A measurement procedure should be equally suitable for any $d \in \mathcal{D}$, that is, no value of D is preferred to another. For a continuous set, only the Lebesgue measure is appropriate in this way, while for a discrete set the counting measure is appropiate.

For convience, in the following only the continuous case is analysed, i.e., only continuous sets and absolute continuous measures are considered.

With

$$|\mathcal{D}| = \int_{\mathcal{D}} \mathrm{d}d\,, \tag{6.33}$$

i.e., the Lebesgue measure of $\mathcal{D}$, let

$$\lambda_{\mathcal{D}}(d) = \frac{1}{|\mathcal{D}|}\mathbb{1}_{\mathcal{D}}(d)\,. \tag{6.34}$$

Then a normed geometrical measure denoted by $\Lambda_{\mathcal{D}}$ on $(\mathcal{D}, \mathbb{B}_{\mathcal{D}})$ is given by:

$$\Lambda_{\mathcal{D}}(\Delta) = \int_{\Delta} \lambda_{\mathcal{D}}(d)\, \mathrm{d}d \quad \text{for any } \Delta \in \mathbb{B}_{\mathcal{D}}\,. \tag{6.35}$$

$\Lambda_{\mathcal{D}}$ may be interpreted as the marginal measure of D of the pair of variables (X, D). Note that $\Lambda_{\mathcal{D}}$ is not at all a probability measure, but only the inverse of a measure of size.

The set of probability distributions $\mathcal{P}_{X,D}$ provides the connection between the measurable sets $(\mathcal{D}, \mathbb{B}_{\mathcal{D}})$ and $(\mathcal{X}(\mathcal{D}), \mathbb{B}_{\mathcal{X}(\mathcal{D})})$. If $\mathrm{P}_{X|\{d\}}(E)$ as a function of $d \in \mathcal{D}$ is measurable for any fixed $E \in \mathbb{B}_{\mathcal{X}(\mathcal{D})}$, then $\mathrm{P}_{X|\{d\}}$ is the conditional probability measure of X under the condition that the value of D is d. This case is denoted by $X|\{d\}$.

Then, the product measure $\Lambda_{\mathcal{D}} \otimes \mathrm{P}_{X|\{d\}}$ on $(\mathcal{U}_{X,D}, \mathbb{B}_{\mathcal{U}_{X,D}})$ is the joint measure of the pair of variables (X, D). Now, let $U \in \mathbb{B}_{\mathcal{U}_{X,D}}$ and its projection into $\mathcal{X}(\mathcal{D})$ with respect to $d \in \mathcal{D}$

$$U_d := \{x \,|\, (d, x) \in U\}\,. \tag{6.36}$$

Therefore, the product measure $\Lambda_{\mathcal{D}} \otimes \mathrm{P}_{X|\{d\}}$ of U is

$$(\Lambda_{\mathcal{D}} \otimes \mathrm{P}_{X|\{d\}})(U) = \int_{\mathcal{D}} \left(\int_{U_d} f_{X|\{d\}}(d)\, \mathrm{d}x \right) \lambda_{\mathcal{D}}(d)\, \mathrm{d}d\,. \tag{6.37}$$

The second possibility is to start with the marginal probability measure of X. The marginal density denoted by $q_{\mathcal{X}(\mathcal{D})}(x)$ of X is

$$q_{\mathcal{X}(\mathcal{D})}(x) \;=\; \int_{\mathcal{D}} f_{X|\{d\}}(x) \lambda_{\mathcal{D}}(d)\, \mathrm{d}d \;=\; \frac{\int_{\mathcal{D}} f_{X|\{d\}}(x)\, \mathrm{d}d}{|\mathcal{D}|} \quad \text{for any } x \in \mathcal{X}(\mathcal{D})\,. \tag{6.38}$$

Then, the marginal probability measure $Q_{\mathcal{X}(\mathcal{D})}$ on $(\mathcal{X}(\mathcal{D}), \mathbb{B}_{\mathcal{X}(\mathcal{D})})$ is

$$Q_{\mathcal{X}(\mathcal{D})}(B) \;=\; \int_{B} q_{\mathcal{X}(\mathcal{D})}(x)\, \mathrm{d}x \quad \text{for any } B \in \mathbb{B}_{\mathcal{X}(\mathcal{D})}\,. \tag{6.39}$$

Defining the *weight function* $w(x)$ of x by

$$w(x) = \int_{\mathcal{D}} f_{X|\{d\}}(x)\, \mathrm{d}d \tag{6.40}$$

we may express $q_{\mathcal{X}(\mathcal{D})}(x)$ by

$$q_{\mathcal{X}(\mathcal{D})}(x) = \frac{w(x)}{|\mathcal{D}|}\,. \tag{6.41}$$

With the marginal measure of X at hand, we obtain a conditional measure of D under the condition that X has the value x, denoted by $D|\{x\}$. The density[2] of $D|\{x\}$ is given for any fixed $x \in \mathcal{X}(\mathcal{D})$ by

$$g_{D|\{x\}}(d) \;=\; \frac{f_{X|\{d\}}(x)\lambda_{\mathcal{D}}(d)}{q_{\mathcal{X}(\mathcal{D})}(x)} \;=\; \frac{f_{X|\{d\}}(x)}{w(x)} \quad \text{for any } d \in \mathcal{D}\,, \tag{6.42}$$

which yields the measure of $D|\{x\}$ on $(\mathcal{D}, \mathbb{B}_{\mathcal{D}})$:

$$G_{D|\{x\}}(\Delta) = \int\limits_{\Delta} g_{D|\{x\}}(d)\;\mathrm{d}d \quad \text{for any } \Delta \in \mathbb{B}_{\mathcal{D}}\,. \tag{6.43}$$

The product measure $Q_{\mathcal{X}(\mathcal{D})} \otimes G_{D|\{x\}}$ is a second joint measure on $(\mathcal{U}_{X,D}, \mathbb{B}_{\mathcal{U}_{X,D}})$ and equal to $\Lambda_{\mathcal{D}} \otimes \mathrm{P}_{X|\{d\}}$. Again, let $U \in \mathbb{B}_{\mathcal{U}_{X,D}}$ and its projection into $\mathcal{D}$ with respect to $x \in \mathcal{X}(\mathcal{D})$

$$U_x := \{d \,|\, (d,x) \in U\}\,. \tag{6.44}$$

Then, the product measure $Q_{\mathcal{X}(\mathcal{D})} \otimes G_{D|\{x\}}$ of U is

$$(Q_{\mathcal{X}(\mathcal{D})} \otimes G_{D|\{x\}})(U) \;=\; \int\limits_{\mathcal{X}(\mathcal{D})} \left(\int\limits_{U_x} g_{D|\{x\}}(d)\;\mathrm{d}d \right) q_{\mathcal{X}(\mathcal{D})}(x)\;\mathrm{d}x\,. \tag{6.45}$$

For any $U \in \mathbb{B}_{\mathcal{U}_{X,D}}$ it follows

$$(\Lambda_{\mathcal{D}} \otimes \mathrm{P}_{X|\{d\}})(U) = (Q_{\mathcal{X}(\mathcal{D})} \otimes G_{D|\{x\}})(U) \tag{6.46}$$

and

$$(\Lambda_{\mathcal{D}} \otimes \mathrm{P}_{X|\{d\}})(\mathcal{U}_{X,D}) = (Q_{\mathcal{X}(\mathcal{D})} \otimes G_{D|\{x\}})(\mathcal{U}_{X,D}) = 1 \tag{6.47}$$

[2] *Density* should be understood here in the broad sense of Measurement Theory and not in the narrow sense of Probability Theory.

6.3.2. Evaluating a β-Measurement Procedure

The quality of a measurement procedure is assessed by its reliability and its precision. The reliability is a predefined requirement leading to β-measurement procedures, while the precision allows to choose among competing β-measurement procedures. The precision of the result of a β-measurement procedure $C_D^{(\beta)}$, i.e., the measurement set $C_D^{(\beta)}(\{x\})$ given the observed value x of X, is naturally quantified by the Lebesgue measure, or the normed geometric volume:

$$V(\{x\}) = \int_{C_D^{(\beta)}(\{x\})} \lambda_{\mathcal{D}}(d)\, \mathrm{d}d \tag{6.48}$$

Since the the volume of $C_D^{(\beta)}(\{x\})$ may vary with the observed value $x \in \mathcal{X}(\mathcal{D})$ it is just as natural to consider the expectation of the random volume $V(\{X\})$, i.e., $\mathrm{E}[V(\{X\})]$, with respect to the marginal distribution measure $Q_{\mathcal{X}(\mathcal{D})}$. The expectation $\mathrm{E}[V(\{X\})]$ is denoted by $V_w(C_D^{(\beta)})$ and called *weighted volume* of the β-measurement procedure. The term *volume* is justified because, $V_w(C_D^{(\beta)})$ represents in fact a weighted volume of the corresponding β-uncertainty space $\mathcal{U}_D^{(\beta)}$. Noting, that the β-uncertainty space $\mathcal{U}_D^{(\beta)}$ of D may be represented in different ways:

$$\mathcal{U}_D^{(\beta)} = \bigcup_{x \in \mathcal{X}(\mathcal{D})} C_D^{(\beta)}(\{x\}) \times \{x\} = \bigcup_{d \in \mathcal{D}} \{d\} \times A_X^{(\beta)}(\{d\}) \tag{6.49}$$

where $A_X^{(\beta)}(\{d\}$ are prediction sets from a suitable β-prediction procedure. Then, with the measures derived in section 6.3.1 we have several ways to represent

$V_w(C_D^{(\beta)})$:

$$V_w(C_D^{(\beta)}) = \int\limits_{\mathcal{X}(\mathcal{D})} \left(\int\limits_{C_D^{(\beta)}(\{x\})} \lambda_{\mathcal{D}}(d)\, \mathrm{d}d \right) q_{\mathcal{X}(\mathcal{D})}(x)\, \mathrm{d}x \tag{6.50}$$

$$= \int\limits_{\mathcal{U}_D^{(\beta)}} q_{\mathcal{X}(\mathcal{D})}(x)\lambda_{\mathcal{D}}(d)\, \mathrm{d}x\, \mathrm{d}d \tag{6.51}$$

$$= \int\limits_{\mathcal{D}} \left(\int\limits_{A_X^{(\beta)}(\{d\})} q_{\mathcal{X}(\mathcal{D})}(x)\, \mathrm{d}x \right) \lambda_{\mathcal{D}}(d)\, \mathrm{d}d \tag{6.52}$$

From the order of calculation in (6.52) it becomes obvious that the (marginal) probability of the prediction sets $A_X^{(\beta)}(\{d\})$ is weighted by the geometrical size of d, given by $\lambda_{\mathcal{D}}(d)$.

Remark: In [77] the derivation of $V_w(C_D^{(\beta)})$ is generalised by introducing an almost arbitrary weight function $w(x) \geq 0$. While $V(\{x\})$ is called *singular precision*,

$$V_w(C_D^{(\beta)}) = \int\limits_{\mathcal{X}(\mathcal{D})} V(\{x\})w(x)\, \mathrm{d}x\,, \tag{6.53}$$

$V_w(C_D^{(\beta)})$ is called *global weighted precision* of the measurement procedure $C_D^{(\beta)}$. With

$$w(x) = \int\limits_{\mathcal{D}} f_{X|\{d\}}(x) \frac{1}{|\mathcal{D}|}\, \mathrm{d}d \tag{6.54}$$

the weighted volume of [20] follows. Other suitable weight functions are introduced and discussed in [77].

The weighted volume now gives rise to the possibility to compare different β-uncertainy spaces of D.

Definition 6.9 (optimal β-measurement procedure) *Let $C_D^{(\beta)}$ and $\overline{C}_D^{(\beta)}$ be two β-measurement procedures for D. Then $C_D^{(\beta)}$ is called more precise than $\overline{C}_D^{(\beta)}$ if $V_w(C_D^{(\beta)}) < V_w(\overline{C}_D^{(\beta)})$. A β-measurement procedure ${}^*C_D^{(\beta)}$ for D is called optimal if $V_w({}^*C_D^{(\beta)}) \leq V_w(C_D^{(\beta)})$ for any other β-measurement procedure $C_D^{(\beta)}$ of D.*

In commemoration and to emphasize that the underlying ideas go back to Jerzy Neyman [57], the authors of [17] name an optimal β-measurement procedure ${}^{*}C_D^{(\beta)}$ *Neyman β-measurement procedure.*

Example 6.10 *We want to calculate the weighted volume $V_w(C_{\mathrm{E}[X]}^{(\beta)})$ of the β-measurement procedures derived in example 6.8. First, we recall the components of the Bernoulli-space $\mathbb{B}_{X,\mathrm{E}[X]}$*

$$\begin{aligned} \mathcal{D} &= \{\mu_1 \,|\, 0.05 \leq \mu_1 \leq 0.45\}\,, \\ \mathcal{X}(\{\mu_1\}) &= \{x \,|\, 0 \leq x \leq 1\}\,, \\ \mathcal{P}(\mu_1) &= Mon(0,1,\mu_1) = Mon(\mu_1)\,. \end{aligned}$$

The calculation should be performed according to equation (6.52)*. The involved functions are*

- *the density functions $f_{X|\{\mu_1\}}$ of $X|\{\mu_1\}$:*

$$f_{X|\{\mu_1\}}(x) = \frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)}-1} e^{\lambda(\mu_1)x} \mathbb{1}_{[0,1]}(x) \tag{6.55}$$

- *the density $\lambda_{\mathcal{D}}$ of the marginal measure $\Lambda_{\mathcal{D}}$ of $\mathrm{E}[X]$:*

$$\lambda_{\mathcal{D}}(\mu_1) = \frac{1}{|\mathcal{D}|}\mathbb{1}_{\mathcal{D}}(\mu_1) = \frac{1}{0.45-0.05}\mathbb{1}_{[0.05,0.45]}(\mu_1) = 2.5 \cdot \mathbb{1}_{[0.05,0.45]}(\mu_1) \tag{6.56}$$

- *the weight function $w(x)$:*

$$w(x) = \int_{\mathcal{D}} f_{X|\{\mu_1\}}(x)\, \mathrm{d}\mu_1 = \int_{0.05}^{0.45} \frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)}-1} e^{\lambda(\mu_1)x} \mathbb{1}_{[0,1]}(x)\, \mathrm{d}\mu_1 \tag{6.57}$$

- *the marginal density $q_{\mathcal{X}(\mathcal{D})}(x)$ of X:*

$$q_{\mathcal{X}(\mathcal{D})}(x) = \frac{w(x)}{|\mathcal{D}|} = 2.5 \cdot \int_{0.05}^{0.45} \frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)}-1} e^{\lambda(\mu_1)x} \mathbb{1}_{[0,1]}(x)\, \mathrm{d}\mu_1 \tag{6.58}$$

Obviously, the integrations cannot be done analytically, but only by numerical methods. Recalling, that we have a functional form of $\mu_1(\lambda)$ and, therefore, its derivative with respect to λ is given by (see proof to proposition 3.3)

$$\frac{\mathrm{d}\mu_1}{\mathrm{d}\lambda} = \frac{1 + e^{2\lambda} - e^{\lambda}(2+\lambda^2)}{(e^{\lambda}-1)^2\lambda^2} \tag{6.59}$$

Now, by substitution of $\mathrm{d}\mu_1$ *in* (6.57) *and adjustment of the integration limits we obtain for the weight function*

$$w(x) = \int_{\lambda(0.05)}^{\lambda(0.45)} \frac{\lambda}{e^\lambda - 1} e^{\lambda x} \cdot \frac{1 + e^{2\lambda} - e^\lambda(2 + \lambda^2)}{(e^\lambda - 1)^2 \lambda^2} \mathbb{1}_{[0,1]}(x) \, \mathrm{d}\lambda \tag{6.60}$$

Even though, the integral still cannot be calculated analytically, the amount for the numerical calculation is reduced through a reduced number of evaluations of $\lambda(\mu_1)$ *in* (6.60)*. Then, we get for the weighted volume* $V_w(C^{(\beta)}_{\mathrm{E}[X]})$ *the following*

$$V_w(C^{(\beta)}_{\mathrm{E}[X]}) = \frac{1}{0.4^2} \int_{0.05}^{0.45} \left(\int_{A_X^{(\beta)}(\{\mu_1\})} \left(\int_{\lambda(0.05)}^{\lambda(0.45)} \frac{\lambda}{e^\lambda - 1} e^{\lambda x} \cdot \right. \right. \\ \left. \left. \cdot \frac{1 + e^{2\lambda} - e^\lambda(2 + \lambda^2)}{(e^\lambda - 1)^2 \lambda^2} \mathbb{1}_{[0,1]}(x) \, \mathrm{d}\lambda \right) \mathrm{d}x \right) \mathrm{d}\mu_1 \tag{6.61}$$

The 0.9*-measurement procedure* $C^{(0.9)}_{\mathrm{E}[X]}$ *based on predictions* $A_X^{(0.9)}(\{\mu_1\}) = [0, z_{\mu_1}(0.9)]$ *for* $\mu_1 \in [0.05, 0.45)$ *and* $A_X^{(0.9)}(\{0.45\}) = [0, 1]$ *for* $\mu_1 = 0.45$ *as displayed in figure 6.7 has a weighted volume of*

$$V_w(C^{(0.9)}_{\mathrm{E}[X]}) = 0.819061\,. \tag{6.62}$$

For the 0.9*-measurement procedure* $\widehat{C}^{(0.9)}_{\mathrm{E}[X]}$ *based on predictions* $A_X^{(0.9)}(\{\mu_1\}) = [z_{\mu_1}(\gamma(\mu_1) - 0.9), z_{\mu_1}(\gamma(\mu_1))]$ *(see figure 6.8) we get a weighted volume of*

$$V_w(\widehat{C}^{(0.9)}_{\mathrm{E}[X]}) = 0.761537\,. \tag{6.63}$$

That is, $\widehat{C}^{(0.9)}_{\mathrm{E}[X]}$ *is more precise than* $C^{(0.9)}_{\mathrm{E}[X]}$*. A comparison of the (normed) geometrical sizes of the* 0.9*-uncertainty spaces in figures 6.7 and 6.8, would mislead*

to just the opposite conclusion:

$$
\begin{aligned}
|C^{(0.9)}_{\mathrm{E}[X]}| &= \frac{1}{0.4}\int_{0.05}^{0.45}\left(\int_{0}^{z_{\mu_1}(0.9)} 1\,\mathrm{d}x\right)\mathrm{d}\mu_1 \\
&= 2.5\cdot\int_{0.05}^{0.45} z_{\mu_1}(0.9)\,\mathrm{d}\mu_1 \qquad (6.64)\\
&= 0.548241 \\
|\widehat{C}^{(0.9)}_{\mathrm{E}[X]}| &= \frac{1}{0.4}\int_{0.05}^{0.45}\left(\int_{z_{\mu_1}(\gamma(\mu_1)-0.9)}^{z_{\mu_1}(\gamma(\mu_1))} 1\,\mathrm{d}x\right)\mathrm{d}\mu_1 \\
&= 2.5\cdot\int_{0.05}^{0.45} z_{\mu_1}(\gamma(\mu_1)) - z_{\mu_1}(\gamma(\mu_1)-0.9)\,\mathrm{d}\mu_1 \qquad (6.65)\\
&= 0.631260
\end{aligned}
$$

6.3.3. Determination of a Neyman β-Measurement Procedure

Again, from (6.52) it becomes clear, that finding an Neyman β-measurement procedure is traced back to finding a β-prediction procedure which yields prediction sets $A^{(\beta)}_X(\{d\})$ meeting three requirements:

- probability $\mathrm{P}_{X|\{d\}}(A^{(\beta)}_X(\{d\})) = \beta$ (*reliability requirement*)
- the union over all $d \in \mathcal{D}$ equals $\mathcal{X}(\mathcal{D})$ (*completeness requirement*)
- minimal marginal probability $Q_{\mathcal{X}(\mathcal{D})}(A^{(\beta)}_X(\{d\}))$ (*precision requirement*)

In [77] we find a generalised theorem to this problem, which we adapt to our case:

Theorem 6.11 (construction of a Neyman β-measurement procedure) *Let ${}^*A^{(\beta)}_X$ be a β-prediction procedure with the following properties: For each $d \in \mathcal{D}$ there is a constant $r^*(d)$ with*

$$
\begin{aligned}
&\mathrm{P}_{X|\{d\}}({}^*A^{(\beta)}_X(\{d\})) = \beta\,, && (6.66)\\
&x \in {}^*A^{(\beta)}_X(\{d\}) \implies f_{X|\{d\}}(x) \geq r^*(d)\cdot w(x)\,, && (6.67)\\
&x \in {}^*A^{(\beta)}_X(\{d\}) \impliedby f_{X|\{d\}}(x) > r^*(d)\cdot w(x)\,. && (6.68)
\end{aligned}
$$

Then, the β-measurement procedure based on ${}^*A_X^{(\beta)}$ is a Neyman β-measurement procedure, i.e.,

$$ {}^*C_D^{(\beta)}(\{x\}) = \{d \,|\, x \in {}^*A_X^{(\beta)}(\{d\})\} \quad \textit{for any } x \in \mathcal{X}(\mathcal{D})\,. \tag{6.69} $$

Proof of 6.11: The proof in [77] works for any weight function $w(x)$ and is given here in a slightly adapted way. It will be shown, that $V_w({}^*C_D^{(\beta)}) \leq V_w(C_D^{(\beta)})$ for any β-measurement procedure $C_D^{(\beta)}$. Since every β-measurement procedure $C_D^{(\beta)}$ also defines a β-prediction procedure $A_X^{(\beta)}$, with (6.52) it is sufficient to show that for all $d \in \mathcal{D}$ the inequality

$$ \int\limits_{{}^*A_X^{(\beta)}(\{d\})} q_{\mathcal{X}(\mathcal{D})}(x)\,\mathrm{d}x \leq \int\limits_{A_X^{(\beta)}(\{d\})} q_{\mathcal{X}(\mathcal{D})}(x)\,\mathrm{d}x \tag{6.70} $$

holds. With $q_{\mathcal{X}(\mathcal{D})}(x) = \frac{w(x)}{|\mathcal{D}|}$, we show the equivalent inequality

$$ \int\limits_{{}^*A_X^{(\beta)}(\{d\})} w(x)\,\mathrm{d}x \leq \int\limits_{A_X^{(\beta)}(\{d\})} w(x)\,\mathrm{d}x\,. \tag{6.71} $$

Since we only consider the absolute continuous case, we have the following two equations

$$ \begin{aligned} \beta &= \int\limits_{{}^*A_X^{(\beta)}(\{d\})} f_{X|\{d\}}(x)\,\mathrm{d}x \\ &= \int\limits_{{}^*A_X^{(\beta)}(\{d\}) \cap A_X^{(\beta)}(\{d\})} f_{X|\{d\}}(x)\,\mathrm{d}x + \int\limits_{{}^*A_X^{(\beta)}(\{d\}) \setminus A_X^{(\beta)}(\{d\})} f_{X|\{d\}}(x)\,\mathrm{d}x \end{aligned} \tag{6.72} $$

and

$$ \begin{aligned} \beta &= \int\limits_{A_X^{(\beta)}(\{d\})} f_{X|\{d\}}(x)\,\mathrm{d}x \\ &= \int\limits_{A_X^{(\beta)}(\{d\}) \cap {}^*A_X^{(\beta)}(\{d\})} f_{X|\{d\}}(x)\,\mathrm{d}x + \int\limits_{A_X^{(\beta)}(\{d\}) \setminus {}^*A_X^{(\beta)}(\{d\})} f_{X|\{d\}}(x)\,\mathrm{d}x \end{aligned} \tag{6.73} $$

which together imply

$$ \int\limits_{{}^*A_X^{(\beta)}(\{d\}) \setminus A_X^{(\beta)}(\{d\})} f_{X|\{d\}}(x)\,\mathrm{d}x = \int\limits_{A_X^{(\beta)}(\{d\}) \setminus {}^*A_X^{(\beta)}(\{d\})} f_{X|\{d\}}(x)\,\mathrm{d}x\,. \tag{6.74} $$

We get

$$r^*(d) \int\limits_{{}^*A_X^{(\beta)}(\{d\})} w(x)\,\mathrm{d}x = r^*(d) \int\limits_{{}^*A_X^{(\beta)}(\{d\})\cap A_X^{(\beta)}(\{d\})} w(x)\,\mathrm{d}x + r^*(d) \int\limits_{{}^*A_X^{(\beta)}(\{d\})\setminus A_X^{(\beta)}(\{d\})} w(x)\,\mathrm{d}x \tag{6.75}$$

$$\overset{(6.67)}{\leq} r^*(d) \int\limits_{{}^*A_X^{(\beta)}(\{d\})\cap A_X^{(\beta)}(\{d\})} w(x)\,\mathrm{d}x + \int\limits_{{}^*A_X^{(\beta)}(\{d\})\setminus A_X^{(\beta)}(\{d\})} f_{X|\{d\}}(x)\,\mathrm{d}x \tag{6.76}$$

$$\overset{(6.74)}{=} r^*(d) \int\limits_{{}^*A_X^{(\beta)}(\{d\})\cap A_X^{(\beta)}(\{d\})} w(x)\,\mathrm{d}x + \int\limits_{A_X^{(\beta)}(\{d\})\setminus {}^*A_X^{(\beta)}(\{d\})} f_{X|\{d\}}(x)\,\mathrm{d}x \tag{6.77}$$

$$\overset{(6.68)}{\leq} r^*(d) \int\limits_{{}^*A_X^{(\beta)}(\{d\})\cap A_X^{(\beta)}(\{d\})} w(x)\,\mathrm{d}x + r^*(d) \int\limits_{A_X^{(\beta)}(\{d\})\setminus {}^*A_X^{(\beta)}(\{d\})} w(x)\,\mathrm{d}x \tag{6.78}$$

$$= r^*(d) \int\limits_{A_X^{(\beta)}(\{d\})} w(x)\,\mathrm{d}x \tag{6.79}$$

and thus, inequality (6.71) follows.

Remarks: In contrast to the properties (6.66)–(6.68) which the prediction procedure and its prediction sets, respectively, have to meet according to theorem 6.11, it is concluded in [17] that the prediction sets are obtained by the rule:

$${}^*A_X^{(\beta)}(\{d\}) = \left\{ x \in \mathcal{X}(\mathcal{D}) \;\middle|\; \frac{f_{X|\{d\}}(x)}{w(x)} \geq r^*(d) \right\} \tag{6.80}$$

with $r^*(d)$ defined by

$$\int\limits_{\left\{x\in\mathcal{X}(\mathcal{D}) \,\middle|\, \frac{f_{X|\{d\}}(x)}{w(x)} \geq r^*(d)\right\}} f_{X|\{d\}}(x)\,\mathrm{d}x = \beta\,. \tag{6.81}$$

If (6.81) does always yield a solution $r^*(d)$, relations (6.67) and (6.68) might be combined into one equivalence relation. But this is not true in general. Consider the case that for some $d \in \mathcal{D}$ we have

$$0 < \mathrm{P}_{X|\{d\}}\left(\left\{ x \;\middle|\; \frac{f_{X|\{d\}}(x)}{w(x)} > r^*(d) \right\} \right) = \gamma < \beta \tag{6.82}$$

and a set $E(\{d\}) = \left\{x \;\middle|\; \frac{f_{X|\{d\}}(x)}{w(x)} = r^*(d)\right\}$ with

$$\mathrm{P}_{X|\{d\}}(E(\{d\})) > \beta - \gamma\,. \tag{6.83}$$

Then it would follow

$$\begin{aligned}\mathrm{P}_{X|\{d\}}\left(\left\{x \;\middle|\; \frac{f_{X|\{d\}}(x)}{w(x)} \geq r^*(d)\right\}\right) = \\ = \mathrm{P}_{X|\{d\}}\left(\left\{x \;\middle|\; \frac{f_{X|\{d\}}(x)}{w(x)} > r^*(d)\right\} \cup E(\{d\})\right) = \\ = \mathrm{P}_{X|\{d\}}\left(\left\{x \;\middle|\; \frac{f_{X|\{d\}}(x)}{w(x)} > r^*(d)\right\}\right) + \mathrm{P}_{X|\{d\}}(E(\{d\})) > \beta\,.\end{aligned} \tag{6.84}$$

To get a probability equal to β it would be necessary to choose some subset $E'(\{d\}) \subset E(\{d\})$ with $\mathrm{P}_{X|\{d\}}(E'(\{d\})) = \beta - \gamma$. In fact, this ambiguity is avoided in theorem 6.11 by the required properties (6.66)–(6.68) for the predicition sets.

As stated in [77], it might happen that the measurement procedure ${}^*C_D^{(\beta)}$ constructed according to theorem 6.11 violates the completeness requirement. It is suggested to proceed in the following way (see [77], p. 17, with adapted notations):

> Map for all x with ${}^*C_D^{(\beta)}(\{x\}) = \emptyset$ some fixed value $d_0 \in \mathcal{D}$. Then the reliability requirement is still fulfilled, because the prediction region [set] $A_X(\{d_0\})$ is not smaller than before. As the Lebegue-measure of ${}^*C_D^{(\beta)}(\{x\}) = \{d_0\}$ equals zero the global precision [weighted volume] V_w remains unchanged, i.e. optimal.

6.3.4. Example

In this section we develop the Neyman β-measurement procedure ${}^*C_{\mathrm{E}[X]}^{(\beta)}$ for the Bernoulli-Space $\mathbb{B}_{X,\mathrm{E}[X]}$. Since the general case will be discussed in detail in Chapter 7, proofs with respect to the courses of the involved functions are omitted.

We choose the same setting as in example 6.10 where we have already derived the necessary functions $q_{\mathcal{X}(\mathcal{D})}$ and $w(x)$, respectively, to determine the β-prediction

procedure ${}^*A_X^{(\beta)}$ which yields the Neyman β-measurement procedure ${}^*C_{\mathrm{E}[X]}^{(\beta)}$. That is, for each value μ_1 of $\mathrm{E}[X]$ we have to identify a constant $r^*(\mu_1)$, so that (6.66)–(6.68) are met. In fact, for the monotonic distribution, relations (6.67) and (6.68) can be replaced by the equivalence relation yielding (6.80).

First, we illustrate the behaviour of the ratio $\frac{f_{X|\{\mu_1\}}(x)}{w(x)}$ for some values μ_1 in figure 6.9. For the smallest value $\mu_1 = 0.05$ of $\mathrm{E}[X]$ the ratio is monotonously decreasing in x, i.e., we get the prediction ${}^*A_X^{(\beta)}(\{0.05\}) = [0, z_{0.05}(\beta)]$, where $z_{\mu_1}(\beta)$ denotes the upper quantile function of $X|\{\mu_1\}$. For larger values μ_1 the ratio becomes unimodal, and monotonously increasing in x from a certain value μ_1 on. For the largest value $\mu_1 = 0.45$ of $\mathrm{E}[X]$ this course yields a prediction ${}^*A_X^{(\beta)}(\{0.45\}) = [z_{0.45}(1-\beta), 1]$. The change of the ratio's course indicates, that both, the upper and lower bound of the predictions are monotonously increasing. The lower bound keeps to be 0 at the beginning and then increases to the final value $z_{0.45}(1-\beta)$, while the upper bound increases from $z_{0.05}(\beta)$ on till it reaches 1:

- ${}^*A_X^{(\beta)}(\{\mu_1\}) = [0, z_{\mu_1}(\beta)]$ for all μ_1 where

$$\frac{f_{X|\{\mu_1\}}(0)}{w(0)} \geq \frac{f_{X|\{\mu_1\}}(z_{\mu_1}(\beta))}{w(z_{\mu_1}(\beta))} \tag{6.85}$$

- ${}^*A_X^{(\beta)}(\{\mu_1\}) = [z_{\mu_1}(1-\beta), 1]$ for all μ_1 where

$$\frac{f_{X|\{\mu_1\}}(z_{\mu_1}(1-\beta))}{w(z_{\mu_1}(1-\beta))} \leq \frac{f_{X|\{\mu_1\}}(1)}{w(1)} \tag{6.86}$$

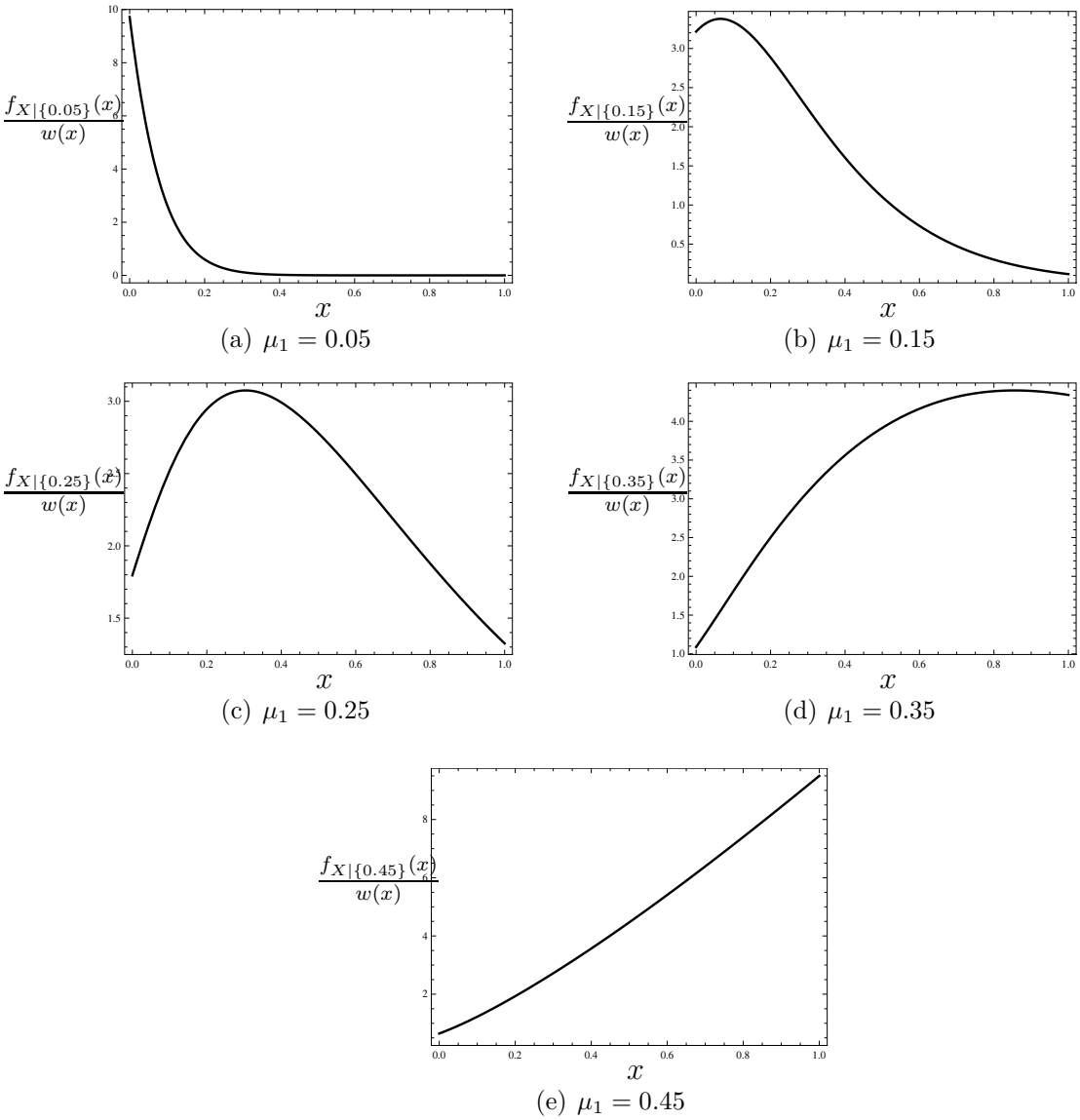
- ${}^*A_X^{(\beta)}(\{\mu_1\}) = [\ell_X(\mu_1), u_X(\mu_1)]$ for all μ_1 where

$$\frac{f_{X|\{\mu_1\}}(\ell_X(\mu_1))}{w(\ell_X(\mu_1))} = \frac{f_{X|\{\mu_1\}}(u_X(\mu_1))}{w(u_X(\mu_1))} \tag{6.87}$$

$$\frac{f_{X|\{\mu_1\}}(0)}{w(0)} < \frac{f_{X|\{\mu_1\}}(\ell_X(\mu_1))}{w(\ell_X(\mu_1))} \tag{6.88}$$

$$\frac{f_{X|\{\mu_1\}}(1)}{w(1)} < \frac{f_{X|\{\mu_1\}}(u_X(\mu_1))}{w(u_X(\mu_1))} \tag{6.89}$$

The difficult part in the determination of the predictions ${}^*A_X^{(\beta)}(\{\mu_1\})$ is represented by the last case ${}^*A_X^{(\beta)}(\{\mu_1\}) = [\ell_X(\mu_1), u_X(\mu_1)]$. Obviously, $\ell_X(\mu_1)$ and

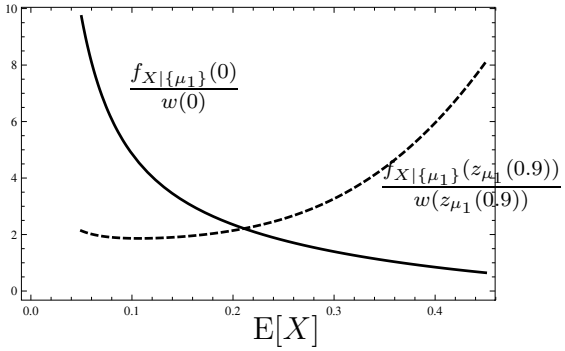


(a) $\mu_1 = 0.05$ (b) $\mu_1 = 0.15$

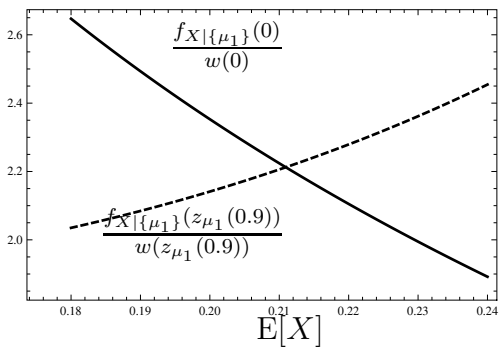
(c) $\mu_1 = 0.25$ (d) $\mu_1 = 0.35$

(e) $\mu_1 = 0.45$

Figure 6.9.: The course of the ratio $\frac{f_{X|\{\mu_1\}}(x)}{w(x)}$ for different values μ_1.



(a) $\mu_1 \in [0.05, 0.45]$.



(b) Detail of 6.10(a) for $\mu_1 \in [0.18, 0.24]$

Figure 6.10.: Comparison of the ratios in relation (6.85). For all $\mu_1 \leq 0.21094$ the ratio's value for $x = 0$ is at least as large as for $z_{\mu_1}(0.9)$.

$u_X(\mu_1)$ cannot be determined analytically and, thus, numerical methods have to be applied. But we can already conclude, that for $\ell_X(\mu_1)$ and $u_X(\mu_1)$, respectively, the following two relations hold

$$0 < \ell_X(\mu_1) < z_{\mu_1}(1-\beta) \tag{6.90}$$

$$z_{\mu_1}(\beta) < u_X(\mu_1) < 1 \tag{6.91}$$

which help to set up the necessary numerical methods.
For $\beta = 0.9$ figures 6.10 and 6.11 illustrate the respective ratios in relations (6.85) and (6.86), respectively.

Numerical analysis yields:

- For all $0.05 \leq \mu_1 \leq 0.21094$ we have (see figure 6.10)

$$\frac{f_{X|\{\mu_1\}}(0)}{w(0)} \geq \frac{f_{X|\{\mu_1\}}(z_{\mu_1}(0.9))}{w(z_{\mu_1}(0.9))} \tag{6.92}$$

  and, thus,

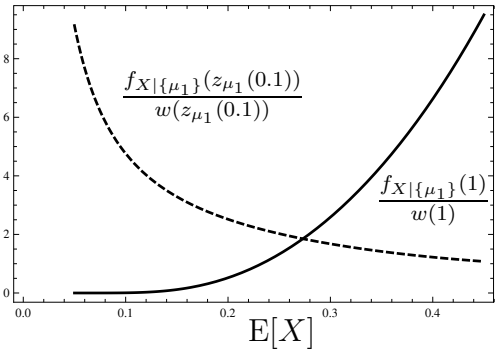
$$^*A_X^{(0.9)}(\{\mu_1\}) = [0, z_{\mu_1}(0.9)] \tag{6.93}$$

- For all $0.27316 \leq \mu_1 \leq 0.45$ we have (see figure 6.11)

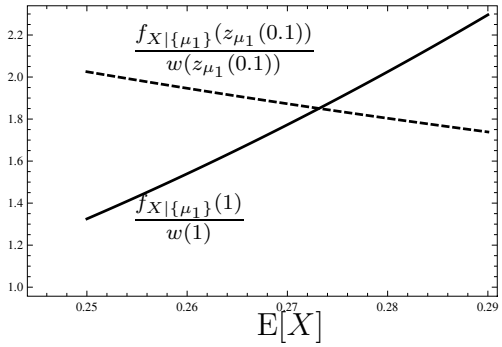
$$\frac{f_{X|\{\mu_1\}}(z_{\mu_1}(0.1))}{w(z_{\mu_1}(0.1))} \leq \frac{f_{X|\{\mu_1\}}(1)}{w(1)} \tag{6.94}$$

  and, thus,

$$^*A_X^{(0.9)}(\{\mu_1\}) = [z_{\mu_1}(0.1), 1] \tag{6.95}$$



(a) $\mu_1 \in [0.05, 0.45]$



(b) Detail of 6.11(a) for $\mu_1 \in [0.25, 0.29]$

Figure 6.11.: Comparison of the ratios in relation (6.86). For all $\mu_1 \geq 0.27316$ the ratio's value for $x = 1$ is at least as large as for $z_{\mu_1}(0.1)$.

- For $0.21094 < \mu_1 < 0.27316$ we cannot give a somehow analytic expression of the predictions ${}^*A_X^{(0.9)}(\{\mu_1\})$. In table 6.1 we find predictions for some values of μ_1.

Finally, we get an optimal 0.9-uncertainty space $\mathcal{U}_X^{(0.9)}$ for X illustrated in figure 6.12 which yields the Neyman 0.9-measurement procedure ${}^*C_{\mathrm{E}[X]}^{(0.9)}$. Calculating the weighted volume yields

$$V_w({}^*C_{\mathrm{E}[X]}^{(0.9)}) = 0.747112 \tag{6.96}$$

which is, of course, smaller than $V_w(\widehat{C}_{\mathrm{E}[X]}^{(0.9)}) = 0.761537$ in example 6.10. It is remarkable, that the geometric volume is 0.651072 – which is even larger than the values for both procedures in example 6.10. To end this example we give measurements ${}^*C_{\mathrm{E}[X]}^{(0.9)}(\{x\}) = [\ell_{\mathrm{E}[X]}(x), u_{\mathrm{E}[X]}(x)] = \{\mu_1 \,|\, x \in {}^*A_X^{(0.9)}(\{\mu_1\})\}$ for some values x in table 6.2.

	${}^*A_X^{(0.9)}(\{\mu_1\}) = [\ell_X(\mu_1), u_X(\mu_1)]$	
μ_1	$\ell_X(\mu_1)$	$u_X(\mu_1)$
0.215	0.001767	0.516281
0.225	0.006198	0.583591
0.235	0.010815	0.656465
0.245	0.015712	0.735626
0.255	0.020998	0.821901
0.265	0.026799	0.916254
0.270	0.029937	0.966807

Table 6.1.: ${}^*A_X^{(0.9)}(\{\mu_1\}) = [\ell(\mu_1), u(\mu_1)]$ for some values μ_1.

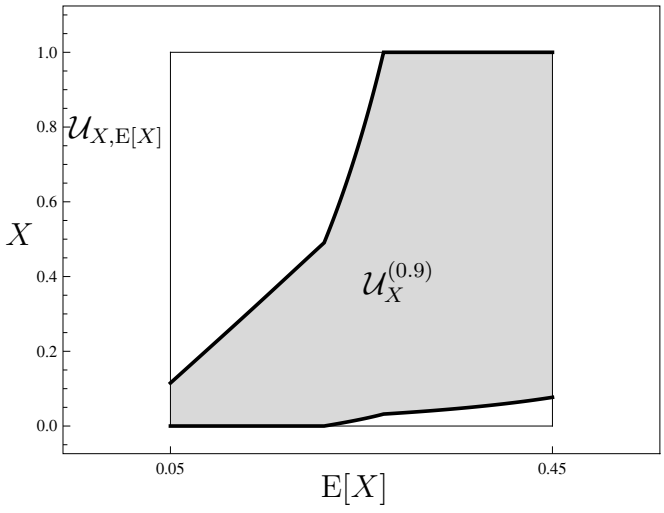


Figure 6.12.: Optimal β-uncertainty space $\mathcal{U}_X^{(0.9)}$ for X with respect to the weighted volume V_w. The corresponding 0.9-measurement procedure ${}^*C_{\mathrm{E}[X]}^{(0.9)}$ is the Neyman 0.9-measurement procedure.

	${}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{x\}) = [\ell_{\mathrm{E}[X]}(x), u_{\mathrm{E}[X]}(x)]$	
x	$\ell_{\mathrm{E}[X]}(x)$	$u_{\mathrm{E}[X]}(x)$
0.025	0.05	0.262
0.05	0.05	0.364
0.1	0.05	0.45
0.2	0.086	0.45
0.4	0.172	0.45
0.6	0.227	0.45
0.8	0.252	0.45

Table 6.2.: Measurements ${}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{x\}) = [\ell_{\mathrm{E}[X]}(x), u_{\mathrm{E}[X]}(x)]$ for some values x.

6.4. Bernoulli-Spaces for Statistical Distributions

In Chapter 4 we have presented the common methods of statistical estimation theory. The main part in traditional estimation is represented by the asssumed probability distribution. The choice is in general not backed by realistic considerations about the random phenomenon, but by the feasibility of the applied methods. The parts which are often not considered are the bounded range of variability and a bounded ignorance space, i.e., available knowledge about the random phenomenon. Additionally, not taking a bounded range of variability into account prevent the derivation of measurement procedures with respect to these bounds.

Nevertheless, the choice of a statistical distribution and an unbounded range of variability might be reasonable in cases where two conditions are fulfilled:

- the statistical distribution is a sufficiently good approximation, and
- the distribution parameter to be measured is almost independent of the range of variability.

In these cases, we may establish a Bernoulli-Space and derive suitable measurement procedures, which also incorporate knowledge about the values of the distribution parameters, i.e., a bounded ignorance space might be considered.

Examples of measurement procedures based on a normal random variable are given in [17] and [77] for measuring the expectation and in [18] for measuring the variance. Throughout Chapter 4 we have illustrated the traditional methods of estimation mostly with the one-parametric exponential distribution $EXP(\vartheta)$. While we have derived methods for estimation of the value ϑ, we want to derive a Neyman measurement procedure for the expectation $\mathrm{E}[X]$ of an exponential distributed random variable X in the remainder.

Let X be a random variable distributed according to the exponential distribution $\sim EXP(\vartheta)$ where ϑ is the reciprocal of the value μ_1 of the expectation of X, i.e., $\mathrm{E}[X] = \mu_1 = \frac{1}{\vartheta}$. We have the pair of variables $(X, \mathrm{E}[X])$ together with the conditional density function

$$f_{X|\{\mu_1\}}(x) = \frac{1}{\mu_1} e^{-x/\mu_1} \mathbb{1}_{[0,+\infty)}(x) \tag{6.97}$$

The range of variability of $X|\{\mu_1\}$ is given by

$$\mathcal{X}(\{\mu_1\}) = \{x \,|\, 0 \leq x < +\infty\} = [0, +\infty)\,, \tag{6.98}$$

which is independent of the value μ_1. Unboundedness of $\mathcal{X}(\{\mu_1\})$ is in fact an extreme violation to reality, but should be considered to be acceptable for deriving a measurement procedure for $\mathrm{E}[X]$.

The deterministic variable is the expectation $\mathrm{E}[X]$ with ignorance space given by

$$\mathcal{D} = \{\mu_1 \,|\, 0 < m_\ell \leq \mu_1 \leq m_u < +\infty\} = [m_\ell, m_u]\,. \tag{6.99}$$

Then the uncertainty space is

$$\mathcal{U}_{X,\mathrm{E}[X]} = [m_\ell, m_u] \times [0, +\infty)\,. \tag{6.100}$$

The weight function $w(x)$ and, thus, the ratio $\frac{f_{X|\{\mu_1\}}(x)}{w(x)}$ cannot be given in an explicit form. For $x \geq 0$ we have:

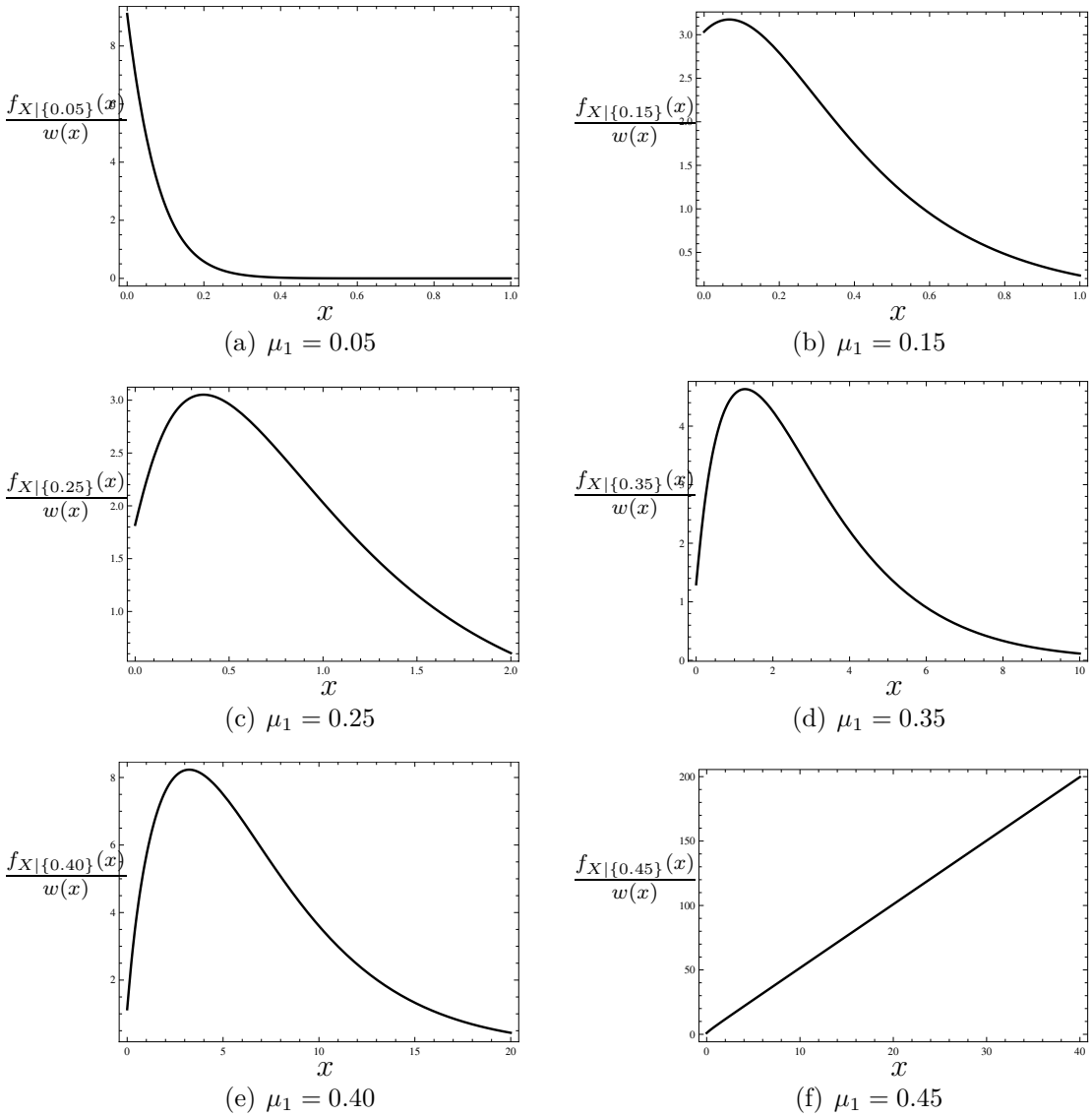
$$w(x) = \int\limits_{m_\ell}^{m_u} f_{X|\{\mu_1\}}(x)\,\mathrm{d}\mu_1 = \int\limits_{x/m_u}^{x/m_\ell} \frac{1}{t} e^{-t}\,\mathrm{d}t \tag{6.101}$$

$$\frac{f_{X|\{\mu_1\}}(x)}{w(x)} = \frac{\frac{1}{\mu_1} e^{-x/\mu_1}}{\int\limits_{x/m_u}^{x/m_\ell} \frac{1}{t} e^{-t}\,\mathrm{d}t} \tag{6.102}$$

For a graphical illustration we choose the same ignorance space as in section 6.3.4, i.e., $m_\ell = 0.05$ and $m_u = 0.45$. The graphics of the ratio $\frac{f_{X|\{\mu_1\}}(x)}{w(x)}$ in figure 6.13 for different values μ_1 are very similar to those in figure 6.9 for the monotonic distribution. However, we have to note the different displayed x-ranges. Determination of the respective predictions works also similar as in section 6.3.4. For $\mathcal{D} = \{\mu_1 \,|\, 0.05 \leq \mu_1 \leq 0.45\}$ and $\beta = 0.9$ we have:

- For all values $0.05 \leq \mu_1 \leq 0.195246$ the ratio's value $\frac{f_{X|\{\mu_1\}}(x)}{w(x)}$ is for $x = 0$ at least as large as for $x = z_{\mu_1}(0.9)$, where $z_{\mu_1}(\gamma)$ is the upper quantile function. Thus, the predictions are

$$A_X^{(0.9)}(\{\mu_1\}) = [0, z_{\mu_1}(0.9)] \tag{6.103}$$



(a) $\mu_1 = 0.05$ (b) $\mu_1 = 0.15$

(c) $\mu_1 = 0.25$ (d) $\mu_1 = 0.35$

(e) $\mu_1 = 0.40$ (f) $\mu_1 = 0.45$

Figure 6.13.: The course of the ratio $\frac{f_{X|\{\mu_1\}}(x)}{w(x)}$ for different values μ_1.

	${}^*A_X^{(0.9)}(\{\mu_1\}) = [\ell_X(\mu_1), u_X(\mu_1)]$	
μ_1	$\ell_X(\mu_1)$	$u_X(\mu_1)$
0.2	0.002367	0.485559
0.25	0.021784	1.025323
0.3	0.031340	2.137088
0.35	0.036876	4.828779
0.375	0.039510	7.893876
0.39	0.041091	11.211820
0.4	0.042144	> 12
0.45	0.047412	$+\infty$

Table 6.3.: ${}^*A_X^{(0.9)}(\{\mu_1\}) = [\ell(\mu_1), u(\mu_1)]$ for some values μ_1.

- For all values $0.195246 < \mu_1 < 0.45$ the predictions are of form

$$A_X^{(0.9)}(\{\mu_1\}) = [\ell(\mu_1), u(\mu_1)] \tag{6.104}$$

  with $0 < \ell(\mu_1) < z_{\mu_1}(0.1)$ and $z_{\mu_1}(0.9) < u(\mu_1) < +\infty$ and, of course, $\mathrm{P}_{X|\{\mu_1\}}(\{x|\ell(\mu_1) \leq x \leq u(\mu_1)\}) = 0.9$.

- For $\mu_1 = m_u = 0.45$ the ratio $\frac{f_{X|\{0.45\}}(x)}{w(x)}$ is monotonously increasing in x. That is, the corresponding prediction $A_X^{(\beta)}(\{0.45\})$ is unbounded to the right:

$$A_X^{(0.9)}(\{0.45\}) = [z_{0.45}(0.1), +\infty) \tag{6.105}$$

In table 6.3 some predictions ${}^*A_X^{(0.9)}(\{\mu_1\})$ for certain values μ_1 are displayed. Figure 6.14 shows the resulting β-uncertainty space $\mathcal{U}_X^{(0.9)}$ for X. For a comparison the β-uncertainty space from section 6.3.4 where $X \sim Mon(\mu_1)$ is also plotted. To complete this example, we present measurements ${}^*C_{\mathrm{E}[X]}^{(0.9)}(\{x\}) = [\ell_{\mathrm{E}[X]}(x), u_{\mathrm{E}[X]}(x)]$ for some values x in table 6.4. In fact, since the exponential distribution serves as an approximation to the monotonic distribution, values $x > 1$ are impossible to appear.

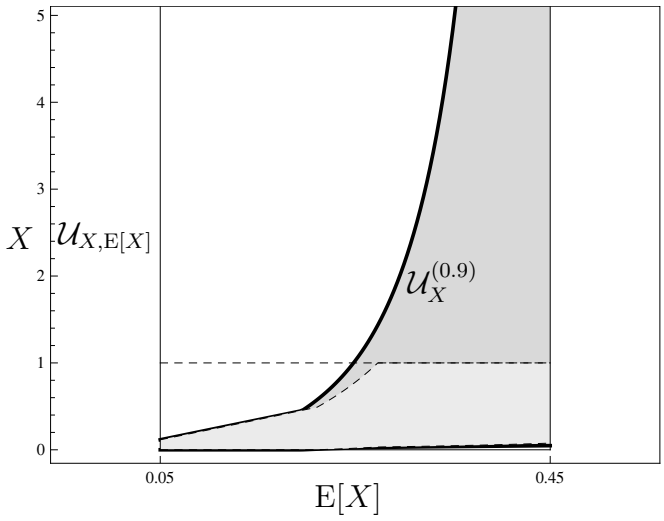


Figure 6.14.: Optimal β-uncertainty space $\mathcal{U}_X^{(0.9)}$ for $X \sim EXP(1/\mu_1)$ with respect to the weighted volume V_w. The corresponding 0.9-measurement procedure ${}^*C_{\mathrm{E}[X]}^{(0.9)}$ is the Neyman 0.9-measurement procedure. The light grey area and the dashed lines indicate the uncertainty space and the optimal β-uncertainty space, respectively, for $X \sim Mon(\mu_1)$.

A comparison to the measurements in table 6.2 shows that the measurements here are larger. That is, if the exponential distribution $EXP(1/\mu_1)$ is considered as an approximation for the monotonic distribution $Mon(\mu_1)$, a measurement procedure based on this approximation yields larger measurements.

	${}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{x\}) = [\ell_{\mathrm{E}[X]}(x), u_{\mathrm{E}[X]}(x)]$	
x	$\ell_{\mathrm{E}[X]}(x)$	$u_{\mathrm{E}[X]}(x)$
0.01	0.05	0.217
0.025	0.05	0.263
0.05	0.05	0.45
0.1	0.05	0.45
0.2	0.086	0.45
0.4	0.173	0.45
0.8	0.239	0.45
2	0.295	0.45
4	0.339	0.45

Table 6.4.: Measurements ${}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{x\}) = [\ell_{\mathrm{E}[X]}(x), u_{\mathrm{E}[X]}(x)]$ for some values x.

7. Neyman Measurement Procedures for $Mon(a, b, \mu_1)$

In section 6.3.4 the Neyman β-measurement procedure for the expectation $\mathrm{E}[X]$ for $X \sim Mon(0, 1, \mu_1)$ and ignorance space $\mathcal{D} = \{\mu_1 \,|\, 0.05 \leq \mu_1 \leq 0.45\}$ is derived. The measurements are based on one realisation x of X, i.e., on a sample of size one. Even though statistical estimation based on small sample sizes is of great importance, the considered sample size is commonly larger than one. Thus, the derivation of Neyman β-measurement procedures ${}^{*}C^{(\beta)}_{\mathrm{E}[X]}$ for $\mathrm{E}[X]$ are extended to the case of an arbitrary sample size $n \geq 1$. For this, we will assume complete knowledge about the range of variability of X, i.e., lower and upper bound are both known. The results of the Neyman β-measurement procedures are compared with those based on the exponential distribution – which could be understood as an approximation of the monotonic distribution. Additionally, we compare the Neyman β-measurement procedures with interval estimations 1) based on the likelihood ratio test statistic (where the knowledge about the values of $\mathrm{E}[X]$ may be incorporated), and 2) based on the normal approximation. Also, both could be understood as approximations.

The chapter concludes with a proposal of an approximative measurement procedure for the upper bound B of the range of variability, where both, the lower bound A and the expectation $\mathrm{E}[X]$ are assumed to be known. In classical statistics, methods for estimating the bounds of a range of variability are on the one hand often unnecessary due to its unboundedness, and on the other hand concentrate on the derivation of point estimators[1]. Therefore, no comparison is at hand and consequently omitted.

[1]For example for the beta distribution, [40] gives an extensive overview of point estimators.

7.1. General Assumptions and Notations

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be an i.i.d. sample for $X \sim Mon(a, b, \mu_1)$. Ignorance about the value $d = (a, b, \mu_1)$ of the deterministic variable $D = (A, B, \mathrm{E}[X])$ can refer to all three, two or only one of its components. Since we also have qualitative knowledge about the monotonicity of the density function, i.e., we know whether it is increasing or decreasing, the values of $(A, B, \mathrm{E}[X])$ are subject to some restrictions. From corollary 3.12 we know, that if $X \sim Mon(a, b, \mu_1)$ then $-X \sim Mon(-b, -a, -\mu_1)$. That is, without loss of generality we may only discuss the case of a monotonously decreasing density function. For each component of the deterministic variable we could assume an individual range:

- The component of the deterministic variable denoted by A stands for the lower bound of the range of variability, the values are denoted by a with $a \in [a_\ell, a_u] =: \mathcal{D}_A$.
- The component of the deterministic variable denoted by B stands for the upper bound of the range of variability, the values are denoted by b with $b \in [b_\ell, b_u] =: \mathcal{D}_B$.
- The component of the deterministic variable denoted by $\mathrm{E}[X]$ is, of course, the expectation of X, the values are denoted by μ_1 with $\mu_1 \in [m_\ell, m_u] =: \mathcal{D}_{\mathrm{E}[X]}$.

The ignorance space $\mathcal{D}$ for D then is not just the cross-product of the individual ranges, but is restricted by two inequalities:

- Trivially, the value a of the lower bound of the range of variability is smaller than the value b of the upper bound: $a < b$
- Since the density function is assumed to be decreasing, we have: $a < \mu_1 < \frac{a+b}{2}$

Therefore, we get as ignorance space

$$\mathcal{D} = \Big\{(a, b, \mu_1) \,|\, a_\ell \leq a \leq a_u,\ b_\ell \leq b \leq b_u,\ m_\ell \leq \mu_1 \leq m_u,\\ a < \mu_1 < \frac{a+b}{2} < b\Big\} \tag{7.1}$$

In the following sections we derive measurement procedures for $\mathrm{E}[X]$ and B, respectively, while the values of the others are assumed to be known.

7.2. Measurement Procedures for $\mathrm{E}[X]$

7.2.1. Neyman Measurement Procedures for $\mathrm{E}[X]$

In this case the value a of the lower bound A and the value b of the upper bound B of the range of variability are known and there is only ignorance about the value μ_1 of the expectation $\mathrm{E}[X]$ of X. Then the deterministic variable can be identified with the expectation, i.e., $D = \mathrm{E}[X]$, with ignorance space:

$$\mathcal{D} = \left\{ \mu_1 \;\middle|\; a < m_\ell \leq \mu_1 \leq m_u < \frac{a+b}{2} \right\} . \tag{7.2}$$

Later, it will be shown, that we may assume $a = 0$ and $b = 1$ without loss of generality, that is, we may assume $X \sim Mon(\mu_1)$. Then we have:

$$\mathcal{D} = \left\{ 0 < \mu_1 \;\middle|\; m_\ell \leq \mu_1 \leq m_u < \frac{1}{2} \right\} . \tag{7.3}$$

According to proposition 6.11 we need the following functions for determining β-prediction procedures ${}^*A^{(\beta)}_{\boldsymbol{X}}$:

- The density functions $f_{X|\{\mu_1\}}$ and $f_{\boldsymbol{X}|\{\mu_1\}}$ of $X|\{\mu_1\}$ and $\boldsymbol{X}|\{\mu_1\}$, respectively:

$$f_{X|\{\mu_1\}}(x) = \frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)} - 1} e^{\lambda(\mu_1)x} \mathbb{1}_{[0,1]}(x) , \tag{7.4}$$

$$f_{\boldsymbol{X}|\{\mu_1\}}(\boldsymbol{x}) = \left(\frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)} - 1} \right)^n e^{\lambda(\mu_1)(\sum_{i=1}^n x_i)} \mathbb{1}_{[0,1]^n}(\boldsymbol{x}) \tag{7.5}$$

- The weight function w (according to example 6.10):

$$w(\boldsymbol{x}) = \int_{m_\ell}^{m_u} \left(\frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)} - 1} \right)^n e^{\lambda(\mu_1)(\sum_{i=1}^n x_i)} \mathbb{1}_{[0,1]^n}(\boldsymbol{x}) \, \mathrm{d}\mu_1 \tag{7.6}$$

$$= \int_{\lambda(m_\ell)}^{\lambda(m_u)} \left(\frac{\lambda}{e^{\lambda} - 1} \right)^n e^{\lambda(\sum_{i=1}^n x_i)} \cdot \frac{1 + e^{2\lambda} - e^{\lambda}(2 + \lambda^2)}{(e^{\lambda} - 1)^2 \lambda^2} \mathbb{1}_{[0,1]^n}(\boldsymbol{x}) \, \mathrm{d}\lambda \tag{7.7}$$

We note, that both functions $f_{\boldsymbol{X}|\{\mu_1\}}(\boldsymbol{x})$ and $w(\boldsymbol{x})$ only depend on the sum $\sum_{i=1}^n x_i$ of the components x_i of $\boldsymbol{x} \in [0,1]^n$. Consequently, for $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^n$ with $\sum_{i=1}^n x_i =$

$\sum_{i=1}^{n} y_i$ we have:

$$f_{\boldsymbol{X}|\{\mu_1\}}(\boldsymbol{x}) = f_{\boldsymbol{X}|\{\mu_1\}}(\boldsymbol{y}), \tag{7.8}$$

$$w(\boldsymbol{x}) = w(\boldsymbol{y}) \tag{7.9}$$

and, thus:

$$\frac{f_{\boldsymbol{X}|\{\mu_1\}}(\boldsymbol{x})}{w(\boldsymbol{x})} = \frac{f_{\boldsymbol{X}|\{\mu_1\}}(\boldsymbol{y})}{w(\boldsymbol{y})} \tag{7.10}$$

Since the monotonic distribution is absolute continuous on the range of variability, we do not have to take into account the case discussed in the remarks to section 6.3.3, and the prediction sets are obtained by the rule (6.80). Defining $t := \sum_{i=1}^{n} x_i$, we conclude, that if one $\boldsymbol{x} \in {}^{*}A_{\boldsymbol{X}}^{(\beta)}(\{\mu_1\}) \subseteq [0,1]^n$ then the whole set:

$$\left\{ \boldsymbol{y} \in [0,1]^n \,\middle|\, \sum_{i=1}^{n} y_i = t) \right\} \tag{7.11}$$

is part of ${}^{*}A_{\boldsymbol{X}}^{(\beta)}(\{\mu_1\})$. In the following, we take $\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$ instead of $\sum_{i=1}^{n} X_i$. This choice enables us to better illustrate the change of the derived procedures with respect to the sample size n, since the range of variability of $\overline{X}$ remains unchanged, i.e., $[0,1]$. The respective prediction sets for $\boldsymbol{X}$ may be transformed to prediction sets for $\overline{X}$. Firstly, we have:

$${}^{*}A_{\boldsymbol{X}}^{(\beta)}(\{\mu_1\}) = \left\{ \boldsymbol{x} \,\middle|\, \frac{f_{\boldsymbol{X}|\{\mu_1\}}(\boldsymbol{x})}{w(\boldsymbol{x})} \geq r^*(\mu_1) \right\} \tag{7.12}$$

$$= \left\{ \boldsymbol{x} \,\middle|\, \frac{\left(\frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)}-1}\right)^n e^{\lambda(\mu_1)(\sum_{i=1}^{n} x_i)}}{\int_{m_\ell}^{m_u} \left(\frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)}-1}\right)^n e^{\lambda(\mu_1)(\sum_{i=1}^{n} x_i)} \,\mathrm{d}\mu_1} \geq r^*(\mu_1) \right\} \tag{7.13}$$

$$= \left\{ \boldsymbol{x} \,\middle|\, \frac{\left(\frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)}-1}\right)^n e^{\lambda(\mu_1) n \overline{x}}}{\int_{m_\ell}^{m_u} \left(\frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)}-1}\right)^n e^{\lambda(\mu_1) n \overline{x}} \,\mathrm{d}\mu_1} \geq r^*(\mu_1) \right\}. \tag{7.14}$$

With the density of $\overline{X}$ given by (3.123) following proposition 3.19 we get with the corresponding weight function:

$$w(\overline{x}) = \int_{m_\ell}^{m_u} n \left(\frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)} - 1} \right)^n e^{\lambda(\mu_1) n \overline{x}} f_{\sum_{i=1}^{n} U_i}(n\overline{x}) \, \mathrm{d}\mu_1 \tag{7.15}$$

$$= n f_{\sum_{i=1}^{n} U_i}(\overline{x}) \int_{m_\ell}^{m_u} \left(\frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)} - 1} \right)^n e^{\lambda(\mu_1) n \overline{x}} \, \mathrm{d}\mu_1 \,, \tag{7.16}$$

yielding the ratio for $\overline{x} \in [0, 1]$:

$$\frac{f_{\overline{X}|\{\mu_1\}}(\overline{x})}{w(\overline{x})} = \frac{n \left(\frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)}-1} \right)^n e^{\lambda(\mu_1) n \overline{x}} f_{\sum_{i=1}^{n} U_i}(n\overline{x})}{n f_{\sum_{i=1}^{n} U_i}(n\overline{x}) \int_{m_\ell}^{m_u} \left(\frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)}-1} \right)^n e^{\lambda(\mu_1) n \overline{x}} \, \mathrm{d}\mu_1} \tag{7.17}$$

$$= \frac{\left(\frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)}-1} \right)^n e^{\lambda(\mu_1) n \overline{x}}}{\int_{m_\ell}^{m_u} \left(\frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)}-1} \right)^n e^{\lambda(\mu_1) n \overline{x}} \, \mathrm{d}\mu_1} \tag{7.18}$$

$$= \frac{f_{\boldsymbol{X}|\{\mu_1\}}(\boldsymbol{x})}{w(\boldsymbol{x})} \,. \tag{7.19}$$

Thus, the prediction sets for $\overline{X}$ are obtained as follows:

$${}^{*}A_{\overline{X}}^{(\beta)}(\{\mu_1\}) = \left\{ \overline{x} \,\middle|\, \frac{f_{\overline{X}|\{\mu_1\}}(\overline{x})}{w(\overline{x})} \geq r^*(\mu_1) \right\} \tag{7.20}$$

$$= \left\{ \overline{x} \,\middle|\, \frac{\left(\frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)}-1} \right)^n e^{\lambda(\mu_1) n \overline{x}}}{\int_{m_\ell}^{m_u} \left(\frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)}-1} \right)^n e^{\lambda(\mu_1) n \overline{x}} \, \mathrm{d}\mu_1} \geq r^*(\mu_1) \right\} \tag{7.21}$$

$$= \left\{ \overline{x} \,\middle|\, \frac{\left(\frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)}-1} \right)^n e^{\lambda(\mu_1) \sum_{i=1}^{n} x_i}}{\int_{m_\ell}^{m_u} \left(\frac{\lambda(\mu_1)}{e^{\lambda(\mu_1)}-1} \right)^n e^{\lambda(\mu_1) \sum_{i=1}^{n} x_i} \, \mathrm{d}\mu_1} \geq r^*(\mu_1) \right\} . \tag{7.22}$$

Consequently, the determination of prediction sets for $\boldsymbol{X}$ is equivalent to the determination of prediction sets for $\overline{X}$. But since, $\overline{X}$ is an one-dimensional random variable, it is much easier to determine the respective prediction set ${}^{*}A_{\overline{X}}^{(\beta)}(\{\mu_1\})$

meeting the reliability requirement:

$$\mathrm{P}_{\overline{X}|\{\mu_1\}}\left({}^{*}A_{\overline{X}}^{(\beta)}(\{\mu_1\})\right) = \beta\,. \tag{7.23}$$

At first, we want to get an impression about the course of the ratio $\frac{f_{\overline{X}|\{\mu_1\}}(\overline{x})}{w(\overline{x})}$ by an example. The proposed settings in the example will be used throughout this chapter.

Example 7.1 *Let $(X_1, \ldots, X_n)$ be an i.i.d. sample for $X \sim Mon(\mu_1)$, where μ_1 is the value of the deterministic variable $D = \mathrm{E}[X]$ with ignorance space given by:*

$$\mathcal{D} = \{\mu_1 \,|\, 0.05 \leq \mu_1 \leq 0.25\}\,. \tag{7.24}$$

Since $X_1, \ldots, X_n$ are i.i.d., the ignorance space for $\mathrm{E}[\overline{X}] = \mathrm{E}[X]$ is the same. The range of variability of $\overline{X}$ is

$$\mathcal{X}(\{\mu_1\}) = [0, 1]\,, \tag{7.25}$$

i.e., independent of μ_1. Thus, the uncertainty space $\mathcal{U}_{\overline{X},\mathrm{E}[X]}$ for the pair of variables $(\overline{X}, \mathrm{E}[X])$ forms an rectangle in the $\mathrm{E}[X]$-$\overline{X}$-plane:

$$\mathcal{U}_{\overline{X},\mathrm{E}[X]} = [0.05, 0.25] \times [0, 1]\,. \tag{7.26}$$

Up to now, the situation displayed in figure 7.1 is quite the same as in examples 6.2 and 6.5. [!htbp] To display the ratio $\frac{f_{\overline{X}|\{\mu_1\}}(\overline{x})}{w(\overline{x})}$ for some values of $\mathrm{E}[X]$, we choose $n = 8$ as fixed sample size.

The ratio's course depending on the value of $\overline{x}$ may be described as follows: For $\mu_1 = m_\ell = 0.05$ the ratio is monotonously decreasing in $\overline{x}$ (figure 7.2(a)). For some m_1 and m_2 with $m_\ell < m_1 \leq \mu_1 \leq m_2 < m_u$, the course is unimodal in the inner of $\mathcal{X}$, i.e., on $(0, 1)$ (figures 7.2(b) - 7.2(h)). For $m_2 < \mu_1$ the course is monotonously increasing in $\overline{x}$ (figure 7.2(i)).

First, the change from monotonously decreasing to unimodality to monotonously increasing of the ratio's course shall be investigated. Defining

$$g(\lambda, \overline{x}) := \left(\frac{\lambda}{e^\lambda - 1}\right)^n e^{\lambda n \overline{x}}\,, \tag{7.27}$$

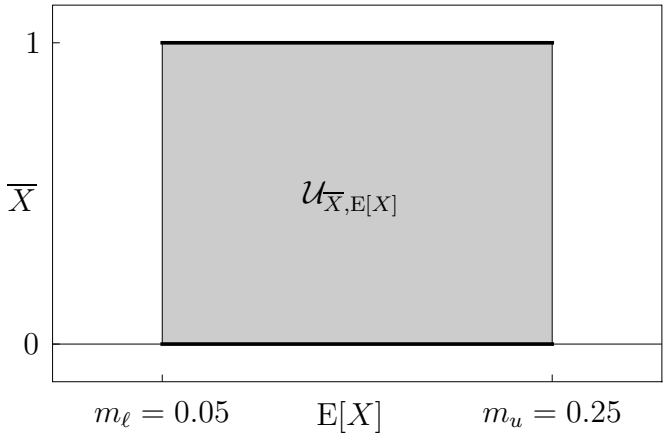


Figure 7.1.: Uncertainty space $\mathcal{U}_{\overline{X},\mathrm{E}[X]}$, where $X|\{\mu_1\} \sim Mon(0,1,\mu_1)$ and $\mu_1 \in \mathcal{D} = \{\mu_1 | m_\ell = 0.05 \leq \mu_1 \leq 0.25 = m_u\}$. Because there is only ignorance about the expectation $\mathrm{E}[X] = \mathrm{E}[\overline{X}]$ of X and $\overline{X}$, respectively, we consider $\mathrm{E}[X]$ as the deterministic variable D.

we obtain:

$$\frac{f_{\overline{X}|\{\mu_1\}}(\overline{x})}{w(\overline{x})} = \frac{g(\lambda,\overline{x})}{w(\overline{x})}\,. \tag{7.28}$$

With $\lambda_0 = \lambda(\mu_1)$ for $\mu_1 \in \mathcal{D} \subset (0, 0.5)$, i.e., $\lambda_0 < 0$, and

$$c(\lambda) := \frac{1 + e^{2\lambda} - e^{\lambda}(2+\lambda^2)}{(e^{\lambda}-1)^2\lambda^2}\,, \tag{7.29}$$

we get

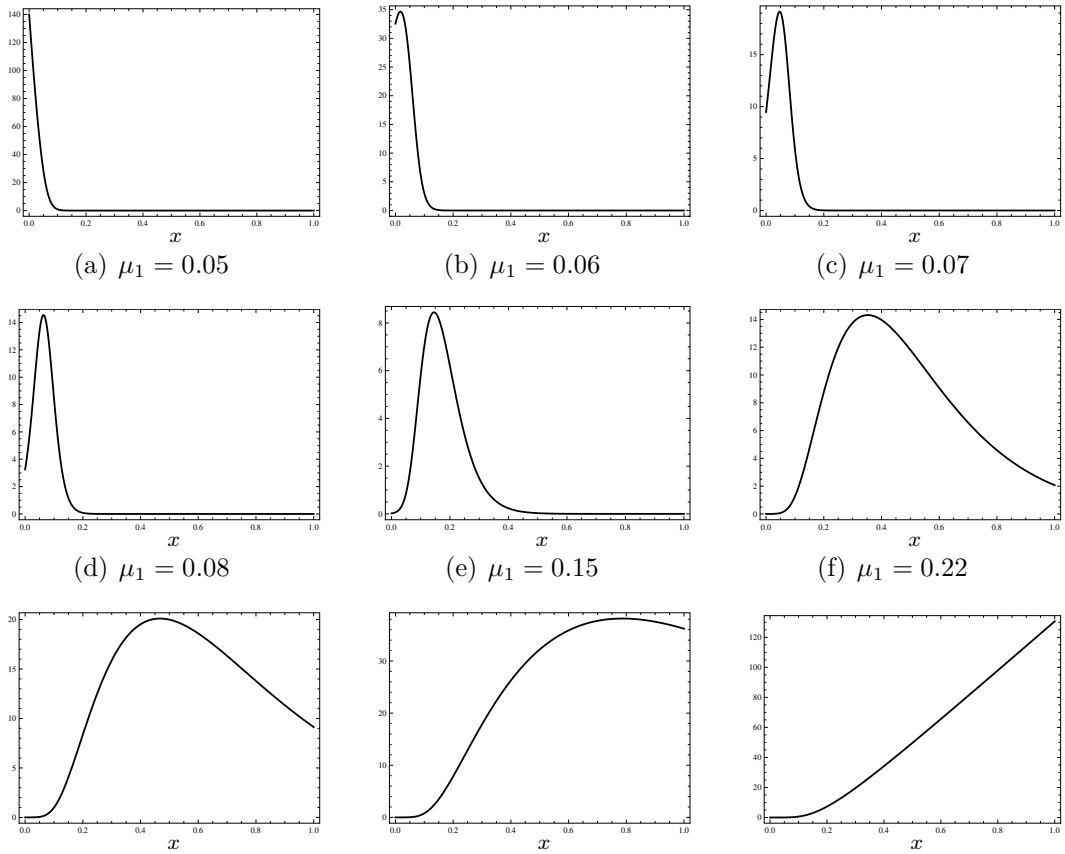
$$\frac{\mathrm{d}}{\mathrm{d}\overline{x}}\frac{g(\lambda_0,\overline{x})}{w(\overline{x})} = \frac{w(\overline{x})\lambda_0 n g(\lambda_0,\overline{x}) - g(\lambda_0,\overline{x})\int_{\lambda(m_\ell)}^{\lambda(m_u)} \lambda n g(\lambda,\overline{x})c(\lambda)\,\mathrm{d}\lambda}{(w(\overline{x}))^2}\,. \tag{7.30}$$

Since $g(\lambda_0,\overline{x}) > 0$ and $w(\overline{x}) > 0$ for all $\overline{x} \in [0,1]$, setting $\frac{\mathrm{d}}{\mathrm{d}\overline{x}}\frac{g(\lambda_0,\overline{x})}{w(\overline{x})} = 0$ we immediately get:

$$\lambda_0 w(\overline{x}) - \int_{\lambda(m_\ell)}^{\lambda(m_u)} \lambda g(\lambda,\overline{x})c(\lambda)\,\mathrm{d}\lambda = 0\,. \tag{7.31}$$

Solving for λ_0 yields:

$$\lambda_0 = \frac{\int_{\lambda(m_\ell)}^{\lambda(m_u)} \lambda g(\lambda,\overline{x})c(\lambda)\,\mathrm{d}\lambda}{w(\overline{x})}\,. \tag{7.32}$$



(a) $\mu_1 = 0.05$ (b) $\mu_1 = 0.06$ (c) $\mu_1 = 0.07$

(d) $\mu_1 = 0.08$ (e) $\mu_1 = 0.15$ (f) $\mu_1 = 0.22$

(g) $\mu_1 = 0.23$ (h) $\mu_1 = 0.24$ (i) $\mu_1 = 0.25$

Figure 7.2.: The course of the ratio $\frac{f_{\overline{X}|\{\mu_1\}}(\overline{x})}{w(\overline{x})}$ for fixed sample size $n = 8$ and different values μ_1.

If there is a value $\overline{x} \in (0,1)$ which satisfies (7.32), the ratio $\frac{f_{\overline{X}|\{\mu_1\}}(\overline{x})}{w(\overline{x})}$ would have a maximum in $\overline{x}$. But, if:

$$\lambda_0 \leq \frac{\int_{\lambda(m_\ell)}^{\lambda(m_u)} \lambda g(\lambda,0)c(\lambda)\,\mathrm{d}\lambda}{w(0)} =: \lambda_{01} \tag{7.33}$$

and

$$\lambda_0 \geq \frac{\int_{\lambda(m_\ell)}^{\lambda(m_u)} \lambda g(\lambda,1)c(\lambda)\,\mathrm{d}\lambda}{w(1)} =: \lambda_{11}\,, \tag{7.34}$$

respectively, the ratio would have no maximum on $(0,1)$, i.e., it is monotonously decreasing or increasing. In these cases the respective optimal prediction sets ${}^*A_{\overline{X}}^{(\beta)}(\{\mu_1\})$ are as follows:

$${}^*A_{\overline{X}}^{(\beta)}(\{\mu_1\}) \;=\; [0, z_{\mu_1}(\beta)] \;\text{ if } \lambda(\mu_1) \leq \lambda_{01} \tag{7.35}$$

and

$${}^*A_{\overline{X}}^{(\beta)}(\{\mu_1\}) \;=\; [z_{\mu_1}(1-\beta), 1] \;\text{ if } \lambda(\mu_1) \geq \lambda_{11}\,, \tag{7.36}$$

where z_{μ_1} is the quantile function of $\overline{X}$. Equality in the right hand sides of (7.35) and (7.36) means that the maximum of the ratio is adopted in $\overline{x} = 0$ and $\overline{x} = 1$, respectively, which yields the same respective optimal prediction sets.

Example 7.2 *For* $\mathcal{D} = \{\mu_1 \,|\, 0.05 \leq \mu_1 \leq 0.25\}$ *and* $n = 8$ *we get*

$$\lambda_{01} = -17.501662 \tag{7.37}$$

and

$$\lambda_{11} = -3.750343 \tag{7.38}$$

That is, for $\lambda_0 < -17.501662 \Leftrightarrow \mu_1 < 0.0571374$ *we have for the prediction sets* ${}^*A_{\overline{X}}^{(\beta)}(\{\mu_1\}) = [0, z_{\mu_1}(\beta)]$, *and for* $\lambda_0 > -3.750343 \Leftrightarrow \mu_1 > 0.242567$ *we have* ${}^*A_{\overline{X}}^{(\beta)}(\{\mu_1\}) = [z_{\mu_1}(1-\beta), 1]$

For $\lambda_{01} < \lambda(\mu_1) < \lambda_{11}$, i.e., for $\mu_{1,01} = \mu_1(\lambda_{01}) < \mu_1 < \mu_{1,11} = \mu_1(\lambda_{11})$, the ratio $\frac{f_{\overline{X}|\{\mu_1\}}(\overline{x})}{w(\overline{x})}$ has a maximum on $(0,1)$.
Nevertheless, even if the ratio is unimodal on $(0,1)$, it is advantageous to compare beforehand the ratio's value in $\overline{x} = 0$ with that in $\overline{x} = z_{\mu_1}(\beta)$ and the ratio's value in $\overline{x} = 1$ with that in $\overline{x} = z_{\mu_1}(1-\beta)$. To make it more precise:

- If
$$\frac{f_{\overline{X}|\{\mu_1\}}(0)}{w(0)} \geq \frac{f_{\overline{X}|\{\mu_1\}}(z_{\mu_1}(\beta))}{w(z_{\mu_1}(\beta))} \tag{7.39}$$
then the optimal prediction set is:
$$^*A_{\overline{X}}^{(\beta)}(\{\mu_1\}) \;=\; [0, z_{\mu_1}(\beta)]\,. \tag{7.40}$$

- If
$$\frac{f_{\overline{X}|\{\mu_1\}}(1)}{w(1)} \geq \frac{f_{\overline{X}|\{\mu_1\}}(z_{\mu_1}(1-\beta))}{w(z_{\mu_1}(1-\beta))} \tag{7.41}$$
then the optimal prediction set is:
$$^*A_{\overline{X}}^{(\beta)}(\{\mu_1\}) \;=\; [z_{\mu_1}(1-\beta), 1]\,. \tag{7.42}$$

We note, that both cases include the situations given in (7.35) and (7.36), but more evaluations of the ratio and additionally of the quantile function are necessary, and thus, the determination of λ_{01} and λ_{11} in general pays.

Through (7.39) and (7.41) two values λ_{02} and λ_{12} for λ (and thus, for $\mu_1 = \mu_1(\lambda)$) are determined so that
$$\frac{f_{\overline{X}|\{\mu_1(\lambda_{02})\}}(0)}{w(0)} \;=\; \frac{f_{\overline{X}|\{\mu_1(\lambda_{02})\}}(z_{\mu_1(\lambda_{02})}(\beta))}{w(z_{\mu_1(\lambda_{02})}(\beta))}\,, \tag{7.43}$$
$$\frac{f_{\overline{X}|\{\mu_1(\lambda_{12})\}}(1)}{w(1)} \;=\; \frac{f_{\overline{X}|\{\mu_1(\lambda_{12})\}}(z_{\mu_1(\lambda_{12})}(1-\beta))}{w(z_{\mu_1(\lambda_{12})}(1-\beta))}\,. \tag{7.44}$$

Example 7.3 *Additionally, let* $\beta = 0.9$. *Then we have for* $\lambda_0 \leq \lambda_{02} = -13.142659 \Leftrightarrow \mu_1 \leq 0.0760861$ *the predictions* $^*A_{\overline{X}}^{(0.9)}(\{\mu_1\}) = [0, z_{\mu_1}(0.9)]$, *and for* $\lambda_0 \geq \lambda_{12} = -4.195361 \Leftrightarrow \mu_1 \geq 0.223063$ *the predictions* $^*A_{\overline{X}}^{(0.9)}(\{\mu_1\}) = [z_{\mu_1}(0.1), 1]$.

Analogously to example 6.3.4, for $\lambda_{02} < \lambda_0 < \lambda_{12}$ and $\mu_1(\lambda_{02}) < \mu_1 < \mu_1(\lambda_{12})$, respectively, we have the case where the prediction sets are of form $^*A_{\overline{X}}^{(\beta)}(\{\mu_1\}) = [\ell_{\overline{X}}(\mu_1), u_{\overline{X}}(\mu_1)]$ with bounds $\ell_{\overline{X}}(\mu_1)$, $u_{\overline{X}}(\mu_1)$ satisfying:
$$\frac{f_{\overline{X}|\{\mu_1\}}(\ell_{\overline{X}}(\mu_1))}{w(\ell_{\overline{X}}(\mu_1))} \;=\; \frac{f_{\overline{X}|\{\mu_1\}}(u_{\overline{X}}(\mu_1))}{w(u_{\overline{X}}(\mu_1))} \tag{7.45}$$
$$\frac{f_{\overline{X}|\{\mu_1\}}(0)}{w(0)} \;<\; \frac{f_{\overline{X}|\{\mu_1\}}(\ell_{\overline{X}}(\mu_1))}{w(\ell_{\overline{X}}(\mu_1))} \tag{7.46}$$
$$\frac{f_{\overline{X}|\{\mu_1\}}(1)}{w(1)} \;<\; \frac{f_{\overline{X}|\{\mu_1\}}(u_{\overline{X}}(\mu_1))}{w(u_{\overline{X}}(\mu_1))} \tag{7.47}$$

	${}^{*}A_{\overline{X}}^{(0.9)}(\{\mu_1\})$	
μ_1	$\ell_{\overline{X}}(\mu_1)$	$u_{\overline{X}}(\mu_1)$
0.0775	0.003726	0.114031
0.08	0.009590	0.117711
0.10	0.037668	0.150456
0.12	0.054381	0.187777
0.14	0.069967	0.228423
0.16	0.086672	0.277276
0.18	0.104272	0.350345
0.20	0.119293	0.498156
0.22	0.132697	0.883714
0.2225	0.134402	0.976479

Table 7.1.: ${}^{*}A_{\overline{X}}^{(0.9)}(\{\mu_1\}) = [\ell_{\overline{X}}(\mu_1), u_{\overline{X}}(\mu_1)]$ for some values μ_1.

Numerical methods have to be applied for the determination of $\ell_{\overline{X}}(\mu_1)$ and $u_{\overline{X}}(\mu_1)$.

Example 7.4 *For some values* $0.0760861 < \mu_1 < 0.223063$ *of* $\mathrm{E}[X]$ *the respective predictions* ${}^{*}A_{\overline{X}}^{(0.9)}(\{\mu_1\}) = [\ell_{\overline{X}}(\mu_1), u_{\overline{X}}(\mu_1)]$ *are given in table 7.1. Figure 7.3 displays the resulting* β*-uncertainty space* $\mathcal{U}_{\overline{X}}^{(0.9)}$ *for* $\overline{X}$*. Table 7.2 contains the measurements* ${}^{*}C_{\mathrm{E}[X]}^{(0.9)}(\{\overline{x}\})$ *for some values* $\overline{x}$*. For the following comparisons, we calculate the weighted volume to be* $V_w({}^{*}C_{\mathrm{E}[X]}^{(0.9)}) = 0.539789$*.*

At the beginning of this section we claimed, that we may assume without loss of generality $X \sim Mon(\mu_1)$ when the range of variability is known. Thus, let $Y \sim Mon(a, b, \mu_1)$ and $\boldsymbol{Y}$ be an i.i.d. sample for Y, then from proposition 3.11 we know that the transformed random variable $X := \frac{X-a}{b-a} \sim Mon(0, 1, \frac{\mu_1 - a}{b-a})$. The corresponding density functions are given by:

$$f_{X|\{\frac{\mu_1-a}{b-a}\}}(x) = (b-a) f_{Y|\{\mu_1\}}((b-a)x + a)\,, \tag{7.48}$$

$$f_{Y|\{\mu_1\}}(y) = \frac{1}{b-a} f_{X|\{\frac{\mu_1-a}{b-a}\}}\left(\frac{y-a}{b-a}\right)\,. \tag{7.49}$$

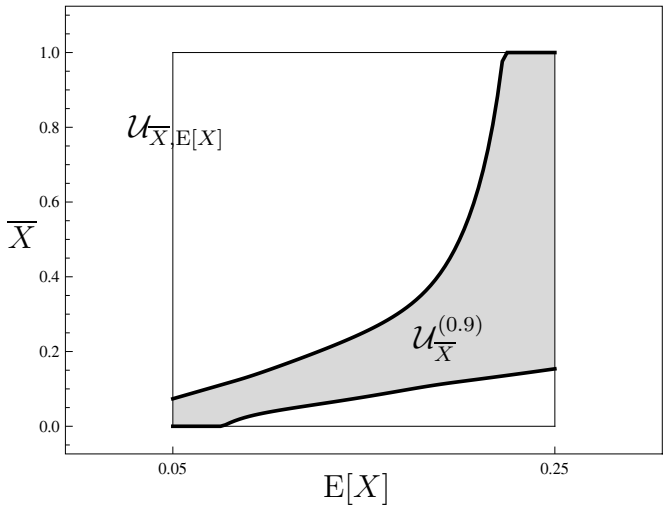


Figure 7.3.: Optimal β-uncertainty space $\mathcal{U}_{\overline{X}}^{(0.9)}$ for $\overline{X}$. The corresponding 0.9-measurement procedure ${}^{*}C_{\mathrm{E}[X]}^{(0.9)}$ is the Neyman 0.9-measurement procedure.

	${}^{*}C_{\mathrm{E}[X]}^{(0.9)}(\{\overline{x}\})$	
$\overline{x}$	$\ell_{\mathrm{E}[X]}(\overline{x})$	$u_{\mathrm{E}[X]}(\overline{x})$
0.025	0.05	0.088810
0.05	0.05	0.114415
0.075	0.050973	0.146223
0.1	0.067964	0.175065
0.125	0.084874	0.208572
0.15	0.099746	0.244846
0.175	0.113307	0.25
0.2	0.126246	0.25
0.4	0.188651	0.25
0.6	0.207734	0.25
0.8	0.217264	0.25
0.9	0.220474	0.25

Table 7.2.: Measurements ${}^{*}C_{\mathrm{E}[X]}^{(0.9)}(\{\overline{x}\}) = [\ell_{\mathrm{E}[X]}(\overline{x}), u_{\mathrm{E}[X]}(\overline{x})]$ for some values $\overline{x}$.

Denoting the weight function with respect to $\boldsymbol{X}$ by $w_{\boldsymbol{X}}$ and to $\boldsymbol{Y}$ by $w_{\boldsymbol{Y}}$, and defining:

$$\mu_1' = \frac{\mu_1 - a}{b - a} \tag{7.50}$$

$$\lambda_X = \lambda \left(\frac{\mu_1 - a}{b - a} \right) \tag{7.51}$$

$$x_i = \frac{y_i - a}{b - a}, \; i = 1, \ldots, n \tag{7.52}$$

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \tag{7.53}$$

$$m_\ell' = \frac{m_\ell - a}{b - a} \tag{7.54}$$

$$m_u' = \frac{m_u - a}{b - a} \tag{7.55}$$

we obtain:

$$\begin{aligned}
w_{\boldsymbol{Y}}(\boldsymbol{y}) &= \int_{m_\ell}^{m_u} \prod_{i=1}^{n} f_{Y|\{\mu_1\}}(y_i) \, \mathrm{d}\mu_1 \\
&= \int_{m_\ell}^{m_u} \prod_{i=1}^{n} \frac{1}{b-a} f_{X|\{\mu_1'\}}(x_i) \, \mathrm{d}\mu_1 \\
&= \int_{m_\ell'}^{m_u'} \frac{1}{(b-a)^n} \left(\frac{\lambda_X}{e^{\lambda_X} - 1} \right)^n e^{\sum_{i=1}^n x_i} (b-a) \, \mathrm{d}\mu_1' \\
&= \frac{1}{(b-a)^{n-1}} \int_{m_\ell'}^{m_u'} \left(\frac{\lambda_X}{e^{\lambda_X} - 1} \right)^n e^{\sum_{i=1}^n x_i} \, \mathrm{d}\mu_1' \\
&= \frac{1}{(b-a)^{n-1}} \, w_{\boldsymbol{X}}(\boldsymbol{x}) \, .
\end{aligned} \tag{7.56}$$

Thus, we get for the ratio $\frac{f_{\boldsymbol{Y}|\{\mu_1\}}(\boldsymbol{y})}{w_{\boldsymbol{Y}}(\boldsymbol{y})}$ the following expression:

$$\begin{aligned}
\frac{f_{\boldsymbol{Y}|\{\mu_1\}}(\boldsymbol{y})}{w_{\boldsymbol{Y}}(\boldsymbol{y})} &= \frac{\frac{1}{(b-a)^n} f_{\boldsymbol{X}|\{\mu_1'\}}(\boldsymbol{x})}{\frac{1}{(b-a)^{n-1}} w_{\boldsymbol{X}}(\boldsymbol{x})} \\
&= \frac{1}{b-a} \frac{f_{\boldsymbol{X}|\{\mu_1'\}}(\boldsymbol{x})}{w_{\boldsymbol{X}}(\boldsymbol{x})} \, .
\end{aligned} \tag{7.57}$$

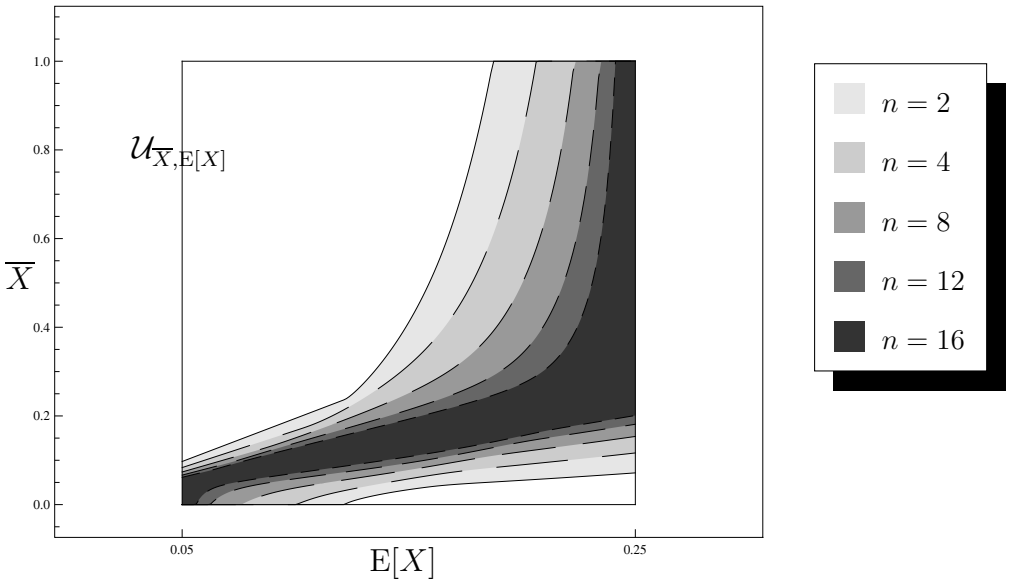


Figure 7.4.: Optimal β-uncertainty spaces $\mathcal{U}_{\overline{X}}^{(0.9)}$ for $\overline{X}$ and different values of the sample size $n = 2, 4, 8, 12, 16$.

That is, the value of the ratio with respect to Y equals the value of the ratio with respect to X scaled with $\frac{1}{b-a}$. With the same arguments which justify the change from $\boldsymbol{X}$ to $\overline{X}$, we conclude that the optimal prediction sets for $\overline{Y}$ are affine transformations of those for $\overline{X}$.

Improving $\mathcal{U}_{\overline{X}}^{(\beta)}$ by increasing n

Commonly known is the fact, that measurement for the unknown value of a deterministic variable may be improved by increasing the sample size n while keeping the reliability β constant. Figure 7.4 shows how the respective β-uncertainty spaces $\mathcal{U}_{\overline{X},\mathrm{E}[X]} = [0, 1] \times \mathcal{D}$ for $n = 2, 4, 8, 12, 16$ become successively smaller, i.e., the corresponding β-measurement procedures yield smaller measurements. The calculations were performed for the ignorance space $\mathcal{D} = \{\mu_1 \,|\, 0.05 \leq \mu_1 \leq 0.25\}$ and $\beta = 0.9$ as in example 7.4.

7.2.2. Approximative Measurement Procedures for $\mathrm{E}[X]$

Approximations for measuring the unknown value μ_1 of $\mathrm{E}[X]$ where X is distributed according to the monotonic distribution $Mon(\mu_1)$ can be understood in different ways. That is, we have four possibilities

(a) approximate the monotonic distribution $Mon(\mu_1)$ by the exponential distribution $EXP(1/\mu_1)$ and derive a Neyman β-measurement procedure ${}^{*a}C^{(\beta)}_{\mathrm{E}[Y]}$

(b) approximate $Mon(\mu_1)$ by $EXP(1/\mu_1)$ and derive a measurement procedure ${}^{HPR}C^{(\beta)}_{\mathrm{E}[Y]}$ based on highest probability regions (HPR)

(c) approximate $Mon(\mu_1)$ by $EXP(1/\mu_1)$ and derive a measurement procedure ${}^{LR}C^{(\beta)}_{\mathrm{E}[Y]}$ based on the likelihood ratio test statistic (LR)

(d) approximate the distribution of $\overline{X}$ by the normal distribution and derive approximative confidence intervals ${}^{N}C_{\mathrm{E}[X]}$ (normal approximation)

Since all measurement procedures, whether based on predictions or not, define respective prediction procedures, we are able to evaluate them by calculating their respective achieved reliabilities – under the condition of the monotonic distribution, i.e., the true distribution. Throughout the following examples we choose the identical settings as in the examples of section 7.2.1, that is,

$$\mathcal{D} = \{\mu_1 \,|\, 0.05 \leq \mu_1 \leq 0.25\}\,, \tag{7.58}$$

$$\beta = 0.9\,, \tag{7.59}$$

$$n = 8\,. \tag{7.60}$$

To distinguish between the true random variable $X|\{\mu_1\} \sim Mon(\mu_1)$, we denote the approximation $Y|\{\mu_1\}$. And thus, the measurement procedure based on the approximation will yield in fact measurements for $\mathrm{E}[Y]$ for every realisation $\overline{x}$ of $\overline{X}|\{\mu_1\}$, i.e., $C^{(\beta)}_{\mathrm{E}[Y]}(\{\overline{x}\})$. Since the true range of variability is $\mathcal{X} = [0, 1]$, i.e., $\overline{X}$ may only adopt values $0 \leq \overline{x} \leq 1$, we only give measurements $C^{(\beta)}_{\mathrm{E}[Y]}(\{\overline{x}\})$ for some values $\overline{x} \in [0, 1]$.

If the reliability of a procedure is at least β, the weighted volumes V_w of the respective measurement procedures $C^{(\beta)}_{\mathrm{E}[Y]}$ will be a second criterion for evaluation the procedure. We will also directly compare lengths of the respective measurements

– which we will realise to be rather misleading, since they should be weighted by the marginal distribution of $\overline{X}$. To overcome this wrong perception, we introduce a new graphical illustration which incorporates the marginal distribution.

To the end of this section, we give some more numerical results for different values n of the sample size and some other ingnorance spaces $\mathcal{D}$.

7.2.2.1. Neyman Measurement Procedures for $\mathrm{E}[Y]$, $X|\{\mu_1\} \sim Mon(\mu_1)$ approximated by $Y|\{\mu_1\} \sim EXP(1/\mu_1)$

We approximate the monotonic distribution $Mon(\mu_1)$ of $X|\{\mu_1\}$ by the exponential distribution $EXP(1/\mu_1)$. That is, we have a random variable $Y|\{\mu_1\}$ with density function

$$f_{Y|\{\mu_1\}}(y) = \frac{1}{\mu_1} e^{-y/\mu_1} \mathbb{1}_{[0,+\infty)}(y) \tag{7.61}$$

and the range of variability

$$\mathcal{Y}(\{\mu_1\}) = \{y \,|\, 0 \leq y < +\infty\} = [0, +\infty)\,. \tag{7.62}$$

The joint density function for an i.i.d. sample $\boldsymbol{Y}|\{\mu_1\}$ of size n for $Y|\{\mu_1\}$ is

$$f_{\boldsymbol{Y}|\{\mu_1\}}(\boldsymbol{y}) = \left(\frac{1}{\mu_1}\right)^n e^{-\sum_{i=1}^n y_i/\mu_1} \mathbb{1}_{[0,+\infty)^n}(\boldsymbol{y}) \tag{7.63}$$

which yields the weight function $w_{\boldsymbol{Y}}$:

$$w_{\boldsymbol{Y}}(\boldsymbol{y}) = \int_{m_\ell}^{m_u} \left(\frac{1}{\mu_1}\right)^n e^{-\sum_{i=1}^n y_i/\mu_1} \mathbb{1}_{[0,+\infty)^n}(\boldsymbol{y})\, \mathrm{d}\mu_1\,. \tag{7.64}$$

In analogy to (7.5) and (7.7), we note that both the joint density function and the weight function in fact only depend on $\sum_{i=1}^n y_i$. Since $T|\{\mu_1\} := \sum_{i=1}^n Y_i|\{\mu_1\} \sim GAM(1/\mu_1, n)$ and, thus,

$$f_{\overline{Y}|\{\mu_1\}}(\overline{y}) = n f_{T|\{\mu_1\}}(n\overline{y}) = n \frac{\left(\frac{1}{\mu_1}\right)^n}{\Gamma(n)} (n\overline{y})^{n-1} e^{-n\overline{y}/\mu_1} \mathbb{1}_{[0,+\infty)}(\overline{y})\,, \tag{7.65}$$

we get the corresponding weight function

$$w_{\overline{Y}}(\overline{y}) = \int_{m_\ell}^{m_u} n \frac{\left(\frac{1}{\mu_1}\right)^n}{\Gamma(n)} (n\overline{y})^{n-1} e^{-n\overline{y}/\mu_1} \mathbb{1}_{[0,+\infty)}(\overline{y})\, \mathrm{d}\mu_1\,. \tag{7.66}$$

Again, for all $\boldsymbol{y}$ with $\sum_{i=1}^{n} y_i = n\overline{y}$ we get equality of the corresponding ratios:

$$\frac{f_{\boldsymbol{Y}|\{\mu_1\}}(\boldsymbol{y})}{w_{\boldsymbol{Y}}(\boldsymbol{y})} = \frac{f_{\overline{Y}|\{\mu_1\}}(\overline{y})}{w_{\overline{Y}}(\overline{y})} \tag{7.67}$$

Consequently, we do not derive optimal prediction sets for $\boldsymbol{Y}|\{\mu_1\}$ but for $\overline{Y}|\{\mu_1\} = \frac{1}{n}\sum_{i=1}^{n} Y_i|\{\mu_1\}$.

Example 7.5 *Since the range of variability* $\mathcal{Y}(\{\mu_1\}) = \{\overline{y}\,|\,0 \leq \overline{y} < +\infty\} = [0, +\infty)$ *of* $\overline{Y}$ *is unbounded we get also an unbounded uncertainty space* $\mathcal{U}_{\overline{Y},\mathrm{E}[Y]}$ *for the pair of variables* $(\overline{Y}, \mathrm{E}[Y])$*:*

$$\mathcal{U}_{\overline{Y},\mathrm{E}[Y]} = [0.05, 0.25] \times [0, +\infty)\,. \tag{7.68}$$

The resulting β*-uncertainty space* $\mathcal{U}_{\overline{Y}}^{(\beta)}$ *for* $\overline{Y}$ *and, thus, the* β*-uncertainty space* $\mathcal{U}_{\mathrm{E}[Y]}^{(\beta)}$ *for* $\mathrm{E}[Y]$ *are subsets of* $\mathcal{U}_{\overline{Y},\mathrm{E}[Y]}$*. The completeness requirement (see section 6.3.3) indicates, that both are unbounded, too. Figure 7.5 displays the course of the ratio* $\frac{f_{\overline{Y}|\{\mu_1\}}(\overline{y})}{w(\overline{y})}$ *for different values* μ_1*. The changes in its courses are very similar to those in example 7.1 and, again, we might identify some value* m*, so that for all* $0.05 = m_\ell \leq \mu_1 \leq m$ *the corresponding optimal predictions have the form*

$${}^{*a}A_{\overline{Y}}^{(\beta)}(\{\mu_1\}) = [0, z_{\mu_1}(\beta)]\,. \tag{7.69}$$

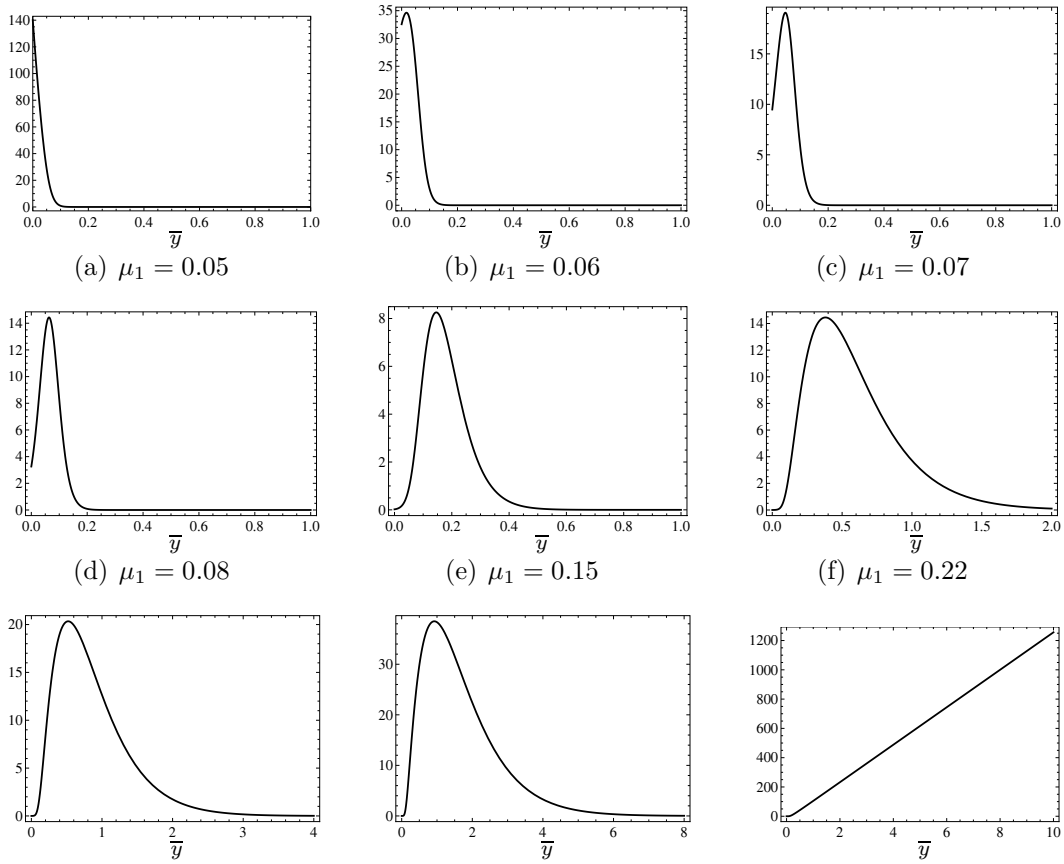
But we note, that for $\mu_1 = m_u = 0.25$ *we get the unbounded prediction*

$${}^{*a}A_{\overline{Y}}^{(\beta)}(\{0.25\}) = [z_{0.25}(1-\beta), +\infty)\,. \tag{7.70}$$

As a side-effect, this guarantees the completeness requirement. Calculating the respective predictions ${}^{*a}A_{\overline{Y}}^{(0.9)}(\{\mu_1\})$ *yields the* 0.9*-uncertainty spaces* $\mathcal{U}_{\overline{Y}}^{(0.9)}$ *for* $\overline{Y}$ *and* $\mathcal{U}_{\mathrm{E}[Y]}^{(0.9)}$ *for* $\mathrm{E}[Y]$*, respectively, displayed in figure 7.6. As mentioned, the first criterion to evaluate the usefulness of the approximation and the resulting measurement procedure, respectively, is to calculate the achieved reliability under the true distribution, i.e., the monotonic distribution. Since we already have derived predictions* ${}^{*a}A_{\overline{Y}}^{(0.9)}(\{\mu_1\})$*, we can directly calculate their probability, i.e.,*

$$\mathrm{P}_{\overline{X}|\{\mu_1\}}({}^{*a}A_{\overline{Y}}^{(0.9)}(\{\mu_1\})) \tag{7.71}$$

In figure 7.7 the probability is displayed as function of the value μ_1 *of* $\mathrm{E}[X]$*. We note, that it never falls below the required reliability of* 0.9 *but exceeds it in*



(a) $\mu_1 = 0.05$ (b) $\mu_1 = 0.06$ (c) $\mu_1 = 0.07$

(d) $\mu_1 = 0.08$ (e) $\mu_1 = 0.15$ (f) $\mu_1 = 0.22$

(g) $\mu_1 = 0.23$ (h) $\mu_1 = 0.24$ (i) $\mu_1 = 0.25$

Figure 7.5.: The course of the ratio $\frac{f_{\overline{Y}|\{\mu_1\}}(\overline{y})}{w(\overline{y})}$ for sample size $n = 8$ and different values μ_1.

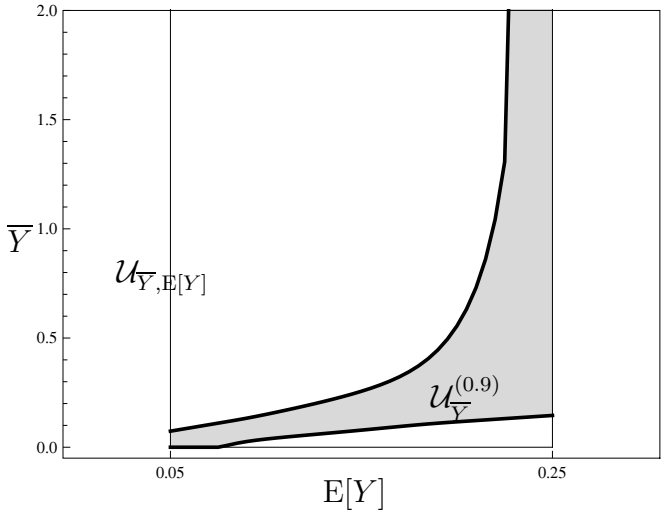


Figure 7.6.: Optimal β-uncertainty space $\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$. The corresponding 0.9-measurement procedure ${}^{*a}C_{\mathrm{E}[Y]}^{(0.9)}$ is the Neyman 0.9-measurement procedure.

particular for large values $\mu_1 \in [0.05, 0.25]$. *That is, the Neyman measurement procedure based on the exponential distribution as approximation for the monotonic disribution meets the required reliability and, thus, would be a reasonable measurement procedure.*

The second criterion is the weighted volume. Its value is given by $V_w({}^{*a}C_{\mathrm{E}[Y]}^{(0.9)}) = 0.550061$, *which is slightly larger than that of the Neyman measurement procedure for the monotonic distribution* $V_w({}^{*}C_{\mathrm{E}[X]}^{(0.9)}) = 0.539789$. *Additionally, table 7.3 lists some measurements* ${}^{*a}C_{\mathrm{E}[Y]}^{(0.9)}(\{\overline{x}\})$ *for some values* $0 \leq \overline{x} \leq 1$, *since for the true random variable* $X|\{\mu_1\}$ *and, thus,* $\overline{X}|\{\mu_1\}$, *only those values are observed.*

We also added the respective measurements ${}^{*}C_{\mathrm{E}[X]}^{(0.9)}(\{\overline{x}\})$ *resulting from the* 0.9*-Neyman measurement procedure for the monotonic distribution* $Mon(\mu_1)$ *(see table 7.2) and calculated the relative difference in their lengths:*

$$\Delta(\overline{x}) := \frac{|{}^{*a}C_{\mathrm{E}[Y]}^{(0.9)}(\{\overline{x}\})| - |{}^{*}C_{\mathrm{E}[X]}^{(0.9)}(\{\overline{x}\})|}{|{}^{*}C_{\mathrm{E}[X]}^{(0.9)}(\{\overline{x}\})|}\,. \tag{7.72}$$

In most of the here considered values $\overline{x}$ *the resulting measurements* ${}^{*a}C_{\mathrm{E}[Y]}^{(0.9)}(\{\overline{x}\})$ *based on the approximation with the exponential distribution are longer then the*

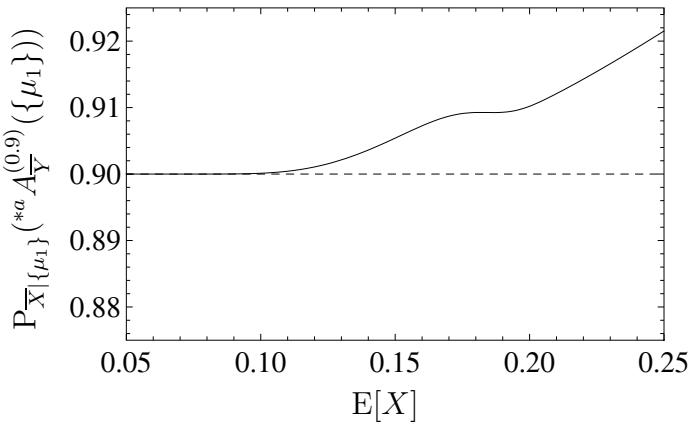


Figure 7.7.: The true probability of the prediction sets ${}^{*a}A_{\overline{Y}}^{(0.9)}(\{\mu_1\})$ based on the exponential distribution $EXP(1/\mu_1)$ as approximation for the monotonic distribution $Mon(\mu_1)$.

	${}^{*a}C_{\mathrm{E}[Y]}^{(0.9)}(\{\overline{x}\})$		${}^{*}C_{\mathrm{E}[X]}^{(0.9)}(\{\overline{x}\})$		
$\overline{x}$	${}^{a}\ell_{\mathrm{E}[Y]}(\overline{x})$	${}^{a}u_{\mathrm{E}[Y]}(\overline{x})$	$\ell_{\mathrm{E}[X]}(\overline{x})$	$u_{\mathrm{E}[X]}(\overline{x})$	$\Delta\cdot$ 100 %
0.025	0.05	0.089168	0.05	0.088810	+0.922
0.05	0.05	0.114446	0.05	0.114415	+0.048
0.075	0.050973	0.146028	0.050973	0.146223	-0.205
0.1	0.067964	0.176208	0.067964	0.175065	+1.067
0.125	0.084888	0.214773	0.084874	0.208572	+5.002
0.15	0.099800	0.25	0.099746	0.244846	+3.515
0.175	0.113212	0.25	0.113307	0.25	+0.069
0.2	0.125765	0.25	0.126246	0.25	+0.389
0.4	0.184178	0.25	0.188651	0.25	+7.291
0.6	0.203079	0.25	0.207734	0.25	+11.014
0.8	0.212879	0.25	0.217264	0.25	+13.395
0.9	0.216275	0.25	0.220474	0.25	+14.221

Table 7.3.: Measurements ${}^{*a}C_{\mathrm{E}[Y]}^{(0.9)}(\{\overline{x}\}) = [{}^{a}\ell_{\mathrm{E}[Y]}(\overline{x}), {}^{a}u_{\mathrm{E}[Y]}(\overline{x})]$ for some values $\overline{x}$. For comparison the measurements ${}^{*}C_{\mathrm{E}[X]}^{(0.9)}(\{\overline{x}\}) = [\ell_{\mathrm{E}[X]}(\overline{x}), u_{\mathrm{E}[X]}(\overline{x})]$ are added based on the monotonic distribution together with the relative difference $\Delta(\overline{x})$ (7.72).

measurements ${}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{\overline{x}\})$ *based on the true monotonic distribution.*

Since the required reliability is met, we may conclude that the measurement procedure based on the approximation is reasonable, but will yield unnecessary large measurements, because of the greater value of the weighted volume V_w.

7.2.2.2. HPR Measurement Procedures for $\mathrm{E}[X]$, $X|\{\mu_1\} \sim Mon(\mu_1)$ approximated by $Y|\{\mu_1\} \sim Exp(1/\mu_1)$

This method to derive prediction regions was introduced in section 4.4.3.1, and in example 4.58 we already have discussed this for the exponential distribution $EXP(\vartheta)$ where the value ϑ of the distribution parameter, i.e., the deterministic variable equals $\frac{1}{\mu_1}$. Since we want to analyse the reparameterised exponential distribution $EXP(1/\mu_1)$, we give a short collection of the previously derived results. The joint density function $f_{\boldsymbol{Y}|\{\mu_1\}}$ of $\boldsymbol{Y}|\{\mu_1\}$ is given by (7.63):

$$f_{\boldsymbol{Y}|\{\mu_1\}}(\boldsymbol{y}) = \left(\frac{1}{\mu_1}\right)^n e^{-\sum_{i=1}^n y_i/\mu_1} \mathbb{1}_{[0,+\infty)^n}(\boldsymbol{y})\,,$$

whose value only depends on $\sum_{i=1}^n y_i$ if $\boldsymbol{y} \in [0,+\infty)^n$. The smaller $\sum_{i=1}^n y_i$ the larger $f_{\boldsymbol{Y}|\{\mu_1\}}(\boldsymbol{y})$, i.e., the β-acceptance regions are of form

$$^{HPR}A^{(\beta)}_{\boldsymbol{Y}}(\{\mu_1\}) = \left\{ \boldsymbol{y} \in [0,+\infty)^n \,\middle|\, \sum_{i=1}^n y_i \le c^* \text{ with } \int\limits_{\sum_{i=1}^n y_i \le c^*} \left(\frac{1}{\mu_1}\right)^n e^{-\sum_{i=1}^n y_i/\mu_1}\, \mathrm{d}\boldsymbol{y} = \beta \right\}. \quad (7.73)$$

The integral

$$\int\limits_{\sum_{i=1}^n y_i \le c^*} \left(\frac{1}{\mu_1}\right)^n e^{-\sum_{i=1}^n y_i/\mu_1}\, \mathrm{d}\boldsymbol{y} \quad (7.74)$$

equals the distribution function $G_{1/\mu_1,n}(c^*)$ of the gamma distribution $GAM(1/\mu_1,n)$[2]. And since $c^*_\beta(\mu_1) = c^*_\beta(1)\cdot\mu_1$, we only need to numerically solve $G_{1,n}(c^*) = \beta$ for $c^* = c^*_\beta(1)$, which is in fact the β-quantile of $GAM(1,n)$[3]. The determination of all β-acceptance regions is already completed with the determination of

[2] This follows from $\sum_{i=1}^n Y_i|\{\mu_1\} \sim GAM(1/\mu_1,n)$

[3] If $Z|\{\mu_1\} \sim GAM(1/\mu_1,n)$, then $(2Z|\{\mu_1\})/\mu_1 \sim \chi_{2n}$, i.e., the chi-square distribution with $2n$ degrees of freedom. This may also be utilised for the determination of c^*.

${}^{HPR}A_{\boldsymbol{Y}}^{(\beta)}(\{1\})$, since all other ${}^{HPR}A_{\boldsymbol{Y}}^{(\beta)}(\{\mu_1\})$ for $\mu_1 \neq 1$ are scaled versions of ${}^{HPR}A_{\boldsymbol{Y}}^{(\beta)}(\{1\})$:

$$
\begin{aligned}
{}^{HPR}A_{\boldsymbol{Y}}^{(\beta)}(\{\mu_1\}) = \Bigg\{ \boldsymbol{y} \in [0,+\infty)^n \,\Bigg|\, & \\
& \sum_{i=1}^{n} y_i \leq c_\beta^*(1) \cdot \mu_1 \text{ with } \int\limits_{\sum_{i=1}^{n} y_i \leq c_\beta^*(1)} e^{-\sum_{i=1}^{n} y_i} \, \mathrm{d}\boldsymbol{y} = \beta \Bigg\}
\end{aligned} \tag{7.75}
$$

Obviously, for $n > 3$, it is impossible to give graphical illustrations of the resulting β-uncertainty spaces ${}^{HPR}\mathcal{U}_{\boldsymbol{Y}}^{(\beta)}$. But since the condition on $\boldsymbol{y}$ only refers to the sum $\sum_{i=1}^{n} y_i$, we may map each ${}^{HPR}A_{\boldsymbol{Y}}^{(\beta)}(\{\mu_1\})$ onto the respective interval $[0, c_\beta^*(\mu_1)]$. That is, with the scaling property we have

$$
S \colon {}^{HPR}\mathcal{U}_{\boldsymbol{Y}}^{(\beta)} \to \mathbb{R} \,, \tag{7.76}
$$

$$
{}^{HPR}A_{\boldsymbol{Y}}^{(\beta)}(\{\mu_1\}) \mapsto S\left({}^{HPR}A_{\boldsymbol{Y}}^{(\beta)}(\{\mu_1\})\right) = [0, c_\beta^*(\mu_1)] = [0, c_\beta^*(1) \cdot \mu_1] \,. \tag{7.77}
$$

Additionally, we define the image of ${}^{HPR}\mathcal{U}_{\boldsymbol{Y}}^{(\beta)}$ under S:

$$
\begin{aligned}
S\left({}^{HPR}\mathcal{U}_{\boldsymbol{Y}}^{(\beta)}\right) &:= \bigcup_{\mu_1 \in \mathcal{D}} \{\mu_1\} \times S\left({}^{HPR}A_{\boldsymbol{Y}}^{(\beta)}(\{\mu_1\})\right) \\
&= \bigcup_{\mu_1 \in \mathcal{D}} \{\mu_1\} \times [0, c_\beta^*(1) \cdot \mu_1]
\end{aligned} \tag{7.78}
$$

Given a bounded ignorance space $\mathcal{D} = \{\mu_1 \,|\, 0 < m_\ell \leq \mu_1 \leq m_u < +\infty\}$ for $\mathrm{E}[Y]$, it is also obvious, that the union of all β-acceptance regions, i.e.,

$$
\bigcup_{\mu_1 \in \mathcal{D}} {}^{HPR}A_{\boldsymbol{Y}}^{(\beta)}(\{\mu_1\}) = {}^{HPR}A_{\boldsymbol{Y}}^{(\beta)}(\mathcal{D}) \tag{7.79}
$$

does not equal the range of variability $\boldsymbol{\mathcal{Y}} = [0,+\infty)^n$ of $\boldsymbol{Y}$ and, therefore, the completeness requirement is violated. Since Y is an approximation for X, this does not need to be a problem, as long as, ${}^{HPR}A_{\boldsymbol{Y}}^{(\beta)}(\mathcal{D})$ covers the true range of variability $\boldsymbol{\mathcal{X}} = [0,1]^n$ of $\boldsymbol{X}$. If this is also not the case, we replace ${}^{HPR}A_{\boldsymbol{Y}}^{(\beta)}(\{m_u\})$, which covers ${}^{HPR}A_{\boldsymbol{Y}}^{(\beta)}(\{\mu_1\})$ for all other values $\mu_1 \in \mathcal{D}$, by $[0,1]^n$. For the above described mapping S (7.77) this is equivalent to replacing $[0, c_\beta^*(\mu_1)]$ with $[0, n]$.

To make β-uncertainty spaces ${}^{HPR}\mathcal{U}_{\boldsymbol{Y}}^{(\beta)}$ and in particular $S({}^{HPR}\mathcal{U}_{\boldsymbol{Y}}^{(\beta)})$ comparable for different values n of the sample size, we scale $[0, c_\beta^*(\mu_1)]$ with $\frac{1}{n}$. That is, the

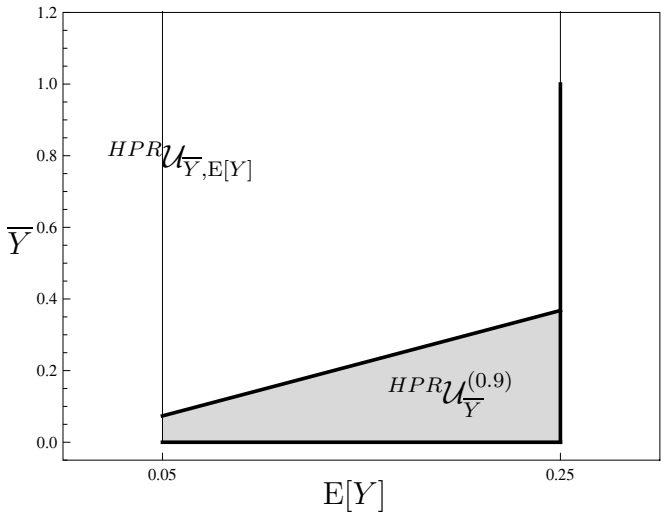


Figure 7.8.: Mapped β-uncertainty space $S(^{HPR}\mathcal{U}^{(0.9)}_{\boldsymbol{Y}})$ for $\boldsymbol{Y}$ and scaled with $\frac{1}{8}$. Note, the prediction region for $\mu_1 = 0.25$ is $[0, 1]$.

mapped prediction regions $[0, c^*_\beta(\mu_1)]$ for $\sum_{i=1}^n Y_i$ become prediction regions for $\frac{1}{n}\sum_{i=1}^n Y_i = \overline{Y}$.

Example 7.6 *Let* $\mathcal{D} = \{\mu_1 \,|\, 0.05 \leq \mu_1 \leq 0.25\}$, $n = 8$ *and* $\beta = 0.9$. *Since* $c^*_{0.9}(1) = 11.770914$, *we get* $c^*_{0.9}(0.25) = 2.942786$, *which is less than* $n = 8$ *and, thus, we need to replace* $[0, c^*_{0.9}(0.25)]$ *with* $[0, 8]$. *The* β*-uncertainty space* $^{HPR}\mathcal{U}^{(0.9)}_{\boldsymbol{Y}}$ *under the mapping* S (7.77) *then is:*

$$S\left(^{HPR}\mathcal{U}^{(0.9)}_{\boldsymbol{Y}}\right) = (\{0.25\} \times [0, 8]) \cup \bigcup_{0.05 \leq \mu_1 < 0.25} \{\mu_1\} \times [0, 11.770914 \cdot \mu_1]\,. \quad (7.80)$$

Figure 7.8 displays the β*-uncertainty space* $S\left(^{HPR}\mathcal{U}^{(0.9)}_{\boldsymbol{Y}}\right)$ *scaled with* $\frac{1}{8}$. *Figure 7.9 displays the true probability of the prediction regions* $^{HPR}A^{(0.9)}_{\boldsymbol{Y}}(\{\mu_1\})$ *and the mapped and scaled prediction regions* $\frac{1}{8}S\left(^{HPR}A^{(0.9)}_{\boldsymbol{Y}}(\{\mu_1\})\right)$, *respectively, under the monotonic distribution* $Mon(\mu_1)$.

To determine the respective measurements $^{HPR}C^{(\beta)}_{\mathrm{E}[Y]}(\{\boldsymbol{x}\})$ based on the original prediction regions $^{HPR}A^{(\beta)}_{\boldsymbol{Y}}(\{\mu_1\})$ we would need to solve

$$\int_{\sum_{i=1}^n t_i \leq \sum_{i=1}^n x_i} \left(\frac{1}{\mu_1}\right) e^{\sum_{i=1}^n t_i/\mu_1}\, \mathrm{d}\boldsymbol{t} = \beta \quad (7.81)$$

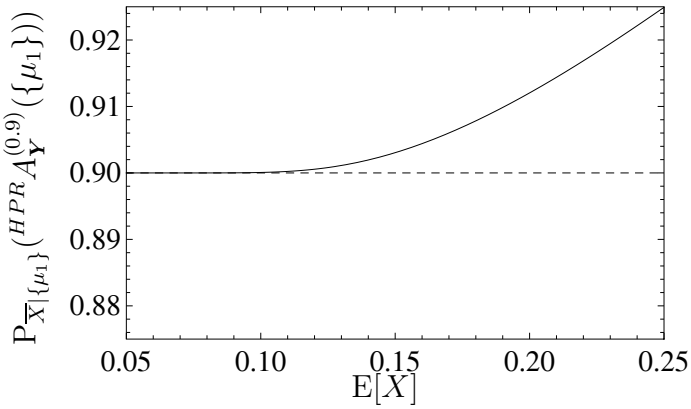


Figure 7.9.: The true probability of the prediction sets ${}^{HPR}A_{\boldsymbol{Y}}^{(0.9)}(\{\mu_1\})$ based on the exponential distribution $EXP(1/\mu_1)$ as approximation for the monotonic distribution $Mon(\mu_1)$. Since the prediction region for $\mu_1 = 0.25$ is $[0, 1]$, its probability equals 1 and is omitted in the graphic.

for μ_1. Again, the integral equals the distribution function $G_{1/\mu_1,n}(\sum_{i=1}^n y_i)$ of the gamma distribution $GAM(1/\mu_1, n)$. Equivalent, is to solve

$$\sum_{i=1}^{n} y_i = c_\beta^*(1) \cdot \mu_1 \tag{7.82}$$

for μ_1 – obviously a lot easier, which results in

$$\mu_1 = \frac{\sum_{i=1}^n y_i}{c_\beta^*(1)}\,. \tag{7.83}$$

Then, we get for the measurement ${}^{HPR}C_{\mathrm{E}[Y]}^{(\beta)}(\{\boldsymbol{x}\})$ the following

$${}^{HPR}C_{\mathrm{E}[Y]}^{(\beta)}(\{\boldsymbol{x}\}) = \left\{ \mu_1 \,\middle|\, \mu_1 \geq \frac{\sum_{i=1}^n x_i}{c_\beta^*(1)} \right\} \tag{7.84}$$

We note, that all $\boldsymbol{x}$ with the same value of $\sum_{i=1}^n x_i$ yield the same measurement. That is, we are able to write ${}^{HPR}C_{\mathrm{E}[Y]}^{(\beta)}(\{\overline{x}\})$, again.
In case of a bounded ignorance space $\mathcal{D} = \{\mu_1 \,|\, 0 < m_\ell \leq \mu_1 \leq m_u < +\infty\}$ problems arises, when $\sum_{i=1}^n x_i > \frac{m_u}{c_\beta^*(1)}$ and $\overline{x} > \frac{m_u}{n c_\beta^*(1)}$. Through the completion of the β-uncertainty space, the resulting measurement for all $\boldsymbol{x}$ with $\sum_{i=1}^n x_i > \frac{m_u}{c_\beta^*(1)}$ then is given by ${}^{HPR}C_{\mathrm{E}[Y]}^{(\beta)}(\{\overline{x}\}) = \{m_u\}$.

	${}^{HPR}C^{(0.9)}_{\mathrm{E}[Y]}(\{\overline{x}\})$		${}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{\overline{x}\})$		
$\overline{x}$	${}^{HPR}\ell_{\mathrm{E}[Y]}(\overline{x})$	${}^{HPR}u_{\mathrm{E}[Y]}(\overline{x})$	$\ell_{\mathrm{E}[X]}(\overline{x})$	$u_{\mathrm{E}[X]}(\overline{x})$	$\Delta\cdot$ 100 %
0.025	0.05	0.25	0.05	0.088810	+415.331
0.05	0.05	0.25	0.05	0.114415	+210.487
0.075	0.050973	0.25	0.050973	0.146223	+108.952
0.1	0.067964	0.25	0.067964	0.175065	+69.967
0.125	0.084955	0.25	0.084874	0.208572	+33.426
0.15	0.101946	0.25	0.099746	0.244846	+2.036
0.175	0.118937	0.25	0.113307	0.25	-4.119
0.2	0.135928	0.25	0.126246	0.25	-7.824
0.4	0.25	0.25	0.188651	0.25	-100
0.6	0.25	0.25	0.207734	0.25	-100
0.8	0.25	0.25	0.217264	0.25	-100
0.9	0.25	0.25	0.220474	0.25	-100

Table 7.4.: Measurements ${}^{HPR}C^{(0.9)}_{\mathrm{E}[Y]}(\{\overline{x}\}) = [{}^{HPR}\ell_{\mathrm{E}[Y]}(\overline{x}), {}^{HPR}u_{\mathrm{E}[Y]}(\overline{x})]$ for some values $\overline{x}$. For comparison the measurements ${}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{\overline{x}\}) = [\ell_{\mathrm{E}[X]}(\overline{x}), u_{\mathrm{E}[X]}(\overline{x})]$ are added based on the monotonic distribution together with the relative difference $\Delta(\overline{x})$ (7.85).

Example 7.7 *The value of the weighted volume is* $V_w({}^{HPR}C^{(0.9)}_{\mathrm{E}[Y]}) = 0.730506$, *which is considerably larger than* $V_w({}^{*}C^{(0.9)}_{\mathrm{E}[X]}) = 0.539789$. *Nevertheless, we present analogue to table 7.3, in table 7.4 some measurements* ${}^{HPR}C^{(\beta)}_{\mathrm{E}[Y]}(\{\overline{x}\})$ *for some values of* $\overline{x}$. *The relative difference* $\Delta(\overline{x})$ *is again calculated, now by*

$$\Delta(\overline{x}) := \frac{|{}^{HPR}C^{(0.9)}_{\mathrm{E}[Y]}(\{\overline{x}\})| - |{}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{\overline{x}\})|}{|{}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{\overline{x}\})|}\,. \tag{7.85}$$

Measurement Procedures for $\mathrm{E}[X]$, $X|\{\mu_1\} \sim Mon(\mu_1)$ approximated by $Y \sim EXP(1/\mu_1)$, based on $2\sum Y_i/\mu_1$: In example 4.59 we already discussed an alternative set-estimation procedure for μ_1, where $Y|\{\mu_1\} \sim EXP(1/\mu_1)$, based on highest probability acceptance regions. There, the acceptance regions,

i.e., the predictions, are not derived on the basis of the joint density of $\boldsymbol{Y}|\{\mu_1\}$, but of the density of $2(\sum_{i=1}^{n} Y_i|\{\mu_1\})/\mu_1 =: Q|\{\mu_1\}$. A slight advantage of considering $Q|\{\mu_1\}$ is, that it is distributed according to the chi-square distribution with $2n$ degrees of freedom, i.e., $Q|\{\mu_1\} \sim \chi^2_{2n}$. Since it is in fact independent of μ_1, it is a pivotal variable. Therefore, we need to find exactly one prediction region ${}^{P}A_{Q}^{(\beta)}$:

$$ {}^{P}A_{Q}^{(\beta)} = \{q \,|\, \chi^2_{2n}(\alpha_1) \leq q \leq \chi^2_{2n}(1-\alpha_2)\} \quad \text{with} \quad 1-(\alpha_1+\alpha_2) = \beta \tag{7.86} $$

By retransformation, we get prediction regions of $\overline{Y}$ for each value μ_1 of $\mathrm{E}[Y]$:

$$ {}^{P}A_{\overline{Y}}^{(\beta)}(\{\mu_1\}) = \left\{ \overline{y} \;\middle|\; \frac{\chi^2_{2n}(\alpha_1)\mu_1}{2n} \leq \overline{y} \leq \frac{\chi^2_{2n}(1-\alpha_2)\mu_1}{2n} \right\} \tag{7.87} $$

By inversion we get measurements of μ_1:

$$ {}^{P}C_{\mathrm{E}[Y]}^{(\beta)}(\{\overline{y}\}) = \left\{ \mu_1 \;\middle|\; \frac{2n\overline{y}}{\chi^2_{2n}(1-\alpha_2)} \leq \mu_1 \leq \frac{2n\overline{y}}{\chi^2_{2n}(\alpha_1)} \right\} \cap \mathcal{D} \tag{7.88} $$

If the measurements (7.88) equal the empty set, then we set ${}^{P}C_{\mathrm{E}[Y]}^{(\beta)}(\{\overline{y}\})$ to be $\{m_\ell\}$ and $\{m_u\}$, respectively.

The way highest probability regions are constructed, leads to minimising the length of the prediction regions[4], that is, minimising $\chi^2_{2n}(1-\alpha_2) - \chi^2_{2n}(\alpha_1)$ with respect to α_1 and α_2 under the condition $1-(\alpha_1+\alpha_2) = \beta$. But in classical estimation theory, an important aim is to minimise the volume (length) of the confidence regions (intervals), i.e., the measurements. Here, it would be actually neccessary to minimise ${}^{P}C_{\mathrm{E}[Y]}^{(\beta)}(\{\overline{y}\})$, but since we understand this procedure as an approximation, we minimise $\frac{1}{\chi^2_{2n}(\alpha_1)} - \frac{1}{\chi^2_{2n}(1-\alpha_2)}$ at first and then make the intersection with $\mathcal{D}$. To distinguish, we denote the in this way minimised measurements ${}^{P}\widetilde{C}_{\mathrm{E}[Y]}^{(\beta)}(\{\overline{y}\})$ and, thus, by inversion, the respective prediction regions ${}^{P}\widetilde{A}_{\overline{Y}}^{(\beta)}(\{\mu_1\})$. The unchanged notation ${}^{P}C_{\mathrm{E}[Y]}^{(\beta)}(\{\overline{y}\})$ should stand for the measurements based on minimised prediction regions.

Example 7.8 *Let* $\mathcal{D} = \{\mu_1 \,|\, 0.05 \leq \mu_1 \leq 0.25\}$, $n = 8$ *and* $\beta = 0.9$. *Minimising the length of* ${}^{P}A_{Q}^{(0.9)}$ *yields* $\alpha_1 = 0.025502$ *and* $\alpha_2 = 0.074498$ *and, therefore,*

[4]The density of the chi-square distribution is unimodal and, thus, the resulting highest probability region is in fact an interval.

$\chi^2_{16}(\alpha_1) = 6.934767$ *and* $\chi^2_{16}(1-\alpha_2) = 24.742360$. *The prediction regions of* $\overline{Y}$ *for each value* μ_1 *follow to be*

$$ {}^{P}A^{(0.9)}_{\overline{Y}}(\{\mu_1\}) = \left\{ \overline{y} \;\middle|\; \frac{6.934767\mu_1}{16} \leq \overline{y} \leq \frac{24.742360\mu_1}{16} \right\} . \tag{7.89} $$

Since

$$ {}^{P}A^{(0.9)}_{\overline{Y}}(\{m_\ell\}) = \{\overline{y} \,|\, 0.021671 \leq \overline{y} \leq 0.077320\} \tag{7.90} $$

and

$$ {}^{P}A^{(0.9)}_{\overline{Y}}(\{m_u\}) = \{\overline{y} \,|\, 0.108356 \leq \overline{y} \leq 0.386599\} \tag{7.91} $$

we note, that the completeness requirement is not met. Therefore, we replace ${}^{P}A^{(0.9)}_{\overline{Y}}(\{m_\ell\})$ *with* $\{\overline{y} \,|\, 0 \leq \overline{y} \leq 0.077320\}$ *and* ${}^{P}A^{(0.9)}_{\overline{Y}}(\{m_u\})$ *with* $\{\overline{y} \,|\, 0.108356 \leq \overline{y} \leq 1\}$, *respectively. Finally, the measurements result to be*

$$ {}^{P}C^{(0.9)}_{\mathrm{E}[Y]}(\{\overline{y}\}) = \left\{ \mu_1 \;\middle|\; \frac{16\overline{y}}{24.742360} \leq \mu_1 \leq \frac{16\overline{y}}{6.934767} \right\} \cap \mathcal{D} . \tag{7.92} $$

The weighted volume of the measurement procedure ${}^{P}C^{(0.9)}_{\mathrm{E}[Y]}$ *has a value of*

$$ V_w({}^{P}C^{(0.9)}_{\mathrm{E}[Y]}) = 0.608841 . \tag{7.93} $$

Now, minimising

$$ \frac{1}{\chi^2_{2n}(\alpha_1)} - \frac{1}{\chi^2_{2n}(1-\alpha_2)} \tag{7.94} $$

yields $\alpha_1 = 0.088435$ *and* $\alpha_2 = 0.011565$ *and measurements*

$$ {}^{P}\widetilde{C}^{(\beta)}_{\mathrm{E}[Y]}(\{\overline{y}\}) = \left\{ \mu_1 \;\middle|\; \frac{16\overline{y}}{31.512560} \leq \mu_1 \leq \frac{16\overline{y}}{9.044611} \right\} \cap \mathcal{D} . \tag{7.95} $$

The prediction regions then are

$$ {}^{P}\widetilde{A}^{(0.9)}_{\overline{Y}}(\{\mu_1\}) = \left\{ \overline{y} \;\middle|\; \frac{9.044611\mu_1}{16} \leq \overline{y} \leq \frac{31.512560\mu_1}{16} \right\} , \tag{7.96} $$

which again do not meet the completeness requirement and we need to replace ${}^{P}\widetilde{A}^{(0.9)}_{\overline{Y}}(\{m_\ell\})$ *with* $\{\overline{y} \,|\, 0 \leq \overline{y} \leq 0.098477\}$ *and* ${}^{P}\widetilde{A}^{(0.9)}_{\overline{Y}}(\{m_u\})$ *with* $\{\overline{y} \,|\, 0.141322 \leq \overline{y} \leq 1\}$. *The value of the weighted volume now is*

$$ V_w({}^{P}\widetilde{C}^{(0.9)}_{\mathrm{E}[Y]}) = 0.606730 \tag{7.97} $$

– *smaller than that of the procedure based on minimised prediction regions.*

Some small calculations show that both meet the requested reliability of 0.9*, but both procedures have a value of the weighted volume larger (*+12.8 *% and* +12.4 *%, respectively) than that of the Neyman measurement procedure* ${}^{*}C^{(0.9)}_{\mathrm{E}[X]}$ *based on the true monotonic distribution, which is* $V_w({}^{*}C^{(0.9)}_{\mathrm{E}[X]}) = 0.539789$*, i.e., they will yield unnecessary large measurements.*

7.2.2.3. LR Measurement Procedures for $\mathrm{E}[Y]$, $X|\{\mu_1\} \sim Mon(\mu_1)$ approximated by $Y \sim Exp(1/\mu_1)$

In section 4.4.3.2 we have introduced one method of inverting acceptance regions, i.e., prediction regions ${}^{LR}A^{(\beta)}_{\boldsymbol{X}}(\vartheta_0)$ which are derived according to the likelihood ratio test statistic

$$ {}^{LR}A^{(\beta)}_{\boldsymbol{X}}(\vartheta_0) = \left\{ \boldsymbol{x} \;\middle|\; \frac{L_{\boldsymbol{x}}(\vartheta_0)}{\sup_{\vartheta\in\Theta} L_{\boldsymbol{x}}(\vartheta)} \geq r \right\} \tag{7.98} $$

with

$$ \mathrm{P}_{\boldsymbol{X}|\{\vartheta_0\}}\left(\boldsymbol{X}|\{\vartheta_0\} \in {}^{LR}A^{(\beta)}_{\boldsymbol{X}}(\vartheta_0) \right) \geq \beta \tag{7.99} $$

where $L_{\boldsymbol{x}}(\vartheta)$ is the likelihood function and Θ the parameter space, i.e., the ignorance space $\mathcal{D} = \Theta$.
With $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ we get for the likelihood ratio test statistic

$$ \frac{L_{\boldsymbol{y}}(\mu_{1,0})}{\sup_{\mu_1\in\mathcal{D}} L_{\boldsymbol{y}}(\mu_1)} = \frac{\left(\frac{1}{\mu_{1,0}}\right)^n e^{-\sum_{i=1}^n y_i/\mu_{1,0}}}{\sup_{\mu_1\in\mathcal{D}} \left(\frac{1}{\mu_1}\right)^n e^{-\sum_{i=1}^n y_i/\mu_1}} . \tag{7.100} $$

The supremum of the likelihood function $L_{\boldsymbol{y}}(\mu_1)$ is achieved if the value μ_1 equals the maximum likelihood estimation $\widehat{\mu}_1(\boldsymbol{y})$. Assuming a bounded ignorance space $\mathcal{D} = \{\mu_1 \,|\, m_\ell \leq \mu_1 \leq m_u\}$ we get

$$ \widehat{\mu}_1(\boldsymbol{y}) = \begin{cases} m_\ell & , \overline{y} < m_\ell \\ \overline{y} & , m_\ell \leq \overline{y} \leq m_u \\ m_u & , \overline{y} > m_u \end{cases} . \tag{7.101} $$

and, thus,

$$\frac{L_{\boldsymbol{y}}(\mu_{1,0})}{\sup_{\mu_1 \in \mathcal{D}} L_{\boldsymbol{y}}(\mu_1)} = \frac{L_{\boldsymbol{y}}(\mu_{1,0})}{L_{\boldsymbol{y}}(\widehat{\mu}_1(\boldsymbol{y}))}$$

$$= \begin{cases} \left(\frac{m_\ell}{\mu_{1,0}}\right)^n e^{-\sum_{i=1}^n y_i((1/\mu_{1,0})-(1/m_\ell))} & , \overline{y} < m_\ell \\ \left(\frac{\frac{1}{n}\sum_{i=1}^n y_i}{\mu_{1,0}}\right)^n e^n e^{-\sum_{i=1}^n y_i/\mu_{1,0}} & , m_\ell \leq \overline{y} \leq m_u \\ \left(\frac{m_u}{\mu_{1,0}}\right)^n e^{-\sum_{i=1}^n y_i((1/\mu_{1,0})-(1/m_u))} & , \overline{y} > m_u \end{cases} \quad (7.102)$$

As in the previously discussed approximated measurement procedures, for fixed value $\mu_{1,0}$ the value of the likelihood ratio test statistic depends on $\boldsymbol{y}$ only via $\sum_{i=1}^n y_i$ or $\overline{y}$, respectively. That is, we actually determine prediction regions for $\sum_{i=1}^n Y_i$ and $\overline{Y}$:

$$^{LR}A_{\overline{Y}}^{(\beta)}(\{\mu_{1,0}\}) = \{\overline{y} \,|\, \ell_{\overline{Y}}(\mu_{1,0}) \leq \overline{y} \leq u_{\overline{Y}}(\mu_{1,0})\} \quad (7.103)$$

with

$$\mathrm{P}_{\overline{Y}|\{\mu_{1,0}\}}\left(\overline{Y}|\{\mu_{1,0}\} \in {}^{LR}A_{\overline{Y}}^{(\beta)}(\mu_{1,0})\right) =$$

$$= \mathrm{P}_{\overline{Y}|\{\mu_{1,0}\}}\left(\ell_{\overline{Y}}(\mu_{1,0}) \leq \overline{Y}|\{\mu_{1,0}\} \leq u_{\overline{Y}}(\mu_{1,0})\right) = \beta \quad (7.104)$$

Since $n \cdot \overline{Y} = \sum_{i=1}^n Y_i \sim GAM(1/\mu_1, n)$, the determination of the prediction regions consists of the determination of quantiles of the gamma distribution.

Two values $\mu_{1,0}$ in (7.102) should be examined more closely, that are $\mu_{1,0} = m_\ell$ and $\mu_{1,0} = m_u$. If $\mu_{1,0} = m_\ell$, the likelihood ratio test statistic equals 1 for all $\overline{y} < m_\ell$, and is decreasing for $\overline{y} \geq m_\ell$. If $\mu_{1,0} = m_u$, the likelihood ratio test statistic is increasing for $\overline{y} \leq m_u$ and equals 1 for all $\overline{y} > m_u$.

Example 7.9 *Let $\mathcal{D} = \{\mu_1 \,|\, 0.05 \leq \mu_1 \leq 0.25\}$, $n = 8$ and $\beta = 0.9$. Figure 7.10 displays the course of the likelihood ratio test statistic for different values $\mu_{1,0}$. The resulting β-uncertainty space $^{LR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ is displayed in figure 7.11. Figure 7.12 displays the true probability of the prediction regions $^{LR}A_{\boldsymbol{Y}}^{(0.9)}(\{\mu_1\})$ and $^{LR}A_{\overline{Y}}^{(0.9)}(\{\mu_1\})$, respectively, under the monotonic distribution $Mon(\mu_1)$. Again, the achieved reliability is at least the requested one of 0.9, which makes the resulting measurement procedure reasonable. The value of the weighted volume now is*

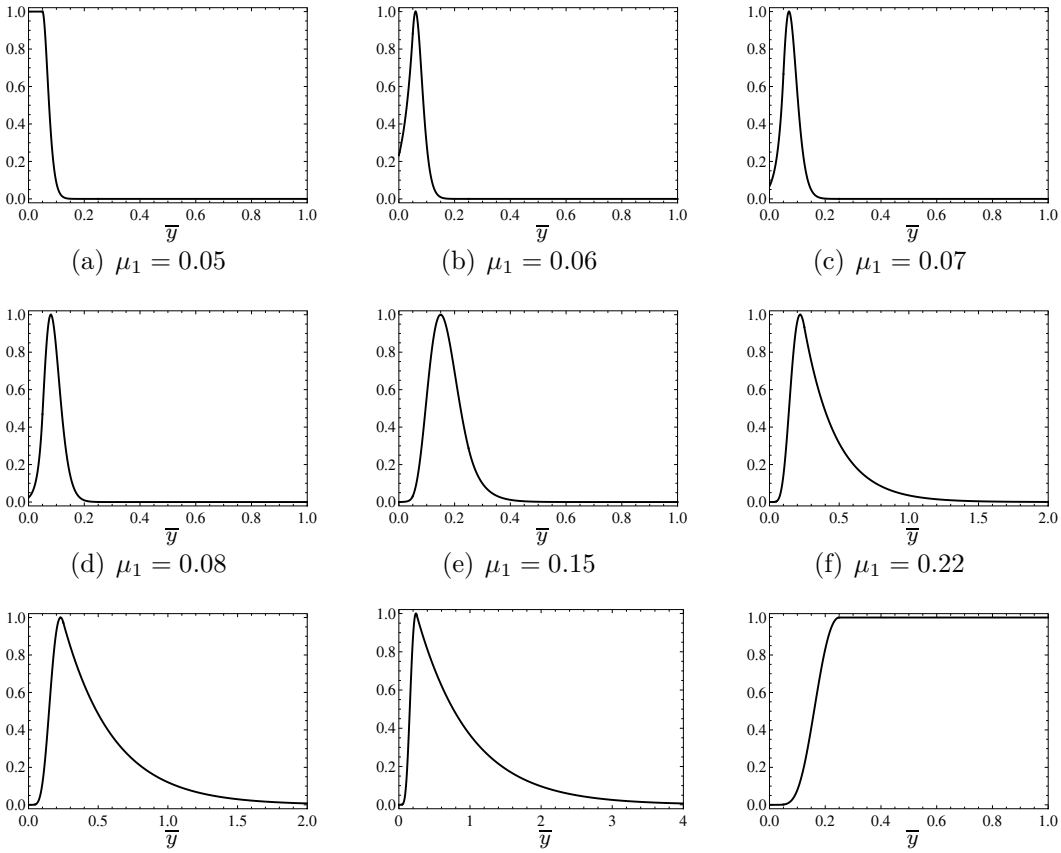


Figure 7.10.: The course of the likelihood ratio test statistic for sample size $n = 8$ and different values $\mu_{1,0}$.

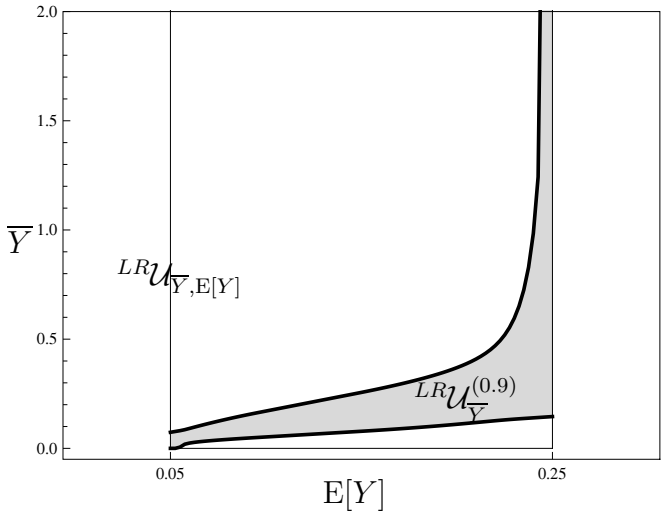


Figure 7.11.: The β-uncertainty space $^{LR}\mathcal{U}^{(0.9)}_{\overline{Y}}$ for $\overline{Y}$.

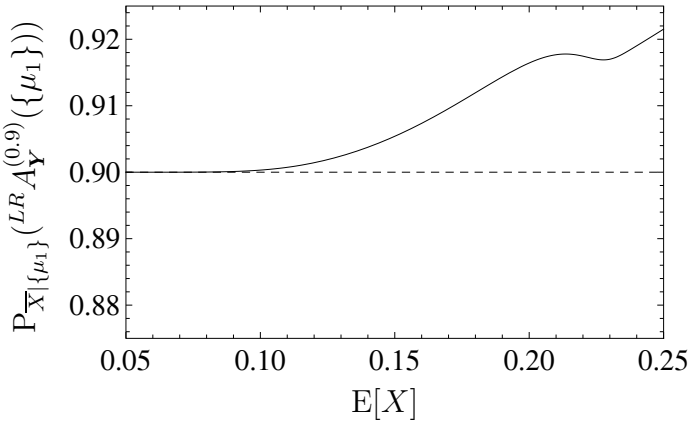


Figure 7.12.: The true probability of the prediction sets $^{LR}A^{(0.9)}_{\overline{Y}}(\{\mu_1\})$ based on the exponential distribution $EXP(1/\mu_1)$ as approximation for the monotonic distribution $Mon(\mu_1)$.

	${}^{LR}C^{(0.9)}_{\mathrm{E}[Y]}(\{\overline{y}\})$		${}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{\overline{y}\})$		
$\overline{y}$	${}^{LR}\ell_{\mathrm{E}[Y]}(\overline{y})$	${}^{LR}u_{\mathrm{E}[Y]}(\overline{y})$	$\ell_{\mathrm{E}[X]}(\overline{y})$	$u_{\mathrm{E}[X]}(\overline{y})$	$\Delta\cdot$ 100 %
0.025	0.05	0.060552	0.05	0.088810	-72.811
0.05	0.05	0.095893	0.05	0.114415	-28.754
0.075	0.050973	0.143838	0.050973	0.146223	-2.504
0.1	0.063754	0.185909	0.067964	0.175065	+14.056
0.125	0.075243	0.218120	0.084874	0.208572	+15.505
0.15	0.088106	0.25	0.099746	0.244846	+11.574
0.175	0.102467	0.25	0.113307	0.25	+7.930
0.2	0.117105	0.25	0.126246	0.25	+7.386
0.4	0.208026	0.25	0.188651	0.25	-31.582
0.6	0.230238	0.25	0.207734	0.25	-53.244
0.8	0.236909	0.25	0.217264	0.25	-60.010
0.9	0.238844	0.25	0.220474	0.25	-62.216

Table 7.5.: Measurements ${}^{LR}C^{(0.9)}_{\mathrm{E}[Y]}(\{\overline{y}\}) = [{}^{LR}\ell_{\mathrm{E}[Y]}(\overline{y}), {}^{LR}u_{\mathrm{E}[Y]}(\overline{y})]$ for some values $\overline{y}$. For comparison the measurements ${}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{\overline{y}\}) = [\ell_{\mathrm{E}[X]}(\overline{y}), u_{\mathrm{E}[X]}(\overline{y})]$ are added based on the monotonic distribution together with the relative difference $\Delta(\overline{y})$ (7.85).

$V_w({}^{LR}C^{(0.9)}_{\mathrm{E}[Y]}) = 0.566800$, *which is only slightly larger than* $V_w({}^{*}C^{(0.9)}_{\mathrm{E}[X]}) = 0.539789$. *For some values* $\overline{y}$ *the measurements* ${}^{LR}C^{(0.9)}_{\mathrm{E}[Y]}(\{\boldsymbol{y}\})$ *are calculated as are the relative differences* $\Delta(\overline{y})$ *(with the same definition* (7.85)*) and presented in table 7.5. We note, that the majority of the here presented measurements are smaller than the ones based on the Neyman measurement procedure for the true monotonic distribution. But since the weighted volume is larger, we have to realise that any comparison of single measurement lengths may be misleading.*

7.2.2.4. Normal Approximation

In the sense of section 4.4.1, we derive an approximative measurement procedure based on the convergence in distribution of

$$\sqrt{nI(\vartheta)}(\widehat{\vartheta}(\boldsymbol{X}) - \vartheta) \tag{7.105}$$

to the standard normal distribution, where $I(\vartheta)$ is the Fisher information defined by (4.76) and $\widehat{\vartheta}(\boldsymbol{X})$ is the MLE of ϑ. For n large the probability

$$\mathrm{P}\left(-z(\frac{1+\beta}{2}) \leq \sqrt{nI(\vartheta)}(\widehat{\vartheta}(\boldsymbol{X}) - \vartheta) \leq z(\frac{1+\beta}{2})\right), \tag{7.106}$$

where $z(\alpha)$ is the quantile of the standard normal distribution, is approximately β. Substituting $I(\vartheta)$ with its estimator $I(\widehat{\vartheta})$ and solving for ϑ yields a confidence interval, i.e., a measurement procedure of ϑ

$$C(\boldsymbol{X}|\{\vartheta\}) = \left[\widehat{\vartheta}(\boldsymbol{X}) - \frac{z(\frac{1+\beta}{2})}{\sqrt{nI(\widehat{\vartheta}(\boldsymbol{X}))}}, \widehat{\vartheta}(\boldsymbol{X}) + \frac{z(\frac{1+\beta}{2})}{\sqrt{nI(\widehat{\vartheta}(\boldsymbol{X}))}}\right] \tag{7.107}$$

with approximative confidence level, i.e., reliability of β.

For $X \sim Mon(\mu_1)$ we already have derived the MLE for $\mathrm{E}[X]$ in section 5.1.2.1. If we assume an ignorance space $\mathcal{D} = \{\mu_1 \,|\, 0 \leq m_\ell < \mu_1 < m_u \leq 0.5\}$, i.e., we have complete knowledge about the range of variability to be $\mathcal{X} = \{x \,|\, 0 \leq x \leq 1\}$ and, thus, that the density function is monotone decreasing on $\mathcal{X}$. Then we have for the MLE

$$\widehat{\mu}_1(\boldsymbol{X}) = \begin{cases} m_\ell & \text{for } \overline{x} \leq m_\ell, \\ \overline{x} & \text{for } m_\ell < \overline{x} < m_u, \\ m_u & \text{for } \overline{x} \geq m_u. \end{cases} \tag{7.108}$$

Even though, this MLE does not really match to the ignorance space, we have to choose $\mathcal{D}$ as an open set to meet the regularity conditions ([13], p. 516), where one condition is, that the true value of $\mathrm{E}[X]$ has to be an inner point of $\mathcal{D}$.

Since the density functions $f_{X|\{\mu_1\}}$ form an exponential family in μ_1, we have the following equation for the Fisher information

$$I(\mu_1) = \mathrm{E}\left[\left(\frac{\partial}{\partial\mu_1} \ln f_{X|\{\mu_1\}}(X)\right)^2\right] = -\mathrm{E}\left[\frac{\partial^2}{\partial\mu_1^2} \ln f_{X|\{\mu_1\}}(X)\right] \tag{7.109}$$

We have

$$\frac{\partial^2}{\partial \mu_1^2} \ln f_{X|\{\mu_1\}}(x) = \frac{\partial^2 \lambda(\mu_1)}{\partial \mu_1^2}(x - \mu_1) - \frac{\partial \lambda(\mu_1)}{\partial \mu_1} \tag{7.110}$$

which yields

$$-\mathrm{E}[\frac{\partial^2}{\partial \mu_1^2} \ln f_{X|\{\mu_1\}}(X)] = \frac{\partial \lambda(\mu_1)}{\partial \mu_1} = \frac{1}{\mu_2 - \mu_1^2}\,. \tag{7.111}$$

Estimating $\mathrm{E}[X]$ with $\widehat{\mu}_1(\boldsymbol{X})$ and $\mathrm{E}[X^2]$ with $\mu_2(\lambda(\widehat{\mu}_1(\boldsymbol{X})))$, we immediatly get an approximative confidence interval estimator (see section 4.4.1) for $\mathrm{E}[X]$:

$$C(\boldsymbol{X}|\{\mu_1\}) = \left[\widehat{\mu}_1(\boldsymbol{X}) - z(\frac{1+\beta}{2})\frac{\sqrt{\mu_2(\lambda(\widehat{\mu}_1(\boldsymbol{X}))) - \widehat{\mu}_1(\boldsymbol{X})^2}}{\sqrt{n}}\,,\right.$$
$$\left.\widehat{\mu}_1(\boldsymbol{X}) + z(\frac{1+\beta}{2})\frac{\sqrt{\mu_2(\lambda(\widehat{\mu}_1(\boldsymbol{X}))) - \widehat{\mu}_1(\boldsymbol{X})^2}}{\sqrt{n}}\right]$$

Obviously, the resulting confidence interval estimates $C(\boldsymbol{x})$ may not be subset of $\mathcal{D}$, but the left bound may be smaller than m_ℓ and/or the right bound may be larger than m_u. That is, to get a reasonable measurement, we have to take the intersection $C(\boldsymbol{x}) \cap \mathcal{D}$, which finally results into an approximative measurement procedure ${}^{N}C_{\mathrm{E}[X]}$ for $\mathrm{E}[X]$:

$${}^{N}C_{\mathrm{E}[X]}(\boldsymbol{X}) = \left[\widehat{\mu}_1(\boldsymbol{X}) - z\left(\frac{1+\beta}{2}\right)\frac{\sqrt{\mu_2(\lambda(\widehat{\mu}_1(\boldsymbol{X}))) - \widehat{\mu}_1(\boldsymbol{X})^2}}{\sqrt{n}}\,,\right. \tag{7.112}$$
$$\left.\widehat{\mu}_1(\boldsymbol{X}) + z\left(\frac{1+\beta}{2}\right)\frac{\sqrt{\mu_2(\lambda(\widehat{\mu}_1(\boldsymbol{X}))) - \widehat{\mu}_1(\boldsymbol{X})^2}}{\sqrt{n}}\right] \cap (m_\ell, m_u)$$
$$\tag{7.113}$$

Since $\widehat{\mu}_1(\boldsymbol{x})$ does in fact depend on $\overline{x}$, the measurements may be equivalently noted as ${}^{N}C_{\mathrm{E}[X]}(\overline{x})$. The corresponding uncertainty space $\mathcal{U}$ results as the union over all measurements ${}^{N}C_{\mathrm{E}[X]}(\overline{x})$:

$$\mathcal{U} = \bigcup_{\overline{x} \in \mathcal{X}} {}^{N}C_{\mathrm{E}[X]}(\overline{x}) \times \{\overline{x}\} \tag{7.114}$$

Intentionally, we have omitted the prefix β for the measurement procedure and the uncertainty space as it is quite unclear whether the demanded reliability is really achieved. At least, the completeness requirement does not pose a problem here, since a measurement exists for each $\overline{x} \in \mathcal{X} = [0, 1]$. We end this section again with an example.

	${}^{N}C_{\mathrm{E}[X]}(\{\overline{x}\})$		${}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{\overline{x}\})$		
$\overline{x}$	${}^{N}\ell_{\mathrm{E}[X]}(\overline{x})$	${}^{N}u_{\mathrm{E}[X]}(\overline{x})$	$\ell_{\mathrm{E}[X]}(\overline{x})$	$u_{\mathrm{E}[X]}(\overline{x})$	$\Delta\cdot$ 100 %
0.025	0.05	0.079077	0.05	0.088810	-25.079
0.05	0.05	0.079077	0.05	0.114415	-54.860
0.075	0.05	0.118611	0.050973	0.146223	-27.969
0.1	0.05	0.158048	0.067964	0.175065	+0.884
0.125	0.052908	0.197092	0.084874	0.208572	+16.561
0.15	0.064623	0.235377	0.099746	0.244846	+17.680
0.175	0.077341	0.25	0.113307	0.25	+26.312
0.2	0.091162	0.25	0.126246	0.25	+28.350
0.4	0.122113	0.25	0.188651	0.25	+108.458
0.6	0.122113	0.25	0.207734	0.25	+202.577
0.8	0.122113	0.25	0.217264	0.25	+290.662
0.9	0.122113	0.25	0.220474	0.25	+333.134

Table 7.6.: Measurements ${}^{N}C_{\mathrm{E}[X]}(\{\overline{x}\}) = [{}^{N}\ell_{\mathrm{E}[X]}(\overline{x}), {}^{N}u_{\mathrm{E}[X]}(\overline{x})]$ for some values $\overline{x}$. For comparison the measurements ${}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{\overline{x}\}) = [\ell_{\mathrm{E}[X]}(\overline{x}), u_{\mathrm{E}[X]}(\overline{x})]$ are added based on the monotonic distribution together with the relative difference $\Delta(\overline{x})$ (7.85).

Example 7.10 *Since $\mathcal{D}$ needs to be an open set, let $\mathcal{D} = \{\mu_1 \,|\, 0.05 < \mu_1 < 0.25\}$, and let $n = 8$ and $\beta = 0.9$. We start the numerical analysis by giving the table 7.6 of measurement results ${}^{N}C_{\mathrm{E}[X]}(\overline{x})$ for different values of $\overline{x}$ – compared to those of the Neyman measurement procedure ${}^{*}C^{(0.9)}_{\mathrm{E}[X]}$. Figure 7.13 displays the uncertainty space $\mathcal{U}_{\overline{X}}$ resulting by inverting the approximative measurements ${}^{N}C_{\mathrm{E}[X]}(\{\overline{x}\})$, i.e., we may call $\mathcal{U}_{\overline{X}}$ approximative uncertainty space. The essential question is whether the demanded reliability is achieved or not. Figure 7.14 displays the true probability of the prediction regions $A^{(0.9)}_{\overline{X}}(\{\mu_1\})$, i.e., the vertical cross-section of $\mathcal{U}_{\overline{X}}$ at μ_1, under the monotonic distribution $Mon(\mu_1)$. For a majority of values μ_1 of $\mathrm{E}[X]$ the probability falls below the demanded reliability of 0.9, which finally disqualifies this procedure. Nevertheless, for the sake of completeness, the weighted volume V_w has an value of $V_w({}^{N}C_{\mathrm{E}[X]}) = 0.606581$, also far larger than*

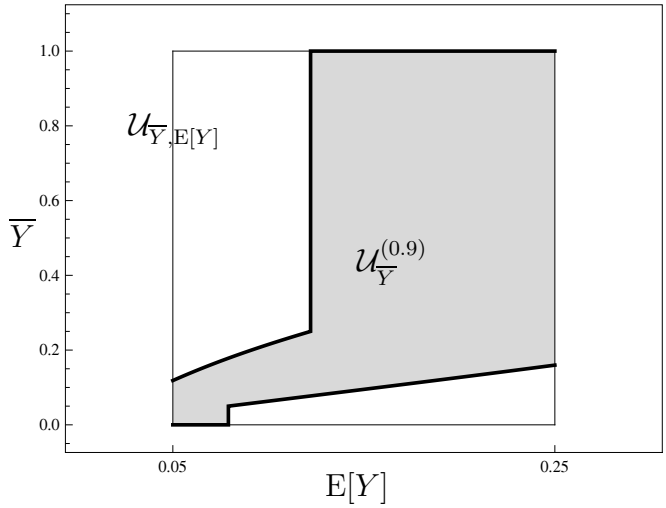


Figure 7.13.: The approximative uncertainty space $\mathcal{U}_{\overline{X}}$ for $\overline{X}$.

$V_w({}^*C_{\mathrm{E}[X]}^{(0.9)}) = 0.539789.$

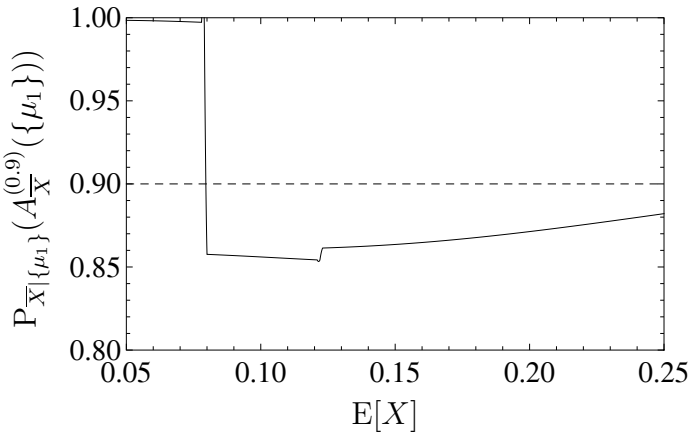


Figure 7.14.: The true probability of the prediction regions $A_{\overline{X}}^{(0.9)}(\{\mu_1\})$ based on the approximative measurements ${}^N C_{\mathrm{E}[X]}(\{\overline{x}\})$.

7.2.3. An Alternative Illustration of $\mathcal{U}_{\overline{X}}^{(\beta)}$

We have already realised, that the comparison of single measurements and of their lengths, respectively, may be misleading – see for instance table 7.4 for measurements based on highest probability regions. Loosley speaking, we may state that values $\overline{x} \geq 0.4$ yield singletons as measurements, which are unbeatable short, while for other values $\overline{x} \leq 0.15$, the resulting measurements are larger. And for values $\overline{x} \leq 0.05$ the measurements even do not reduce the ignorance space $\mathcal{D}$. Nevertheless, the weighted volume for the measurement procedure based on the highest probability regions is $V_w\left({}^{HPR}C_{\mathrm{E}[Y]}^{(0.9)}\right) = 0.730506$ and, thus, by far larger than $V_w\left({}^{*}C_{\mathrm{E}[X]}^{(0.9)}\right) = 0.539789$. The situation when comparing measurements based on the likelihood ratio test statistic is even more misleading, since for the majority of values $\overline{x}$ the measurements ${}^{LR}C_{\mathrm{E}[Y]}^{(0.9)}(\{\overline{x}\})$ are shorter than those from the optimal procedure, and only for some values $\overline{x}$ the measurements are a bit larger. Figure 7.15 shows the respective β-uncertainty spaces ${}^{LR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ and $\mathcal{U}_{\overline{X}}^{(0.9)}$. The procedure based on the likelihood ratio test statistic ${}^{LR}C_{\mathrm{E}[Y]}^{(0.9)}$ appears to be superior to ${}^{*}C_{\mathrm{E}[X]}^{(0.9)}$ – but only in a geometric sense.

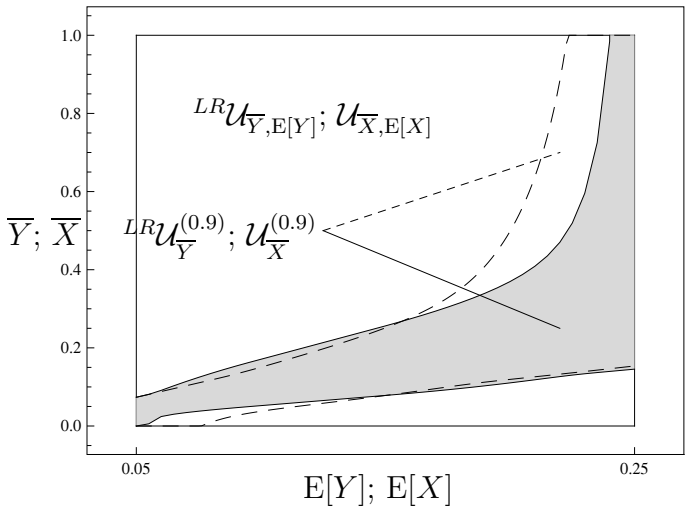


Figure 7.15.: The β-uncertainty space ${}^{LR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ and $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ for $n = 8$ based on the likelihood ration test statistic and the optimal one $\mathcal{U}_{\overline{X}}^{(0.9)}$ (dashed shape).

Of course, the key to a correct interpretation is to take the marginal distribution

of $\overline{X}$ into account, i.e., the marginal density (see (6.38))

$$q_{\mathcal{X}(\mathcal{D})}(\overline{x}) = \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} f_{\overline{X}|\{\mu_1\}}(\overline{x}) \, \mathrm{d}\mu_1 \tag{7.115}$$

and the marginal distribution function (see (6.39)), respectively,

$$Q_{\mathcal{X}(\mathcal{D})}(\overline{x}) := Q_{\mathcal{X}(\mathcal{D})}\left(\{\overline{y} \,|\, 0 \leq \overline{y} \leq \overline{x}\}\right) \tag{7.116}$$

$$= \int_0^{\overline{x}} q_{\mathcal{X}(\mathcal{D})}(\overline{y}) \, \mathrm{d}\overline{y} \,. \tag{7.117}$$

Exactly this is done for calculating the weighted volume V_w of a measurement procedure $C_D^{(\beta)}$ in (6.50):

$$V_w(C_D^{(\beta)}) = \int_{\mathcal{X}(\mathcal{D})} \left(\int_{C_D^{(\beta)}(\{x\})} \lambda_{\mathcal{D}}(d) \, \mathrm{d}d \right) q_{\mathcal{X}(\mathcal{D})}(x) \, \mathrm{d}x \tag{7.118}$$

For a graphical illustration, this is done by a rescaling of the $\overline{x}$-axis with the marginal distribution function of $\overline{X}$, which results into a dilation of those parts of the $\overline{x}$-axis with high probability while less probable parts are shrinked. Applying this transformation to figure 7.15 results into figure 7.16. The former seeming superiority of ${}^{LR}C_{\mathrm{E}[Y]}^{(0.9)}$ to ${}^{*}C_{\mathrm{E}[X]}^{(0.9)}$ is now corrected and reflects the right proportion of the weighted volume's values $V_w\left({}^{LR}C_{\mathrm{E}[Y]}^{(0.9)}\right) = 0.566800$ and $V_w\left({}^{*}C_{\mathrm{E}[X]}^{(0.9)}\right) = 0.539789$. In contrast to the unscaled plots of β-uncertainty spaces, it is not possible to compare the probability scaled plots of uncertainty spaces for different values of n, since the marginal probability distribution changes with n.

7.2.4. Numerical Comparisons with respect to n and $\mathcal{D}$

In this section we present some more examples,to illustrate the dependency of the different derived procedures with respect to the sample size n and the ignorance space $\mathcal{D}$. Common to all approximative procedures is that $X|\{\mu_1\} \sim Mon(\mu_1)$ is approximated by $Y|\{\mu_1\} \sim EXP(1/\mu_1)$. The procedures which should be compared to the Neyman measurement procedure for $\mathrm{E}[X]$ are

- the Neyman measurement procedure for $\mathrm{E}[Y]$,

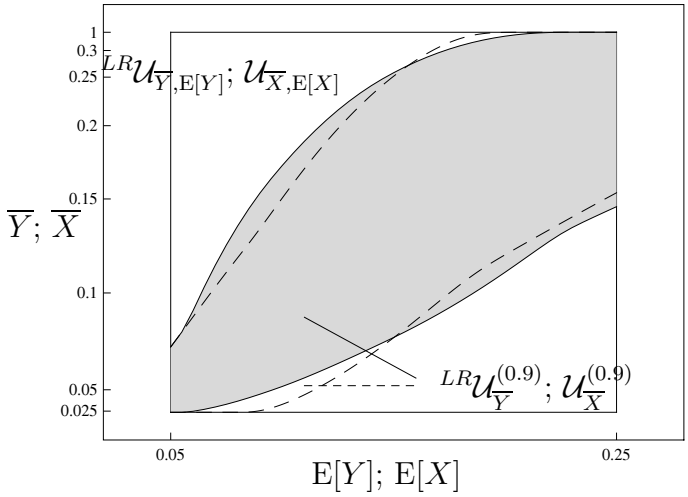


Figure 7.16.: The β-uncertainty space $^{LR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ and $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ for $n = 8$ based on the likelihood ration test statistic and the optimal one $\mathcal{U}_{\overline{X}}^{(0.9)}$ (dashed shape). The $\overline{x}$-axis is rescaled by the marginal probability distribution of $\overline{X}$.

- the LR measurement procedure for $\mathrm{E}[Y]$ and
- the HPR measurement procedure for $\mathrm{E}[Y]$ based on $2\sum Y_i/\mu_1$.

The examples are organised in the way, that we assume a certain ignorance space $\mathcal{D} \subset \{\mu_1 \,|\, 0 < \mu_1 < 0.5\}$, display the different β-ignorance spaces for different sample sizes $n \in \{2, 4, 8, 12, 16\}$ and give graphical comparisons of the approximative ones with the optimal. Throughout the examples we choose the reliability $\beta = 0.9$.

The first focus lies on the question whether the reliability is achieved. As we have seen in the previous sections, single measurements and the relative difference of their lengths give a somehow misleading impression and, thus, are omitted here. Instead, the weighted volume V_w will be calculated and illustrated by the probability scaled β-uncertainty spaces according to section 7.2.3. Of course, the Neyman measurement procedures will have minimum weighted volume by construction, but for the others it may serve as the second criterion for choosing between competing ones.

Since we have seen already numerous graphs showing (one or more) β-uncertainty spaces, we will ommit extensive labellings within the graphics in the following and give necessary descriptions in the figures captions.

7.2.4.1. $\mathcal{D} = \{\mu_1 \,|\, 0.01 \leq \mu_1 \leq 0.05\}$

In relation to the range of variability $\mathcal{X} = \{x \,|\, 0 \leq x \leq 1\}$ the possible values μ_1 of $\mathrm{E}[X]$ are very near to the lower bound. We will see, that the boundedness of the range of variability plays a minor role and, thus, that the approximative procedures yield quite good results. Since this observation is already clearly visible for sample sizes of $n = 2, 4$, we will restrict attention to these values.

Neyman measurement procedure for $\mathrm{E}[X]$, $X|\{\mu_1\} \sim Mon(\mu_1)$

The Neyman measurement procedure ${}^*C^{(0.9)}_{\mathrm{E}[X]}$ for $\mathrm{E}[X]$ is an optimal measurement procedure in the sense that it has smallest value of the weighted volume V_w. Figure 7.17 displays the two β-uncertainty spaces $\mathcal{U}^{(0.9)}_{\overline{X}}$ for $\overline{X}$ for $n = 2$ and $n = 4$. As we have seen before, the geometric size of $\mathcal{U}^{(0.9)}_{\overline{X}}$ becomes smaller with larger value of n, i.e., resulting measurements will become shorter intervals. The weighted volume $V_w({}^*C^{(0.9)}_{\mathrm{E}[X]})$ has a value of 0.746186 for $n = 2$ and of 0.655538 for $n = 4$.

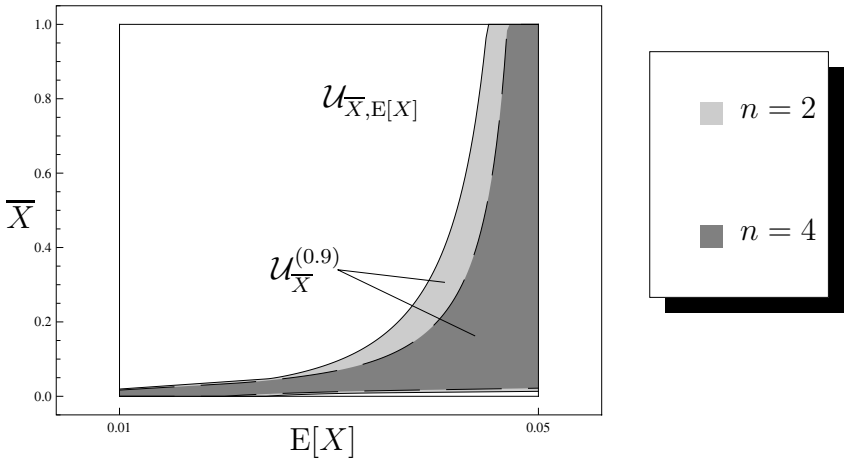
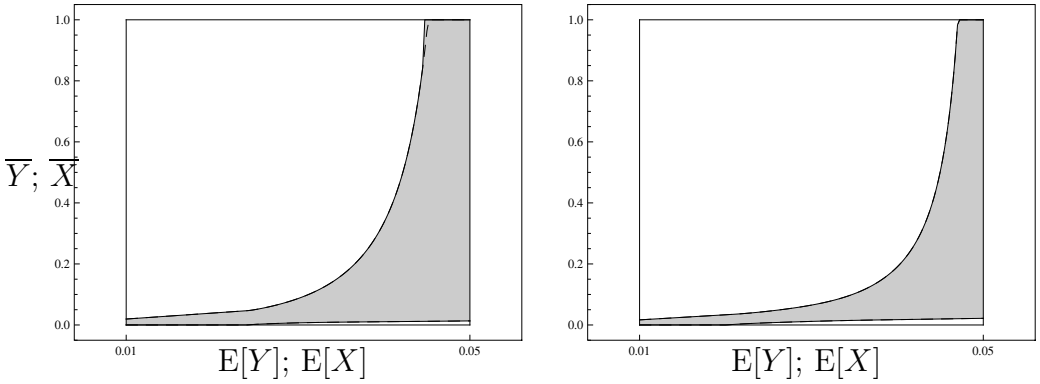


Figure 7.17.: The optimal β-uncertainty spaces $\mathcal{U}^{(0.9)}_{\overline{X}}$ for $\overline{X}$ and $X|\{\mu_1\} \sim Mon(\mu_1)$ for $n = 2$ and $n = 4$.



(a) $n = 2$ (b) $n = 4$

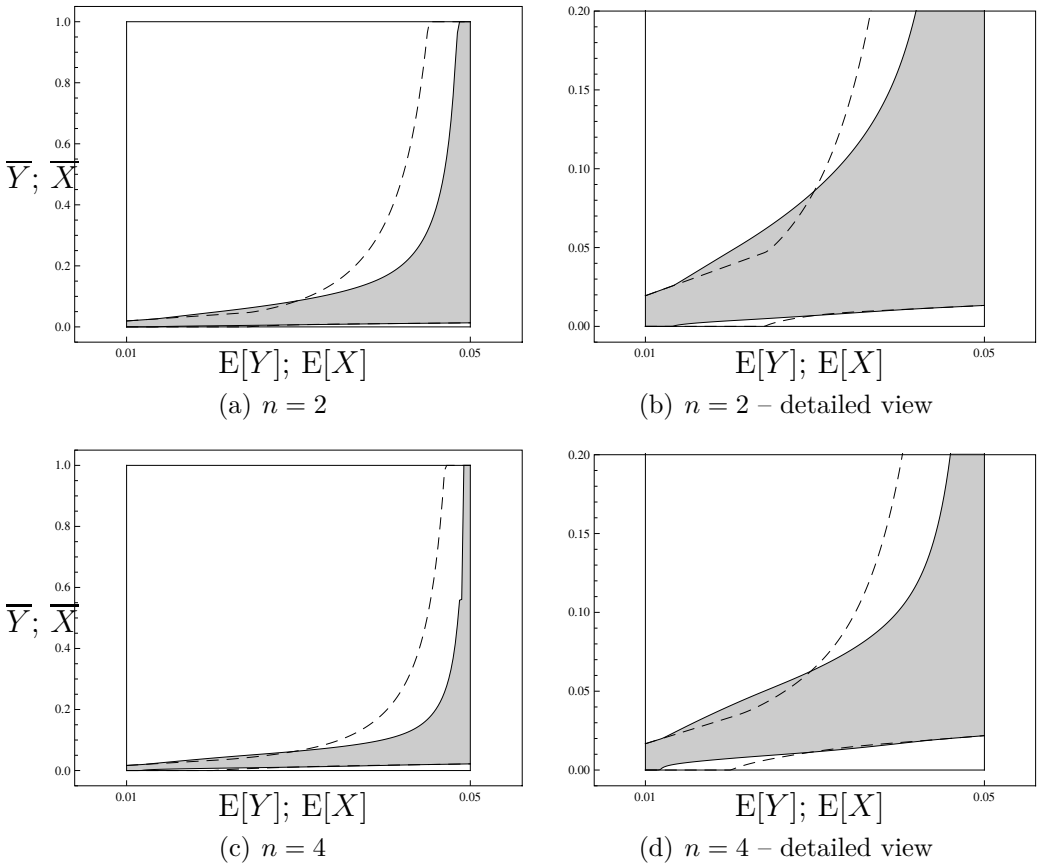
Figure 7.18.: The optimal β-uncertainty spaces $\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ and $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ for $n = 2$ and $n = 4$. In both cases compared to the optimal ones $\mathcal{U}_{\overline{X}}^{(0.9)}$ whose shapes are indicated by the (hardly visible) dashed lines.

Neyman measurement procedure for E[Y], $Y|\{\mu_1\} \sim EXP(1/\mu_1)$

In figure 7.18 the optimal β-uncertainty spaces $\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ are displayed for $n = 2$ and $n = 4$ directly in comparison to the optimal ones $\mathcal{U}_{\overline{X}}^{(0.9)}$. There are almost no visible differences, and in fact, the reliability is perfectly met and the values of the weighted volume are identical with respect to the numerical accuracy of the calculations. Thus, we may conclude, that both Neyman measurement procedures ${}^{*}C_{\mathrm{E}[X]}^{(0.9)}$ and ${}^{*}C_{\mathrm{E}[Y]}^{(0.9)}$ are equal – why we also omit to display the probability scaled plots of the respective β-uncertainty spaces.

LR measurement procedure for E[Y], $Y|\{\mu_1\} \sim EXP(1/\mu_1)$

Figure 7.19 displays the β-uncertainty spaces ${}^{LR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $n = 2$ and $n = 4$ in comparision to the optimal ones. The differences are pretty obvious in the upper right part of the graphics, where the optimal uncertainty space is a lot larger. The consequence is, that for relatively large realisations of $\overline{X}$ we would get longer measurements from the Neyman measurement procedure than from the approximative LR measurement procedure. In the lower left part of the graphics, the differences are well visible only in the detailed views. For small realisations the Neyman measurement procedure results into measurements shorter than the ones obtained from the approximative LR measurement procedure. Hardly visible



(a) $n = 2$ (b) $n = 2$ – detailed view

(c) $n = 4$ (d) $n = 4$ – detailed view

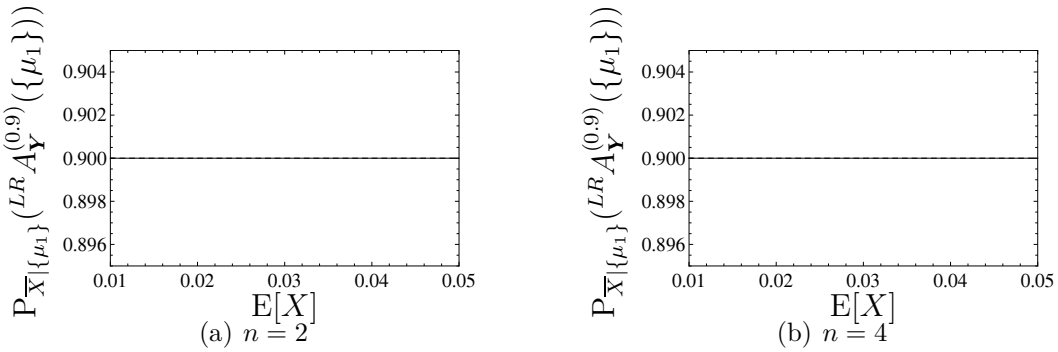
Figure 7.19.: The β-uncertainty space ${}^{LR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ and $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ for $n = 2$ and $n = 4$, respectively, based on the likelihood ratio test statistic and the optimal one $\mathcal{U}_{\overline{X}}^{(0.9)}$ (dashed shape).

also in the detailed views is the difference at the bottom of the graphics. For very small realisations the Neyman measurement procedure would yield longer measurements than the approximative LR measurement procedure.

The analysis of the actually reached reliability, that is the true probability of the prediction sets ${}^{LR}A_{\boldsymbol{Y}}^{(0.9)}(\{\mu_1\})$, shows no visible difference (see figure 7.20).

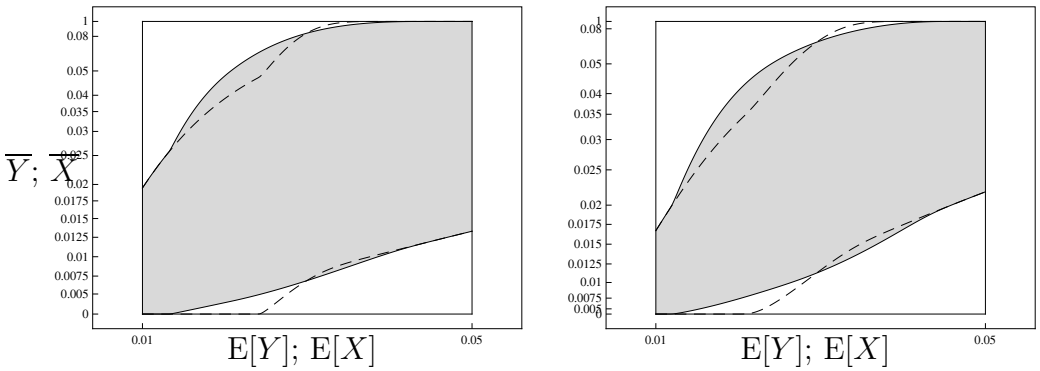
But calculating the values of the weighted volume yields $V_w\left({}^{LR}C_{\mathrm{E}[Y]}^{(0.9)}\right) = 0.755936$ for $n = 2$ and $V_w\left({}^{LR}C_{\mathrm{E}[Y]}^{(0.9)}\right) = 0.672171$ for $n = 4$, which are a little larger than 0.746186 and 0.655538, respectively. Here, it is worthwhile to take a look at the probability scaled plots of the β-uncertainty spaces in figure 7.21.

We realise again how misleading it would be to only compare the length of the



(a) $n = 2$ (b) $n = 4$

Figure 7.20.: The true probability of the prediction sets ${}^{LR}A_{\boldsymbol{Y}}^{(0.9)}(\{\mu_1\})$ from the LR measurement procedure for $n = 2$ and $n = 4$.



(a) $n = 2$ (b) $n = 4$

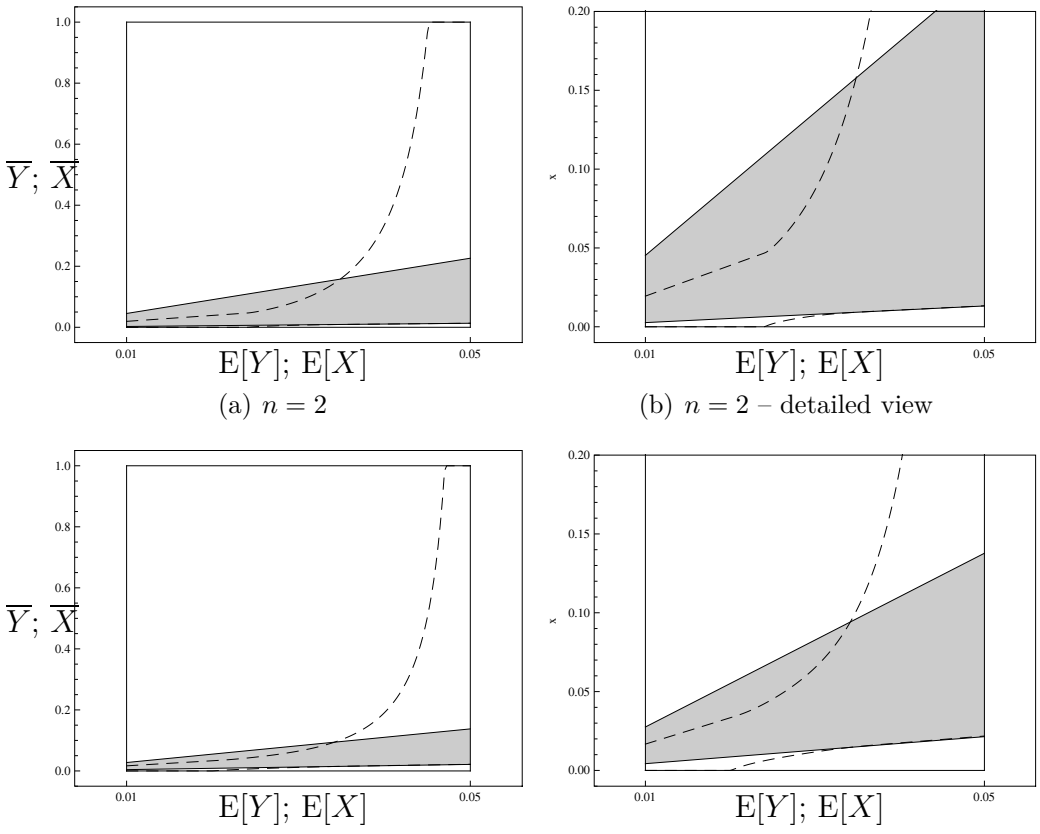
Figure 7.21.: The probability scaled plots of the β-uncertainty space ${}^{LR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ and $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ for $n = 2$ and $n = 4$, respectively, based on the likelihood ration test statistic and the optimal one $\mathcal{U}_{\overline{X}}^{(0.9)}$ (dashed shape).

resulting measurements without taking the underlying probability structure into account, which is done in the calculation of the weighted volume and the probability scaled plots.

HPR measurement procedure for E[Y], $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ based on $2\sum Y_i/\mu_1$

The comparison of the β-uncertainty spaces ${}^{HPR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ with the optimal ones shows really huge differences (figure 7.22). The misleading character of the pure analysis of the geometric sizes of measurements becomes even more evi-

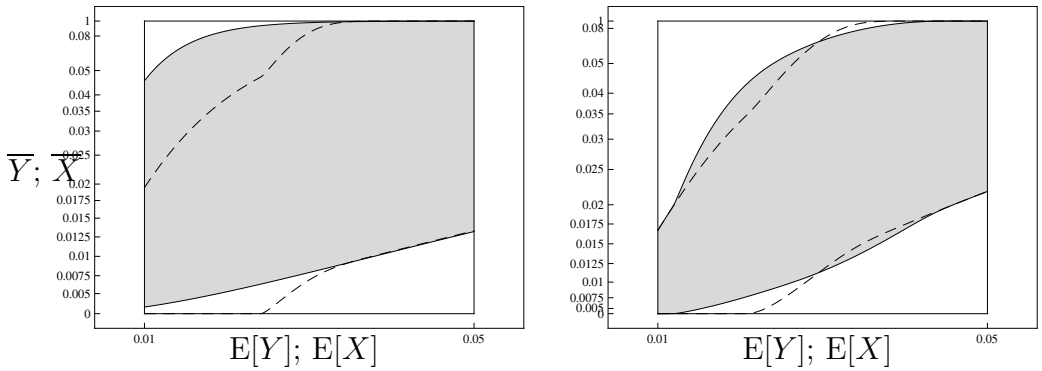
dent: While the measurements resulting from the HPR measurement procedure would be singletons for the vast majority of realisations of $\overline{X}$, i.e., either $\{m_\ell\}$ or (mainly) $\{m_u\}$, its weighted volume has a value of 0.827792 for $n = 2$ and 0.732585 for $n = 4$, respectively, both considerably larger than 0.746186 and 0.655538, respectively. Figure 7.23 displays this in the probability scaled plot of the respective β-uncertainty spaces. Interesting is the fact, that the demanded reliability is still met (figure 7.24).



(a) $n = 2$ (b) $n = 2$ – detailed view

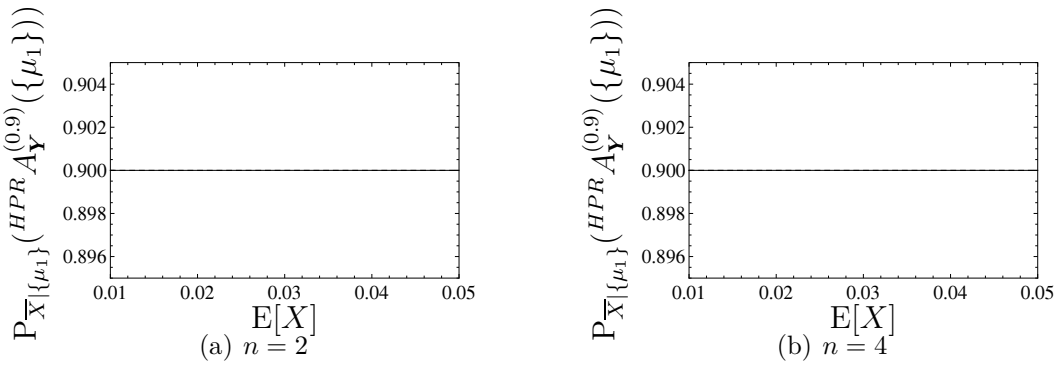
(c) $n = 4$ (d) $n = 4$ – detailed view

Figure 7.22.: The β-uncertainty space ${}^{HPR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ and $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ for $n = 2$ and $n = 4$, respectively, based on the likelihood ratio test statistic and the optimal one $\mathcal{U}_{\overline{X}}^{(0.9)}$ (dashed shape).



(a) $n = 2$ (b) $n = 4$

Figure 7.23.: The probability scaled plots of the β-uncertainty space ${}^{HPR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ and $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ for $n = 2$ and $n = 4$, respectively, based on the likelihood ratio test statistic and the optimal one $\mathcal{U}_{\overline{X}}^{(0.9)}$ (dashed shape).



(a) $n = 2$ (b) $n = 4$

Figure 7.24.: The true probability of the prediction sets ${}^{HPR}A_{\boldsymbol{Y}}^{(0.9)}(\{\mu_1\})$ from the HPR measurement procedure for $n = 2$ and $n = 4$.

7.2.4.2. $\mathcal{D} = \{\mu_1 \,|\, 0.05 \leq \mu_1 \leq 0.25\}$

Since the Neyman measurement procedure for E[X] and the considered approximations are already fully explained, we present subsequently only the respective coverage probabilities and weighted volumes. The uncertainty spaces in the unscaled and scaled form are displayed in the appendix B. The calculations here are performed for sample sizes $n = 2, 4, 8, 12, 16$.

Coverage Probabilities The graphical illustrations in figures 7.25, 7.26 and 7.27 of the respective coverage probabilities of the approximation procedures all show, that the required reliability $\beta = 0.9$ is met. We note that in all cases the exceedance is increasing with the true value μ_1 approaching the upper bound $m_u = 0.25$ and with greater sample sizes n. We conclude that all are reasonable approximations, but, at the same time, they will yield some larger measurements, i.e., longer measurement intervals.

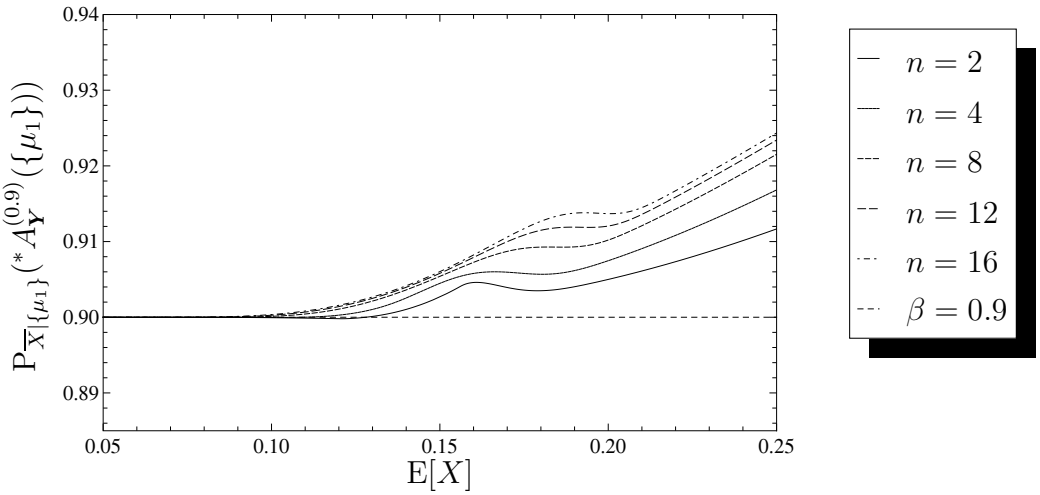


Figure 7.25.: The true probability of the prediction sets ${}^*A_{\boldsymbol{Y}}^{(0.9)}(\{\mu_1\})$ from the corresponding optimal measurement procedures for E[Y] with $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ for $n = 2, 4, 8, 12, 16$.

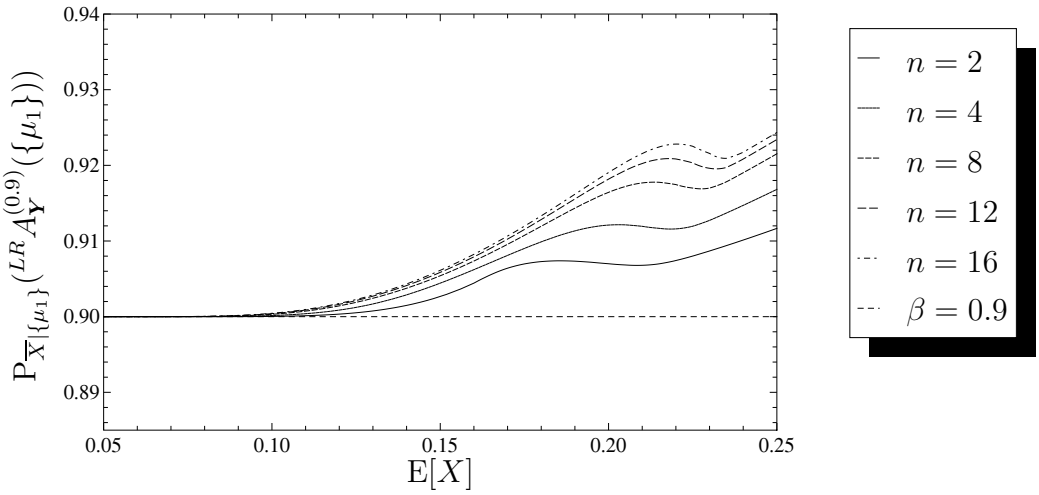


Figure 7.26.: The true probability of the prediction sets ${}^{LR}A_{\boldsymbol{Y}}^{(0.9)}(\{\mu_1\})$ from the corresponding measurement procedures for E[Y] with $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ based on the likelihhod ratio test statistic for $n = 2, 4, 8, 12, 16$.

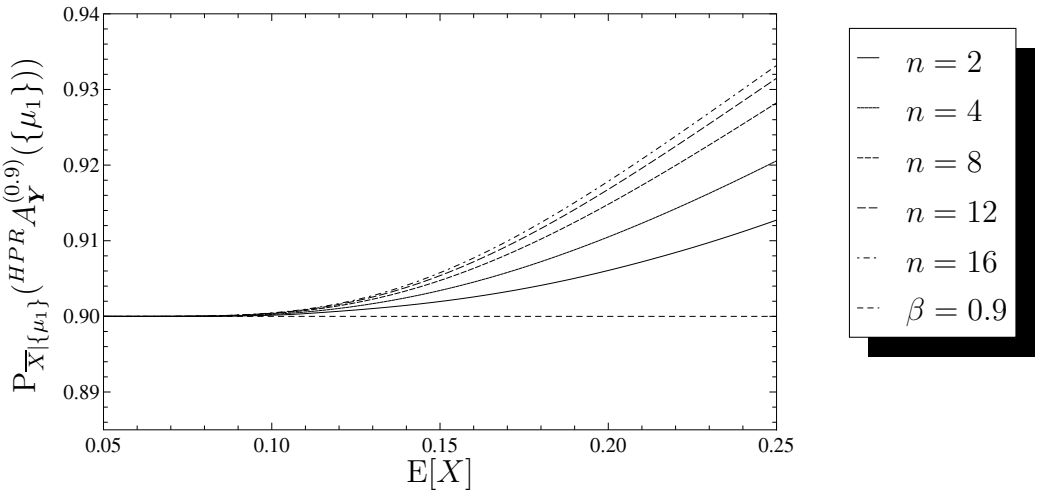


Figure 7.27.: The true probability of the prediction sets ${}^{HPR}A_{\boldsymbol{Y}}^{(0.9)}(\{\mu_1\})$ from the corresponding measurement procedures for E[Y] with $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ based on highest probability regions for $n = 2, 4, 8, 12, 16$.

Weighted Volumes Table 7.7 contains the respective weighted volumes V_w in comparison to the optimal Neyman measurement procedure: 1) For all procedures, the weighted volume decreases with increasing sample size, i.e., the accuracy increases. 2) Compared to the optimal Neyman procedure, the exceedance of V_w for the Neyman procedure w.r.t. $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ is considerably small, but increases with greater sample size. The LRT procedure shows some more exceedance, whilst the HPR procedure has the largest one. In the latter two cases, the exceedances increase at first, but then decrease with increasing sample size.

n	2	4	8	12	16
$V_w({}^{*}C^{(0.9)}_{\mathrm{E}[X]})$	0.73956	0.64674	0.53979	0.47650	0.43258
$V_w({}^{*}C^{(0.9)}_{\mathrm{E}[Y]})$	0.74533	0.65546	0.55007	0.48698	0.44293
	+0.78%	+1.35%	+1.90%	+2.20%	+2.39%
$V_w({}^{LR}C^{(0.9)}_{\mathrm{E}[Y]})$	0.75623	0.67323	0.56685	0.50011	0.45355
	+2.25%	+4.10%	+5.01%	+4.95%	+4.85%
$V_w({}^{HPR}C^{(0.9)}_{\mathrm{E}[Y]})$	0.83054	0.73553	0.60673	0.52869	0.47568
	+12.30%	+13.73%	+12.40%	+10.95%	+9.96%

Table 7.7.: Values of the weighted volume $V_w({}^{*}C^{(0.9)}_{\mathrm{E}[X]})$ of the optimal measurement procedure compared to those of the approximations for sample sizes $n = 2, 4, 8, 12, 16$.

7.2.4.3. $\mathcal{D} = \{\mu_1 \,|\, 0.40 \leq \mu_1 \leq 0.45\}$

As in the previous section, we present only the coverage probabilites and the weighted volumes of the respective approximations and compare them to the optimal Neyman procedure. Further graphics can again be found in the appendix B.

Coverage Probabilities As can be seen in figures 7.28, 7.29 and 7.30, the approximating procedures all meet the required reliability $\beta = 0.9$, but exceed it by far. That is, we could expect, that the procedure yields a correct measurement result. Actually, this is achieved more often than demanded and – presumably – by longer measurement results.

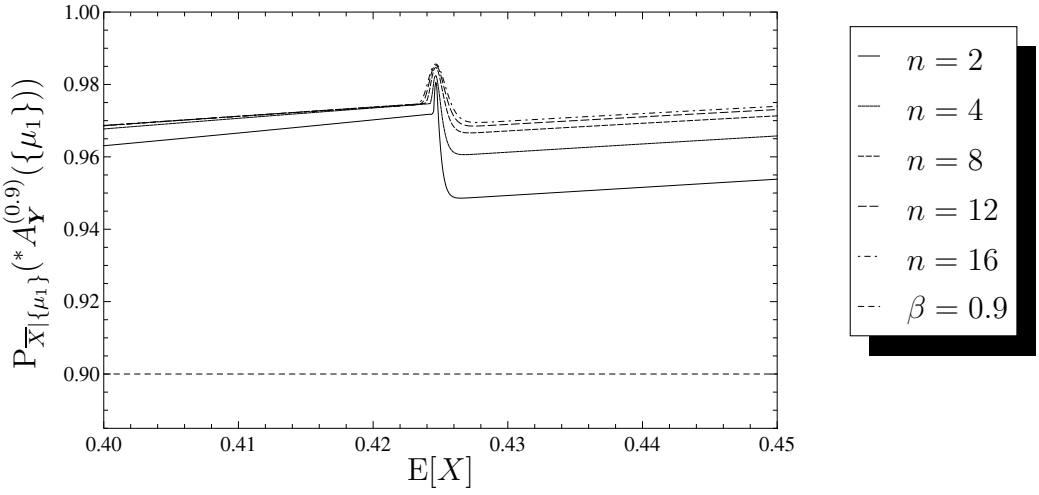


Figure 7.28.: The true probability of the prediction sets ${}^*A_{\boldsymbol{Y}}^{(0.9)}(\{\mu_1\})$ from the corresponding optimal measurement procedures for $\mathrm{E}[Y]$ with $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ for $n = 2, 4, 8, 12, 16$.

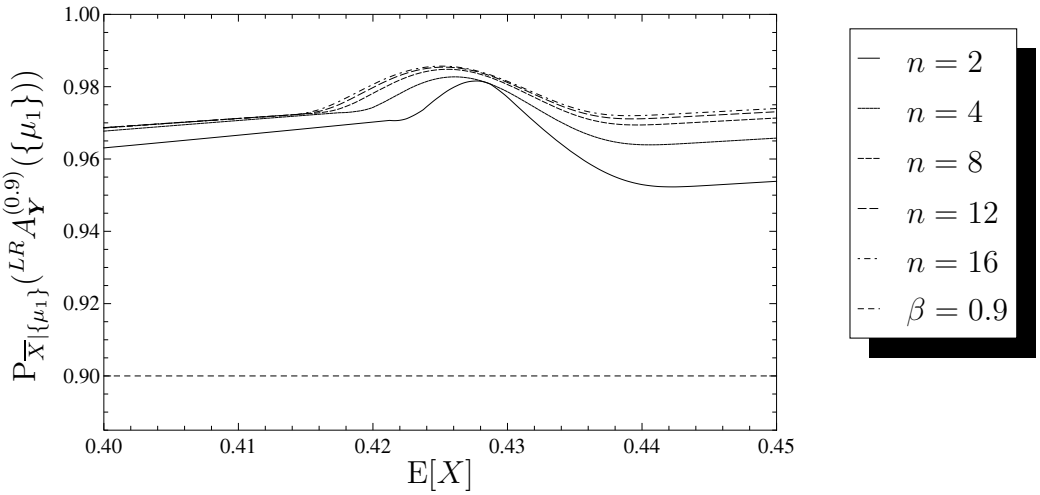


Figure 7.29.: The true probability of the prediction sets ${}^{LR}A_{\boldsymbol{Y}}^{(0.9)}(\{\mu_1\})$ from the corresponding measurement procedures for $\mathrm{E}[Y]$ with $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ based on the likelihhod ratio test statistic for $n = 2, 4, 8, 12, 16$.

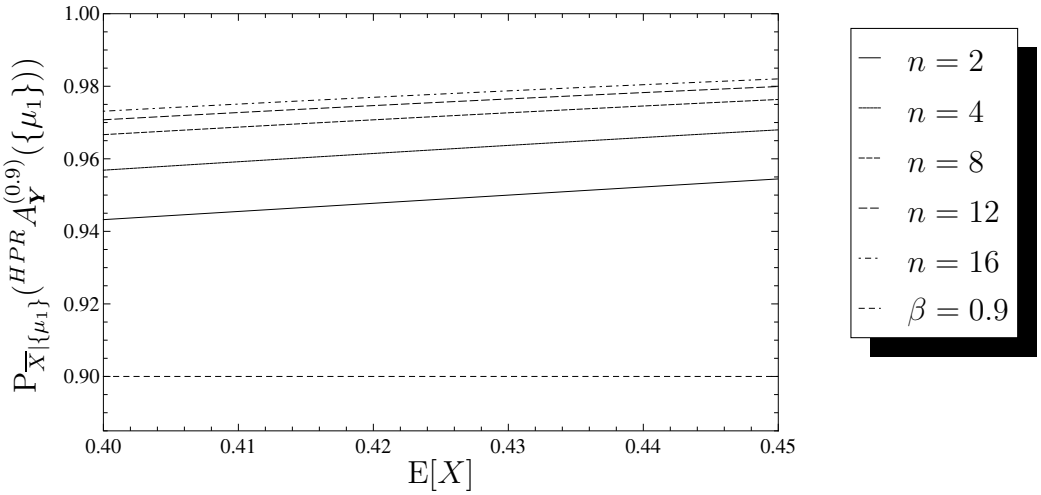


Figure 7.30.: The true probability of the prediction sets ${}^{HPR}A_{\boldsymbol{Y}}^{(0.9)}(\{\mu_1\})$ from the corresponding measurement procedures for $\mathrm{E}[Y]$ with $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ based on highest probability regions for $n = 2, 4, 8, 12, 16$.

Weighted Volumes Consistent with the exceedance in the coverage probability, the weighted volumes of the approximating procedures all exceed by similar amounts the one of the optimal Neyman procedure. Furthermore, this exceedance increases with increasing sample size n.

n	2	4	8	12	16
$V_w({}^{*}C^{(0.9)}_{\mathrm{E}[X]})$	0.88808	0.88218	0.87324	0.86581	0.85916
$V_w({}^{*}C^{(0.9)}_{\mathrm{E}[Y]})$	0.95357	0.95961	0.95932	0.95707	0.95453
	+7.37%	+8.78%	+9.86%	+10.54%	+11.10%
$V_w({}^{LR}C^{(0.9)}_{\mathrm{E}[Y]})$	0.95945	0.96369	0.96308	0.96084	0.95841
	+8.04%	+9.24%	+10.29%	+10.98%	+11.55%
$V_w({}^{HPR}C^{(0.9)}_{\mathrm{E}[Y]})$	0.94854	0.96176	0.97002	0.97297	0.97426
	+6.81%	+9.02%	+11.08%	12.38%	+13.40%

Table 7.8.: Values of the weighted volume $V_w({}^{*}C^{(0.9)}_{\mathrm{E}[X]})$ of the optimal measurement procedure compared to those of the approximations for sample sizes $n = 2, 4, 8, 12, 16$.

7.2.4.4. Concluding notes on the Measurement of $\mathrm{E}[X]$

We have discussed three different cases of ignorance spaces $\mathcal{D}$ with respect to $\mathrm{E}[X]$. We have compared the optimal Neyman measurement procedure to possible approximations.

For $\mathcal{D} = \{\mu_1 \,|\, 0.01 \leq \mu_1 \leq 0.05\}$, we notice almost no difference in the respective results: 1) The reliability is perfectly met and 2) the weighted volumes only exceed the one of the optimal procedure only by small amounts. These results were expectable: Since the ignorance space $\mathcal{D}$ contains only small values, the vast amount of the probability is concentrated on the left side of the range of variability $\mathcal{X} = [0, 1]$. That is, the boundedness to the right plays a very minor role.

The situation changes for $\mathcal{D} = \{\mu_1 \,|\, 0.05 \leq \mu_1 \leq 0.25\}$: 1) The coverage probabilities now exceed the required reliability and even increase when the true value

μ_1 approaches the right bound of $\mathcal{D}$, and 2) the weighted volumes exceeds the optimal one more distinctly. This is due to an ignorance space which causes that a certain amount of probability is allocated near the right bound of the range of variability $\mathcal{X} = [0, 1]$ – but not above as in the approximations.

The greatest impact can be observed for the ignorance space $\mathcal{D} = \{\mu_1 \,|\, 0.40 \leq \mu_1 \leq 0.45\}$. The assumed possible values of $\mathrm{E}[X]$ cause the respective density functions still to be monotone decreasing, but, at the same time, a considerable large amount of probability is allocated in the right half of $\mathcal{X} = [0, 1]$. The weighted volume of the optimal Neyman measurement procedure, which is the only one incorporating the boundedness of $\mathcal{X} = [0, 1]$, is by far the smallest. We also note, that the relative differences in the weighted volumes increase with increasing sample size n. Nevertheless, the approximations still meet the required reliability.

7.3. Measurement Procedures for B

We assume that there is only ignorance about the value b of the upper bound B of the range of variability $\mathcal{X}$ of $X|\{(a,b,\mu_1)\} \sim Mon(a,b,\mu_1)$, while the value μ_1 of the expectation $\mathrm{E}[X]$ and the value a of the lower bound A of $\mathcal{X}$ are known. Without loss of generality we may then assume $a = 0$, since by proposition 3.11 we get for $X|\{(a,b,\mu_1)\} \sim Mon(a,b,\mu_1)$ that $X|\{(a,b,\mu_1)\} - a = Y|\{(0,b-a,\mu_1 - a)\} \sim Mon(0,b-a,\mu_1-a)$. Furthermore, we want to assume that the density function $f_{X|\{(0,b,\mu_1)\}}$ is monotonously decreasing, i.e., $\mu_1 < \frac{b}{2}$. The deterministic variable then can be identified by the upper bound of the range of variability, that is $D = B$ with ignorance space:

$$\mathcal{D} \;=\; \{b \,|\, 2\mu_1 < b_\ell \leq b \leq b_u\} \tag{7.119}$$

Even though, the deterministic variable consists only of the one component B while A and $\mathrm{E}[X]$ have fixed values $a = 0$ and μ_1, we keep the notation $X|\{(b,\mu_1)\}$ to omit confusions.

As in section 7.2.1 and according to proposition 6.11 we collect the functions necessary for determining the β-prediction procedures ${}^*A^{(\beta)}_{\boldsymbol{X}}$:

- The density functions $f_{X|\{(b,\mu_1)\}}$ and $f_{\boldsymbol{X}|\{(b,\mu_1)\}}$ of $X|\{(b,\mu_1)\}$ and $\boldsymbol{X}|\{(b,\mu_1)\}$, respectively:

$$f_{X|\{(b,\mu_1)\}}(x) \;=\; \frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)}-1} e^{\lambda(b,\mu_1)x} \mathbb{1}_{[0,b]}(x)\,, \tag{7.120}$$

$$f_{\boldsymbol{X}|\{(b,\mu_1)\}}(\boldsymbol{x}) \;=\; \left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)}-1}\right)^n e^{\lambda(b,\mu_1)(\sum_{i=1}^n x_i)} \prod_{i=1}^n \mathbb{1}_{[0,b]}(x_i) \tag{7.121}$$

- The weight function w:

$$w(\boldsymbol{x}) \;=\; \int_{b_\ell}^{b_u} \left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)}-1}\right)^n e^{\lambda(b,\mu_1)(\sum_{i=1}^n x_i)} \prod_{i=1}^n \mathbb{1}_{[0,b]}(x_i)\, \mathrm{d}b \tag{7.122}$$

The density functions are the same as in the aforementioned section 7.2.1, and thus, we note the dependency on the sum $\sum_{i=1}^n x_i$. Since the aim is to derive a measurement procedure for B we rewrite the indicator function part of

$f_{\boldsymbol{X}|\{(b,\mu_1)\}}(\boldsymbol{x})$ to get a function in b:

$$\prod_{i=1}^{n} \mathbb{1}_{[0,b]}(x_i) \;=\; \mathbb{1}_{[\max x_i,+\infty)}(b) \cdot \mathbb{1}_{(-\infty,\min x_i]}(0) \tag{7.123}$$

While the factor $\mathbb{1}_{(-\infty,\min x_i]}(0)$ yields always 1 and may be neglected, the factor $\mathbb{1}_{[\max x_i,+\infty)}(b)$ reflects the connection between $\max x_i$ and b. Thus, inserting this into the weight function (7.122) yields

$$w(\boldsymbol{x}) \;=\; \int_{b_\ell}^{b_u} \left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)}-1}\right)^n e^{\lambda(b,\mu_1)(\sum_{i=1}^{n} x_i)} \mathbb{1}_{[\max x_i,+\infty)}(b)\, \mathrm{d}b \tag{7.124}$$

$$= \int_{\max\{\max x_i, b_\ell\}}^{b_u} \left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)}-1}\right)^n e^{\lambda(b,\mu_1)(\sum_{i=1}^{n} x_i)}\, \mathrm{d}b \tag{7.125}$$

Given $\boldsymbol{x}, \boldsymbol{y} \in [0,b]^n$ for some $b \in \mathcal{D}$ with $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$ and $\max x_i = \max y_i$ yields

$$f_{\boldsymbol{X}|\{(b,\mu_1)\}}(\boldsymbol{x}) \;=\; f_{\boldsymbol{X}|\{(b,\mu_1)\}}(\boldsymbol{y}) \tag{7.126}$$

$$w(\boldsymbol{x}) \;=\; w(\boldsymbol{y}) \tag{7.127}$$

and, thus:

$$\frac{f_{\boldsymbol{X}|\{(b,\mu_1)\}}(\boldsymbol{x})}{w(\boldsymbol{x})} = \frac{f_{\boldsymbol{X}|\{(b,\mu_1)\}}(\boldsymbol{y})}{w(\boldsymbol{y})} \tag{7.128}$$

Consequently, if one $\boldsymbol{x} \in {}^{*}A_{\boldsymbol{X}}^{(\beta)}(\{b\}) \subseteq [0,b]^n$ with $s := \sum_{i=1}^{n} x_i$ and $t := \max x_i$, then

$$\left\{\boldsymbol{y} \in [0,b]^n \middle| \sum_{i=1}^{n} y_i = s \text{ and } \max x_i = t\right\} \subset {}^{*}A_{\boldsymbol{X}}^{(\beta)}(\{b\})\,. \tag{7.129}$$

Now, it is very nearby to conjecture that a prediction procedure for $\boldsymbol{X}$ is equivalent to a prediction procedure for

$$(X_{(n)}, K_n)|\{(b,\mu_1)\} \;:=\; (\max X_i|\{(b,\mu_1)\}, \sum_{i=1}^{n} X_i|\{(b,\mu_1)\})\,.$$

In corollary 3.21 we have derived the joint density of $(X_{(n)}, K_n)|\{(b,\mu_1)\}$:

$$f_{(X_{(n)},K_n)|\{(b,\mu_1)\}}(t,s) =$$

$$\begin{cases} \left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)b}-1}\right)^n e^{\lambda(b,\mu_1)s} \sum_{m=1}^{k} \frac{n(n-1)}{(m-1)!(n-m)!}(s-mt)^{n-2}(-1)^{m+1} \\ \qquad\qquad \frac{1}{1+k}s \le t \le \frac{1}{k}s, \\ \qquad\qquad \text{for} \quad 0 \le t \le b, \qquad k = 1,\ldots,n-1, \\ \qquad\qquad 0 \le s \le nb, \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad \text{else.} \end{cases} \tag{7.130}$$

The summation part in (7.130) is independent of b and, thus, is cancelled in the ratio

$$\frac{f_{(X_{(n)},K_n)|\{(b,\mu_1)\}}(t,s)}{w(t,s)} = \frac{\left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)b}-1}\right)^n e^{\lambda(b,\mu_1)s}}{\int\limits_{\max\{t,b_\ell\}}^{b_u} \left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)}-1}\right)^n e^{\lambda(b,\mu_1)s}\,\mathrm{d}b}\,. \tag{7.131}$$

Since for all $\boldsymbol{x}$ with $t = \max x_i$ and $s = \sum_{i=1}^n x_i$ we get the same ratio

$$\frac{f_{\boldsymbol{X}|\{(b,\mu_1)\}}(\boldsymbol{x})}{w(\boldsymbol{x})} = \frac{\left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)b}-1}\right)^n e^{\lambda(b,\mu_1)s}}{\int\limits_{\max\{t,b_\ell\}}^{b_u} \left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)}-1}\right)^n e^{\lambda(b,\mu_1)s}\,\mathrm{d}b}\,, \tag{7.132}$$

we may conclude that both prediction procedures for $\boldsymbol{X}|\{(b,\mu_1)\}$ and for $(X_{(n)}, K_n)|\{(b,\mu_1)\}$, respectively, yield equivalent prediction sets.

We first want to get an impression of the uncertainty space $\mathcal{U}_{(X_{(n)},K_n)}$, of which the β-predictions ${}^*A^{(\beta)}_{(X_{(n)},K_n)}(\{b\})$ are subsets. Of course, $\mathcal{U}_{(X_{(n)},K_n)} \subset \mathbb{R}^3$ and for some fixed $b \in \mathcal{D}$ we have two inequalities defining the range of variability of $(X_{(n)}, K_n)|\{(b,\mu_1)\}$:

$$0 \le \max x_i \le b \tag{7.133}$$

$$\max x_i \le \sum x_i \le n \max x_i \tag{7.134}$$

Writing $t = \max x_i$ and $s = \sum x_i$ for the realisations of $(X_{(n)}, K_n)|\{(b,\mu_1)\}$ we get

$$\mathcal{U}_{(X_{(n)},K_n)} = \left\{(t,s,b) \,|\, b \in \mathcal{D}, 0 \le t \le b, t \le s \le nt\right\} \tag{7.135}$$

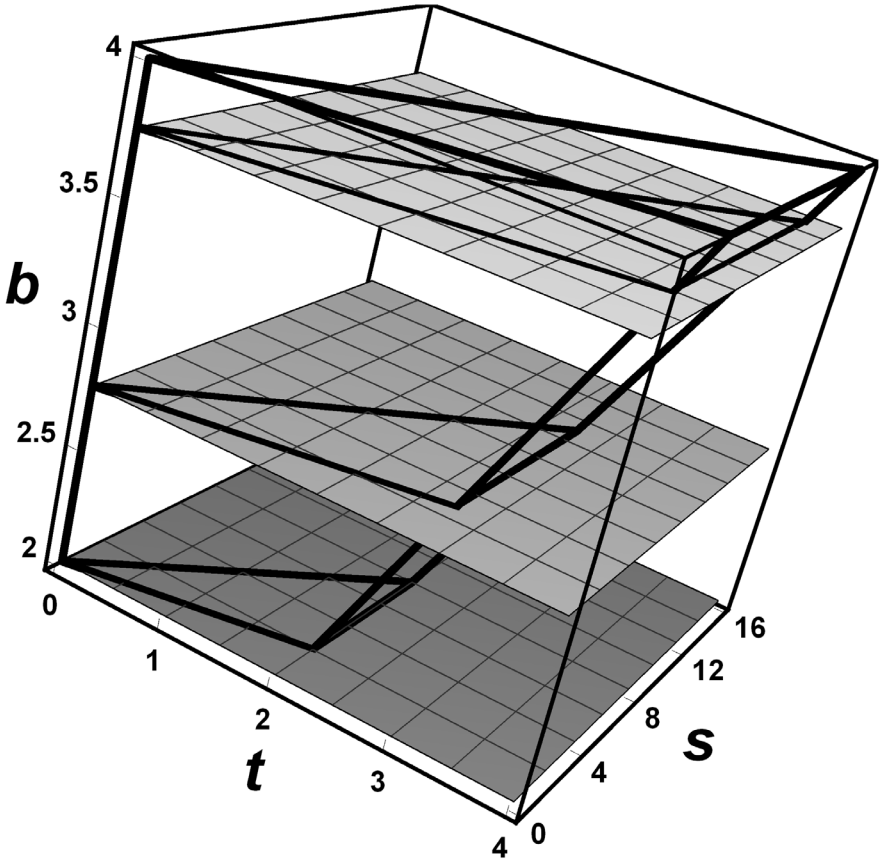


Figure 7.31.: Uncertainty space $\mathcal{U}_{(X_{(4)},K_4)}$ of $(X_{(4)}, K_4) = (\max X_i, \sum_{i=1}^{4} X_i)$ with $\mathcal{D} = \{b \,|\, b_\ell = 2 \leq b \leq 4 = b_u\}$. The intersections for certain values b show the respective ranges of variability of $(X_{(4)}, K_4)|\{(b, \mu_1)\}$.

Assuming $n = 4$ and $b \in \mathcal{D} = \{b \,|\, 2 \leq b \leq 4\}$, figure 7.31 displays $\mathcal{U}_{(X_{(n)},K_n)}$ together with some intersections for certain values b. The joint density $f_{(X_{(n)},K_n)|\{(b,\mu_1)\}}$ for $n = 4$ is displayed in figures 7.32 and 7.33 for $\mu_1 = 0.4$ and $\mu_1 = 0.9$, respectively, and values $b = 2, 3, 4$. The value $\mu_1 = 0.4$ has the effect, that most of the probability of $X|\{(b, 0.4)\}$ is concentrated to the left of $\mathcal{X}(\{(b, 0.4)\}) = [0, b]$ for all values b (the course of the density is rather steeply decreasing), which translates into the range of variability of $(X_{(n)}, K_n)|\{(b, \mu_1)\}$ in the way, that most of the probability is concentrated in the lower left corner. The value b has almost no effect on the course of the density function.

For $\mu_1 = 0.9$, the probability of $X|\{(b, 0.9)\}$ is more evenly distributed on

$\mathcal{X}(\{(b, 0.9)\}) = [0, b]$ for small values $b \in \mathcal{D}$ (the density is rather weakly decreasing) and, thus, the probability is mostly concentrated in the center of the range of variability of $(X_{(n)}, K_n)|\{(b, \mu_1)\}$. To make the graphics more comparable, the displayed underlying range of variability is always the one for $b = 4$ (in blue), i.e., the maximal value of B. We have to keep in mind that the actual ones are cut to the right by the actual value b.

Again, for each value $b \in \mathcal{D}$ the prediction sets for $(X_{(n)}, K_n)|\{(b, \mu_1)\}$ are obtained as follows:

$$ {}^{*}A^{(\beta)}_{(X_{(n)},K_n)}(\{b\}) = \left\{ (t,s) \;\middle|\; \frac{f_{(X_{(n)},K_n)|\{(b,\mu_1)\}}(t,s)}{w(t,s)} \geq r^*(b) \right\} \tag{7.136} $$

$$ = \left\{ (t,s) \;\middle|\; \frac{\left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)b}-1}\right)^n e^{\lambda(b,\mu_1)s}}{\int\limits_{\max\{t,b_\ell\}}^{b_u} \left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)}-1}\right)^n e^{\lambda(b,\mu_1)s}\,\mathrm{d}b} \geq r^*(b) \right\} \tag{7.137} $$

$$ \text{with } \mathrm{P}_{(X_{(n)},K_n)|\{(b,\mu_1)\}}({}^{*}A^{(\beta)}_{(X_{(n)},K_n)}(\{b\})) = \beta \tag{7.138} $$

Therefore, we should take some closer looks onto the ratio

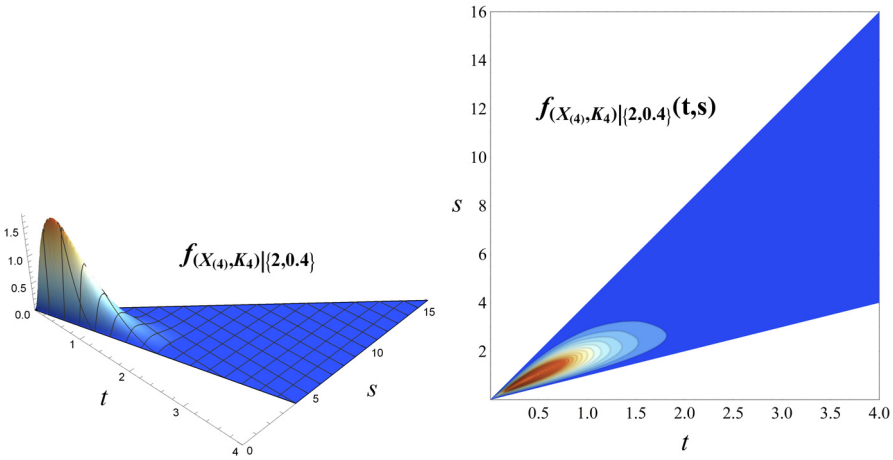
$$ r_b^{(\mu_1,n)}(t,s) \;:=\; \frac{f_{(X_{(n)},K_n)|\{(b,\mu_1)\}}(t,s)}{w(t,s)}\,. \tag{7.139} $$

With the afore assumed sample size $n = 4$, $\mathcal{D} = \{b \,|\, b_\ell = 2 \leq b \leq 4 = b_u\}$ for the ignorance space of B and the two values $\mu_1 = 0.4$ and $\mu_1 = 0.9$ for the expectation, figure 7.34 displays the ratio's course arranged in the same way as figures 7.32 and 7.33.

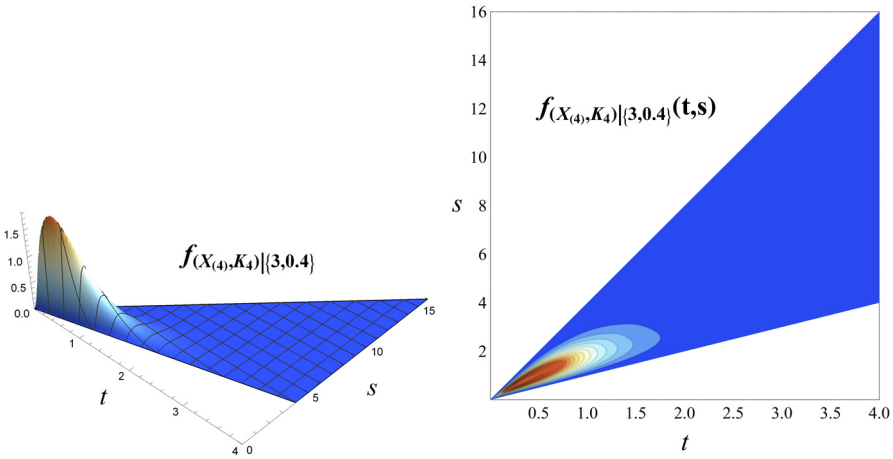
Obviously, we need to distinguish the two cases $t \leq b_\ell$ and $t > b_\ell$.
For all $t \leq b_\ell$ we have $\max\{t, b_\ell\} = b_\ell$ and the ratio's value only depends on s:

$$ r_b^{(\mu_1,n)}(t,s) = \frac{f_{(X_{(n)},K_n)|\{(b,\mu_1)\}}(t,s)}{w(t,s)} = \frac{\left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)b}-1}\right)^n e^{\lambda(b,\mu_1)s}}{\int\limits_{\max\{t,b_\ell\}}^{b_u} \left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)}-1}\right)^n e^{\lambda(b,\mu_1)s}\,\mathrm{d}b} \tag{7.140} $$

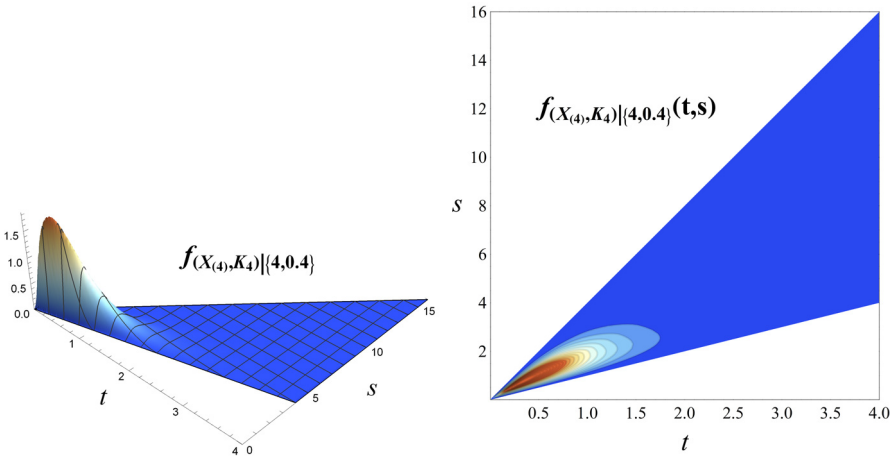
$$ = \frac{\left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)b}-1}\right)^n e^{\lambda(b,\mu_1)s}}{\int\limits_{b_\ell}^{b_u} \left(\frac{\lambda(b,\mu_1)}{e^{\lambda(b,\mu_1)}-1}\right)^n e^{\lambda(b,\mu_1)s}\,\mathrm{d}b} \tag{7.141} $$



(a) $\mu_1 = 0.4$, $b = 2$

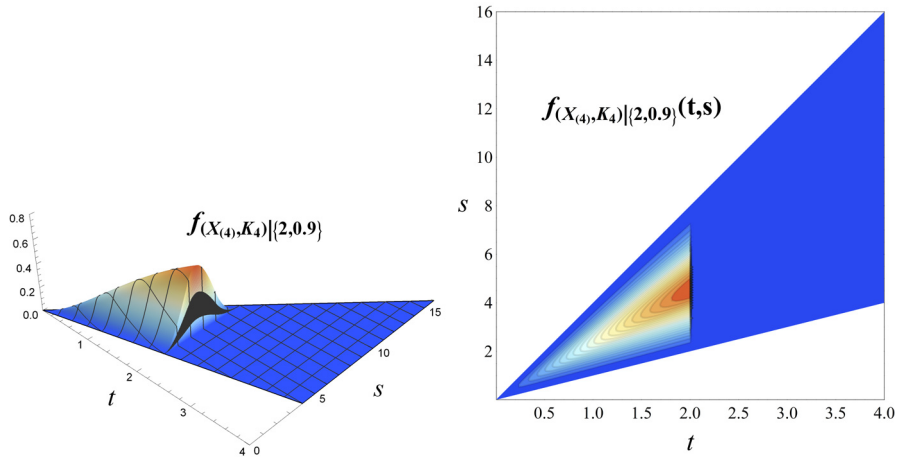


(b) $\mu_1 = 0.4$, $b = 3$

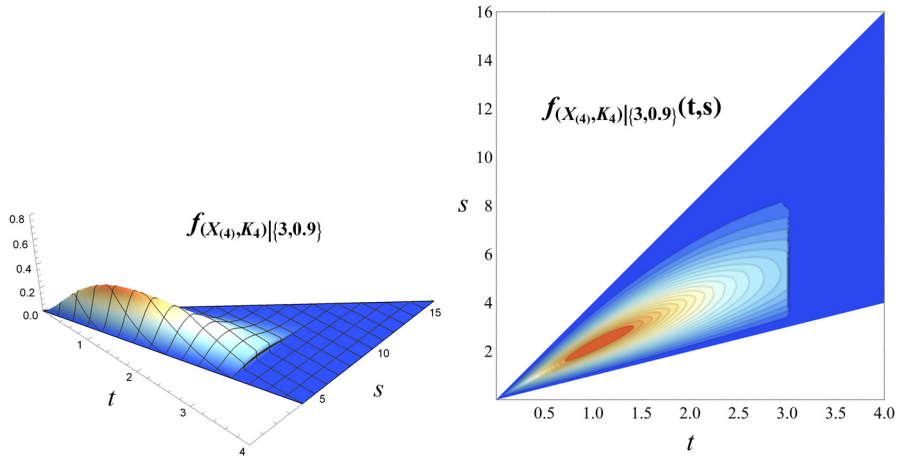


(c) $\mu_1 = 0.4$, $b = 4$

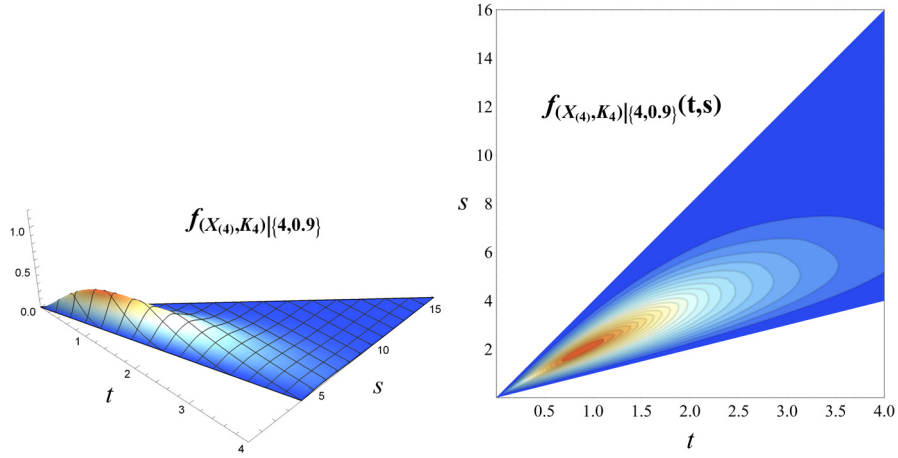
Figure 7.32.: 3-dimensional and contour plots of the joint density of $(X_{(4)}, K_4)|\{(b, \mu_1)\}$ for $\mu_1 = 0.4$ and values $b = 2, 3, 4$ for the upper bound B. The blue region in all graphics shows the maximal range of variability for $b = 4$.



(a) $\mu_1 = 0.9$, $b = 2$

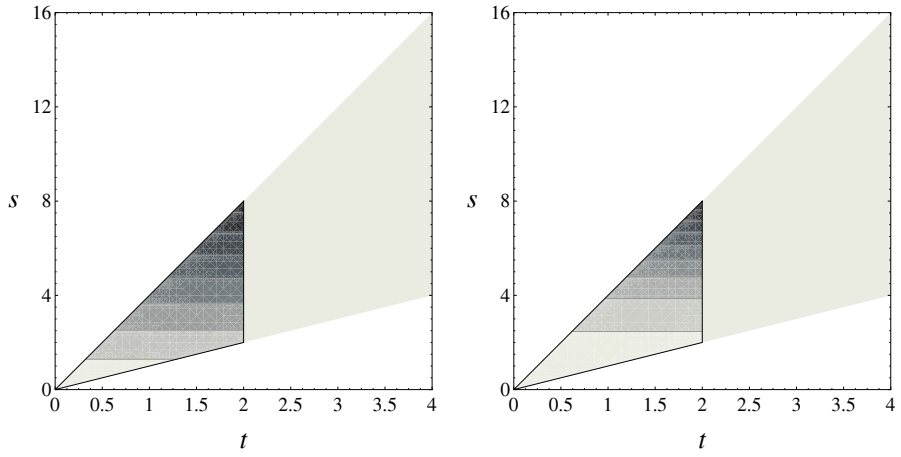


(b) $\mu_1 = 0.9$, $b = 3$



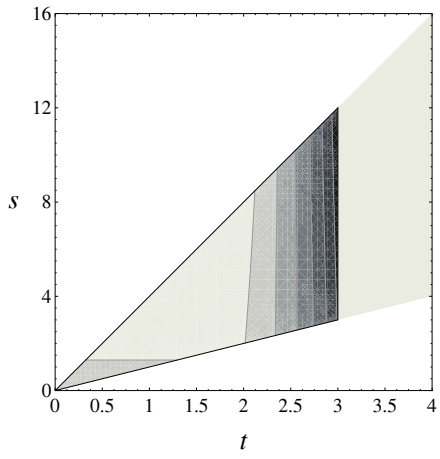
(c) $\mu_1 = 0.9$, $b = 4$

Figure 7.33.: 3-dimensional and contour plots of the joint density of $(X_{(4)}, K_4)|\{(b, \mu_1)\}$ for $\mu_1 = 0.9$ and values $b = 2, 3, 4$ for the upper bound B. The blue region in all graphics shows the maximal range of variability for $b = 4$.

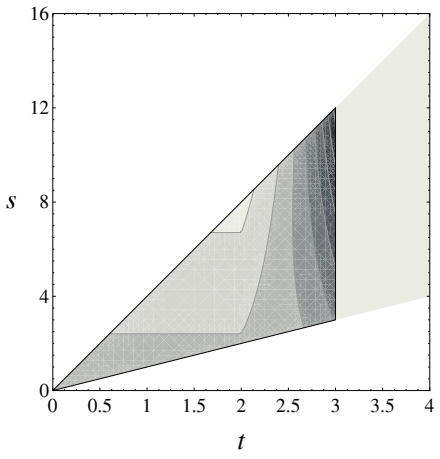


(a) $\mu_1 = 0.4$, $b = 2$

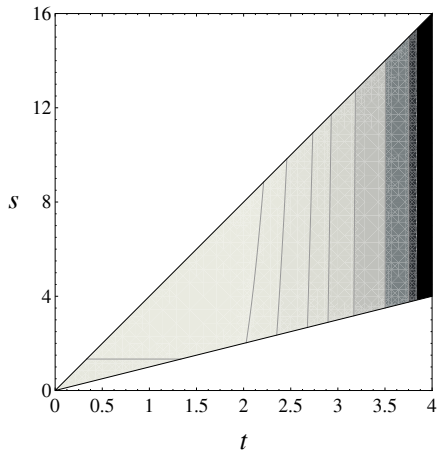
(b) $\mu_1 = 0.9$, $b = 2$



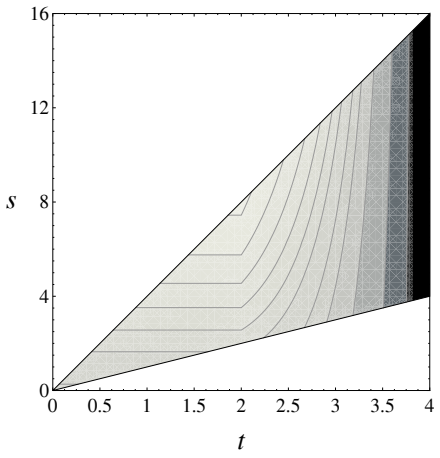
(c) $\mu_1 = 0.4$, $b = 3$



(d) $\mu_1 = 0.9$, $b = 3$



(e) $\mu_1 = 0.4$, $b = 4$

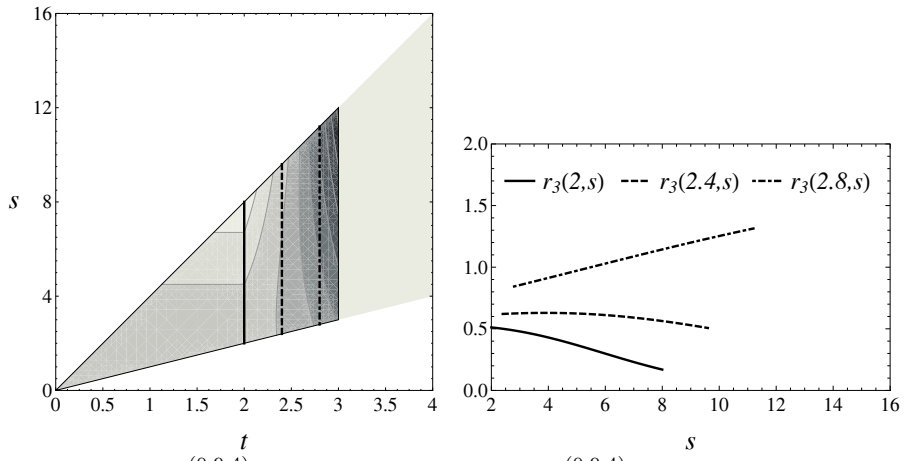


(f) $\mu_1 = 0.9$, $b = 4$

Figure 7.34.: Contour plots of the ratio $\frac{f_{(X_{(4)},K_4)|\{(b,\mu_1)\}}(t,s)}{w(t,s)}$ for $\mu_1 = 0.4$ (left column) and $\mu_1 = 0.9$ (right) and values $b = 2, 3, 4$ for the upper bound B.

That is, keeping s fixed the ratio adopts the same value for all t with $\frac{s}{n} \leq t \leq \min\{s, b_\ell\}$. In figure 7.34 this behaviour is clearly visible in the graphics 7.34(d) and 7.34(f) for $\mu_1 = 0.9$ and $b = 3, 4$, respectively, i.e., the contour lines are parallel to the t-axis. For $t > b_\ell$, i.e., the lower integration limit is $\max\{t, b_\ell\} = t$. Then, with some fixed s, the ratio $r_b^{(\mu_1,n)}(t, s)$ is strictly increasing in t, since the denominator, i.e., the integral, is getting smaller with increasing t.

On the other hand, keeping t fixed the course of the ratio $r_b^{(\mu_1,n)}(t, s)$ changes from strictly decreasing to unimodal to strictly increasing in s depending on t illustrated in figure 7.35. Since the ratio $r_b^{(\mu_1,n)}(t, s)$ cannot be expressed explicitly



(a) $r_3^{(0.9,4)}(t, s)$ (b) $r_3^{(0.9,4)}(t, s)$, $t = 2, 2.4, 2.8$

Figure 7.35.: Contour plot of the ratio $r_3^{(0.9,4)}(t, s)$ and the course for some fixed values t.

and the course eludes some clear rules, it seems to be unpromising to tighten the characterisation of $r_b^{(\mu_1,n)}(t, s)$. Instead, we propose an approximation for the prediction sets ${}^*A^{(\beta)}_{(X_{(n)},K_n)}(\{b\})$:

- Let $\mathcal{U}_{((X_{(n)},K_n),b_0)}$ be the uncertainty space of $(X_{(n)}, K_n)|\{(\mu_1, b_0)\}$ for $b = b_0$.
- Partionate $\mathcal{U}_{((X_{(n)},K_n),b_0)}$ into polygons $Poly_i$.
- Calculate $r_{b_0}(t, s)$ for each corner point of each $Poly_i$. The maximum of the values for each polygon $Poly_i$ defines r_i.

- Define the approximative β-prediction set ${}^{P}A^{(\beta)}_{(X_{(n)},K_n)}(\{b_0\})$ as follows:

$$ {}^{P}A^{(\beta)}_{(X_{(n)},K_n)}(\{b_0\}) = \bigcup_{\{i|r_i \geq r^*\}} Poly_i \tag{7.142}$$

with

$$ \mathrm{P}({}^{P}A^{(\beta)}_{(X_{(n)},K_n)}(\{b_0\})) = \mathrm{P}(\bigcup_{\{i|r_i \geq r^*\}} Poly_i) = \sum_{\{i|r_i \geq r^*\}} \mathrm{P}(Poly_i) = \beta \tag{7.143}$$

Figure 7.36 displays the resulting approximative β-prediction set ${}^{P}A^{(\beta)}_{(X_{(n)},K_n)}(\{b_0\})$ for $b_0 = 2.75 \in \mathcal{D} = \{b \,|\, 2 \leq b \leq 4\}$ with $\beta = 0.9$ and $n = 4$. Calculating approx-

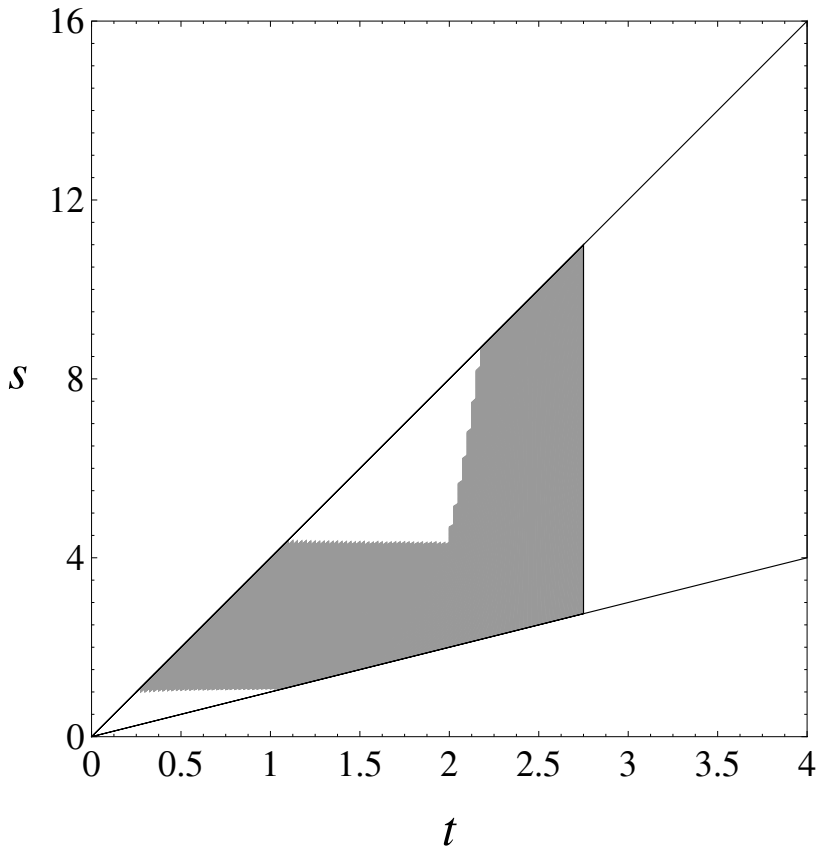


Figure 7.36.: Approximative β-prediction set ${}^{P}A^{(0.9)}_{(X_{(4)},K_4)}(\{2.75\})$ with $\mathcal{D} = \{b \,|\, 2 \leq b \leq 4\}$.

imative β-prediction sets ${}^{P}A^{(\beta)}_{(X_{(n)},K_n)}(\{b_0\})$ for a considerable number of $b_0 \in \mathcal{D}$ yields an approximative β-uncertainty space ${}^{P}\mathcal{U}^{(\beta)}_{B}$ and, thus, an approximative β-measurement procedure ${}^{P}C^{(\beta)}_{B}$ for B, respectively.

7.3.1. Proposal of an approximative measurement for B

To achieve an approximative measurement ${}^{P}C_{B}^{(\beta)}(t,s)$ for a realisation (t,s) of $(X_{(n)}, K_n)$:

- The inversion of the approximative β-prediction sets ${}^{P}A^{(\beta)}_{(X_{(n)},K_n)}(\{b_0\})$ yields a discrete set $H := \{b_0 \,|\, (t,s) \in {}^{P}A^{(\beta)}_{(X_{(n)},K_n)}(\{b_0\})\}$. Define $h_\ell := \min H$ and $h_u := \max H$. Obviously, it holds that $h_\ell \geq b_\ell$ and $h_u \leq b_u$.
- Those b_0 for which $(t,s) \notin {}^{P}A^{(\beta)}_{(X_{(n)},K_n)}(\{b_0\})\}$ may be excluded to be part of the desired measurement. In the most general case, these values b_0 may be devided into to distinguishable subsets: into one, say E_ℓ, which consists of b_0 smaller than h_ℓ and one, say E_u, which consists of b_0 larger than h_u. If $h_\ell = b_\ell$, then E_ℓ would be empty and will not be part of further considerations. Analogue, if $h_u = b_u$, then E_u would be empty and is excluded from further considerations. It may happen that both $h_\ell = b_\ell$ and $h_u = b_u$, then the preliminary measurement would be $\mathcal{D}$, i.e., no improvement by reducing $\mathcal{D}$ may be achieved.
- Define $e_\ell = \max E_\ell$ and $e_u = \min E_u$.
- For all $b_0 \leq e_\ell$ we have $(t,s) \notin {}^{P}A^{(\beta)}_{(X_{(n)},K_n)}(\{b_0\})\}$ and $(t,s) \in {}^{P}A^{(\beta)}_{(X_{(n)},K_n)}(\{h_\ell\})\}$.
- For all $b_0 \geq e_u$ we have $(t,s) \notin {}^{P}A^{(\beta)}_{(X_{(n)},K_n)}(\{b_0\})\}$ and $(t,s) \in {}^{P}A^{(\beta)}_{(X_{(n)},K_n)}(\{h_u\})\}$.
- A realised maximum t of the sample indicates that the real value b of B equals at least t, i.e., a lower bound of ${}^{P}C_{B}^{(\beta)}(t,s)$ is at least t if $t > b_\ell$.

Putting these considerations together we get an approximative measurement ${}^{P}C_{B}^{(\beta)}(t,s) = [c_\ell, c_u]$ by:

- the lower bound c_ℓ equals
  - b_ℓ if $t \leq b_\ell$ and $h_\ell = b_\ell$
  - e_ℓ if $t \leq e_\ell$ and $h_\ell > b_\ell$
  - t if $t > e_\ell$
- the upper bound c_u equals
  - b_u if $h_u = b_u$
  - e_u if $h_u < b_u$

At least, two open question appear: On the one hand whether the set $H := \{b_0 | (t,s) \in {}^{P}A^{(\beta)}_{(X_{(n)},K_n)}(\{b_0\})\}$ is always non-empty and, on the other hand, consists of successive values b_0 in the choosen division. All calculated examples suggest that both may be answered with 'yes'.

For $b_0 = 2, 2.2, \ldots, 3.8, 4$ with $\beta = 0.9$ and $n = 4$ figure 7.37 displays the approximative β-uncertainty space ${}^{P}\mathcal{U}^{(\beta)}_{B}$ together with the realisation $(t,s) = (2.1, 5)$. For values $b_0 \in \{2.2, 2.4, \ldots, 3.2\}$ we have $(2.1, 5) \in {}^{P}A^{(0.9)}_{(X_{(4)},K_4)}(\{b_0\})$. Since $t = 2.1 > b_\ell = 2$, we get an approximative measurement ${}^{P}C^{(0.9)}_{B}(2.1, 5) = [2.1, 3.4)$. Having identified e_ℓ, h_ℓ and e_u, h_u, respectively, we may do a kind of step-wise refinement to get a more accurate measurement.

7.3.2. Conclusion and Outlook

We have derived an approximate but nevertheless reliable measurement procedure for the upper bound B under prior knowledge for the monotonic distribution. In contrast to classical statistics the measurement procedure does without the derivation of a point estimator at first and developing from this an approximative set estimator. We admit, that the concrete implementation would be an laborious work and, thus, more research with respect to the occuring functions and the numerical issues should be done.

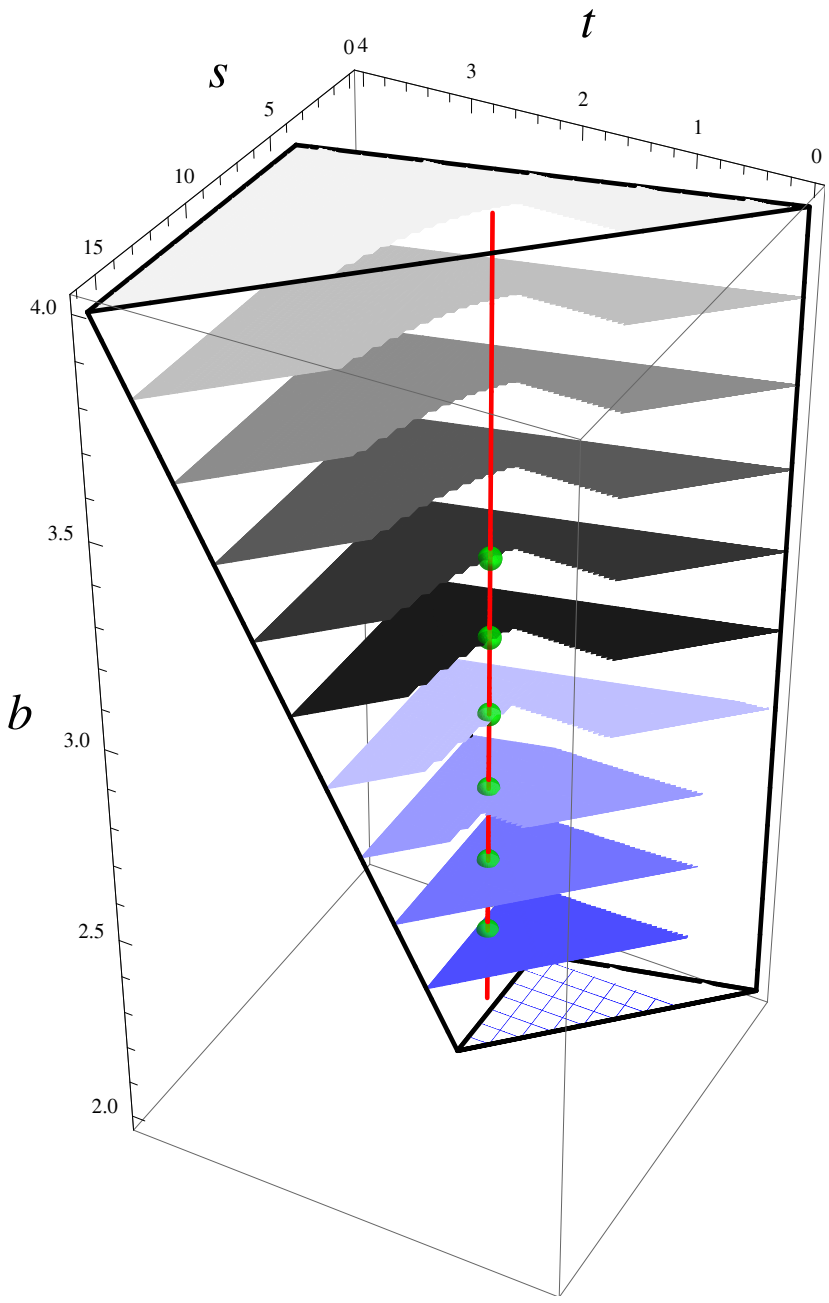


Figure 7.37.: Approximative β-uncertainty space ${}^{P}\mathcal{U}_{B}^{(\beta)}$ with $\mathcal{D} = \{b \,|\, 2 \leq b \leq 4\}$ for $\beta = 0.9$ and $n = 4$. The realisation $(t, s) = (2.1, 5)$ is element of ${}^{P}A^{(\beta)}_{(X_{(n)},K_n)}(\{b_0\})\}$ for some values b_0 (marked by the balls).

8. Two-sided Confidence Intervals for the Mean of Zero-Inflated Populations under Prior Information

This Chapter is based on a submitted article Göb & Sans (2018) [31].

8.1. Introduction

The study of zero-inflated populations has substantially been influenced by the statistical theory of industrial auditing.

Industrial auditing tests the *conformance* of monetary *book values* U kept in lists and databases on items like accounts, articles in an inventory, transactions, with the corresponding *de facto values* or *audit values* W of the items in reality. The degree of misstatement of U on W can be measured by the *tainting ratio* $Y = (U - W)/U$, i.e., the deviation of the book value U from the de facto value W relative to the stipulated book value U. In many contexts, misstatements tend to be *overstatements* where $0 \leq W < U$. Misstatement by overstatement is the dominant error mode particularly in asset accounts, accounts receivable and revenue accounts, see the empirical studies by Ramage et al. (1979) [64], Johnson et al. (1981) [38], Ham et al. (1985) [33], Icerman & Hillison (1990) [36].

Under the error mode overstatement, the tainting ratio ranges on the support $[0, 1]$, where $Y = 0$ represents a correct statement, and $Y = 1$ represents the case that a de facto value $W = 0$ is erroneously stated by a positive account entry $U > 0$.

In regular business practice, most book entries are correct, and small deviations

are more frequent than large deviations. Hence a common type of the distribution of the tainting ratio Y under overstatement has three properties: i) highly right skewed, ii) zero inflation, i.e., large probability point mass at 0, iii) PDF decreasing on $[0, 1]$. This pattern is often addressed as an "inverse J-shape", see Commission on Physical Sciences, Mathematics, and Applications (1988) [23]. The same pattern occurs in many other areas, e. g., accident costs in insurance, contamination, seismic analysis. In particular, zero inflation can be a consequence of measurement imprecision where small signals are cumulated on zero.

Confidence intervals for the mean μ_Y are of major interest in the analysis of zero-inflated random variables Y. Various authors, e. g., Kaplan (1973) [42, 43], Teitlebaum & Robinson (1975) [72], Neter & Loebbecke (1977) [55], have demonstrated that the familiar confidence limits based on Gauss or t statistics cannot guarantee the prescribed confidence level when sampling from zero-inflated distributions. The first approach to confidence limits specifically appropriate for inverse J-shaped distributions was suggested by Stringer (1963) [71] in the context of audit sampling. The formal development of the bound nowadays associated with Stringer's name is mainly due to studies of Meikle (1972) [53], Anderson & Teitlebaum (1973) [1], and Goodfellow et al. (1974) [32]. The intuitive basis of Stringer's approach in a multinomial error model is nicely described by Pap & van Zuijlen (1996) [59].

Various authors have studied the actual coverage of Stringer type intervals under prescribed nominal confidence levels. Throughout, simulation studies show that the Sringer bound is overconservative under reasonable nominal levels $0.5 \leq \gamma < 1$, see Burdick & Reneau (1978) [7], Reneau (1978) [65], Leitch et al. (1982) [47], Plante et al. (1985) [61], Tsui et al. (1985) [73]. Bickel (1992) [5] initiated the study of the asymptotic behaviour of the bound for large sample size n. Pap & vanZuijlen (1996) [59] showed that the bound is asymptotically conservative for confidence levels $0.5 \leq \gamma < 1$.

The conservatism of the Stringer bound has motivated various attempts to construct tighter limits for μ_Y by exploiting prior information on the distribution of Y, particularly in auditing, where sources of prior knowledge are historical auditing records and auditing procedures which precede sampling inspection in

the test of details context, namely risk assessment procedures, test of controls, substantive analytical procedures. Bayesian approaches to limits for the mean of zero-inflated random variables Y can be classified according to the respective ways of exploiting prior information, see Godfrey & Neter (1984) [30] into four categories: i) prior information only on the error rate $\mathrm{P}(Y > 0)$, see Vanecek (1978) [74]; ii) prior information only on the conditional distribution $\mathrm{P}(\cdot|Y > 0)$ of Y under $Y > 0$, e. g., Garstka (1977) [28]; iii) prior information on the compound distribution of Y, e. g., McCray (1984) [52]; iv) prior information both on the error rate $\mathrm{P}(Y > 0)$ and on the conditional distribution of Y under $Y > 0$, e. g., Cox & Snell (1979) [24], Godfrey & Neter (1984) [30], Neter & Godfrey (1985) [54].

The majority of studies on confidence intervals for zero-inflated populations has been concentrating on the one-sided intervals with upper bounds. However, two-sided bounds are important in many fields of application, particularly in auditing. In view of financial, administrative, and legal consequences, the auditor's primary interest is to restrict the type 1 risk of incorrectly not detecting an excessive mean tainting in the accounts. This interest has lead research to focus on upper confidence bounds for the mean tainting. However, the type 2 risk of incorrectly assuming an excessive mean tainting also has considerable impact on the auditor and the auditee. In particular, the supposition of misstatements will increase subsequent audit efforts, to the disadvantage of both the auditor and the auditee. Accordingly, the Commission on Physical Sciences, Mathematics, and Applications (1988) [23] considered lower confidence bounds as "an area of considerable importance where research is needed". In spite of this encouraging suggestion, lower confidence limits haven't received much attention. A lower Stringer bound can simply be obtained by applying the Stringer method to the observations $1 - Y$, but these intervals use to be wide. Plante, Neter, and Leitch (1984) [60] developed a lower bound by the multinomial method. Tsui et al. (1985) [73] take a Bayes approach by assuming a Dirichlet prior for the multinomial probability parameters. The simulation study by Matsumura et al. (1991) [51] shows that for both methods the actual confidence level is reasonably close to the prespecified nominal level.

Relative to the relevant literature, the present study has three distinctive features:

1) The intervals obtained are frequentist confidence intervals under prior information, not Bayesian credible intervals.

2) We use prior information on the conditional distribution $\mathrm{P}(\cdot|Y > 0)$ expressed via the monotonic distribution $Mon(\mu_X)$ with support $[0, 1]$, which reverts the choice of the conditional distribution to a substantiated principle instead of the convenience of mathematical analysis. At the same time, an extremely simple and convenient specification of the prior information is enabled, since the conditional distributions are uniquely determined by the support and the first moment.

3) We consider two-sided confidence intervals for the mean μ_Y.

The chapter is organised in the following way: After introducing a stochastic model for zero-inflated distributions on $[0, 1]$ in section 8.2, we define two two-sided confidence intervals for the tainting mean μ_Y in section 8.3. The first is a two-sided interval of Stringer type, the other a composition of a confidence interval for the proportion of overstatement and a confidence interval for the conditional mean under overstatement. It is shown, that the proposed composed interval meets a nominal confidence level γ. Section 8.4 introduces a stochastic model for the conditional distribution of overstatement by means of the principle of minimum information and maximum entropy, respectively, by exploiting only information about the support, i.e., the unit interval, and the conditional mean under overstatement. The monotonic distributions $Mon(\mu_X)$ meets all these requested properties. Consequently, the Neyman β-measurement procedure for $\mathrm{E}[X]$ of $X \sim Mon(\mu_X)$ derived in Chapter 7 is used to measure the conditional mean by means of confidence intervals under prior information. A simulation study and its results are outlined in the following sections 8.5 and 8.6: The results for the proposed two-sided confidence interval are compared to those from the two-sided Stringer interval. Finally, two sensitivity examinations are conducted in sections 8.7 and 8.8: one with respect to a near misstatement of prior information about the conditional mean, the other with respect to a misstatement of the conditional distribution under overstatement.

8.2. Zero inflated Distributions on the Unit Interval

A random variable Y with a potentially zero-inflated distribution on $[0,1]$ can be modelled as

$$Y = ZX, \qquad \text{with } Z,\ X \text{ independent, } Z \sim Bi(1,p),\ \mathrm{P}(0 \le X \le 1) = 1 \tag{8.1}$$

where X is an absolutely continuous random variable with support $[0,1]$. The distribution of Y is strictly zero-inflated iff $1-p = \mathrm{P}(Y=0) > 0$. By (8.1), the mean μ_Y and the CDF F_Y can be decomposed as

$$\mu_Y = \mathrm{E}[Y] = \mathrm{E}[ZX] = \mathrm{E}[Z]\mathrm{E}[X] = p\mu_X\,, \tag{8.2}$$

$$F_Y(y) = \begin{cases} 1-p, & \text{for } y = 0, \\ 1-p+pF_X(X \le y) = 1-p+p\int\limits_0^y f_X(x)\,\mathrm{d}x, & \text{for } y > 0. \end{cases} \tag{8.3}$$

In the auditing context, Y is the tainting ratio $Y = (U-W)/U$ where U is the book value and W is the associated de facto value. The tainting ratio Y satisfies the model (8.1) in an overstatement framework where $0 \le W \le U$ with probability one.

8.3. Two-sided confidence intervals for μ_Y

In this section we will 1) recapitulate a two-sided version of the Stringer bounds to derive confidence intervals for μ_Y, and 2) introduce a procedure based on model (8.1) which utilises the decomposition (8.3).

Let $Y_1, \ldots, Y_n$ be an i.i.d. sample of variables of type $Y_i = Z_iX_i$ as decomposed by (8.1), arranged in the sample vector $\boldsymbol{Y} = (Y_1, ..., Y_n)^\top$. Let $0 \le Y_{(1,n)} \le \ldots \le Y_{(n,n)}$ be the corresponding ordered sample. Let $K = \sum_{i=1}^n \mathbb{1}_{(0,1]}(Y_i) = \sum_{i=1}^n Z_i$ be the number of non-zero $Y_i = X_i$ in the sample, and let

$$\boldsymbol{X}_K = \begin{cases} (Y_{(n-K+1,n)}, \ldots, Y_{(n,n)})^\top & \text{if } K > 0, \\ \emptyset & \text{if } K = 0 \end{cases} \tag{8.4}$$

be the vector of the corresponding non-zero sample components. K is distributed by the binomial distribution $Bi(n,p)$.

Both types of intervals for μ_Y considered subsequently are based on the two-sided Clopper & Pearson [14] confidence bounds $0 \leq p_{L,\beta}(m) < p_{U,\beta}(m) \leq 1$ for some success probability p in a series of Bernoulli trials of size n if $0 \leq m \leq n$ successes are observed. Under nominal level $0 < \beta < 1$ the Clopper-Pearson bounds are defined as the solutions of the equations

$$\sum_{j=m}^{n} \binom{n}{j} p_{L,\beta}(m)^j [1 - p_{L,\beta}(m)]^{n-j} \stackrel{!}{=} \frac{1-\beta}{2} \stackrel{!}{=} \sum_{j=0}^{m} \binom{n}{j} p_{U,\beta}(m)^j [1 - p_{U,\beta}(m)]^{n-j} . \tag{8.5}$$

Stringer's method can easily be adapted to obtain two-sided intervals for μ_Y, see Lurz (2015) [49].

Definition 8.1 (two-sided interval of Stringer type) *Let the nominal confidence level $\beta \in (0;1)$ be prescribed. Then the two-sided confidence interval $C^{St}_{\mu_Y}(\boldsymbol{Y}) := [\mu^{St}_{Y,L}, \mu^{St}_{Y,U}]$ of Stringer type for μ_Y at nominal level β is defined by*

$$\begin{aligned} \mu^{St}_{Y,L} &= p_{L,\beta}(0) + \textstyle\sum_{m=1}^{n} \left(p_{L,\beta}(m) - p_{L,\beta}(m-1)\right) Y_{(n-m+1,n)} , \\ \mu^{St}_{Y,U} &= p_{U,\beta}(0) + \textstyle\sum_{m=1}^{n} \left(p_{U,\beta}(m) - p_{U,\beta}(m-1)\right) Y_{(n-m+1,n)} . \end{aligned} \tag{8.6}$$

In an alternative approach, we establish separately a Clopper-Pearson confidence interval C_p for the probability $p = \mathrm{P}(Y > 0) = \mathrm{P}(Z = 1)$ and a confidence interval C_{μ_X} for the conditional mean $\mu_X = \mu_{Y|Y>0}$, and combine these intervals using the decomposition (8.2). The following proposition 8.2 states the coverage properties of the combined interval.

Proposition 8.2 (combined confidence interval) *Let $0 < \beta_1, \beta_2 < 1$. Let $p_{L,\beta_1}(K) < p_{U,\beta_1}(K)$ be the Clopper-Pearson bounds defined by (8.5) at the nominal level β_1, and let $C_{\mu_X}(\boldsymbol{X}_K) = [\mu_{X,L}(\boldsymbol{X}_K), \mu_{X,U}(\boldsymbol{X}_K)]$ be a level β_2 confidence interval for μ_X. Then*

$$C^{de}_{\mu_Y}(\boldsymbol{Y}) = [p_L(K) \cdot \mu_{X,L}(\boldsymbol{X}_K),\ p_U(K) \cdot \mu_{X,U}(\boldsymbol{X}_K)] \tag{8.7}$$

is a confidence interval for μ_Y at the actual level $\beta_1 + \beta_2 - 1$, i.e., $\mathrm{P}_{\mu_Y}(\mu_Y \in C^{de}_{\mu_Y}(\boldsymbol{Y})) \geq \beta_1 + \beta_2 - 1$.

Proof of 8.2: The proof is a simple application of the Bonferroni inequality. Although we assume independence of Z and X in model (8.1), the number $K = \sum_{i=1}^{n} Z_i$ of non-zeros and $\boldsymbol{X}_K$ are clearly not independent and, thus, the respective confidence intervals also not. But by Bonferroni inequality we get

$$\begin{aligned}
&\mathrm{P}_{\mu_Y}(\mu_Y \in C^{\mathrm{de}}_{\mu_Y}(\boldsymbol{Y})) = \\
&\quad = \mathrm{P}_{p\mu_X}(p\mu_X \in [p_L(K) \cdot \mu_{X,L}(\boldsymbol{X}_K),\ p_U(K) \cdot \mu_{X,U}(\boldsymbol{X}_K)]) \\
&\quad \geq \mathrm{P}_{p\mu_X}(p \in [p_L(K),\ p_U(K)], \mu_X \in [\mu_{X,L}(\boldsymbol{X}_K),\ \mu_{X,U}(\boldsymbol{X}_K)]) \\
&\quad \geq \mathrm{P}_p(p \in [p_L(K),\ p_U(K)]) + \mathrm{P}_{\mu_X}(\mu_X \in C_{\mu_X}(\boldsymbol{X}_K)) - 1 = \beta_1 + \beta_2 - 1\,.
\end{aligned}$$

•

Any combination $\beta_1 + \beta_2 - 1 = \beta$ provides a confidence interval of level β for μ_Y, in particular, the standard choice $\beta_1 = (1+\beta)/2 = \beta_2$.

Subsequently, we use the decomposition interval established by proposition 8.2 with the prior information based interval

$$C_{\mu_X}(\boldsymbol{X}_K) \;=\; \begin{cases} \mathcal{D} & \text{if } K = 0, \\ C^{\star}_{\mu_X}\left(\frac{1}{K}\sum_{\ell=n-K+1}^{n} Y_\ell\right) & \text{if } K > 0 \end{cases} \tag{8.8}$$

for $\mu_X = \mu_{Y|Y>0}$ where prior information on μ_X is expressed by a subinterval $\mathcal{D} \subset (0.0, 0.5)$ which contains μ_X with certainty. The interval $C^{\star}_{\mu_X}$ is established in the subsequent section 8.4.

8.4. Prior information model

The proposition 8.2 decomposes the confidence interval μ_Y into a confidence interval C_p for $p = \mathrm{P}(Y > 0)$ and a confidence interval C_{μ_X} for the conditional mean $\mu_X = \mu_{Y|Y>0}$. In many applications, particularly in auditing, at least some vague prior knowldege on the range of μ_X is available. The present section expresses such prior information in a stochastic model for the PDF $f_X = f_{Y|Y>0}$.

We focus on the case of a decreasing PDF f_X. The latter assumption is particularly warranted in the auditing context where inverse J-shaped tainting distributions widely occur. For the choice of a class of model densities with support $[0, 1]$,

two requirements are paramount: 1) A sparse parametrisation with an explicit link to the mean $\mu = \mu_X = \mathrm{E}[X]$ as a parameter so as to enhance acceptance and understanding by the user community. 2) Substantiating the choice of the model class by transparent principles so as to avoid arbitrariness of the model approach.

As discussed in Chapter 2, both requests 1) and 2) are guaranteed by the class of *minimum information representatives*, or, equivalently *maximum entropy distributions* (see section 2.3.2). In particular, the monotonic distribution $Mon(\mu)$ meets these requirements. That is, the PDF $f_{X|\{\mu\}}$ of $X|\{\mu\}$ is

$$f_{X|\{\mu\}}(x) = \begin{cases} \frac{e^{\lambda(\mu)x}}{\int_0^1 e^{\lambda(\mu)x}\,\mathrm{d}x} = \frac{\lambda(\mu)}{e^{\lambda(\mu)}-1}e^{\lambda(\mu)x} & \text{for } x \in [0,1] \\ 0 & \text{else} \end{cases} \tag{8.9}$$

where $\lambda(\mu)$ is the solution of

$$\mu = \mathrm{E}[X|\{\mu\}] \stackrel{!}{=} \frac{1-\lambda-e^{-\lambda}}{\lambda(e^{-\lambda}-1)}. \tag{8.10}$$

The class of monotonic distributions $(f_{X|\{\mu\}})_{0<\mu<1}$ satisfies the request of a sparse parametrisation with an explicit link to the mean $\mu = \mu_1 = \mu_X$ as a parameter. The requirement of a decreasing PDF is equivalent to restricting to $\mu \in (0, 0.5]$ where $\mu = 0.5$ amounts to the uniform distribution on $[0,1]$. A more precise prior knowledge can be expressed by a prior information interval $\mathcal{D} = [\underline{\mu}_X, \overline{\mu}_X] \subset (0, 0.5)$.

A confidence interval for μ_X of the monotonic distribution $Mon(\mu_X)$ was already derived in section 7.2.1 by the Neyman β-measurement procedure ${}^*C^{(\beta)}_{\mathrm{E}[X]}$. It was shown, that the procedure need not be based on the whole sample vector $\boldsymbol{X}_K$, but on $\overline{X}$. Since the number K of non-zero $Y_i = X_i$ in the sample is random, we write $\overline{X}_K := \frac{1}{K}\sum_{\ell=n-K+1}^{n} Y_\ell$ and get

$$C_{\mu_X}(\boldsymbol{X}_K) \;=\; \begin{cases} \mathcal{D} & \text{if } K = 0, \\ {}^*C^{(\beta)}_{\mathrm{E}[X]}\left(\overline{X}_K\right) & \text{if } K > 0 \end{cases} \tag{8.11}$$

8.5. Objectives, Design, and Numerics of Simulation Study

Analytical finite sample results are difficult to obtain both for the two-sided Stringer intervals introduced by definition 8.1 and for the decomposition interval established by proposition 8.2. We use simulation to study two crucial issues: a) Comparing the two approaches with respect to the expected length and the actual pointwise coverage of the intervals. b) Analyse the sensitivity of the decomposition interval with respect to the assumed prior information.

Consider a sample size n, N independent simulation runs resulting into N i.i.d. simulation vectors $(\boldsymbol{Y}_i)_{i=1,...,N}$ where the components $Y_{i1}, ..., Y_{in}$ of the vectors $\boldsymbol{Y}_i$ are of type $Y_{ij} = Z_{ij}X_{ij}$ as decomposed by (8.1) with fixed $p = \mathrm{P}(Y_{ij} > 0) = \mathrm{P}(Z_{ij} = 1)$.

We study the expectation of two statistics $T = T(C(\boldsymbol{Y}))$ of the confidence intervals $C(\boldsymbol{Y}) = [\mu_{Y,L}, \mu_{Y,U}]$ for μ_Y:

1) the empirical coverage

$$Q \quad = \quad \mathbb{1}_{C(\boldsymbol{Y})}(\mu_Y) \tag{8.12}$$

pointwise in the true value μ_Y.

2) the interval length

$$H \quad = \quad \mu_{Y,U} - \mu_{Y,L}. \tag{8.13}$$

The expectation $\mu_{Q|\mu_Y} = \mathrm{P}_{\mu_Y}(\mu_Y \in C(\boldsymbol{Y}))$ is the coverage probability of the confidence interval pointwise as a function of μ_Y. The expectation $\mu_{H|\mu_Y}$ is the expected length of the confidence interval pointwise as a function of μ_Y. A precise direct estimation of the means μ_T of these statistics is difficult for small values $p = \mathrm{P}(Y > 0)$ where there are few samples only with a relatively large number k of non-zeros. For instance, under $p = 0.05$ and $n = 100$ we get $\mathrm{P}(k > 15) \approx 3.7 \times 10^{-5}$, and even for $N = 10^6$ we would have on the average only 37 samples with a number of non-zeros greater than 15. To account for the variability under larger values k requires enormous simulation sizes N. Instead,

we base the estimation of the mean μ_T via simulation on the decomposition

$$\mu_T \;=\; \sum_{k=0}^{n} \mu_{T|K=k} \mathrm{P}(K=k) \;=\; \sum_{k=0}^{n} \mu_{T|K=k} \binom{n}{k} p^k (1-p)^{n-k} \tag{8.14}$$

relative to the values $K = 0, ..., n$ of the random number K of nonzero components in the sample vector $\boldsymbol{Y}$. For prescribed probability $p = \mathrm{P}(Y > 0) = \mathrm{P}(Z = 1)$ and $\mu = \mu_X = \mu_{Y|Y>0}$, the estimation algorithm proceeds in the following steps:

i) For each $K = k > 0$:

   i.1) simulate N vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ of k i.i.d. components from $Mon(\mu)$;

   i.2) form the total sample vectors $\boldsymbol{Y}_i$ by joining $n - k$ zeroes to $\boldsymbol{X}_i$;

   i.3) estimate $\mu_{T|K=k}$ by the average $\widehat{\mu}_{T|K=k} = \sum_{i=1}^{N} T_i/N$ over the observations under $K = k$.

ii) adopting the decomposition (8.14), estimate μ_T by

$$\widehat{\mu}_T = \sum_{k=0}^{n} \widehat{\mu}_{T|K=k} \binom{n}{k} p^k (1-p)^{n-k} \,. \tag{8.15}$$

We choose the following parameters for the simulation study:

- The probability p ranges over $\{0.025, 0.050, 0.075, 0.100\}$.
- The mean μ_X adopts the values 0.1 and 0.2
- The sample size n ranges over $\{25, 50, 75, 100, 150, 200, 250\}$
- $N = 500$ simulation runs for each value $K = k \in \{1, \ldots, 250\}$ ($k = 0$ needs no simulation).
- Nominal confidence levels $\beta_1 = \beta_2 = 0.975$ for the decomposition interval introduced by proposition 8.2. The corresponding total confidence level $\beta = 2 \cdot 0.975 - 1 = 0.95$ is used for the two-sided Stringer intervals as introduced by definition 8.1.
- For simulation data with $\mu_X = 0.1$ the prior information intervals are $\mathcal{D}_{Ia} = [0.01, 0.49]$, $\mathcal{D}_{Ib} = [0.01, 0.30]$, $\mathcal{D}_{Ic} = [0.05, 0.20]$.
- For simulation data with $\mu_X = 0.2$ the prior information intervals are $\mathcal{D}_{IIa} = [0.01, 0.49]$, $\mathcal{D}_{IIb} = [0.05, 0.35]$, $\mathcal{D}_{IIc} = [0.10, 0.25]$

For $\mu_X = 0.1$, the considered range of expected values $\mu_Y = 0.1p$ is 0.0025, 0.005, 0.0075, 0.01. For $\mu_X = 0.2$, the considered range of expected values $\mu_Y = 0.2p$ is 0.005, 0.01, 0.015, 0.02.

Determining an optimal β-uncertainty space $\mathcal{U}_{\overline{X}_k}^{(\beta)}$ for each $k = 1, \ldots, n$ is a prerequisite for obtaining the optimum β-measurement procedure ${}^*C_{\mathrm{E}[X]}^{(\beta)}$. For large k, obtaining $\mathcal{U}_{\overline{X}_k}^{(\beta_1)}$ is laborious. To save time we calculate β-uncertainty spaces only for k in the range $1, 2, \ldots, 50, 60, 70, \ldots, 240$. For k outside the latter range we choose the largest element $k' \in \{1, 2, \ldots, 50, 60, 70, \ldots, 240\}$ smaller than k and use in the decomposition interval by proposition 8.2 the optimum β-measurement ${}^*C_{\mathrm{E}[X]}^{(\beta)}(\overline{X}_{k'})$ calculated from a vector $\boldsymbol{X}_{k'}$ of length k' where the mean $\overline{X}_{k'}$ should be the mean $\overline{X}_k$ of $\boldsymbol{X}_k$. For instance, for $k = 77$ we use the β-uncertainty space $\mathcal{U}_{\overline{X}_{k'}}^{(\beta)}$ for $k' = 70$. To indicate the potential difference between ${}^*C_{\mathrm{E}[X]}^{(\beta)}(\overline{X}_k)$ (nominal sample size k equals the length of $\boldsymbol{X}_k$) and ${}^*C_{\mathrm{E}[X]}^{(\beta)}(\overline{X}_{k'})$, we write ${}^*\widetilde{C}_{\mathrm{E}[X]}^{(\beta)}(\overline{X}_{k'})$.

The effects of the numerical policy on k are checked by a supplementary simulation study. For $k \in \{20, \ldots, 50\}$ we calculate β-uncertainty spaces $\mathcal{U}_{\overline{X}_k}^{(\beta)}$ with $\beta = 0.975$ and prior information $\mathcal{D} = [0.01, 0.49]$. For $X \sim Mon(0.2)$ we simulate 2000 sample vectors $\boldsymbol{X}_{k,i}$ for each k and obtain the optimum intervals ${}^*C_{\mathrm{E}[X]}^{(\beta)}(\overline{X}_{k,i})$ based on the corresponding β-uncertainty space. For the samples of size 50 we consider the intervals ${}^*\widetilde{C}_{\mathrm{E}[X]}^{(\beta)}(\overline{X}_{k',i})$ from β-uncertainty spaces based on smaller nominal sample sizes $k' \in \{50, 49, \ldots, 20\}$. From the simulation data, we estimate the coverage relative to the true value $\mu_X = 0.2$ and the expected interval length. Figure 8.1 displays the estimators. The coverage increases if intervals ${}^*\widetilde{C}_{\mathrm{E}[X]}^{(\beta)}(\overline{X}_{k',i})$ are obtained from a β-uncertainty space of smaller nominal sample size k', i.e., the latter intervals comply with the prescribed level $\beta = 0.975$. The mean length of ${}^*\widetilde{C}_{\mathrm{E}[X]}^{(\beta)}(\overline{X}_{k',i})$ only slightly increases.

8.6. Findings from Simulation Study

The estimators of the actual pointwise coverage and of the expected interval length obtained from the simulation study outlined in the preceding section 8.5 are displayed by figures 8.2 and 8.3. Subsequently we review the results in detail.

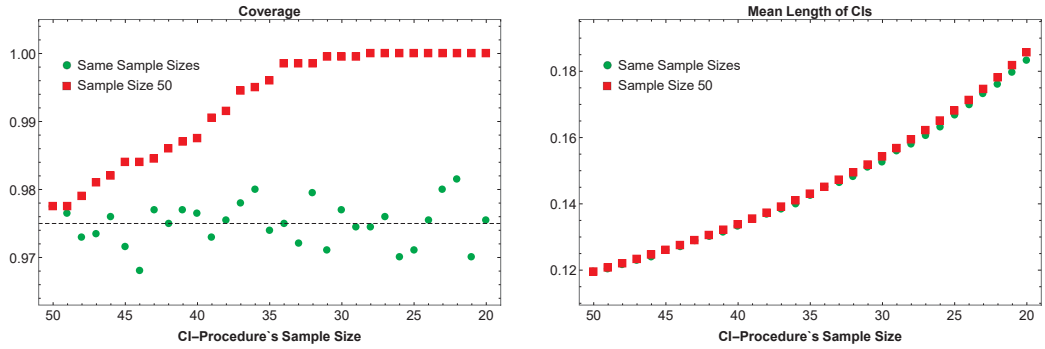


Figure 8.1.: Coverage and expected length of confidence intervals ${}^*C^{(\beta)}_{\mathrm{E}[X]}(\overline{X}_k)$ for samples of size $k \in \{50, 49, \ldots, 20\}$ derived from corresponding β-uncertainty spaces (green), and for intervals ${}^*\widetilde{C}^{(\beta)}_{\mathrm{E}[X]}(\overline{X}_{k'})$ for samples of size 50 derived from β-uncertainty spaces with $\beta = 0.975$ based on smaller nominal sample sizes k' (red).

8.6.1. Actual coverage

In both cases $\mu_X = 0.1$ and $\mu_X = 0.2$ and for all prior information intervals $\mathcal{D}$, all p and all sample sizes n, the mean coverage exceeds the requested confidence level of 0.95 by far. In particular, the combined confidence intervals $C^{\mathrm{de}}_{\mu_Y}$ even more than the two-sided Stringer intervals $C^{st}_{\mu_Y}$. That is, the combined confidence interval seems to be conservative with respect to the coverage probability.

8.6.2. Expected interval length

For both intervals the expected length is decreasing with increasing sample size.

The expected interval length of the decomposition interval

In both cases $\mu_X = 0.1$ and $\mu_X = 0.2$ the (absolute) reduction of the mean length of $C^{\mathrm{de}}_{\mu_Y}$ due to the prior knowledge becomes smaller with increasing sample size. Consider $\mu_X = 0.1$, $p = 0.1$ and $n = 25$: starting from a mean length of 0.11268 for $\mathcal{D}_{Ia} = [0.01, 0.49]$, more prior knowledge leads to 0.082884 for $\mathcal{D}_{Ib} = [0.01, 0.30]$ and 0.058809 for $\mathcal{D}_{Ic} = [0.05, 0.20]$. The latter amounts to a relative reduction of 26.4% and 47.8%, respectively. For $n = 250$ the mean length

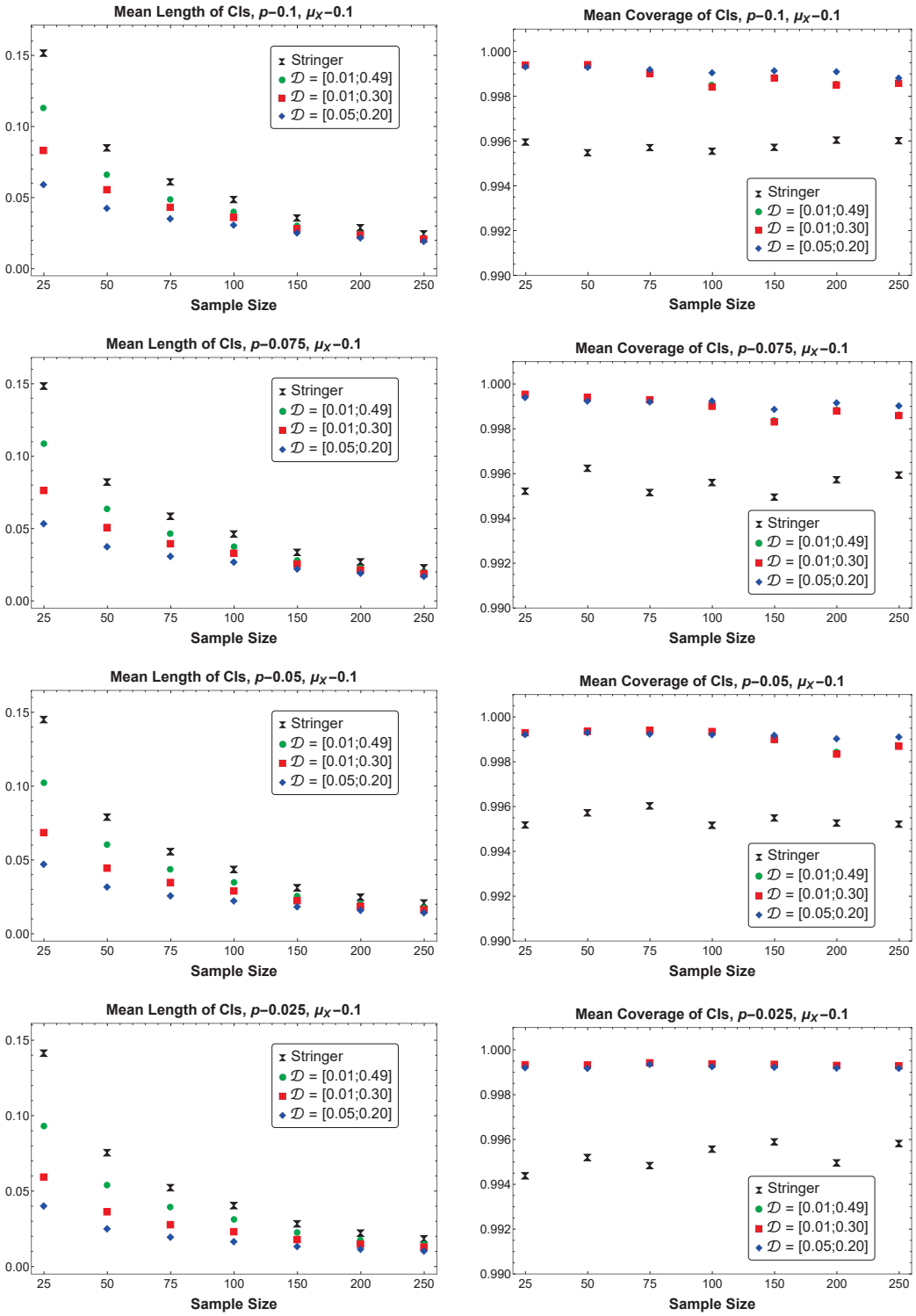


Figure 8.2.: Mean lengths and mean coverages of confidence intervals $C^{\mathrm{de}}_{\mu_Y}$ and $C^{st}_{\mu_Y}$. True values of $\mathrm{E}[Y]$ are $\mu_Y = p \cdot \mu_X = p \cdot 0.1$ with $p \in \{0.1, 0.075, 0.05, 0.025\}$.

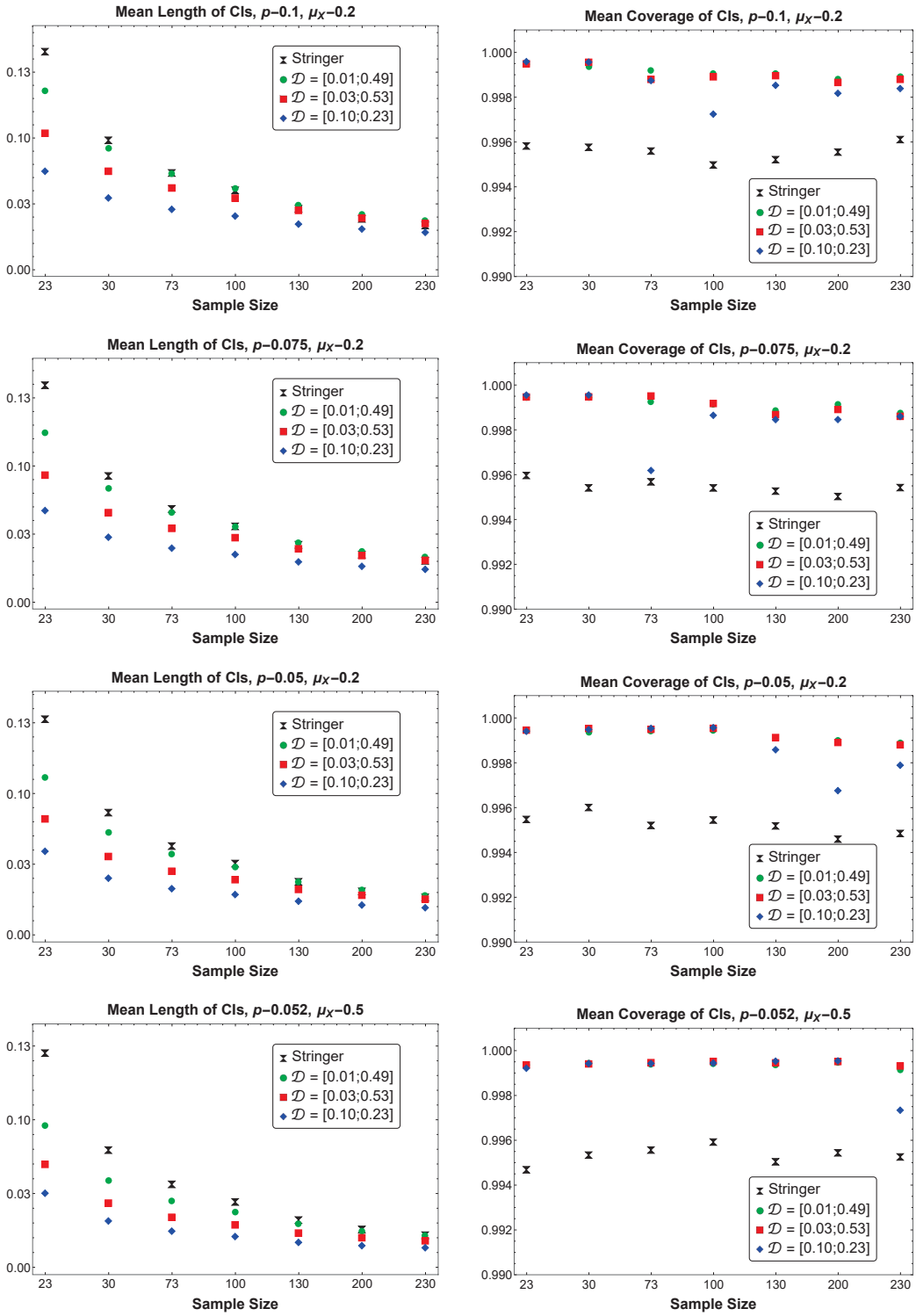


Figure 8.3.: Mean lengths and mean coverages of confidence intervals $C^{\mathrm{de}}_{\mu_Y}$ and $C^{st}_{\mu_Y}$. True values of $\mathrm{E}[Y]$ are $\mu_Y = p \cdot \mu_X = p \cdot 0.2$ with $p \in \{0.1, 0.075, 0.05, 0.025\}$.

for $\mathcal{D}_{Ia}$ is 0.020950 and reduces to 0.020499 (-2.2%) and 0.018939 (-9.6%). Smaller values of p lead to larger relative reductions when more precise prior knowledge is available[1]: for $n = 25$ we have 0.092817, 0.059072 (-36.4%) and 0.39877 (-57.0%), and for $n = 250$ the respective lengths are 0.014683, 0.012684 (-13.6%) and 0.010028 (-31.7%).
For $\mu_X = 0.2$, $p = 0.1$ and $n = 25$ the mean length reduces from 0.13565 for $\mathcal{D}_{IIa} = [0.01, 0.49]$ to 0.10340 (-23.8%) for $\mathcal{D}_{IIb} = [0.05, 0.35]$ and 0.074569 (-45.0%) for $\mathcal{D}_{IIc} = [0.10, 0.25]$. For $n = 250$ the reduction is less severely: from 0.037038 to 0.034909 (-5.7%) and 0.028172 (-23.9%), respectively.

Comparison of the expected interval length of the decomposition interval and the Stringer interval

In the case $\mu_X = 0.1$, the Stringer intervals have a larger expected length for all considered values of p and n and all prior information intervals, even though, the differences diminish with greater n. For example, for $n = 25$ the mean length of $C^{st}_{\mu_Y}$ is 0.15131 which is underrun by 0.11268, i.e., by -25.5%, for the combined confidence interval $C^{\text{de}}_{\mu_Y}$ under the weak prior information $\mathcal{D}_{Ia} = [0.01, 0.49]$. For $n = 250$ the effect for the same prior information is less pronounced, but still yields a reduction from 0.024235 to 0.020950 (-13.6%). More accurate prior information, i.e., narrower $\mathcal{D}$, leads to stronger reductions. The effect of smaller considered values of p with respect to the mean length lies in a general reduction of all combined confidence and two-sided Stringer intervals, respectively. The situation changes for $\mu_X = 0.2$ and depends on the values of both p and n. For $p = 0.1$ the length of $C^{st}_{\mu_Y}$ is greater for small n but underruns the length of $C^{\text{de}}_{\mu_Y}$ from $n = 75$ on for $\mathcal{D}_{IIa} = [0.01, 0.49]$; for $n = 250$ only $C^{\text{de}}_{\mu_Y}$ for $\mathcal{D}_{IIc} = [0.10, 0.25]$ is still smaller than $C^{st}_{\mu_Y}$. Smaller values of p delays this effect to greater sample sizes n, and only for $p = 0.025$ all $C^{\text{de}}_{\mu_Y}$ shows a smaller mean length than $C^{st}_{\mu_Y}$.

Sample size comparison

Under prescribed expected interval length, the decomposition interval enables

[1] This effect is expectable, since a small value of p leads to high probabilities of small numbers of tainting ratios > 0, even for large sample sizes n. Thus, the effect of prior information becomes considerable.

considerable reductions in sample size in comparison with the Stringer interval. Consider $\mu_X = 0.1$ and $p = 0.1$ and $\mathcal{D}_{Ib} = [0.01, 0.30]$. For $n = 25$ we have a mean length of 0.082884 of $C^{\text{de}}_{\mu_Y}$ which is comparable to the length 0.084698 of $C^{st}_{\mu_Y}$ for $n = 50$, that is, the necessary sample size is twice as large. For $p = 0.025$ the effect is more distinct: for the same prior information and $n = 25$ we have 0.059072 of $C^{\text{de}}_{\mu_Y}$, whilst $C^{st}_{\mu_Y}$ adopts a mean length of 0.052030 for $n = 75$, thus, we would need a sample size almost three times as large. Almost the same may be observed for $\mu_X = 0.2$: Consider $p = 0.1$ and $\mathcal{D}_{IIb} = [0.05, 0.35]$ yielding a mean length of 0.10340 of $C^{\text{de}}_{\mu_Y}$ for $n = 25$, whereas the length 0.091919 of $C^{st}_{\mu_Y}$ is achieved for $n = 50$. For more prior information $\mathcal{D}_{IIc} = [0.10, 0.25]$ and $n = 100$ we have 0.040562 of $C^{\text{de}}_{\mu_Y}$, whilst the length 0.041852 of $C^{st}_{\mu_Y}$ would need a sample size $n = 200$. Again, smaller values of p lead to an increase of the described savings in n.

Remark on the validity of the simulation We made two additional simulation runs each of size $N = 500$ as described above. For $p = 0.1$, $\mathcal{D} = [0.01, 0.30]$ table 8.1 displays for each run the mean lengths of $C^{st}_{\mu_Y}$ and of $C^{\text{de}}_{mu_Y}$, respectively. We note, that there are only minor variations, and conclude, that the choosen simulation size $N = 500$ is large enough to make valid statements.

8.7. Sensitivity in the Prior Information Interval

From the construction of the decomposition interval $C^{\text{de}}_{\mu_Y}(\boldsymbol{Y})$ in proposition 8.2 and from the definition (8.11) of the underlying confidence interval ${}^*C^{(\beta)}_{\text{E}[X]}(\overline{X}_k)$ it is clear that under a totally misspecified prior information interval $\mathcal{D}$ with $\mu_X \notin \mathcal{D}$, the coverage at the true value $\mu_Y = p\mu_X$ is zero. In practice, $\mathcal{D}$ should be chosen in a very conservative way. The prior information interval $\mathcal{D} = [0.01, 0.49]$ considered in the simulation study outlined in section 8.5 reflects almost no prior knowledge about μ_X, but still enables a shorter length than the one of the Stringer interval for small to medium sample sizes.

To study the effect of nearly misspecified prior information we consider two exemplary cases, both with $p = 0.1$, where the true value of μ_X equals one of the

	$\lvert C^{st}_{\mu_Y} \rvert$			$\lvert C^{\mathrm{de}}_{\mu_Y} \rvert$		
n	Run 1	Run 2	Run 3	Run 1	Run 2	Run 3
25	0.15131	0.15126	0.15124	0.082884	0.082965	0.082867
50	0.084698	0.084810	0.084661	0.055275	0.055462	0.055221
75	0.060779	0.060936	0.060768	0.042917	0.043109	0.042877
100	0.048365	0.048452	0.048333	0.035797	0.035895	0.035704
150	0.035404	0.035381	0.035331	0.027855	0.027816	0.027713
200	0.028552	0.028517	0.028542	0.023430	0.023369	0.023409
250	0.024235	0.024228	0.024282	0.020499	0.020465	0.020557

Table 8.1.: Mean length of confidence intervals $C^{\mathrm{de}}_{\mu_Y}$ and $C^{st}_{\mu_Y}$ for 3 simulation runs. Considered values are $p = 0.1$, $\mu_X = 0.1$ and $\mathcal{D} = [0.01, 0.30]$.

bounds of of a prior information interval more restrictive than $\mathcal{D} = [0.01, 0.49]$. Figure 8.4 displays the case $\mu_X = 0.2$ and $\mathcal{D} = [0.05, 0.20]$. The mean length of $C^{\mathrm{de}}_{\mu_Y}$ is always smaller than that of $C^{st}_{\mu_Y}$, e. g., for $n = 25$ it is 0.059891 compared to 0.16531 – less than half of the two-sided Stringer interval. For $n = 75$ the lengths are 0.036596 and 0.073264, again less than half. The requested confidence level of 0.95 is also met by $C^{\mathrm{de}}_{\mu_Y}$.

Similar observations are made for $\mu_X = 0.1$ and $\mathcal{D} = [0.10, 0.25]$: the two-sided Stringer intervals $C^{st}_{\mu_Y}$ are larger than $C^{\mathrm{de}}_{\mu_Y}$ for all sample sizes n, and $C^{\mathrm{de}}_{\mu_Y}$ meets the requested confidence level 0.95 (see figure 8.5).

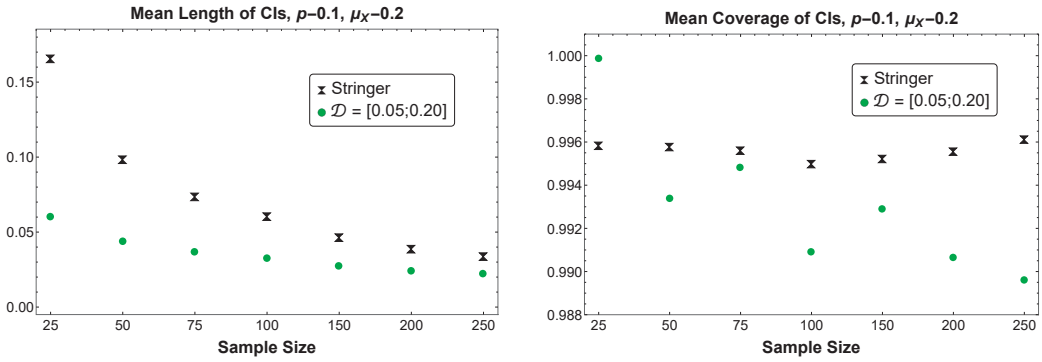


Figure 8.4.: Mean lengths and mean coverages of confidence intervals $C^{\mathrm{de}}_{\mu_Y}$ and $C^{st}_{\mu_Y}$. Considered values are $p = 0.1$ and $\mu_X = 0.2$. The prior information interval is $\mathcal{D} = [0.05, 0.20]$. That is, the true value $\mu_X = 0.2$ lies on the right edge of $\mathcal{D}$.

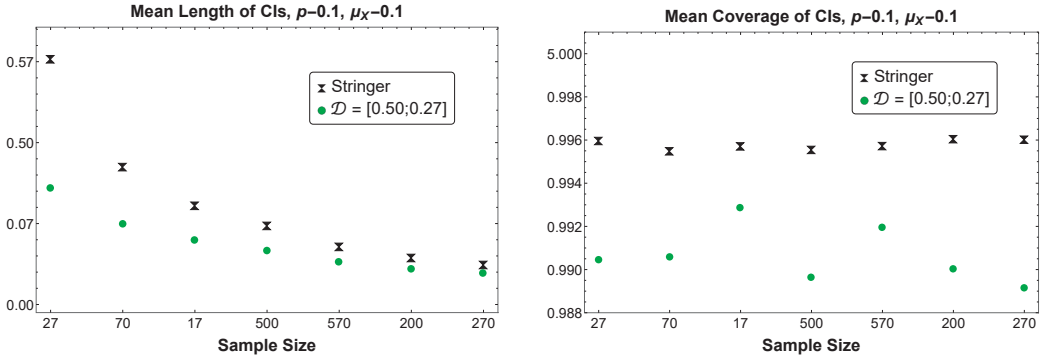


Figure 8.5.: Mean lengths and mean coverages of confidence intervals $C^{\mathrm{de}}_{\mu_Y}$ and $C^{st}_{\mu_Y}$. Considered values are $p = 0.1$ and $\mu_X = 0.1$. The prior information interval is $\mathcal{D} = [0.10, 0.25]$. That is, the true value $\mu_X = 0.1$ lies on the left edge of $\mathcal{D}$.

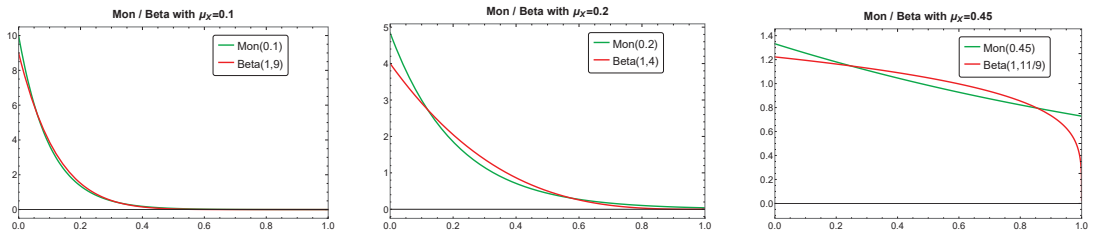


Figure 8.6.: Simulation distributions for tainting ratios $Y > 0$. Beta distribution $Beta(1, b)$ compared to $Mon(\mu_X)$, where b is choosen in the way, that the beta distribution has the same expectation $\mu_X \in \{0.1, 0.2, 0.45\}$.

8.8. Sensitivity in the Class of Prior Information Distributions

How does the decomposition interval compare to the two-sided Stringer intervals if the underlying PDF of $X = Y|Y > 0$ is still decreasing on $(0, 1]$, but not from the class $Mon(\mu_X)$?

We consider a beta distribution $Beta(a, b)$ with $a = 1$, i.e., the PDF $f_{1,b}(x)$:

$$f_{1,b}(x) = \begin{cases} \frac{(1-x)^{b-1}}{B(1,b)} & \text{for } x \in [0;1] \\ 0 & \text{else} \end{cases} \tag{8.16}$$

The parameter b is defined by the considered value of the expectation μ_X, that is, $\mu_X = \frac{1}{1+b} \Leftrightarrow b = \frac{1}{\mu_X} - 1$. Figure 8.6 displays three beta distributions in comparison to a monotonic distribution with the same expectation μ_X.

Each simulation for $\mu_X \in \{0.10, 0.20, 0.45\}$ has a size of $N = 2000$. In each case we consider the prior information $\mathcal{D} = [0.01, 0.49]$, i.e., the least amount of prior information which reflects the monotone decreasing behaviour. Finally, for the calculations of the mean lengths and coverages we consider $p = 0.1$.

Figure 8.7 displays the respective results. In all cases, the coverage excels the requested confidence level of 0.95. For $\mu_X = 0.1$ the mean lengths of the two-sided Stringer intervals $C^{st}_{\mu_Y}$ are larger than those of $C^{\text{de}}_{\mu_Y}$: for $n = 25$ we have 0.15113 compared to 0.11369 (-24.8%), while the difference reduces to 0.023954 compared to 0.020975 (-12.4%) for $n = 250$. For $\mu_X = 0.2$ the situation changes

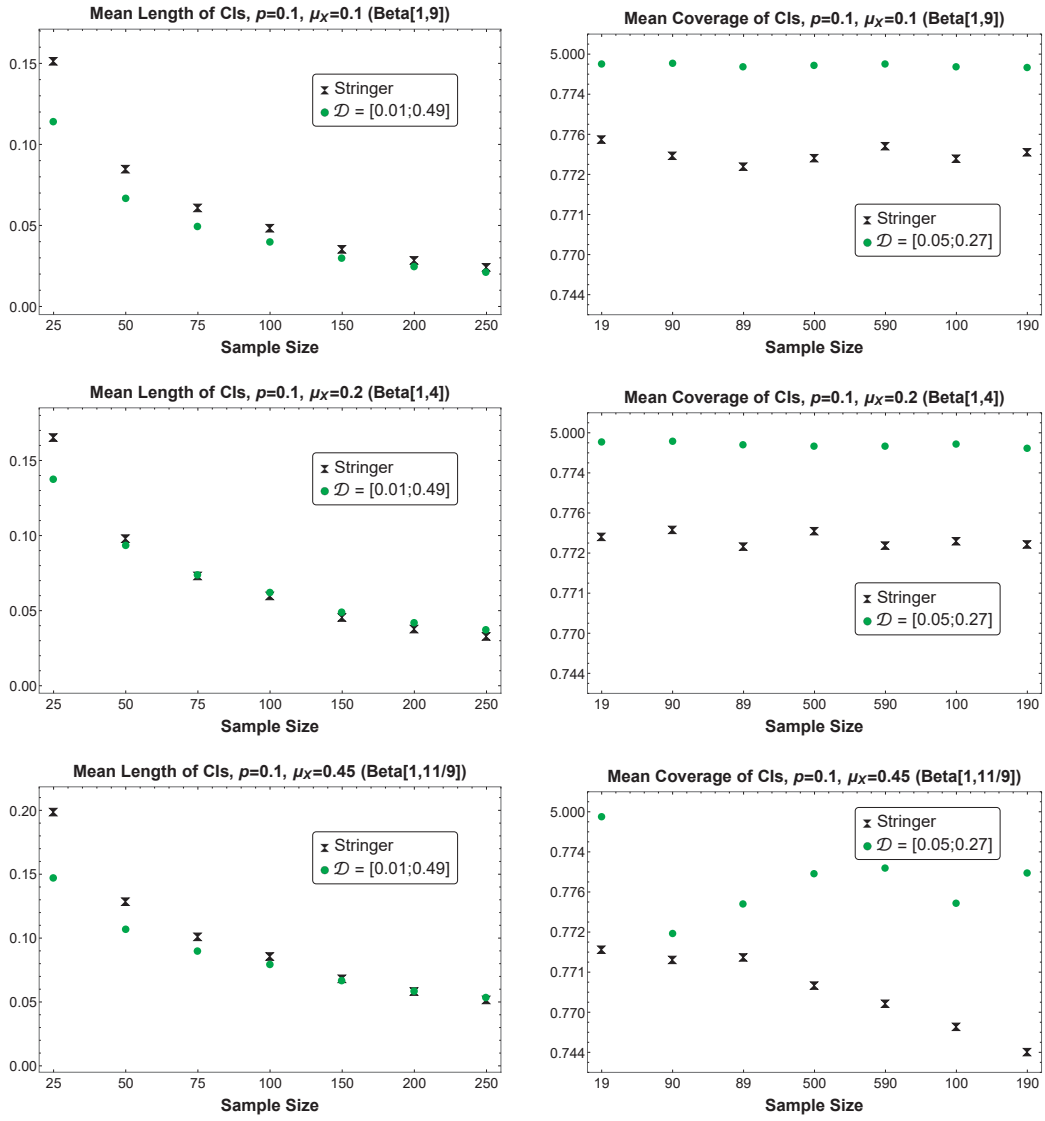


Figure 8.7.: Mean lengths and mean coverages of confidence intervals C_{μ_Y} and $C^{st}_{\mu_Y}$. Simulation Distributions are Beta distributions with expected values $\mu_X = 0.1, 0.2, 0.45$.

	$C^{st}_{\mu_Y}$		$C^{\text{de}}_{\mu_Y}$		
n		$\mathcal{D}$	$[0.01, 0.49]$	$[0.01, 0.30]$	$[0.05, 0.20]$
25	$[0, 0.137185]$		$[0, 0.078788]$	$[0, 0.048234]$	$[0, 0.032156]$
50	$[0, 0.071122]$		$[0, 0.041116]$	$[0, 0.025173]$	$[0, 0.016782]$

Table 8.2.: Two-sided Stringer intervals $C^{st}_{\mu_Y}$ and decomposition confidence intervals $C^{\text{de}}_{\mu_Y}$ when no overstatement is observed, nominal confidence level $\gamma = 0.95$.

in the way, that for $n = 25$ $C^{\text{de}}_{\mu_Y}$ is smaller than $C^{st}_{\mu_Y}$, but for $n = 50, 75, 100$ we have comparable lengths, and for greater n $C^{\text{de}}_{\mu_Y}$ is longer than $C^{st}_{\mu_Y}$. The maximum difference is reached for $n = 250$ with 0.032632 compared to 0.037157 (+13.9%). For $\mu_X = 0.45$ the situation is as follows: $C^{\text{de}}_{\mu_Y}$ is shorter than $C^{st}_{\mu_Y}$ from $n = 25$ to $n = 100$, where the difference reduces from -25.9% to -7.5%. For sample sizes $n = 150, 200, 250$ we observe comparable lengths for both $C^{\text{de}}_{\mu_Y}$ and $C^{st}_{\mu_Y}$.

8.9. Illustrative examples

In the previous sections, we have compared the coverage and the expected lengths of two-sided Stringer intervals $C^{st}_{\mu_Y}$ and decomposition confidence intervals $C^{\text{de}}_{\mu_Y}$, respectively. Here, we present some individual confidence intervals for some realisations of $\boldsymbol{Y}$ with sample sizes $n = 25, 50$.

The special case of no item having overstatement in the sample yields for the two-sided Stringer intervals $C^{st}_{\mu_Y}$ always the same interval, which is the Clopper-Pearson confidence interval for p. The same is true for the decomposition confidence intervals $C^{\text{de}}_{\mu_Y}$, but it combines the slightly longer Clopper-Pearson CI for p (due to a larger confidence level, see 8.2) with the actual prior information interval $\mathcal{D}$. Table 8.2 shows the respective intervals.

Consider the probability for overstatement $p = 0.1$, then for a sample size of 25

and 50 we would expect a mean number of 2.5 and 5 overstatements, respectively. Tables 8.3 and 8.4 show some individual samples and resulting CIs.

$C^{st}_{\mu_Y}$	$C^{\text{de}}_{\mu_Y}$		
$\mathcal{D}$	$[0.01, 0.49]$	$[0.01, 0.30]$	$[0.05, 0.20]$
$k = 2$: $\boldsymbol{y} = \{0.058345, 0.135846, 0, \ldots, 0\}$			
0.000653	0.000237	0.000237	0.000340
0.149509	0.135764	0.086260	0.057507
$k = 3$: $\boldsymbol{y} = \{0.008413, 0.138172, 0.169193, 0, \ldots, 0\}$			
0.001522	0.000855	0.000855	0.000979
0.156691	0.130818	0.102061	0.068041
$k = 4$: $\boldsymbol{y} = \{0.113103, 0.170762, 0.319875, 0.930446, 0, \ldots, 0\}$			
0.008686	0.006395	0.005715	0.004921
0.231430	0.190738	0.116779	0.077852

Table 8.3.: $n = 25$: Two-sided Stringer intervals $C^{st}_{\mu_Y}$ and decomposition confidence intervals $C^{\text{de}}_{\mu_Y}$ when k overstatements are observed, nominal confidence level $\gamma = 0.95$.

8.10. Conclusion and Outlook

We have introduced a decomposition model for the tainting ratio, which enabled us to model the conditional distribution under overstatement seperatly. This is done in a rather simple, but nevertheless highly intuitive way by the monotonic distribution $Mon(\mu_X)$ which is solely based on its decreasing nature, the support and the conditional mean. Using an interval for the prior information about the conditional mean is also nearby. This prior information is incorporated into a procedure yielding confidence intervals for the conditional mean. The simulation

$C^{st}_{\mu_Y}$	$C^{\text{de}}_{\mu_Y}$		
$\mathcal{D}$	$[0.01, 0.49]$	$[0.01, 0.30]$	$[0.05, 0.20]$
$k = 4$: $\boldsymbol{y} = \{0.113103, 0.170762, 0.319875, 0.930446, 0, \ldots, 0\}$			
0.004275	0.003125	0.002793	0.002405
0.121699	0.102727	0.062894	0.041930
$k = 5$: $\boldsymbol{y} = \{0.030344, 0.067421, 0.098682, 0.290637, 0.318480, 0 \ldots, 0\}$			
0.003177	0.002133	0.002122	0.002188
0.096683	0.091229	0.070810	0.047207
$k = 6$: $\boldsymbol{y} = \{0.028034, 0.086625, 0.104616, 0.201071, 0.227915, 0.467075, 0 \ldots, 0\}$			
0.005083	0.003597	0.003572	0.003707
0.106065	0.103497	0.078448	0.052299

Table 8.4.: $n = 50$: Two-sided Stringer intervals $C^{st}_{\mu_Y}$ and decomposition confidence intervals $C^{\text{de}}_{\mu_Y}$ when k overstatements are observed, nominal confidence level $\gamma = 0.95$.

study illustrated, that both, the proposed composed confidence interval and the two-sided interval of Stringer type meet the nominal confidence level, whereas the composed confidence interval is superior to the one of Stringer type with respect to the expected length – particularly for small sample sizes. If the prior information interval for the conditional mean is nearly misspecified, the nominal level is still met, and the composed interval is also shorter. Even if the true conditional distribution is a beta distribution (which shares same support, monotonicity, and mean of the considered monotonic distribution), and under weak prior information, the composed confidence interval meets the required confidence level, and yields intervals comparable to those of Stringer type.

Expansions to the here proposed approach could incoporate prior information about the proportion of overstatement, like proposed by Lurz (2015) [49] who

generalised an approach of von Collani & Dräger (2001) [15]. Additionally, the prior information about the conditional mean may also be considered as a distribution over a certain range of possible values and incorporated into the confidence interval procedure. Some motivation to, and interpretations of the latter can be found in Weigand (2009) [77].

Besides these expansions, further studies could be conducted with respect to the sensitivity in the class of prior information distributions.

9. Comparing Neyman Measurement Procedure with Bayes Measurement Procedure for $\mathrm{E}[X]$ for $X \sim EXP(1/\mu_1)$

The Neyman measurement procedure seems to have some similarities with the Bayes procedures, that are

- pre-available knowledge about the parameter to be measured is taken into account by a set of possible values $\mathcal{D}$,
- functional terms involved in the calculations looks identical at first sight.

This chapter is organised in the way, that we firstly sum up the Neyman measurement procedure for $\mathrm{E}[X]$ for $X \sim EXP(1/\mu_1)$ together with an example. Second, we derive the respective Bayes measurement procedure and, at last, work out the parallels and compare both procedures. Not surprisingly, we will realise, that both procedures are based on totally different (philosophical) principles and, therefore, are in fact not comparable.

To exemplify the respective procedures we choose throughout this chapter the following settings:

- $X|\{\mu_1\} \sim EXP(1/\mu_1)$
- $\mu_1 \in \mathcal{D} = \{m \,|\, 0.05 \leq m \leq 0.25\}$
- $\boldsymbol{X}|\{\mu_1\} = (X_1|\{\mu_1\}, X_2|\{\mu_1\}, \ldots, X_8|\{\mu_1\})$ be an i.i.d. sample for $X|\{\mu_1\}$ of size $n = 8$
- reliability $\beta = 0.9$

9.1. Recap of Neyman Measurement Procedure for $\mathrm{E}[X]$ for $X \sim EXP(1/\mu_1)$

The pre-available knowledge about $\mathrm{E}[X]$ is represented by $\mathcal{D} = \{\mu_1 \,|\, 0 < m_\ell \leq \mu_1 \leq m_u\}$. Its interpretation is, that all other values μ_1 except those in $\mathcal{D}$ can be excluded from consideration. Since no value in $\mathcal{D}$ is favoured over another, they are all equitable. This is formulated by the measure on $\mathcal{D}$ given by its density[1]

$$\lambda_{\mathcal{D}}(\mu_1) = \frac{\mathbb{1}_{[m_\ell, m_u]}(\mu_1)}{m_u - m_\ell} \,. \tag{9.1}$$

The Neyman measurement procedure ${}^*C^{(\beta)}_{\mathrm{E}[X]}$ for $\mathrm{E}[X]$ was already derived in section 7.2.2.1. In general, the aim is to get a β-measurement procedure ${}^*C^{(\beta)}_{\mathrm{E}[X]}$ for $\mathrm{E}[X]$ with minimal weighted volume V_w, calculated by

$$V_w({}^*C^{(\beta)}_{\mathrm{E}[X]}) = \int\limits_{\mathcal{X}(\mathcal{D})} \left(\frac{1}{|\mathcal{D}|} \int\limits_{{}^*C^{(\beta)}_{\mathrm{E}[X]}(\{\boldsymbol{x}\})} \mathrm{d}\mu_1 \right) \frac{\int_{\mathcal{D}} f_{\boldsymbol{X}|\{\mu_1\}}(\boldsymbol{x}) \,\mathrm{d}\mu_1}{|\mathcal{D}|} \,\mathrm{d}\boldsymbol{x} \tag{9.2}$$

$$= \int\limits_{\mathcal{D}} \frac{1}{|\mathcal{D}|} \left(\int\limits_{{}^*A^{(\beta)}_{\boldsymbol{X}}(\{\mu_1\})} \frac{w(\boldsymbol{x})}{|\mathcal{D}|} \,\mathrm{d}\boldsymbol{x} \right) \mathrm{d}\mu_1 \,. \tag{9.3}$$

Then, the optimal prediction sets ${}^*A^{(\beta)}_{\boldsymbol{X}}(\{\mu_1\})$ have to be constructed according to

$${}^*A^{(\beta)}_{\boldsymbol{X}}(\{\mu_1\}) = \left\{ \boldsymbol{x} \;\middle|\; \frac{f_{\boldsymbol{X}|\{\mu_1\}}(\boldsymbol{x})}{\int_{\mathcal{D}} f_{\boldsymbol{X}|\{\tilde{\mu}_1\}}(\boldsymbol{x}) \,\mathrm{d}\tilde{\mu}_1} \geq r^*(\mu_1) \right\} \tag{9.4}$$

with $r^*(\mu_1)$ defined by

$$\beta = \int\limits_{\left\{\boldsymbol{x} \,\middle|\, \frac{f_{\boldsymbol{X}|\{\mu_1\}}(\boldsymbol{x})}{\int_{\mathcal{D}} f_{\boldsymbol{X}|\{\tilde{\mu}_1\}}(\boldsymbol{x}) \,\mathrm{d}\tilde{\mu}_1} \geq r^*(\mu_1)\right\}} f_{\boldsymbol{X}|\{\mu_1\}}(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \,. \tag{9.5}$$

[1] "Density" here should be understood in terms of measurement theory.

We have also seen, that the ratio in $\boldsymbol{x}$ can be transformed into one in $\overline{x}$:

$$\frac{f_{\boldsymbol{X}|\{\mu_1\}}(\boldsymbol{x})}{w_{\boldsymbol{X}}(\boldsymbol{x})} = \frac{f_{\boldsymbol{X}|\{\mu_1\}}(\boldsymbol{x})}{\int\limits_{m_\ell}^{m_u} f_{\boldsymbol{X}|\{\tilde{\mu}_1\}}(\boldsymbol{x})\,\mathrm{d}\tilde{\mu}_1} \tag{9.6}$$

$$= \frac{\left(\frac{1}{\mu_1}\right)^n e^{-\sum_{i=1}^n x_i/\mu_1}\mathbb{1}_{[0,+\infty)^n}(\boldsymbol{x})}{\int\limits_{m_\ell}^{m_u}\left(\frac{1}{\tilde{\mu}_1}\right)^n e^{-\sum_{i=1}^n x_i/\tilde{\mu}_1}\mathbb{1}_{[0,+\infty)^n}(\boldsymbol{y})\,\mathrm{d}\tilde{\mu}_1} \tag{9.7}$$

$$= \frac{n\frac{(n\overline{x})^{n-1}}{\Gamma(n)}\left(\frac{1}{\mu_1}\right)^n e^{-n\overline{x}/\mu_1}\mathbb{1}_{[0,+\infty)}(\overline{x})}{\int\limits_{m_\ell}^{m_u} n\frac{(n\overline{x})^{n-1}}{\Gamma(n)}\left(\frac{1}{\tilde{\mu}_1}\right)^n e^{-n\overline{x}/\tilde{\mu}_1}\mathbb{1}_{[0,+\infty)}(\overline{x})\,\mathrm{d}\tilde{\mu}_1} \tag{9.8}$$

$$= \frac{f_{\overline{X}|\{\mu_1\}}(\overline{x})}{\int\limits_{m_\ell}^{m_u} f_{\overline{X}|\{\tilde{\mu}_1\}}(\overline{x})\,\mathrm{d}\tilde{\mu}_1} \tag{9.9}$$

$$= \frac{f_{\overline{X}|\{\mu_1\}}(\overline{x})}{w_{\overline{X}}(\overline{x})} \tag{9.10}$$

That is, we may base the Neyman measurement procedure ${}^*C^{(\beta)}_{\mathrm{E}[X]}$ for $\mathrm{E}[X]$ on prediction sets for $\overline{X}|\{\mu_1\} = \frac{1}{n}\sum_{i=1}^n X_i|\{\mu_1\}$. As an side-effect this yields the possibility of graphical illustrations.

Example 9.1 *Under the given numerical settings we calculate the prediction sets* ${}^*A^{(0.9)}_{\overline{X}}(\{\mu_1\})$, $\mu_1 \in \mathcal{D}$, *and, thus, the* 0.9*-uncertainty space* $\mathcal{U}^{(0.9)}_{\mathrm{E}[X]}$ *for* $\mathrm{E}[X]$. *By construction, the weighted volume is minimal and has a value of* $V_w({}^*C^{(0.9)}_{\mathrm{E}[X]}) = 0.549124$. *Figure 9.1 displays* $\mathcal{U}^{(0.9)}_{\mathrm{E}[X]}$. *Again, we realise how advantageous it is to scale the graphic with the marginal distribution of* $\overline{X}$*: in this way the part of* $\mathcal{X} = [0, +\infty)$ *where it is most likely to observe a realisation of* $\overline{X}$ *is dilated while the other parts which are unlikely are shrinked.*

Since we aim at the comparison with a Bayes measurement procedure, it is noteworthy to recall the interpretation of the result of a classical measurement procedure in general and a Neyman measurement procedure in particular. A measurement procedure C for the unknown but fixed value d of a parameter D meeting the reliability βrepresents a random set $C(\boldsymbol{X})$ which covers d with a probability of at least β, i.e., $\mathrm{P}_{\boldsymbol{X}|\{d\}}(d \in C(\boldsymbol{X})) \geq \beta$ for all values $d \in \mathcal{D}$.

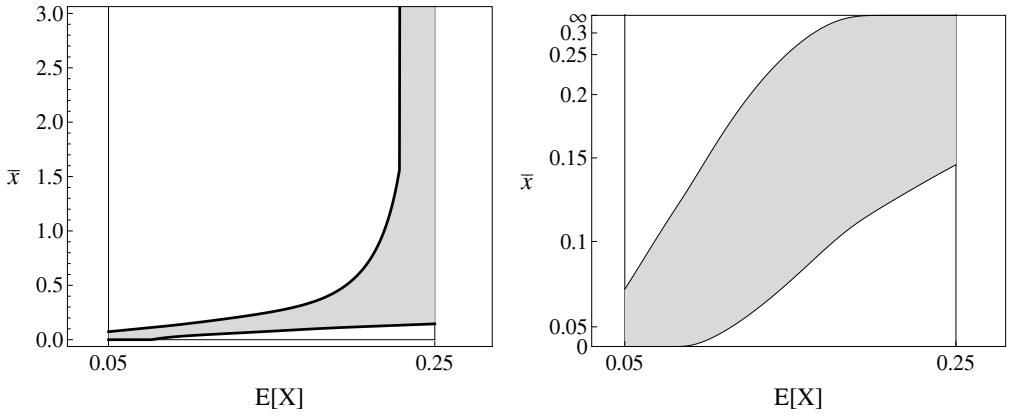


Figure 9.1.: Optimal 0.9-uncertainty space $\mathcal{U}^{(0.9)}_{\mathrm{E}[X]}$ for $\mathrm{E}[X]$ in the unscaled version on the left and scaled with the marginal distribution of $\overline{X}$ on the right. Note that the unscaled uncertainty space is in fact unbounded and only cut at $\overline{x} = 3$ while the scaled version really covers the whole range of variability $\mathcal{X} = [0, +\infty)$.

9.2. Bayes Measurement Procedure for $\mathrm{E}[X]$ for $X \sim EXP(1/\mu_1)$

At first, we want shortly describe in general the Bayesian measurement procedure, denoted $B^{(\beta)}_D$, for a parameter D with values $d \in \mathcal{D} \subseteq \mathbb{R}^k$. Therefore, let $\boldsymbol{X}|\{d\}$ be a random variable with density function $f_{\boldsymbol{X}|\{d\}}(\boldsymbol{x})$. d is assumed to be a value of the random variable D with some prior distribution $\pi_D(d)$ on $\mathcal{D}$. Given a realisation $\boldsymbol{x} \in \mathbb{R}^n$ of $\boldsymbol{X}|\{d\}$, the posterior distribution $\pi_D(d|\boldsymbol{x})$ is given by

$$\pi_D(d|\boldsymbol{x}) = \frac{f_{\boldsymbol{X}|\{d\}}(\boldsymbol{x})\pi_D(d)}{\int_{\mathcal{D}} f_{\boldsymbol{X}|\{\tilde{d}\}}(\boldsymbol{x})\pi_D(\tilde{d})\,\mathrm{d}\tilde{d}}\,. \tag{9.11}$$

The posterior distribution $\pi_D(d|\boldsymbol{x})$ at hand, a credible set for D is then defined as any set $B \subseteq \mathcal{D}$ with a posterior probability (also called credible probability or (rarely) credibility) given by

$$\mathrm{P}_D(D \in B|\boldsymbol{x}) = \int_B \pi_D(d|\boldsymbol{x})\,\mathrm{d}d\,. \tag{9.12}$$

For a prespecified posterior probability level β, β-credible sets $B_D^{(\beta)}(\boldsymbol{x})$ are often constructed as highest posterior density sets[2], i.e.,

$$B_D^{(\beta)}(\boldsymbol{x}) = \{d \in \mathcal{D} \,|\, \pi_D(d|\boldsymbol{x}) \geq p^*(\boldsymbol{x})\}\,, \tag{9.13}$$

where $p^*(\boldsymbol{x})$ is defined by

$$\beta = \int\limits_{\{d|\pi_D(d\,|\,\boldsymbol{x}) \geq p^*(\boldsymbol{x})\}} \pi_D(d|\boldsymbol{x})\;\mathrm{d}d\,. \tag{9.14}$$

The interpretation of β-credible sets is quite different to the one of confidence sets: Under the given knowledge represented by the prior distribution of D on $\mathcal{D}$ and after the observation $\boldsymbol{x}$, a posterior probability of β is the probability that the parameter D, which is still a random variable, is element of the β-credible set $B_D^{(\beta)}(\boldsymbol{x})$.

Now, let $\boldsymbol{X}|\{\mu_1\}$ be an i.i.d. sample for $X|\{\mu_1\} \sim EXP(1/\mu_1)$, i.e., the density function is

$$f_{\boldsymbol{X}|\{\mu_1\}}(\boldsymbol{x}) = \left(\frac{1}{\mu_1}\right)^n \exp(-\sum x_i/\mu_1)\mathbb{1}_{[0,+\infty)^n}(\boldsymbol{x})\,. \tag{9.15}$$

Since our aim is to compare the Neyman procedure with the Bayes one, it is reasonable to choose as prior distribution $\pi_{\mathrm{E}[X]}(\mu_1)$ the uniform distribution on $\mathcal{D} = \{\mu_1 \,|\, 0 < m_\ell \leq \mu_1 \leq m_u\}$, i.e.,

$$\pi_{\mathrm{E}[X]}(\mu_1) = \frac{\mathbb{1}_{[m_\ell, m_u]}(\mu_1)}{m_u - m_\ell}\,. \tag{9.16}$$

Given the realisation $\boldsymbol{x}$, the posterior distribution $\pi_{\mathrm{E}[X]}(\mu_1|\boldsymbol{x})$ is as follows:

$$\pi_{\mathrm{E}[X]}(\mu_1|\boldsymbol{x}) = \frac{\left(\frac{1}{\mu_1}\right)^n \exp(-\sum x_i/\mu_1)\mathbb{1}_{[m_\ell,m_u]}(\mu_1)}{\int\limits_{m_\ell}^{m_u} \left(\frac{1}{\tilde{\mu}_1}\right)^n \exp(-\sum x_i/\tilde{\mu}_1)\;\mathrm{d}\tilde{\mu}_1}\,. \tag{9.17}$$

Obviously, the dependence of $\pi_{\mathrm{E}[X]}(\mu_1|\boldsymbol{x})$ on $\boldsymbol{x}$ is only via $\sum x_i = n\overline{x}$ and, thus, we may base the Bayes measurement procedure on $\overline{X}|\{\mu_1\}$. In fact, we may have directly started with $\overline{X}|\{\mu_1\}$ and its distribution and eventually would got the same result[3], i.e., $\pi_{\mathrm{E}[X]}(\mu_1|\boldsymbol{x}) = \pi_{\mathrm{E}[X]}(\mu_1|\overline{x})$.

[2] In the case that D is univariate and $\pi_D(d|\boldsymbol{x})$ is unimodal or strictly in- or decreasing, a β-credible set (9.13) results to be an interval.

[3] Analogue to the equality $f_{\boldsymbol{X}|\{\mu_1\}}(\boldsymbol{x})/w_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\overline{X}|\{\mu_1\}}(\overline{x})/w_{\overline{X}}(\overline{x})$

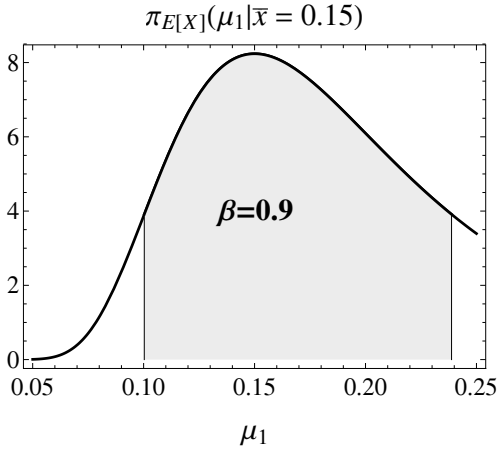


Figure 9.2.: Posterior distribution $\pi_{\mathrm{E}[X]}(\mu_1|\overline{x})$ for $\overline{x} = 0.15$ and resulting 0.9-credible set $B^{(0.9)}_{\mathrm{E}[X]}(0.15)$.

To analyse the course of $\pi_{\mathrm{E}[X]}(\mu_1|\overline{x})$ we only have to look onto the nominator, since the denominator is independent of μ_1. That is, we take the derivative of $\left(\frac{1}{\mu_1}\right)^n \exp(-n\overline{x}/\mu_1)$ with respect to μ_1 and get

$$\frac{\mathrm{d}}{\mathrm{d}\mu_1}\left(\frac{1}{\mu_1}\right)^n \exp(-n\overline{x}/\mu_1) = n\left(\frac{1}{\mu_1}\right)^{n+1}\left(\frac{\overline{x}}{\mu_1} - 1\right)\exp(-n\overline{x}/\mu_1)\,. \tag{9.18}$$

Consequently, we get for $\mu_1 \in \mathcal{D}$

- $\pi_{\mathrm{E}[X]}(\mu_1|\overline{x})$ is strictly decreasing in μ_1 if $\overline{x} \leq m_\ell$,
- $\pi_{\mathrm{E}[X]}(\mu_1|\overline{x})$ is unimodal with maximum at $\mu_1 = \overline{x}$ if $m_\ell < \overline{x} < m_u$, and
- $\pi_{\mathrm{E}[X]}(\mu_1|\overline{x})$ is strictly increasing in μ_1 if $\overline{x} \geq m_u$.

Example 9.2 *With the numerical assumptions and* $\overline{x} = 0.15$ *the posterior distribution* $\pi_{\mathrm{E}[X]}(\mu_1|0.15)$ *is unimodal. Calculating the* 0.9*-credible sets* $B^{(0.9)}_{\mathrm{E}[X]}(0.15)$ *according to 9.13 yields*

$$B^{(0.9)}_{\mathrm{E}[X]}(0.15) = [0.1003, 0.2388]\,. \tag{9.19}$$

Figure 9.2 illustrates the procedure.

Finally, we define a "Bayes β-Uncertainty Space" $\mathcal{B}^{(\beta)}_D$ by

$$\mathcal{B}^{(\beta)}_D := \bigcup_{\boldsymbol{x}\in\mathcal{X}} B^{(\beta)}_D(\boldsymbol{x}) \times \{\boldsymbol{x}\}\,. \tag{9.20}$$

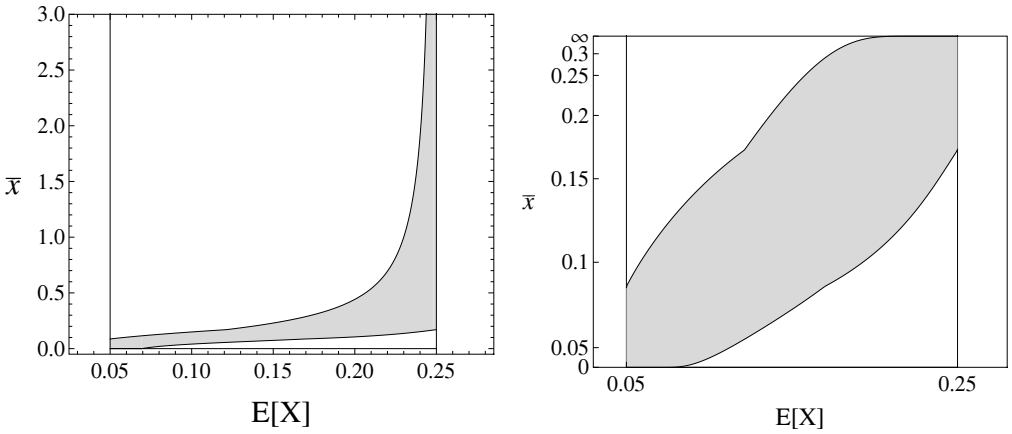


Figure 9.3.: Bayes 0.9-uncertainty space $\mathcal{B}^{(0.9)}_{\mathrm{E}[X]}$ for $\mathrm{E}[X]$ in the unscaled version on the left and scaled with the marginal distribution of $\overline{X}$ on the right.

For the exponential distribution we then have $\mathcal{B}^{(\beta)}_{\mathrm{E}[X]} = \bigcup_{\overline{x}\geq 0} B^{(\beta)}_{\mathrm{E}[X]}(\overline{x}) \times \{\overline{x}\}$.

Example 9.3 *In figure 9.3 the Bayes* 0.9*-uncertainty space* $\mathcal{B}^{(0.9)}_{\mathrm{E}[X]}$ *for* $\mathrm{E}[X]$ *is displayed. Since it is possible, we also use the scaling with the marginal distribution of* $\overline{X}$ *– which proofs again to be benefiting for the perception. Table 9.1 displays some* 0.9*-credible sets* $B^{(0.9)}_{\mathrm{E}[X]}(\overline{x})$ *for some values* $\overline{x}$*. E. g. with the assumed knowledge represented by the uniform prior distribution* $\pi_{\mathrm{E}[X]}(\mu_1)$ *on* $\mathcal{D} = \{\mu_1 \,|\, 0.05 \leq \mu_1 \leq 0.25\}$ *and an observed value* $\overline{x} = 0.2$*, the value of* $\mathrm{E}[X]$ *is element of the interval* $[0.1368, 0.25]$ *with a posterior probability of* 0.9*.*

9.3. Comparing Neyman and Bayes Measurement Procedure

If possible, a comparison of two measurement procedures should include the respective (mean) length of resulting measurements. For some values $\overline{x}$ this is done in table 9.1. For the majority of the given values $\overline{x}$ the Bayes measurements are shorter than the Neyman measurements. But since we have realised in chapter 7 that the comparison of the pure length for some values $\overline{x}$ is misleading, we have to take the marginal distribution of $\overline{X}$ into account. That consideration

$\overline{x}$	$B^{(0.9)}_{\mathrm{E}[X]}(\overline{x})$	${}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{\overline{x}\})$	shorter
0.025	$[0.05, 0.08408]$	$[0.05, 0.08915]$	B
0.075	$[0.05, 0.1518]$	$[0.05097, 0.1460]$	N
0.1	$[0.05834, 0.1928]$	$[0.06796, 0.1762]$	N
0.15	$[0.1003, 0.2388]$	$[0.09980, 0.25]$	B
0.175	$[0.1239, 0.25]$	$[0.1132, 0.25]$	B
0.2	$[0.1368, 0.25]$	$[0.1258, 0.25]$	B
1.5	$[0.2371, 0.25]$	$[0.2283, 0.25]$	B

Table 9.1.: 0.9-credible sets $B^{(0.9)}_{\mathrm{E}[X]}(\overline{x})$ (Bayes measurements) and 0.9-Neyman measurement sets ${}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{\overline{x}\})$ for some values $\overline{x}$. The last column indicates which measurement set (interval) is shorter.

eventually led to the weighted volume V_w and, as a graphical representation, to the scaled β-uncertainty space. Defining Bayes β-prediction sets as the inversion of the β-credible sets

$$ {}^{B}A^{(\beta)}_{\boldsymbol{X}}(\{d\}) := \{\boldsymbol{x} \,|\, d \in B^{(\beta)}_{D}(\boldsymbol{x})\} \tag{9.21} $$

the weighted volume of the Bayes measurement procedure $B^{(\beta)}_{D}$ given an uniform prior on $\mathcal{D}$ then is

$$ V_w(B^{(\beta)}_{D}) = \int\limits_{\mathcal{X}(\mathcal{D})} \left(\frac{1}{|\mathcal{D}|} \int\limits_{B^{(\beta)}_{D}(\{\boldsymbol{x}\})} \mathrm{d}d \right) \frac{\int_{\mathcal{D}} f_{\boldsymbol{X}|\{d\}}(\boldsymbol{x})\, \mathrm{d}d}{|\mathcal{D}|}\, \mathrm{d}\boldsymbol{x} \tag{9.22} $$

$$ = \int\limits_{\mathcal{D}} \frac{1}{|\mathcal{D}|} \left(\int\limits_{{}^{B}A^{(\beta)}_{\boldsymbol{X}}(\{d\})} \frac{\int_{\mathcal{D}} f_{\boldsymbol{X}|\{d\}}(\boldsymbol{x})\, \mathrm{d}d}{|\mathcal{D}|}\, \mathrm{d}\boldsymbol{x} \right) \mathrm{d}d\,. \tag{9.23} $$

Example 9.4 *Calculating the weighted volume V_w for the Bayes 0.9-measurement procedure $B^{(0.9)}_{\mathrm{E}[X]}$ yields $V_w(B^{(0.9)}_{\mathrm{E}[X]}) = 0.553454$ which is slighlty larger than the weighted volume of the 0.9-Neyman measurement procedure with value $V_w({}^{*}C^{(0.9)}_{\mathrm{E}[X]}) = 0.549124$. That is, in the mean the measurements yielding the Neyman procedure are in fact shorter than the ones yielding the Bayes procedure. Figure 9.4 displays the respective scaled 0.9-uncertainty spaces $\mathcal{U}^{(0.9)}_{\mathrm{E}[X]}$ and $\mathcal{B}^{(0.9)}_{\mathrm{E}[X]}$.*

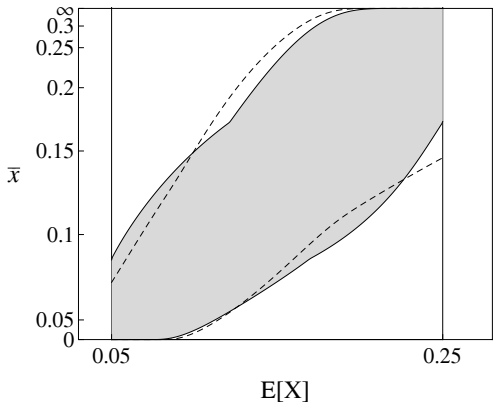


Figure 9.4.: The scaled 0.9-uncertainty spaces $\mathcal{U}^{(0.9)}_{\mathrm{E}[X]}$ (dashed) and $\mathcal{B}^{(0.9)}_{\mathrm{E}[X]}$ (solid).

From the frequentist point of view, one of the most important features of a measurement procedure is whether it meets the demanded reliability in the sense of coverage probability. On the other hand, the Bayesians ask that the demanded reliability in the sense of credible probability is met. Answering this question, we realise what the real difference between the Neyman approach and the Bayes approach is. Nevertheless, we calculate the posterior probability of the β-Neyman measurements ${}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{\overline{x}\})$ and the coverage probability of Bayes β-prediction sets ${}^{B}A^{(0.9)}_{\overline{X}}(\{\mu_1\})$:

- posterior probability of the β-Neyman measurements:

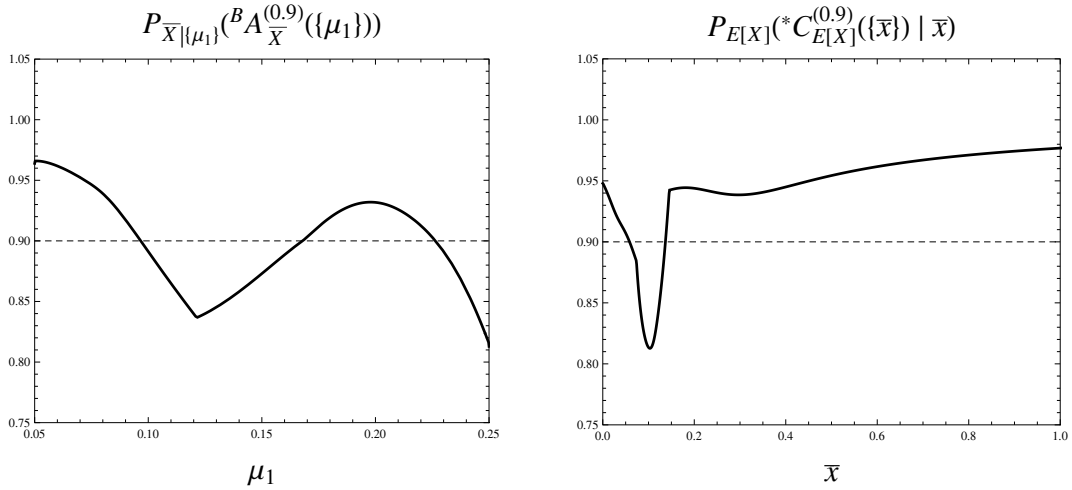
$$\mathrm{P}_D({}^{*}C^{(\beta)}_D(\{\boldsymbol{x}\})|\boldsymbol{x}) \tag{9.24}$$

- coverage probability of Bayes β-prediction sets:

$$\mathrm{P}_{\boldsymbol{X}|\{d\}}({}^{B}A^{(\beta)}_{\boldsymbol{X}}(\{d\})) \tag{9.25}$$

Example 9.5 *Calculating the posterior probability* $\mathrm{P}_{\mathrm{E}[X]}({}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{\overline{x}\})|\overline{x})$ *for* $\overline{x} \in [0,1]$ *and the coverage propability* $\mathrm{P}_{\overline{X}|\{\mu_1\}}({}^{B}A^{(0.9)}_{\overline{X}}(\{\mu_1\}))$ *for* $\mu_1 \in \mathcal{D}$ *yields figure 9.5. We realise, that both procedures violate the demanded reliability* $\beta = 0.9$ *if interpreted in the sense of each other.*

Conclusion The Neyman (classical or frequentist) approach and the Bayes approach represent two different concepts of statistical set estimation theory! Inter-



(a) Coverage probability of the Bayes 0.9-prediction sets ${}^{B}A^{(0.9)}_{\overline{X}}(\{\mu_1\})$.

(b) Credible probability of the 0.9-Neyman measurement sets ${}^{*}C^{(0.9)}_{\mathrm{E}[X]}(\{\overline{x}\})$.

Figure 9.5.: Neyman and Bayes procedures interpreted in the sense of each other.

preting the results of the one in the sense of the other is – at least – disputable, if not meaningless.

Similarities of these two concepts arise through the commonality of utilising pre-available knowledge about the possible values d of the to be estimated parameter D. Assuming the uniform prior on $\mathcal{D}$ in the Bayes approach yields for the observation $\boldsymbol{x}$ the posterior distribution

$$\pi_D(d|\boldsymbol{x}) = \frac{f_{\boldsymbol{X}|\{d\}}(\boldsymbol{x})\mathbb{1}_D(d)}{\int\limits_{\mathcal{D}} f_{\boldsymbol{X}|\{\tilde{d}\}}(\boldsymbol{x})\mathbb{1}_D(\tilde{d})\,\mathrm{d}\tilde{d}}\,. \tag{9.26}$$

The posterior distribution $\pi_D(d|\boldsymbol{x})$ is a function in d given a value $\boldsymbol{x} \in \mathcal{X}$, which serves as both, the construction criterion when highest posterior probability β-credible sets are seeked, and as the side-condition for the demanded reliability β. The same ratio also appears in the Neyman approach as

$$\frac{f_{\boldsymbol{X}|\{d\}}(\boldsymbol{x})\mathbb{1}_D(d)}{\int\limits_{\mathcal{D}} f_{\boldsymbol{X}|\{\tilde{d}\}}(\boldsymbol{x})\mathbb{1}_D(\tilde{d})\,\mathrm{d}\tilde{d}} = \frac{f_{\boldsymbol{X}|\{d\}}(\boldsymbol{x})\mathbb{1}_D(d)}{w(\boldsymbol{x})} \tag{9.27}$$

but now this is a function in $\boldsymbol{x}$ given a value $d \in \mathcal{D}$. The ratio serves only as the construction criterion for the β-prediction sets, while the side-condition for the demanded reliability β is provided through the probability distribution $\mathrm{P}_{\boldsymbol{X}|\{d\}}$.

10. Summary and Outlook

This thesis discussed various approaches to deal with uncertainty due to randomness and ignorance when prior information is available. To reduce ignorance measurement procedures were derived.

The first part of this thesis dealt with the question how a reasonable probability distribution should be choosen in order to meet requested qualitative properties. Those qualitative properties constitute a first form of prior information, and were translated into quantitative properties. Under the minimum information principle, a universal form of PDFs was proposed which was defined by the support and a sequence of moments – which was named *Minimum Information Representative* (MIR). The main feature of the proposed probability distributions was the boundedness of their support (range of variability). The latter feature, though very realistic, is rarely adequately considered in stochastic modelling. One of such MIR is the Monotonic Distribution, uniquely defined by a bounded interval as the support and the first moment. An extensive characterisation of the monotonic distribution was presented and many of its properties were derived.

The second part of this thesis was dedicated to the reduction of ignorance through measurement procedures. From the metrological view, such procedures should meet two requirements, namely *accuracy* and *reliability*. Concepts of estimation in classical statistics were outlined and discussed with respect to those requirements. It was argued that point estimators do not meet these requirements, and therefore should be considered as unsuitable in the metrological framework.

Set estimators conform to the metrological requirements. The reliability requirement is represented by the confidence level, and accuracy by the size of the resulting estimates. However, it was outlined that classical statistics only provides a point-wise measure of accuracy for the set estimates, but not an overall measure for the set estimator. The duality of acceptance (prediction) regions and measurement (confidence) regions, is well-known in classical statistic. This concept was formulated by Neyman (1935, 1937) [56, 57], and enables the derivation

of confidence regions by inverting acceptance regions. It is remarkable, that this concept attracted so little interest, altough it would make it possible to take prior information into consideration.

The classical concepts of inference were applied for the monotonic distribution. When the first moment is subject to measurement, point estimation by means of moment estimators, and ML estimators were feasible. Since the latter proved to be biased, further evaluations became needless. Measuring the upper bound of the range of variability posed many difficulties, and particularly the ML estimator revealed some unwelcome behaviour.

von Collani [19, 20] set up a framework called Bernoulli-Space. Embedded into this framework are the Neyman measurement procedures (earlier defined in [16, 17]) which constitute a different approach to estimation. Based as well on the duality of prediction and measurement regions, the procedures incorporate prior information and aim at minimising an weighted volume of the measurement regions from the beginning. In the metrological context, the weighted volume could be interpreted as the accuracy. The minimisation is eventually achieved by prediction regions consisting of points with maximum likelihood ratio[1]. Prior information is incorporated in a rather simple but appealing way by considering every possible value of the parameter likewise. That is, for a one-dimensional parameter a reasonable form of prior information would be an interval. In its generality, the approach can be applied to every statistical distribution. Weigand (2009) [77] has derived some generalisation of the Neyman measurement procedures by considering other definitions of accuracy, and weight functions. In this way, he justified to define a prior distribution on the set of possible values of the parameter to be measured as a form of prior information. At the same time he stressed that this distribution must not be confused with a probability distribution, as it would be interpreted in Bayesian statistic. With respect to the parameter p of the binomial distribution $Bi(n, p)$, Lurz (2015) [49] has derived Neyman measurement procedures under different prior distributions of p, mainly Beta distributions.

With respect to the monotonic distribution, Neyman measurement procedures

[1] Not to be confused with the maximum likelihood test statistic!

were derived for the first moment $\mathrm{E}[X]$ of $X \sim Mon(\mu_1)$ under prior information in form of intervals. Comparisons with classical procedures were performed. These comparisons include the approximation of the monotonic distribution by the exponential distribution. The approximations yielded comparable results with respect to the coverage probability and the weighted volume, when the ignorance space $\mathcal{D}$ contained only small values μ_1. That is, the actual boundedness of the range of variability played only a minor role. The situation changed, when larger values μ_1 constituted $\mathcal{D}$. The approximations exceeded the requested reliability level and yielded larger weighted volumes. That is, approximations of the monotonic distribution were indeed reasonable, but only if the boundedness was insignificant. Noteworthy was the observation, that the differences in the weighted volumes even increased with increasing sample size n in the example for $\mathcal{D} = \{\mu_1 \,|\, 0.40 \leq \mu_1 \leq 0.45\}$.

Deriving a Neyman measurement procedure for the upper bound B of the range of variability of the monotonic distribution constitutes a complex task. Mainly this is due to the necessity to take both the sample sum and the sample maximum into account, which amounted to two-dimensional prediction regions. By a proposed discretisation of the uncertainty space, an approximation was achieved. The approximation turned out to be rather rough and the calculations were rather straightforward than elaborated, and thus further work should be invested with respect to faster algorithms and smoother prediction regions.

The monotonic distribution on the unit interval was used to set up a decomposition model for the tainting ratio in audit sampling under overstatement and zero-inflation. This model gave rise to a combined confidence interval, for which the Neyman measurement procedure contributed the confidence interval for the conditional mean tainting ratio under overstatement. Comparisons with a two-sided interval of Stringer type were performed by a simulation study. Even if the amount of prior information is considerably weak in the sense that it almost only expresses the monotonic decreasing nature of the considered monotonic distribution, reductions in the mean length were observed. In particular, this was the case for small sample sizes. Considering stronger prior information led to further reductions in the mean lengths on the one hand, and on the other hand,

the possibility of reducing the sample size to achieve similar mean lengths of the combined confidence interval as of the two-sided Stringer interval. In most cases, these reductions were up to 50%, but even greater ones were observed. Prior information was taken into account only for the conditional mean, thus, it would be reasonable to also consider prior information for the proportion of overstatements. The latter was solely considered in the aforementioned work of Lurz (2015) [49]. Since the decomposition model enables separate measurements, it will be straightforward to combine both procedures in a future study.

A. More about the Monotonic Distribution

A.1. Relations between μ_1, λ and b

Proposition 3.3 together with proposition 3.4 in section 3.2.1 shows that λ is the unique solution of

$$\mu_1 = \frac{(-1+\lambda b)e^{\lambda b} - (-1+\lambda a)e^{\lambda a}}{\lambda(e^{\lambda b} - e^{\lambda a})} \tag{A.1}$$

Therefore, we might have considered λ as a function of (a, b, μ_1), i.e., $\lambda(a, b, \mu_1)$. Now, let a be fixed, thus with proposition 3.4 we may assume without loss of generality $a = 0$. Then, for given μ_1 and b, $\lambda(b, \mu_1)$ is the unique solution of

$$\mu_1 = \frac{(-1+\lambda b)e^{\lambda b} + 1}{\lambda(e^{\lambda b} - 1)}\,. \tag{A.2}$$

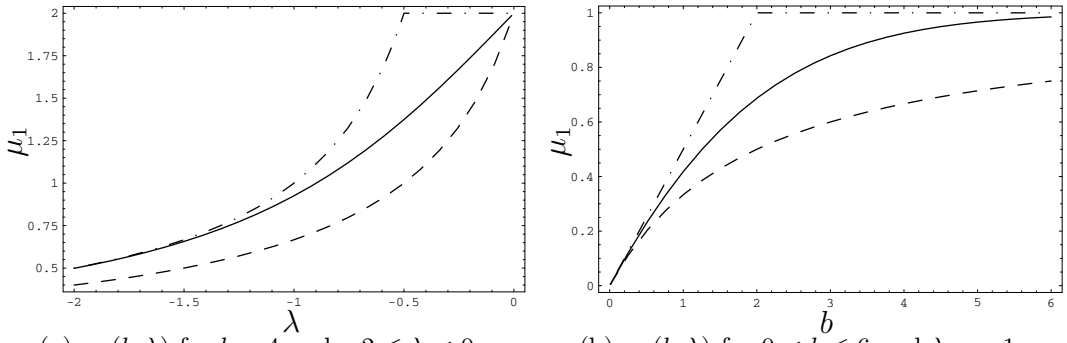
Furthermore we restrict our analysis to the case of strictly monotone decreasing propability density function on the range of variability $\mathcal{X} = [0, b]$, which is equivalent to $\lambda < 0$ and $b > 2\mu_1 > 0$, respectively.

In general, equation (A.2) provides a unique solution for one of the three parameters μ_1, b and λ, when the other two are given, i.e., one parameter may be considered as a function of the other two parameters.

A.1.1. $\mu_1(b, \lambda)$

Only for $\mu_1(b, \lambda)$ we are able to give an explicit function – which is just (A.2). Besides $\lambda < 0$ and $b > 0$, there are no further restrictions, i.e., $(b, \lambda) \in (0, +\infty) \times (-\infty, 0)$. In the proof of proposition 3.3 it is shown, that $\mu_1(b, \lambda)$ for arbitrary but fixed $b > 0$ is increasing in $\lambda \in (-\infty, 0)$. On the other hand, for arbitrary but fixed $\lambda < 0$, $\mu_1(b, \lambda)$ is increasing in $b \in (0, +\infty)$, since $e^{\lambda b} \geq \lambda b + 1$:

$$\frac{\partial}{\partial b}\mu_1(b, \lambda) = \frac{e^{\lambda b}\big(e^{\lambda b} - (\lambda b + 1)\big)}{\big(e^{\lambda b} - 1\big)^2} > 0\,. \tag{A.3}$$



(a) $\mu_1(b,\lambda)$ for $b = 4$ and $-2 \leq \lambda < 0$.

(b) $\mu_1(b,\lambda)$ for $0 < b \leq 6$ and $\lambda = -1$.

Figure A.1.: $\mu_1(b,\lambda)$ as function of b and λ, respectively, with upper bound $\min\{\frac{b}{2}, -\frac{1}{\lambda}\}$ (dot-dashed line) and lower bound $\frac{b}{-\lambda b+2}$ (dashed line).

Although the function $\mu_1(b,\lambda)$ is explicitly known, we want to give a lower and an upper bound for it. With corollary 3.9 and $\mu_1 < \frac{b}{2}$ we have:

$$\frac{b}{-\lambda b + 2} < \mu_1(b,\lambda) < \min\{\frac{b}{2}, -\frac{1}{\lambda}\}\,. \tag{A.4}$$

The limits of $\mu_1(b,\lambda)$ and the corresponding limiting distributions are as follows:

$$\lim_{b\to 0} \mu_1(b,\lambda) = 0 \qquad \text{(one-point distribution } \mathrm{P}(X=0)=1) \tag{A.5}$$

$$\lim_{b\to +\infty} \mu_1(b,\lambda) = -\frac{1}{\lambda} \qquad \text{(exponential distribution } EXP(1/\mu_1)) \tag{A.6}$$

$$\lim_{\lambda\to 0} \mu_1(b,\lambda) = \frac{b}{2} \qquad \text{(uniform distribution } \mathrm{U}(0,b)) \tag{A.7}$$

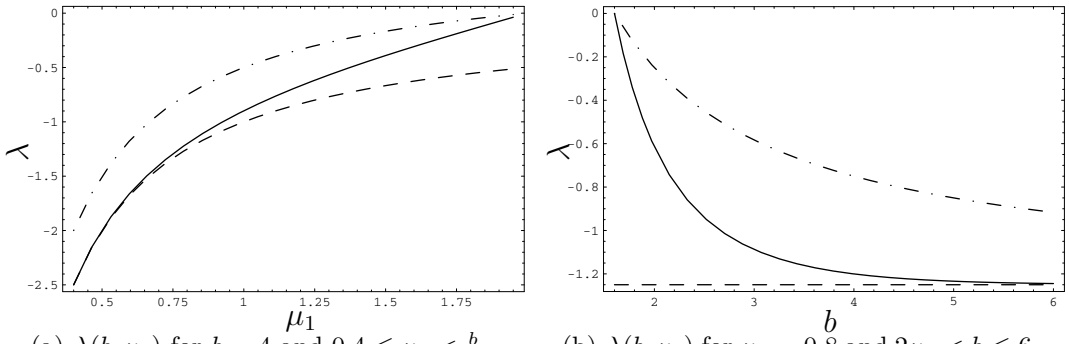
$$\lim_{\lambda\to -\infty} \mu_1(b,\lambda) = 0 \qquad \text{(one-point distribution } \mathrm{P}(X=0)=1) \tag{A.8}$$

Figure A.1 illustrates the dependency of $\mu_1(b,\lambda)$ on b and λ, respectively, together with the upper bound $\min\{\frac{b}{2}, -\frac{1}{\lambda}\}$ and lower bound $\frac{b}{-\lambda b+2}$. We note in figure A.1(a) how $\mu_1(4,\lambda)$ approaches $-\frac{1}{\lambda}$ for decreasing λ, and that $\mu_1(b,-1)$ in figure A.1(b) fastly approaches $1 = -\frac{1}{-1}$ for increasing b.
Proposition 3.4 shows, that $\lambda(b,\mu_1)b = \lambda(\frac{\mu_1}{b})$ holds. Equivalent is the following transformation formula for the function $\mu_1(b,\lambda)$:

$$\mu_1(b,\lambda) = \frac{(-1+\lambda b)e^{\lambda b}+1}{\lambda(e^{\lambda b}-1)} = b\frac{(-1+\lambda b)e^{\lambda b}+1}{\lambda b(e^{\lambda b}-1)} = b\mu_1(1,\lambda b) \tag{A.9}$$

A.1.2. $\lambda(b, \mu_1)$

As derived in section 3.2.1, $\lambda(b, \mu_1)$ is uniquely defined by (A.2). We have shown with proposition 3.3 and (A.3), that $\mu_1(b, \lambda)$ is increasing in both b and λ. Thus, increasing b while μ_1 is kept constant results into a decrease of λ, i.e., $\lambda(b, \mu_1)$ is decreasing in $b > 2\mu_1$ for arbitrary μ_1. Conversely, increasing $\mu_1 < \frac{b}{2}$ with arbitrary b yields an increase of λ, i.e., $\lambda(b, \mu_1)$ is increasing in $\mu_1 < \frac{b}{2}$ for arbitrary b.



(a) $\lambda(b, \mu_1)$ for $b = 4$ and $0.4 \leq \mu_1 < \frac{b}{2}$. (b) $\lambda(b, \mu_1)$ for $\mu_1 = 0.8$ and $2\mu_1 < b \leq 6$.

Figure A.2.: $\lambda(b, \mu_1)$ as function of b and μ_1, respectively, with upper bound $-\frac{1}{\mu_1} + \frac{2}{b}$ (dot-dashed line) and lower bound $-\frac{1}{\mu_1}$ (dashed line).

Figure A.2 illustrates the dependency of $\lambda(b, \mu_1)$ on b and μ_1, respectively, together with the upper bound $-\frac{1}{\mu_1} + \frac{2}{b}$ and lower bound $-\frac{1}{\mu_1}$ derived in corollary 3.9. We also note that $\lambda(4, \mu_1)$ in figure A.2(a) fastly approaches $-\frac{1}{\mu_1}$ for decreasing μ_1, and in figure A.2(b) how $\lambda(b, 0.8)$ approaches $-1.25 = -\frac{1}{0.8}$ for increasing b.

Additionally we explicitly calculate the derivative of $\lambda(b, \mu_1)$ with respect to b for an arbitrary $\mu_1 < \frac{b}{2}$. With $b\lambda(b, \mu_1) = \lambda(\frac{\mu_1}{b})$ we first get

$$\frac{\partial}{\partial b}\lambda(b, \mu_1) = \frac{\partial}{\partial b}\left(\lambda\left(\frac{\mu_1}{b}\right)\frac{1}{b}\right) = \left(\frac{\partial\lambda(m)}{\partial m}\right)\Big|_{m=\frac{\mu_1}{b}} \cdot \frac{\partial\frac{\mu_1}{b}}{\partial b} \cdot \frac{1}{b} + \lambda\left(\frac{\mu_1}{b}\right) \cdot \frac{\partial\frac{1}{b}}{\partial b}. \quad \text{(A.10)}$$

The derivative of $\lambda(\mu_1) = \lambda(1, \mu_1)$ with respect to μ_1 is already known from the proof to proposition 3.8. Considering the second moment μ_2 also as a function

of $\lambda(\mu_1)$ we get

$$\left(\frac{\partial\lambda(m)}{\partial m}\right)\Big|_{m=\frac{\mu_1}{b}} = \left(\frac{1}{\mu_2(\lambda(m)) - (\mu_1(\lambda(m)))^2}\right)\Big|_{m=\frac{\mu_1}{b}} \tag{A.11}$$

$$= \frac{1}{\mu_2(\lambda(\frac{\mu_1}{b})) - \left(\frac{\mu_1}{b}\right)^2} \tag{A.12}$$

$$= \frac{b^2}{\mu_2(b, \lambda(b, \mu_1)) - \mu_1^2} \tag{A.13}$$

For equation (A.13) we have used the following transformation of $\mu_2(b, \lambda(b, \mu_1))$:

$$\mu_2\big(b, \lambda(b, \mu_1)\big) = \frac{\Big(2 - 2\lambda(b, \mu_1)b + \big(\lambda(b, \mu_1)\big)^2 b^2\Big)e^{\lambda(b,\mu_1)b} - 2}{\big(\lambda(b, \mu_1)\big)^2\big(e^{\lambda(b,\mu_1)b} - 1\big)} \tag{A.14}$$

$$= b^2\frac{\Big(2 - 2\big(\lambda(b, \mu_1)b\big) + \big(\lambda(b, \mu_1)b\big)^2\Big)e^{\lambda(b,\mu_1)b} - 2}{\big(\lambda(b, \mu_1)b\big)^2\big(e^{\lambda(b,\mu_1)b} - 1\big)} \tag{A.15}$$

$$= b^2\mu_2\big(1, \lambda(b, \mu_1)b\big) =: b^2\mu_2\big(\lambda(b, \mu_1)b\big) \tag{A.16}$$

Then, for (A.10) we get

$$\frac{\partial}{\partial b}\lambda(b, \mu_1) = -\frac{\mu_1}{b^3}\cdot\frac{b^2}{\mu_2\big(b, \lambda(b, \mu_1)\big) - \mu_1^2} - \lambda(b, \mu_1)\frac{1}{b} \tag{A.17}$$

$$= \frac{1}{b}\left(-\mu_1\frac{1}{\mu_2\big(b, \lambda(b, \mu_1)\big) - \mu_1^2} - \lambda(b, \mu_1)\right). \tag{A.18}$$

At the beginning of this paragraph we derived that for arbitrary μ_1, $\lambda(b, \mu_1)$ is decreasing in $b > 2\mu_1$, i.e., $\frac{\partial}{\partial b}\lambda(b, \mu_1) < 0$, and so

$$-\mu_1\frac{1}{\mu_2(b, \lambda(b, \mu_1)) - \mu_1^2} - \lambda(b, \mu_1) < 0\,. \tag{A.19}$$

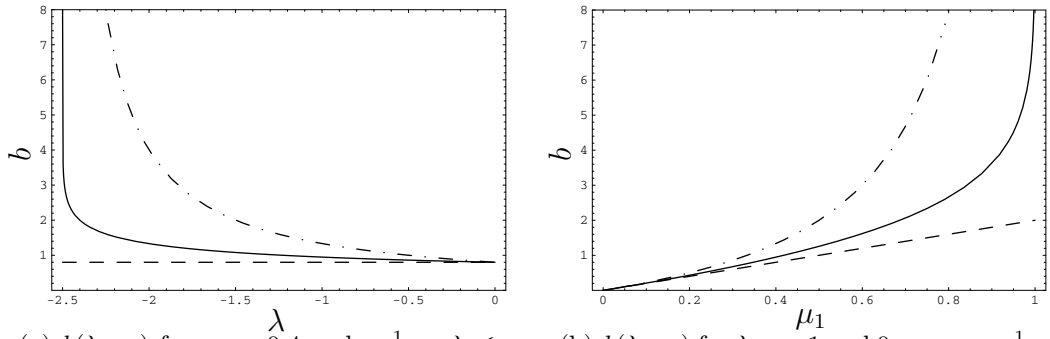
A.1.3. $b(\lambda, \mu_1)$

Although, the dependency of b on λ and μ_1 is entirely obtained by $\mu_1(b, \lambda)$ and $\lambda(b, \mu_1)$, we want to take a closer look for which combinations of values of λ and μ_1 the function $b(\lambda, \mu_1)$ is defined, so that the three values determine the monotonic probability distribution $Mon(b, \mu_1)$.

Solving the inequality in corollary 3.9 for b and considering the case of a monotone decreasing density yields

$$2\mu_1 < b(\lambda, \mu_1) < \frac{2\mu_1}{\lambda\mu_1 + 1}\,. \tag{A.20}$$

A condition which is implicitly part of inequality (A.20) implies $\lambda\mu_1 > -1$. It follows, that given some arbitrary $\mu_1 > 0$, the admissible values of λ have to meet $-\frac{1}{\mu_1} < \lambda < 0$. Then, $b(\lambda, \mu_1)$ is decreasing in $\lambda < 0$ and converges towards $2\mu_1$ for $\lambda \to 0$ and tends to $+\infty$ for $\lambda \to -\frac{1}{\mu_1}$, respectively. On the other hand, given some arbitrary $\lambda < 0$, only values of μ_1 which meet $0 < \mu_1 < -\frac{1}{\lambda}$ are allowed, and $b(\lambda, \mu_1)$ is increasing in μ_1 and converges to 0 for $\mu_1 \to 0$ and tends to $+\infty$ for $\mu_1 \to -\frac{1}{\lambda}$. Figure A.3 is illustrating this dependency together with the upper and lower bound functions.



(a) $b(\lambda, \mu_1)$ for $\mu_1 = 0.4$ and $-\frac{1}{\mu_1} < \lambda \leq 0$.

(b) $b(\lambda, \mu_1)$ for $\lambda = -1$ and $0 < \mu_1 < -\frac{1}{\lambda}$.

Figure A.3.: $b(\lambda, \mu_1)$ as function of λ and μ_1, respectively, with upper bound $\frac{2\mu_1}{\lambda\mu_1+1}$ (dot-dashed line) and lower bound $2\mu_1$ (dashed line).

B. Uncertainty Spaces of Neyman Procedures and Approximations

B.1. Uncertainty spaces for $\mathcal{D} = \{\mu_1 \,|\, 0.05 \leq \mu_1 \leq 0.25\}$

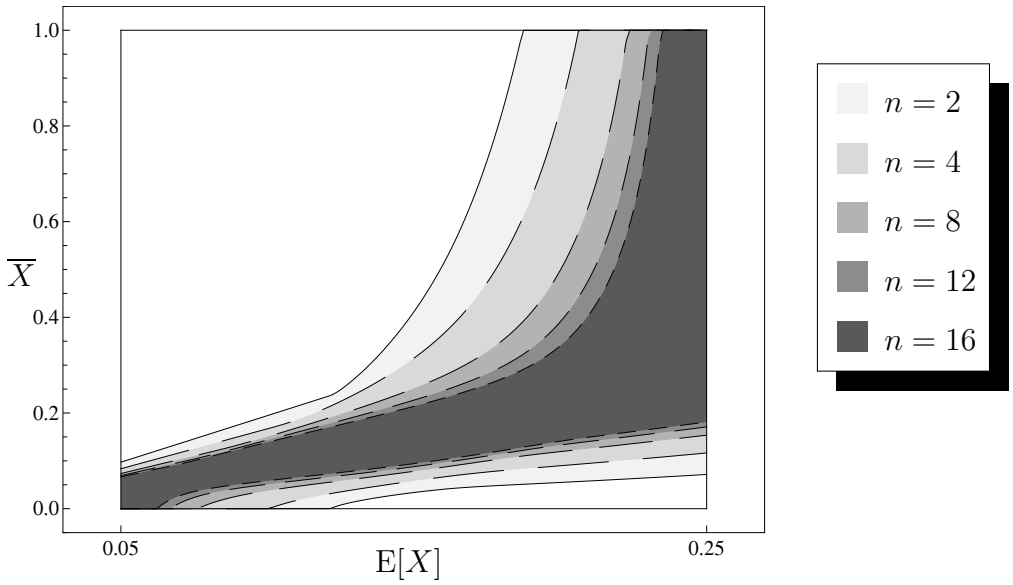


Figure B.1.: Optimal β-uncertainty spaces $\mathcal{U}_{\overline{X}}^{(0.9)}$ for $\overline{X}$ and different values of the sample size $n = 2, 4, 8, 12, 16$.

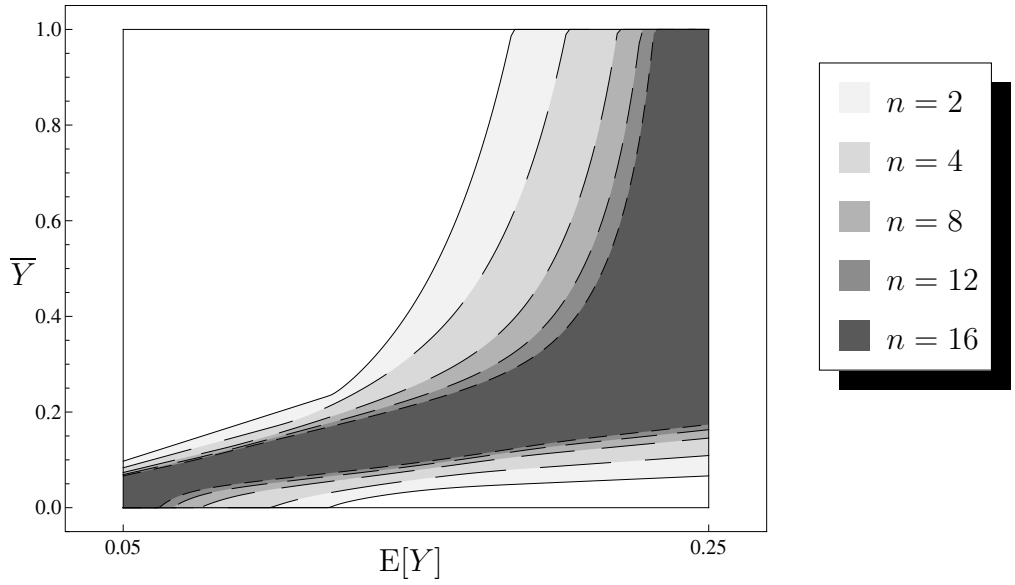


Figure B.2.: Optimal β-uncertainty spaces $\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ and $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ and different values of the sample size $n = 2, 4, 8, 12, 16$.

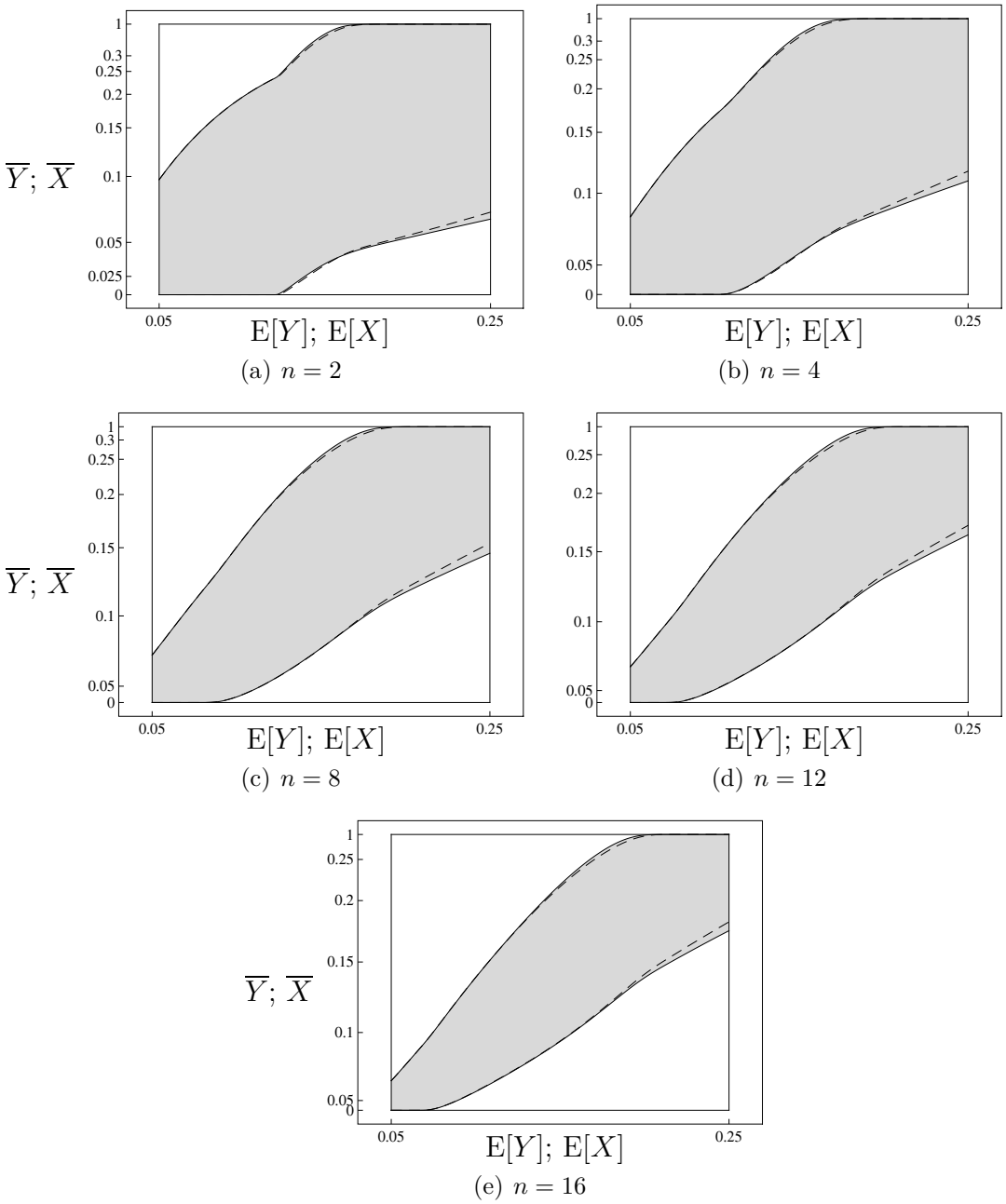


Figure B.3.: The probability scaled plots of the optimal β-uncertainty spaces $\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ with $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ and the optimal ones $\mathcal{U}_{\overline{X}}^{(0.9)}$ (dashed shape).

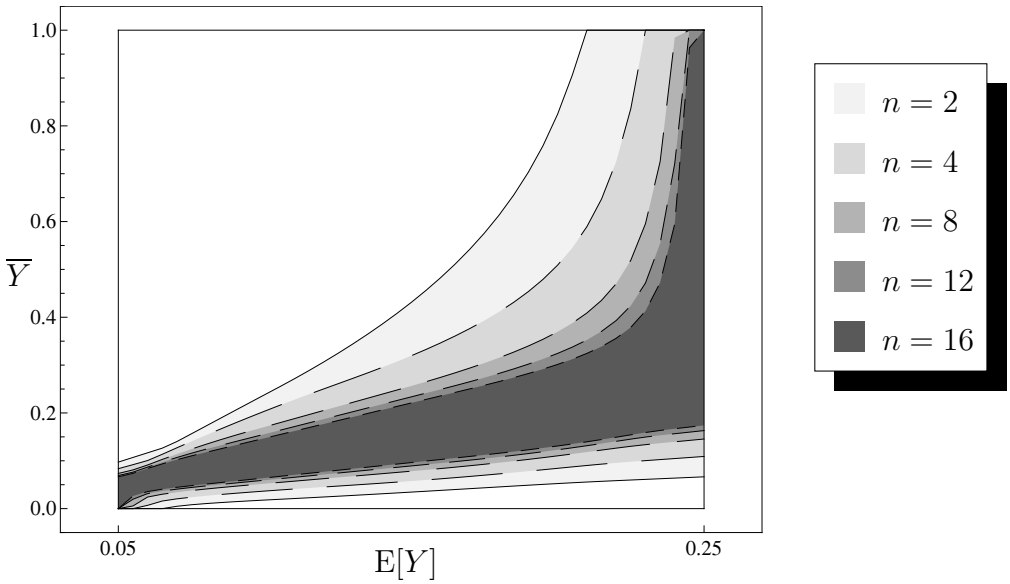
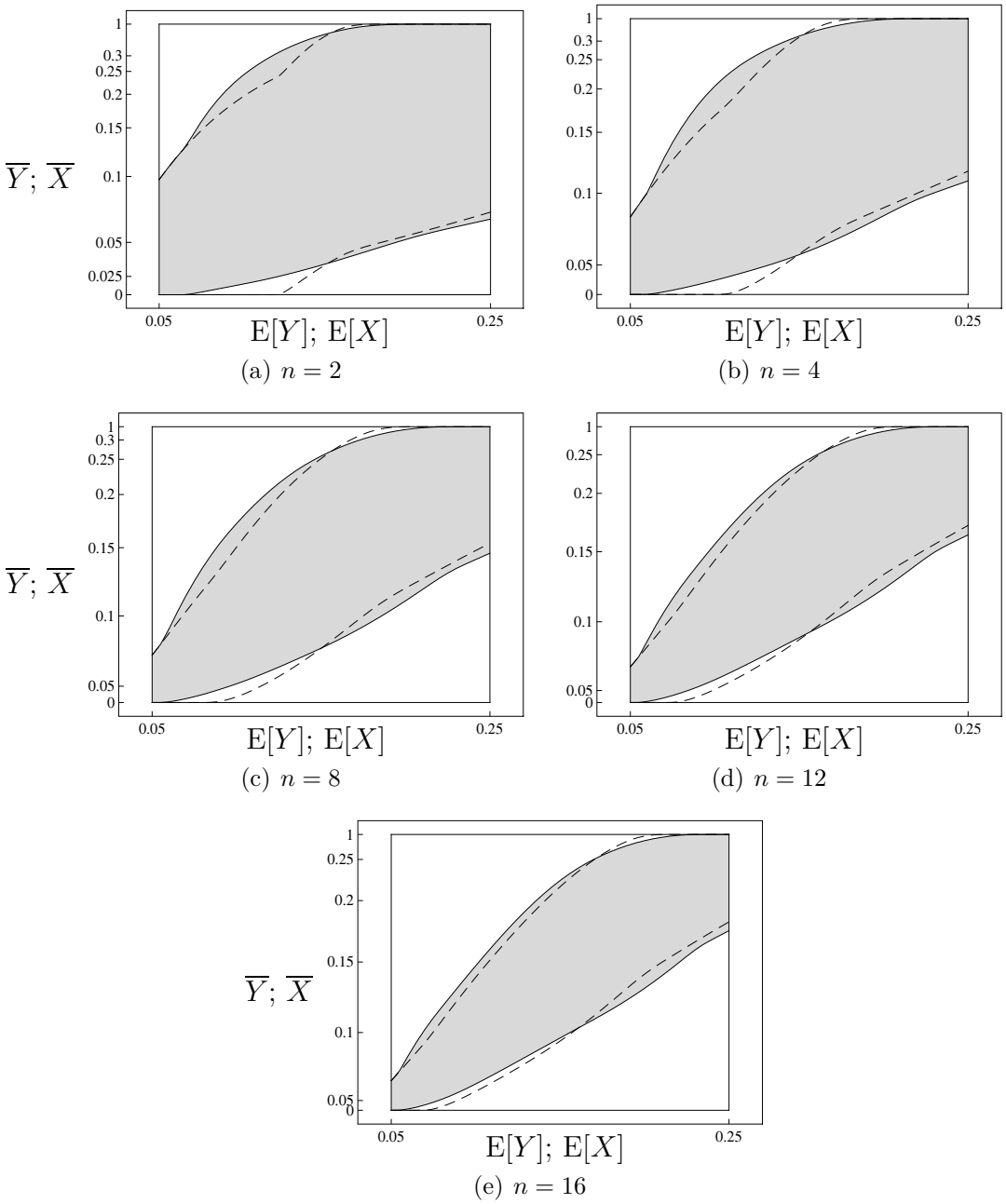


Figure B.4.: β-uncertainty spaces ${}^{LR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ and $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ based on the likelihood ratio test statistic and different values of the sample size $n = 2, 4, 8, 12, 16$.



(a) $n = 2$ (b) $n = 4$

(c) $n = 8$ (d) $n = 12$

(e) $n = 16$

Figure B.5.: The probability scaled plots of the β-uncertainty spaces ${}^{LR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ with $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ based on the likelihood test statistic and the optimal ones $\mathcal{U}_{\overline{X}}^{(0.9)}$ (dashed shape).

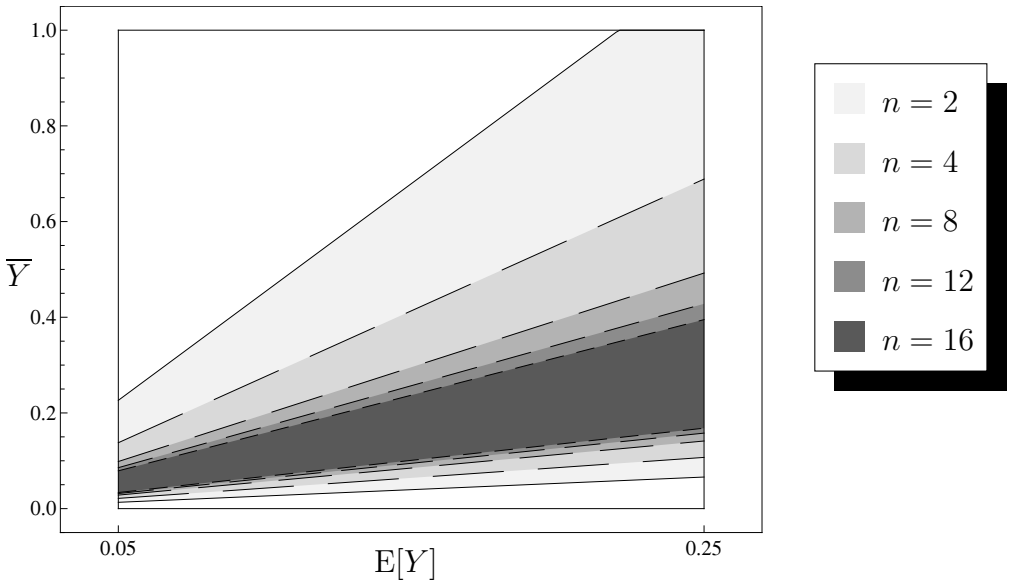
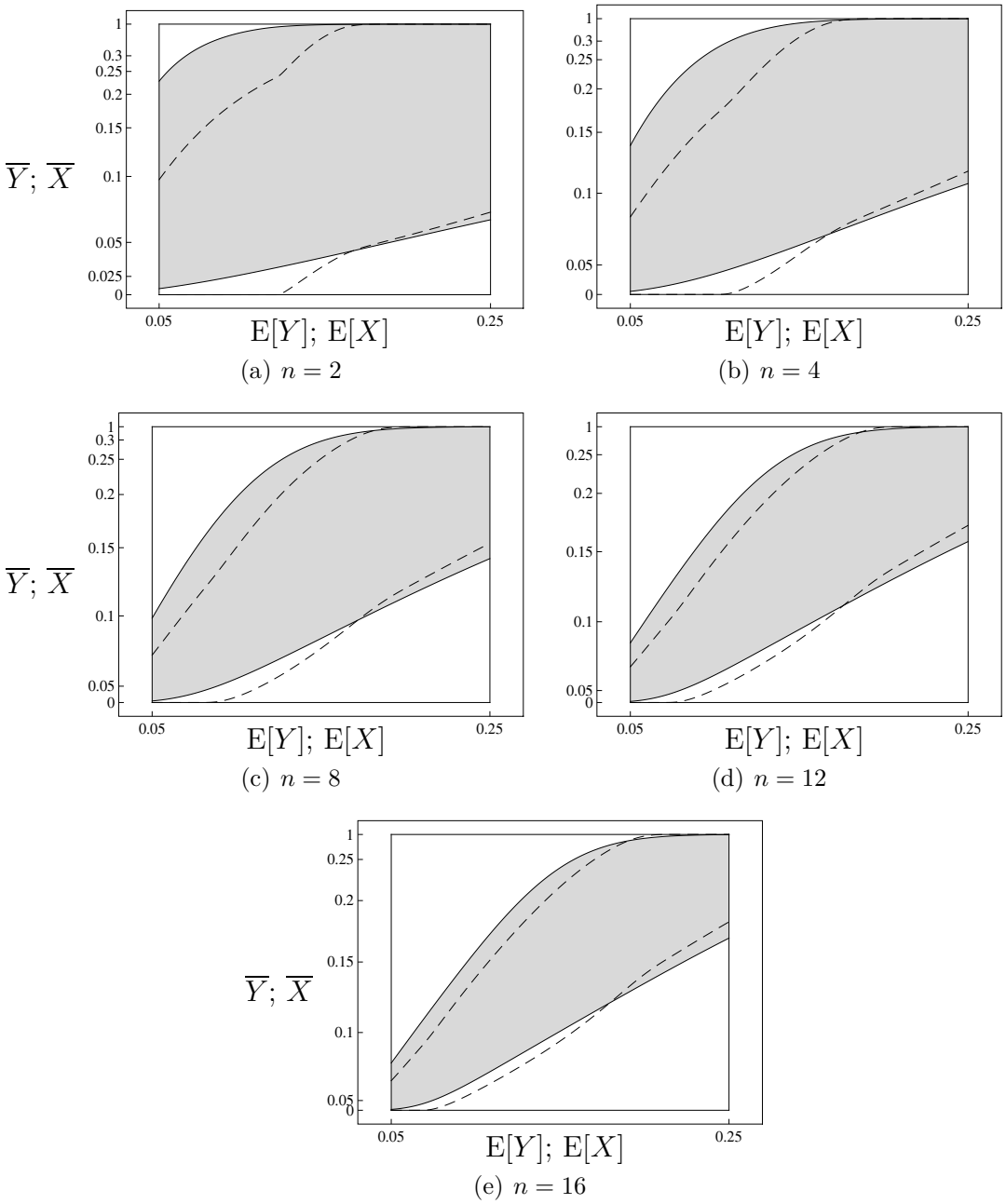


Figure B.6.: β-uncertainty spaces ${}^{HPR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ and $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ based on highest probability regions and different values of the sample size $n = 2, 4, 8, 12, 16$.



(a) $n = 2$ (b) $n = 4$

(c) $n = 8$ (d) $n = 12$

(e) $n = 16$

Figure B.7.: The probability scaled plots of the β-uncertainty spaces ${}^{HPR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ with $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ based on highest probability regions and the optimal ones $\mathcal{U}_{\overline{X}}^{(0.9)}$ (dashed shape).

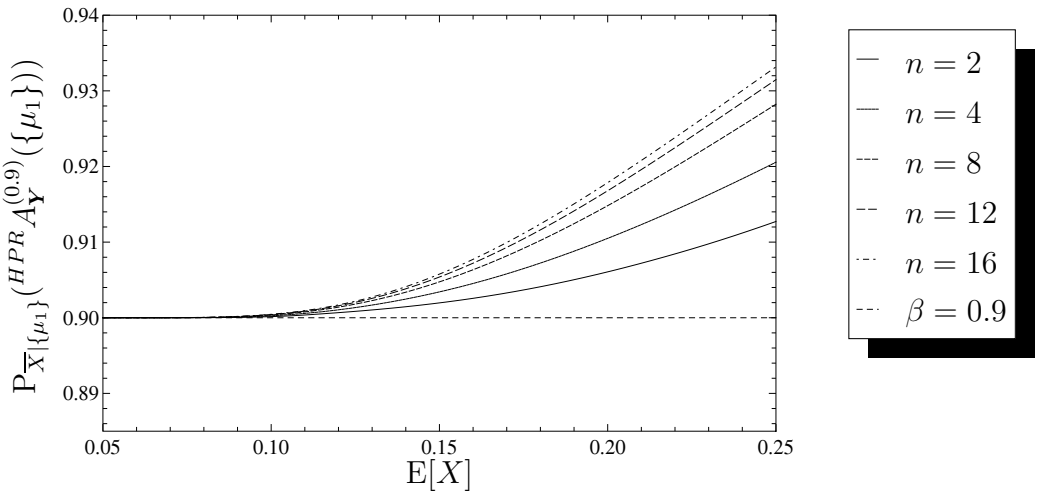


Figure B.8.: The true probability of the prediction sets ${}^{HPR}A_{\boldsymbol{Y}}^{(0.9)}(\{\mu_1\})$ from the corresponding measurement procedures for $\mathrm{E}[Y]$ with $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ based on highest probability regions for $n = 2, 4, 8, 12, 16$.

B.2. Uncertainty spaces for $\mathcal{D} = \{\mu_1 \,|\, 0.40 \leq \mu_1 \leq 0.45\}$

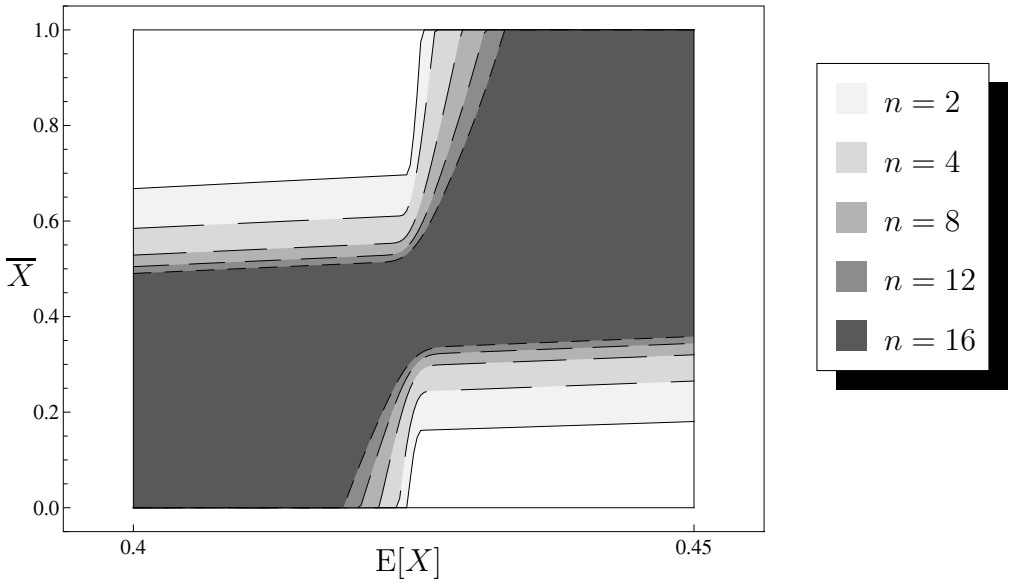


Figure B.9.: Optimal β-uncertainty spaces $\mathcal{U}_{\overline{X}}^{(0.9)}$ for $\overline{X}$ and different values of the sample size $n = 2, 4, 8, 12, 16$.

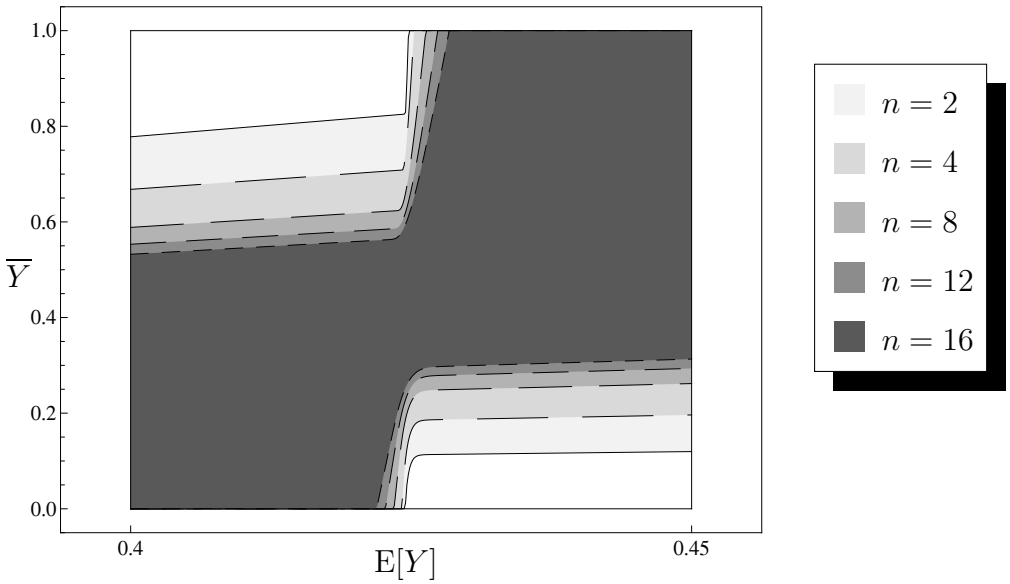
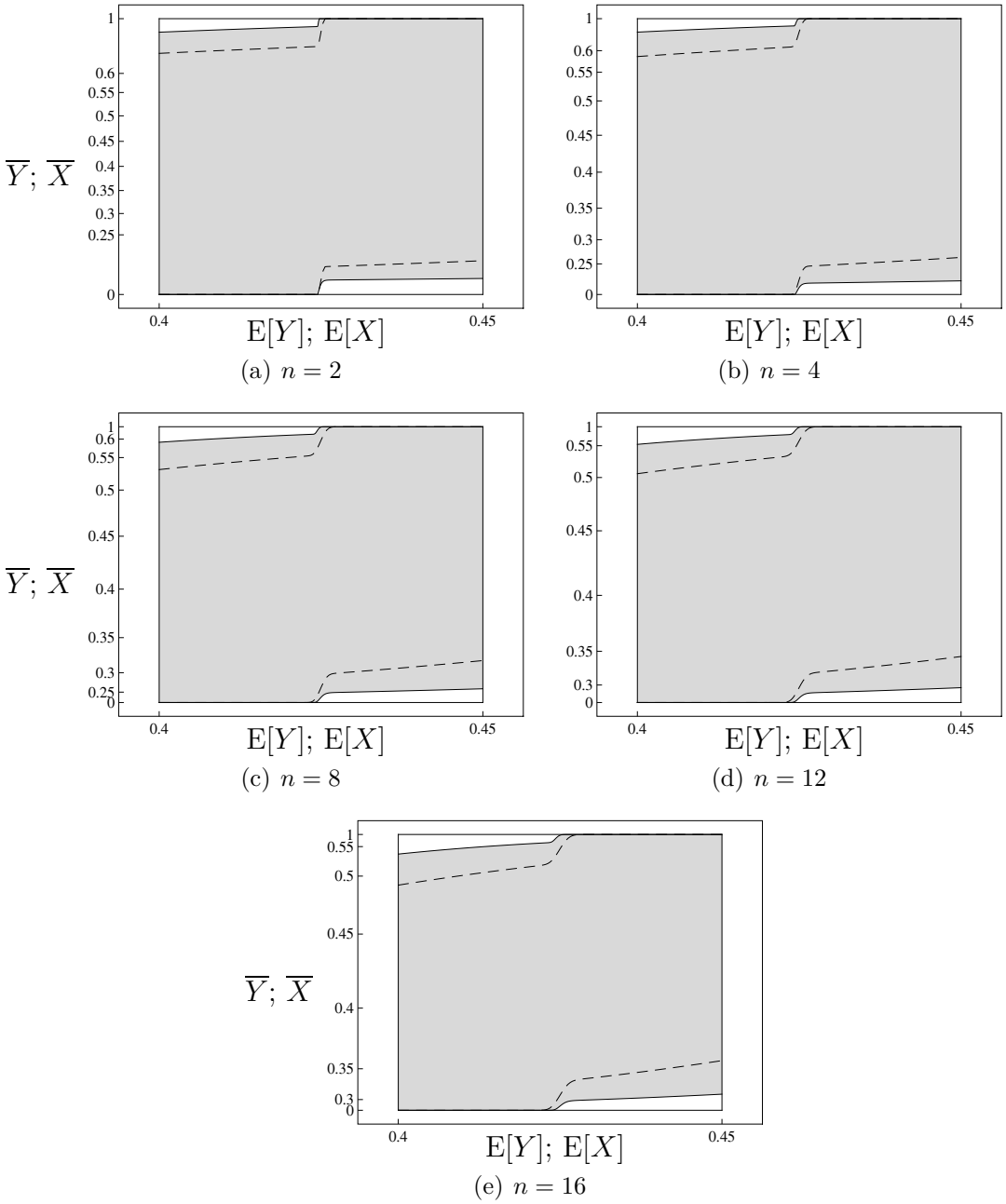


Figure B.10.: Optimal β-uncertainty spaces $\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ and $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ and different values of the sample size $n = 2, 4, 8, 12, 16$.



(a) $n = 2$ (b) $n = 4$

(c) $n = 8$ (d) $n = 12$

(e) $n = 16$

Figure B.11.: The probability scaled plots of the optimal β-uncertainty spaces $\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ with $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ and the optimal ones $\mathcal{U}_{\overline{X}}^{(0.9)}$ (dashed shape).

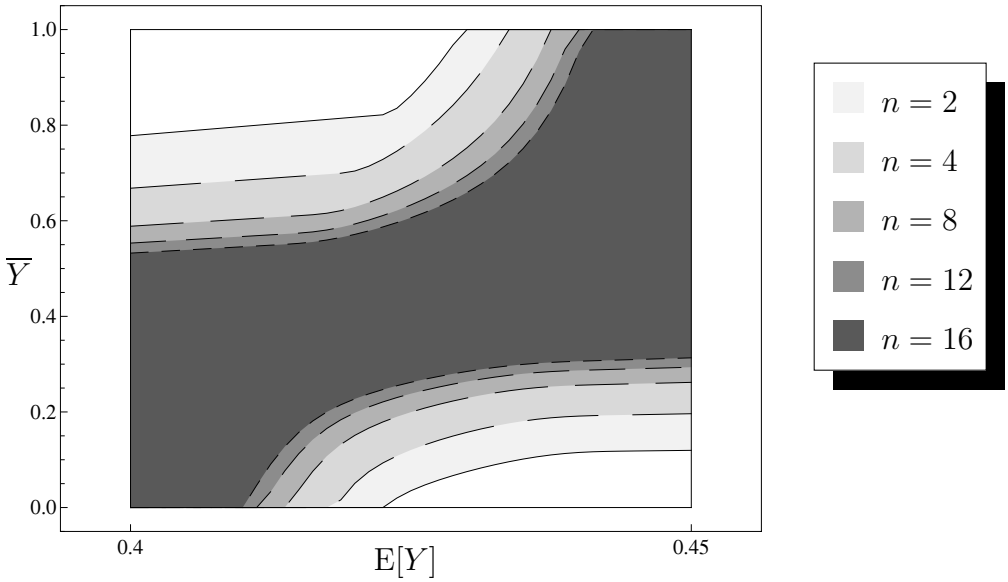
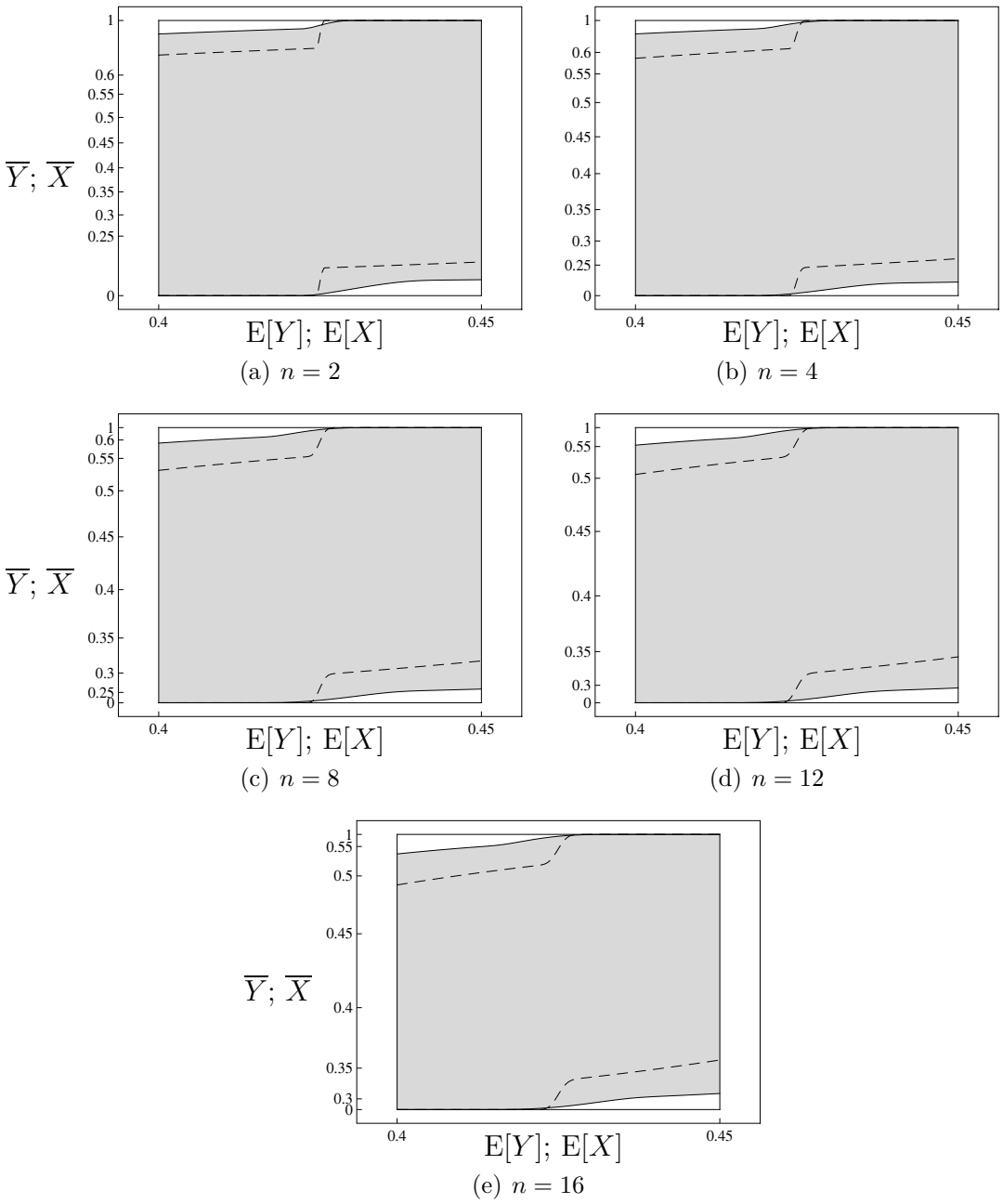


Figure B.12.: β-uncertainty spaces ${}^{LR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ and $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ based on the likelihood ratio test statistic and different values of the sample size $n = 2, 4, 8, 12, 16$.



(a) $n = 2$ (b) $n = 4$

(c) $n = 8$ (d) $n = 12$

(e) $n = 16$

Figure B.13.: The probability scaled plots of the β-uncertainty spaces ${}^{LR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ with $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ based on the likelihood test statistic and the optimal ones $\mathcal{U}_{\overline{X}}^{(0.9)}$ (dashed shape).

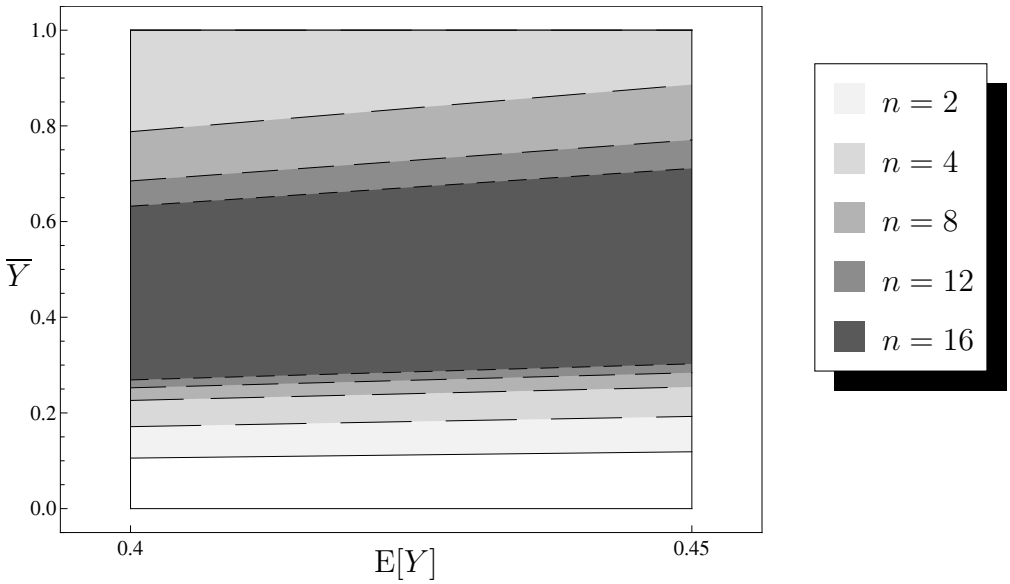


Figure B.14.: β-uncertainty spaces ${}^{HPR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ and $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ based on highest probability regions and different values of the sample size $n = 2, 4, 8, 12, 16$.

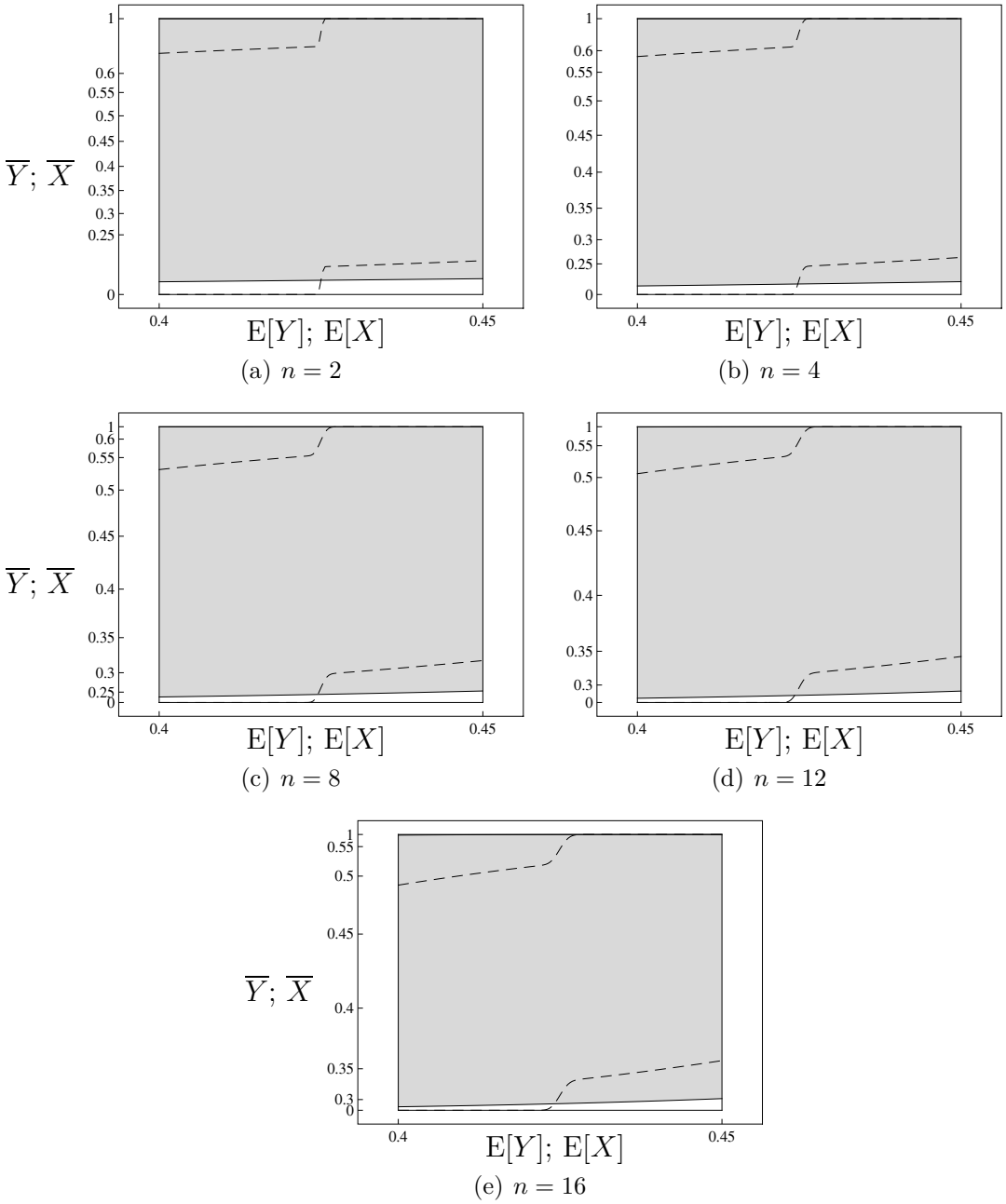


Figure B.15.: The probability scaled plots of the β-uncertainty spaces ${}^{HPR}\mathcal{U}_{\overline{Y}}^{(0.9)}$ for $\overline{Y}$ with $Y|\{\mu_1\} \sim EXP(1/\mu_1)$ based on highest probability regions and the optimal ones $\mathcal{U}_{\overline{X}}^{(0.9)}$ (dashed shape).

Bibliography

[1] Anderson, R. J., and Teitlebaum, D. A. (1973). *Dollar-Unit Sampling.* CA Magazine, April 1973, pp. 30-39.

[2] Behnen, K. and Neuhaus, G. (1995). *Grundkurs Stochastik.* 3rd Edition. Teubner, Stuttgart.

[3] Bernoulli, J. (1713). *Ars conjectandi, opus posthumum. Accedit Tractatus de seriebus infinitis, et epistola gallicé scripta de ludo pilae reticularis.* Thurneysen Brothers, Basel.

[4] Bich, W., Cox, M.G., and Harris, P.M. (2006). *Evolution of the 'Guide to the Expression of Uncertainty in Measurement'.* Metrologia 43, S161-S166

[5] Bickel, P. J. (1992). *Inference and Auditing: The Stringer Bound.* International Statistical Review, Vol. 60, pp. 197-209.

[6] Bickel, P. J., and Doksum, K.A. (2001). *Mathematical Statistics: Basic Ideas and Selected Topics* Volume I. 2nd Edition. Prentice Hall, Upper Saddle River, NJ.

[7] Burdick, R. K., and Reneau, J. H. (1978). *The impact of different error distribution on the performance of selected sampling estimators in accounting populations.* In: Proceedings of the Business and Economical Statistics Section, 1978, pp. 779-781.

[8] Bureau International des Poids et Mesures (2004). *What is metrology?* Available: http://www.bipm.org/en/convention/wmd/2004/

[9] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML (2008). *Guide to the expression of uncertainty in measurement* (GUM). JCGM 100:2008, GUM 1995 with minor corrections. Available: http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

[10] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML (2008). *Supplement 1 to the 'Guide to the expression of uncertainty in measurement' – Propagation of distributions using a*

Monte Carlo method (GUM-S1). JCGM 101:2008. Available: http://www.bipm.org/utils/common/documents/jcgm/JCGM_101_2008_E.pdf

[11] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML (2011). *Supplement 2 to the 'Guide to the expression of uncertainty in measurement' – Extension to any number of output quantities* (GUM-S2). JCGM 102:2011. Available: http://www.bipm.org/utils/common/documents/jcgm/JCGM_102_2011_E.pdf

[12] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML (2012). *International vocabulary of metrology – Basic and general concepts and associated terms* (VIM). JCGM 200:2012. Available: http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf

[13] Casella, G. and Berger, R.L. (2002). *Statistical Inference*. 2nd Edition. Duxbury, Pacific Grove, CA.

[14] Clopper, C. J., and Pearson, E. S. (1934). *The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial*. Biometrika, Vol. 26, pp. 404-413.

[15] von Collani, E., and Dräger, K. (2001). *Binomial Distribution Handbook for Scientists and Engineers*. Birkhäuser, Boston.

[16] von Collani, E., and Dumitrescu, M. (2001). *Complete Neyman measurement procedure*. Metrika 54, pp. 111-130.

[17] von Collani, E., Dumitrescu, M. and Lepenis, R. (2001). *Neyman Measurement and Prediction Procedures*. Economic Quality Control 16, pp. 109-132.

[18] von Collani, E., Dumitrescu, M. and Panaite, V. (2002). *Measurement Procedures for the Variance of a Normal Distribution. Economic Quality Control* 17, pp. 155-176.

[19] von Collani, E. (2004). *Theoretical Stochastics*. In: v. Collani, E. (ed.), Defining the Science of Stochastics. Heldermann Verlag, Lemgo, pp. 147-174.

[20] von Collani, E. (2004). *Empirical Stochastics*. In: v. Collani, E. (ed.),

Defining the Science of Stochastics. Heldermann Verlag, Lemgo, pp. 175-214.

[21] von Collani, E. (2006). *Jacob Bernoulli Deciphered*. Bernoulli News – Newsletter of the Bernoulli Society for Mathematical Statistics and Probability. Vol 13, No 2.

[22] von Collani, E. (2008). *Review of Statistical Standards and Specifications*. Economic Quality Control 23, pp. 123-149.

[23] Commission on Physical Sciences, Mathematics, and Applications (CPSMA) (1988). *Statistical Models and Analysis in Auditing: A Study of Statistical Models and Methods for Analyzing Nonstandard Mixtures of Distributions in Auditing*. National Academy Press, Washington, D. C.

[24] Cox, D. R., and Snell, E. J. (1979). *On Sampling and the Estimation of Rare Errors*. Biometrika, Vol. 66, No. 1, pp. 125-132.

[25] Dowson, D. C., and Wragg, A. 1973. *Maximum-Entropy Distributions Having Prescribed First and Second Moments*. IEEE Transactions on Information Theory, IT-19, pp. 689-693.

[26] EURAMET e. V. (2008). *Metrology – In Short*. 3rd Edition. Available: http://www.euramet.org/

[27] Gamboa, F., Loubes, J.-M., and Rochet, P. (2009). *Maximum Entropy Estimation for Survey sampling*. arXiv:0909.4046v1.

[28] Garstka, S. J. (1977). *Models for Computing Upper Error Limits in Dollar-Unit Sampling*. Journal of Accounting Research, Vol. 15, Issue 2, pp. 179-192.

[29] Gnedenko, B.V. and Kolmogorov, A.N., translated by Chung, K.L. (1968). *Limit Distributions for Sums of Independent Random Variables*. Revised Edition. Addison-Wesley, Reading.

[30] Godfrey, J., and Neter, J. (1984). *Bayesian Bounds for Monetary Unit Sampling in Accounting and Auditing*. Journal of Accounting Research, Vol. 22, Issue 2, pp. 497-525.

[31] Göb, R., and Sans, W. (2018). *Two-sided confidence intervals for the mean of zero inflated populations under prior information. Submitted*

[32] Goodfellow, J. L., Loebbecke, J. K., and Neter, J. (1974). *Some Perspectives on CAV Sampling Plans*. CA Magazine, Part I, October 1974, pp. 23-30, Part II, November 1974, pp. 46-53.

[33] Ham, J., Losell, D., and Smieliauskas, W. (1985). *An Empirical Study of Error Characteristics in Accounting Populations*. The Accounting Review, Vol. 60, No. 3, pp. 387-406.

[34] Hausdorff, F. (1921). *Summationsmethoden und Momentfolgen. I. Mathematische Zeitschrift* 9, pp. 74-109.

[35] Hausdorff, F. (1921). *Summationsmethoden und Momentfolgen. II. Mathematische Zeitschrift* 9, pp. 280-299.

[36] Icerman, R., and Hillison, W. (1990). *Distributions of Audit-Detected Errors Partitioned by Internal Control*. Journal of Accounting, Auditing and Finance, Vol. 5, No. 4, pp. 527-543.

[37] Jaynes, E. T. (1957). *Information Theory and Statistical Mechanics*. Physical Review, Vol. 106, No. 4, pp. 620-630.

[38] Johnson, J. R., Leitch, R. A., and Neter, J. (1981). *Characteristics of Errors in Accounts Receivable and Inventory Audits*. The Accounting Review, Vol. 56, No. 2, pp. 270-293.

[39] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions, Volume 1*. 2nd edition. *John Wiley & Sons, New York.*

[40] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions, Volume 2*. 2nd edition. *John Wiley & Sons, New York.*

[41] Kacker, R. and Jones, A. (2003). *On use of Bayesian statistics to make the* Guide to the Expression of Uncertainty in Measurement *consistent*. Metrologia 40, pp. 235-248.

[42] Kaplan, R. S. (1973a). *A Stochastic Model for Auditing*. Journal of Accounting Research, Vol. 11, Issue 1, pp. 38-46.

[43] Kaplan, R. S. (1973b). *Statistical Sampling in Auditing with Auxiliary Information Estimators*. Journal of Accounting Research, Vol. 11, Issue 2, pp. 238-258.

[44] Killmann, F. and v. Collani, E. (2001). *A Note on the Convolution of the Uniform and Related Distributions and their Use in Quality Control*. Economic Quality Control 16, pp. 17-41.

[45] Kullback, S., and Leibler,R. A. (1951). *On information and sufficiency*. Ann. Math. Statist. 22, pp. 79-86.

[46] Lazo, A., and Rathe, P. (1978). *On the Entropy of Continuous Probability Distributions*. IEEE Transactions on Information Theory 24, Issue 1, pp. 120-122.

[47] Leitch, R. A., Neter, J., Plante, R., and Sinha, P. (1982). *Modified multinomial bound for larger number of errors in audits*. The Accounting Review, Vol. 57, No. 2, pp. 384-400.

[48] Lira, I. (2008). *The generalized maximum entropy trapezoidal probability density function*. Metrologia 45, pp. L17-L20.

[49] Lurz, K. (2015). *Confidence and Prediction under Covariates and Prior Information*. URL: http://nbn-resolving.org/urn:nbn:de:bvb:20-opus-122748

[50] De Martino, A. and De Martino, D. (2018). *An introduction to the maximum entropy approach and its application to inference problems in biology*. Heliyon 4 (2018) e00596.

[51] Matsumura, E. M., Plante, R., Tsui, K. W., and Kannan, P. (1991). *Comparative Performance of Two Multinomial-Based Methods for Obtaining Lower Bounds on the Total Overstatement Error in Accounting Populations*. Journal of Business & Economic Statistics, Vol. 9, No. 4, pp. 423-429.

[52] McCray, J. H. (1984). *A Quasi-Bayesian Audit Risk Model for Dollar Unit Sampling*. The Accounting Review, Vol. 59, Issue 1, pp. 35-51.

[53] Meikle, G. R. (1972). *Sampling in an Audit Context*. The Canadian Institute of Chartered Accountants, Toronto.

[54] Neter, J., and Godfrey, J. (1985). *Robust Bayesian Bounds for Monetary-Unit Sampling in Auditing*. Applied Statistics, Vol. 34, pp. 157-168.

[55] Neter, J., and Loebbecke, J. K. (1977). *On the Behaviour of Statistical Estimators When Sampling Accounting Populations*. Journal of the American Statistical Association, Vol. 72, No. 359, pp. 501-507.

[56] Neyman, J. (1935). On the problem of confidence intervals. *Ann. Math. Statist.* 6, 111-116.

[57] Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Phil. Trans. Roy. Soc. A* 236, 333-380.

[58] Pap, G., and van Zuijlen, M. C. A. (1995). *The Stringer Bound in Case of Uniform Taintings*. Computers & Mathematics with Applications, Vol. 29, No. 10, pp. 51-59.

[59] Pap, G., and van Zuijlen, M. C. A. (1996). *On the asymptotic behaviour of the Stringer bound*. Computers & Mathematics with Applications, Vol. 33, No. 5, pp. 37-54.

[60] Plante, R., Neter, J., and Leitch, R. A. (1984). *A Lower Multinomial Bound for the Total Overstatement Error in Accounting Populations*. Management Science, Vol. 30, No. 1, pp. 37-50.

[61] Plante, R., Neter, J., and Leitch, R. A. (1985). *Comparative performance of multinomial, cell and Stringer bounds*. Auditing: a journal of practice and theory, Vol. 5, No. 1, pp. 40-56.

[62] Qeadan, F., Kozubowski, T. J., Panorska, A. K. (2012). *The Joint Distribution of the Sum and the Maximum of IID Exponential Random Variables*. Communications in Statistics – Theory and Methods, 41:3, pp. 544-569.

[63] Rajan, A. et al. (2017). *Moments and maximum entropy method for expanded uncertainty estimation in measurements*. 2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 1-6.

[64] Ramage, J. G., Krieger, A. M., and Spero, L. L. (1979). *An Empirical Study of Error Characteristics in Audit Populations*. Journal of Accounting Research, Vol. 17, Supplement, pp. 72-102.

[65] Reneau, J. H. (1978). *CAV Bounds in Dollar Unit Sampling: Some Simulation Results*. The Accounting Review, Vol. 53, No. 3, pp. 669-680.

[66] Rényi, A. (1970). *Probability Theory*. North-Holland, Amsterdam.

[67] Sans, W. and v. Collani, E. (2003). *A Note on Monotonic Probability Distributions*. Economic Quality Control 18(1), pp. 67-81.

[68] Sans, W. (2011). *Sampling Methods and Market Surveillance*. Economic Quality Control, 26(2), pp. 107-119.

[69] Sans, W. (2013). *A Critical Review of Statistical Methods Used in Quality Control*. Economic Quality Control 27(2), pp. 97-142.

[70] Shannon, C.E. (1948). *A Mathematical Theory of Communication*. The Bell System Technical Journal 27, pp. 379-423, and pp. 623-656.

[71] Stringer, K. W. (1963). *Practical Aspects of Sampling in Auditing*. Proceedings of the Business and Economic Statistics Section of the American Statistical Association, December 1963, pp. 406-411.

[72] Teitlebaum, A. D., and Robinson, C. F. (1975). *The real risks in audit sampling*. Journal of Accounting Research, Vol. 13, pp. 70-91.

[73] Tsui, K. W., Matsumura, E. M., and Tsui, K. L. (1985). *Multinomial-Dirichlet bounds for Dollar-Unit Sampling in Auditing*. Accounting Review, Vol60, pp. 76-96.

[74] Vanecek, M. T. (1978). *An overview of decision theory for management accountants*. Management Accounting, Vol. 59, No. 7, pp. 41-48.

[75] *Virtual Laboratories in Probability and Statistics*, University of Alabama in Huntsville, Department of Mathematical Science, http://www.math.uah.edu/stat/

[76] Eric W. Weisstein. *Weierstrass Approximation Theorem.* From *MathWorld* – A Wolfram Web Resource. Available: http://mathworld.wolfram.com/WeierstrassApproximationTheorem.html

[77] Weigand, C. (2009). *Defining Precision for Reliable Measurement and Estimation Procedures*. Economic Quality Control 24(1), pp. 5-33.

[78] Wöger, W. (2003). *Information-based Probability Density Functions for a Quantity*. Measurement Techniques, Vol. 46, No. 9, pp. 815-823.

[79] Wolfram Research, Inc. Mathematica. Champaign, IL.