E-LEARNING PROGRAM STOCHASTKION MAGISTER

www.stochastikon.com

COURSE 2:

APPLICATION

Xiaomin Zhai & Elart von Collani
Stochastikon GmbH
Schießhausstr. 15, D-97072 Würzburg, Germany

Contents

Preface: Handling of Uncertainty

The first course of Stochastikon Magister is devoted to quantification and modeling of uncertainty which both are necessary for handling appropriately uncertainty. Handling uncertainty means to exploit or apply the stochastic model for developing procedures to solve the uncertainty related problems. Recently, uncertainty is vividly discussed in many conferences all over the world and many new "uncertainty theories" are developed. However, these discussions are not based on an unambiguous explanation about the nature of uncertainty. Consequently, an adequate quantification is not possible, and the proposed models have no unique interpretation.

As already stated in the first course there are two classes of procedures namely one which aims at throwing light on the future and the other one which aims to detecting unknown quantified facts. The second aim is dealt with in metrology, i.e., the science of measurement. Performing measurements is one of the most important activity in any human civilization and it is therefore not surprising that measurements belong to the sovereign tasks of states, where the national bodies of metrology supervise all measurement activities.

In Germany, for example, the *Physikalisch-Technische Bundesanstalt* (PTB) is the national metrology institute which provides scientific and technical services for questions releted to measurements. The PTB claims to measure with the "highest accuracy and reliability", but according to von Collani[1] fails to define both accuracy and reliability in a comprehensible way.

One of the main reasons for the shortcomings of conventional metrology, is the fact that randomness has not been taken account in the International System of Units (SI) that is used for quantification and measurement. As a direct consequence, measurement uncertainty cannot be dealt with in a scientific way. The International System of Units has been devised particularly by the *Bureau International des Poids et Mesures* (BIPM) in Paris. About the development of the SI is described by the BIPM as follows:

Following an international inquiry by the BIPM, which began in 1948, the 10th CGPM[2], in 1954, approved the introduction of the ampere, the kelvin and the candela as base units, respectively, for electric current, thermodynamic temperature and luminous intensity. The name International System of Units, with the abbreviation SI, was given to the system by the 11th CGPM in 1960. At the 14th CGPM in 1971, after lengthy discussions between physicists and chemists, the current version of the SI was completed by adding the mole as the base unit for amount of substance, bringing the total number of base units to seven.

Figure 1: The BIPM courtyard in Sèvres, France.

Unfortunately, randomness was not considered as a physical characteristics, although it can be ob-

---

[1]Collani, E.v. (2010): A Note on the XIX IMEKO World Congress Fundamental and Applied Metrology. *EQC* 24, 287 – 307.

[2]Conférence Générale des Poids et Mesures (General Conference on Weights and Measures).

served during any physical experiment and measurement process. This is surprising, since randomness was already quantified by Jakob Bernoulli 300 years ago, when there were no units for many of the characteristics that are included in the SI. As a consequence of neglecting randomness, the so-called measurement uncertainty could not be adequately taken into account.

Only in 1993, a preliminary guide[3] for handling measurement uncertainty was released by the ISO[4] after many years of preparation. Since then some revisions have been made, but still the provisional guide has not been acknowledged as an authoritative standard. The reasons are many weaknesses of the guide which lead to doubtful results and cause a lot of disputes. The here developed measurement procedures may be looked upon as a first but decisive step in the development of a sustainable stochastic metrology.

Actually, it is planned to further develop Stochastikon Magister and add in near future a third course which will be devoted to a stochastic metrology.

---

[3]Guide to the Expression of Uncertainty of Measurement.
[4]ISO = International Organization of Standardization.

Course 2: Application

Content and Aim of the Course *Application*

In Course 1 "Modeling Uncertainty" the general stochastic model of a real-word process as a part of evolution is introduced and named Bernoulli Space. The stochastic model covers ignorance as characteristic feature of mankind and randomness as characteristic feature of evolution. Of course, the derivation of the Bernoulli Space should not end in itself, but it should be applied for solving problems. However, applying an intricate mathematical model like a Bernoulli Space assumes extensive support by elaborated procedures, algorithms and software programs. The content of the second course of Stochastikon Magister consists of the procedures necessary for applying, i.e., utilizing a Bernoulli Space.

The possible problems related to uncertainty are identified in the course and adequate solution procedures are established. As a matter of fact, the two sources of uncertainty, i.e., ignorance and randomness define two types or classes of problems and, hence, two types of solution procedures. However, surprisingly, it turns out that any stochastic procedure is based on a prediction procedure, which, therefore, constitute the core of stochastics and of this entire course.

The course aims at enabling the learners in case of a given problem to identify the appropriate stochastic procedure and to find the necessary algorithmic support. It is assumed that the situation has been quantified by a Bernoulli Space as described in Course 1. A successful participation in this course will lead to a better understanding and handling of uncertainty. As an immediate consequence the decision making process is improved, which will ultimately have an effect in any field of the human society.

Unfortunately the conventional statistical software packages can hardly be used for solving adequately problems which involve uncertainty. In particular the important prediction problems cannot be handled by the offered algorithms, because statistics does not cover the stochastic model and consequently no prediction procedures can be derived. This fact can be viewed as the main difference between statistics and stochastics. Bernoulli Stochastics aims at predictions taking into account adequately the inherent variability, while statistics tries to remove or at least to reduce variability in order to make statements about the past.

The past is dealt with in Bernoulli Stochastics by reliable and accurate measurement procedures, which are developed based on suitable prediction procedures. Thus, Bernoulli Stochastics is developed as a unified scientific approach based on the Bernoulli Space and prediction procedures.

This second course dealing with stochastic procedures has not been completed so far, since the procedures to be used for verifying a given Bernoulli Space have not been developed. Verification procedures are therefore only indicated here. The corresponding modules and learning units will be added in a future version.

Module 2.1: Stochastic Procedures

Content and Aim of the Module *Stochastic Procedure*

This module explains the general concept of stochastic procedures and distinguishes important classes of procedures. The classes refer to the aim, which might be to look into the future or to look into the past. The former means to predict the future event, while the latter means to identify an event which has already happened.

Moreover, the question about the relevant features of a stochastic procedure is discussed. Stochastic procedures are looked upon as products, which should meet certain specifications. The answer to the above question allows to specify meaningful requirements for stochastic procedures, which make sure that the respective goal is reached.
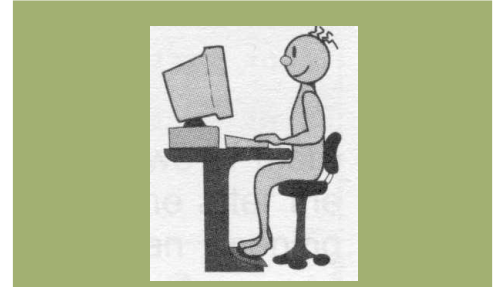
The module's aim is to give a preliminary idea of stochastic procedures and to prepare the reader for the subsequent modules which introduce explicitly a number of stochastic procedures.

This module should not only lead to a better understanding of stochastic procedures, but should also enable to evaluate systematically any other solution procedure, for example, statistical procedures. As illustrated in Course 1, Bernoulli Stochastics may be looked at as an alternative to statistics and, one of the aims of this module is to make the learner competent in order to judge statistical procedures.

Unit 2.1.1: General Classes of Procedures

TARGET

The main goal of Learning Unit 2.1.1 is to show that the Bernoulli Space allows to investigate the indeterminate future as well as the determinate past and to identify three classes of procedures.

CONTENT

Introduction

The classes mentioned in the title refer to different types of stochastic procedures, which may be derived from the different categories involved. The uncompleted evolution constitutes the first category and the second one is the completed evolution. The Bernoulli Space itself yields the description of the connection between these two categories and, thus, represents a third one.

The target of a procedure with respect to the above given categories is used as classification criterion. One class consists of procedures centered on the uncompleted evolution, another one on the completed evolution and, finally, a third one on the Bernoulli Space.

Note that any procedure is mathematically represented by a function, where the codomain determines the direction or the target and the domain the starting point.

Future and Past or Uncompleted and Completed Evolution

The Bernoulli Space $\mathbb{B}_{X,D}$ of the pair of variables (X, D) describes mathematically the relation between the completed evolution (past) represented by the deterministic variable D and the uncompleted evolution (future) represented by random variable X. Although, the relation is not at all one-to-one, i.e., deterministic, it nevertheless allows based on D to have a glance into the future represented by X or to look from the future X back into the past D.

Procedures which are based on the existing knowledge and aiming at reducing the uncertainty about the uncompleted evolution build up the

- *Prediction Class.*

On the other hand, procedures which are based on the future outcome attempting to reduce the ignorance about the determinate past, i.e., about facts build up the

- *Measurement Class.*

The word "measurement" is traditionally used for a procedure aiming at determining a fixed, but unknown value of a quantified characteristic. Note that quantification is the first necessary step, while measuring is only the second one. Without quantification there is no measurement.

Finally, procedures used for examining the Bernoulli Space itself build up the

- *Verification Class.*

A Bernoulli Space is the result of various exclusion procedures with respect to facts representing the relevant initial conditions. An erroneous exclusion of an actual fact leads to a wrong Bernoulli Space. A wrong Bernoulli Space yields results which are wrong with certainty. Therefore, a verification procedure aims at assuring that the Bernoulli Space covers the actual facts.

In contrast to the procedures of the prediction or measurement class, a verification procedure does not aim at controlling uncertainty or reducing ignorance. The only aim is to ensure that the Bernoulli Space covers the given situation.

Any procedure of the prediction class is completely based on a specified Bernoulli Space. No further experiments or observations are necessary. The entire relevant knowledge is used for developing the Bernoulli Space, which forms the basis for deriving the desired procedures of the prediction class.

Any procedure of the measurement class is used to reduce ignorance about facts. Reducing ignorance is tantamount to learning, which means that a learning process has to be performed. Thus, in contrast to procedures of the prediction class any procedure of the measurement class is based on a learning or measuring experiment, which provides new insight and enables to reduce the ignorance space by excluding further elements.

Analogously as in the case of the measurement class, the procedures of the verification class require additional experiments, too. Because, the verification deals with the Bernoulli Space, which describes facts, one could consider the verification class as part of the measurement class. However, the goals are different and, therefore, it makes sense to distinguish a third class.

EXAMPLES

1. **Prediction Class**

   The prediction class refers to the future and contains two types procedures. The first one determines a future event which will occur with a specified probability, while the second one determines the probability of a specified future event.

   (a) Consider the light bulbs of a given production process. The marketing department wants to make a TV advertisement for the bulbs by specifying its lifetime. Therefore, the variable of interest is given by the following random variable:

   $$X = \text{ lifetime of a light bulb} \tag{1}$$

   For developing the TV advertisement, a statement about a guaranteed lifetime of the bulbs is needed, which can be obtained by a stochastic prediction procedure.

   (b) For having a sound basis for their decision, the European Central Bank would like to be able to exclude the event of an inflation rate during the next year larger than 2%. The inflation rate is represented by the random variable X and of interest is the future event $E = \{x \mid x > 2\%\}$. What is needed is the probability of the event E. The question refers to a future development and, therefore, the procedure for determining the probability of the given event belongs to the prediction class.

2. **Measurement Class**

The measurement class refers to the past and consists of procedures which are used for reducing the ignorance space, i.e., for learning about the past.

(a) Consider you want to determine the actual weight. The weight is fixed and, therefore, a fact, which can be represented by a deterministic variable:

$$D = \text{ weight at a given time point} \tag{2}$$

The actual value of D can be determined by means of a measurement process in conjunction with a stochastic measurement procedure.

(b) Assume that you are the guide of a group of people and responsible that nobody gets lost during the various activities. Therefore, after each activity the actual number of people is of importance and must be determined. In this case the deterministic variable D is given by:

$$D = \text{ number of people after a specified activity} \tag{3}$$

The counting process constitutes a measurement procedure.

(c) Consider a police speed control:

The police checks the actual speed of the cars at a certain location on the highway with a speed limit. The deterministic variable D is given by:

$$D = \text{ actual speed of a specific car}$$

The actual speed of a car is a fact and, therefore a measurement procedure is needed.

Figure 1: Speed measurement device of the police.

(d) A pharmaceutical firm wants to prove that the efficiency of a new drug is higher than the efficiency of an old one. The effect of the new drug is represented by a random variable X_{new}, while the effect of the old drug is represented by X_{old}. Let the efficiency of a drug be defined by its first moment, which is commonly called expectation. Therefor we have to consider the following two deterministic variables:

$$\begin{aligned} D_1 &= E[X_{new}] \\ D_2 &= E[X_{old}] \end{aligned} \tag{4}$$

The problem is to show that the actual value of D_1 is larger than the actual value of D_2. Therefore, the problem refers to facts and the corresponding procedure belongs to the measurement class.

Unit 2.1.2: The Prediction Class

The main goal of Learning Unit 2.1.2 is to prepare the learner for the more mathematical treatment of prediction procedures in the subsequent module. As will become clear later prediction procedures constitute the unifying element in Bernoulli Stochastics and understanding the concept of prediction it therefore of utmost importance for understanding the entire stochastic approach.

Content

Introduction

The procedures of the prediction class aim at reducing uncertainty about the future development and, therefore, refer to the random variable X. They are based on the Bernoulli Space, i. e., on the entire knowledge about the initial conditions.

From a more practical point of view a user may want to know what will occur or what will not occur with respect to the random variable X, based on some knowledge or assumptions about the initial conditions and on experience about the corresponding future development.

In Course 1, the first step of quantification consisted of representing each of the characteristics of interest or relevance by a variable. The second step of quantification referred to the relations between past and future, and the appropriate means for representing them were identified as functions.

Application means to utilize the relations between past and future, which again is done by functions.

Prediction Procedures as Functions

In this introductory module the functions representing stochastic procedures are not explicitly introduced. However, their main characteristics are briefly outlined for facilitating the understanding of the more mathematical introductions in the subsequent modules.

A prediction procedure is a function that aims at forecasting the future development based on some knowledge about the initial conditions. Any level of knowledge is represented by a subset of the ignorance space $\mathcal{D}$ specified within the Bernoulli Space. Thus, the domain of any element of the prediction class is given by a suitable system of subsets $\mathcal{T}_D(\mathcal{D})$ of the ignorance space, where the symbol $\mathcal{T}$ denotes throughout this course a function assigning to a given set a suitable systems of subsets.

Note that the suitability of the system of subsets should be looked upon from a practical view and not at all from a mathematical view. Therefore, the closesure of the system with respect to certain mathematical operations is of no relevance at all. The only important factor for selecting the different levels of knowledge and, thus, the system of subsets should refer to the underlying aim of applying the procedure.

The variability function $\mathcal{X}$ yields the entire variability of X for any given level of knowledge, say $\mathcal{D}_0 \in \mathcal{T}_D(\mathcal{D})$. Therefore, any meaningful forecast must necessarily be a subset of the range of variability given by $\mathcal{X}_X(\mathcal{D}_0)$, implying that the codomain of any element of the prediction class must necessarily be a system of subsets $\mathcal{T}_X\big(\mathcal{X}(\mathcal{D}_0)\big)$. Analogously as in the case of the selection of $\mathcal{T}_X(\mathcal{D})$ no mathematical arguments should be used for specifying $\mathcal{T}_X\big(\mathcal{X}(\mathcal{D}_0)\big)$. The elements should be selected in a way that they represent that type of event, which is most suitable to reach the underlying aim.

Preliminary Result

Any prediction procedure is a function defined on a system of subsets $\mathcal{T}_D(\mathcal{D})$ of the ignorance space, where each element of the system represents a special level of knowledge or equivalently a level of ignorance. The codomain of the function is a system of subsets $\mathcal{T}_X\big(\mathcal{X}(\mathcal{D})\big)$ of the corresponding range of variability of X, where the elements of the system represent the possible forecasts.

The diversity of prediction procedures is very limited, as the future with respect to a prediction procedure is represented solely by the random variable X. Thus, there are only two types of prediction procedures, which from a mathematical point of view are equivalent. The first type answers the question

  • What will happen with regard to X?

and the second type answers the question

  • What will not happen with regard to X?

The two questions are answered by subsets of the given range of variability of X, which represent the predictions. Note that in case of the first question a small event is preferred to a large one. In contrast the second question is answered in a better way by a large event.

EXAMPLE

1. **Production process**

    Consider a production process with the number of nonconforming items of one day shift being of interest. Assume that there is a Bernoulli Space describing the productions process with respect to the number of nonconforming items of one day shift:

    • pair of variables:

$$\begin{aligned}
X &= \quad \text{number of nonconforming items of one day shift} \\
D &= (D_n, D_p) \\
&\quad D_n = \text{number of produced item per shift} \\
&\quad D_p = \text{nonconformity probability}
\end{aligned}$$

    • ignorance space:
$$\mathcal{D} = \{(n,p) \mid \underline{n} \leq n \leq \overline{n}, \underline{p}(n) \leq p \leq \overline{p}(n)\}$$

    • variability function:
$$\mathcal{X}(\{(n,p)\}) = \{0, 1, \ldots, n\}$$

    • random structure function:
$$\mathcal{P}(\{(n,p)\}) = \{P_{X|\{(n,p)\}}\} \quad \text{with } X|\{(n,p)\} \sim Bi(n,p)$$

On the basis of the Bernoulli Space it is possible to make a reliable prediction about the number of nonconforming items produced during a day shift. Moreover, it becomes possible to meet any reliability requirement with respect to the predicted event.

2. **Tensile strength**

Consider the tensile strength of so-called M5 alloy tubes, which are used in the nuclear fuel production. The aspect of interest is the tensile strength of the produced tubes yielding the following Bernoulli Space:

- pair of variables:

$$
\begin{aligned}
X &= \quad \text{tensile strength of a tube at a given temperature } t_0 \\
D &= (D_n, D_p) \\
&\quad D_{\mu_1} = \text{ first moment of } X \\
&\quad D_{\mu_2} = \text{ second moment of } X
\end{aligned}
$$

- ignorance space:

$$
\mathcal{D} = \{(\mu_1, \mu_2) \,|\, \underline{\mu}_1 \leq \mu_1 \leq \overline{\mu}_1, \underline{\mu}_2(\mu_1) \leq \mu_2 \leq \overline{\mu}_2(\mu_1)\}
$$

- variability function:

$$
\mathcal{X}(\{(\mu_1, \mu_2)\}) = \{x \,|\, \underline{x}(\mu_1, \mu_2) \leq x \leq \overline{x}(\mu_1, \mu_2)\}
$$

- random structure function:

$$
\mathcal{P}(\{(\mu_1, \mu_2)\}) = P_{X|\{(\mu_1, \mu_2)\}} \quad \text{being a uni-model distribution}
$$

In this case one could make a reliable prediction with respect to the tensile strength of the tubes.

Unit 2.1.3: The Measurement Class

TARGET

Traditionally, measurements are looked upon as the core of quantified sciences and actually metrology, the science of measurement, is a key discipline for any human civilization. This learning unit shall introduce the basic concepts of stochastic measurements. Moreover, the deficiencies of many conventional measurement procedure shall be revealed.

CONTENT

Introduction

Making measurements possible constitutes the most striking result of quantification. Only measurements guarantee unique and controllable statements about facts and, thus, measurements form a prerequisite of any science, which wants to be more than a debating society.

Measurements refer to the past or in other words to the fixed values of certain characteristics, which have been quantified. In the framework of a Bernoulli Space, measurement procedures are related to the initial conditions, i. e., the actual value of the deterministic variable D, which determines essentially the three components of the corresponding Bernoulli Space.

It should be reemphasized that a measurement with respect to a given characteristic, often called measurand in metrology, requires quantification. Moreover, any measurement is necessarily based on a process that depend on the unknown value of the measurand. The history of mankind is tightly associated with the progress in measurement methods. In fact, measurement procedures were and still are the basis of any technological progress.

Nowadays, almost everything in our daily life has been quantified and accurate measurement devices have been developed and, therefore, quantification seems to be often a part of nature. However, this is not at all true. Quantification is a genuine achievement of mankind based on mathematics resulting in technology. Note that a majority of technical breakthroughs do not date back very much. For instance, the first thermometer, which actually worked sufficiently well, was developed only in the 18th century by the physicist Gabriel Fahrenheit (1686-1724).

Measurement Procedures as Functions

A measurement procedure aims at determining the actual value of a deterministic variable D or a component of it, where "determination" means to exclude all those values of D which are not consistent with the unknown value of the measurand. The consistency is determined by a measurement process, which depends more or less closely on the unknown value d to be determined.

The measurement process is modelled by a Bernoulli-Space, with the random variable X representing the future outcome of the process and the deterministic variable D containing the variable to be measured.

A measurement procedure assigns to each possible outcome of the measurement experiment a

subset of the ignorance space as the measurement result. Thus, any measurement procedure can be mathematically described by a function denote by C_D.

A measurement function C_D is defined on the observable outcomes of the measurement experiment, which is given by a suitable system of subsets $\mathcal{T}_X\big(\mathcal{X}(\mathcal{D})\big)$ of the corresponding range of variability of X. If the observations consist of single values, then $\mathcal{T}_X\big(\mathcal{X}(\mathcal{D})\big)$ is selected as the system of singletons, each standing for an observed outcome of the measurement process.

The ignorance space $\mathcal{D}$ specifies the potential values of D. Therefore, any meaningful measurement result must necessarily be a subset of the ignorance space implying that the codomain of any meaningful measurement function C_D must necessarily be a system of subsets of $\mathcal{D}$ denoted by $\mathcal{T}_D(\mathcal{D})$. The elements of the system should be selected in a way that they represent that type of measurement, which is most suitable to reach the underlying aim. In the univariate case the measurements are represented by intervals.

Preliminary Result

Any measurement procedure is defined as a function C_D with domain being the system of subsets $\mathcal{T}_X\big(\mathcal{X}(\mathcal{D})\big)$ of the range of variability of $X|\mathcal{D}$, where each element of the system represents an observed outcome of the measurement process. The codomain of a measurement function is a system of subsets $\mathcal{T}_D(\mathcal{D})$ of the ignorance space, where each elements represents a possible measurement result.

Types of Measurement Procedures

In contrast to the prediction class, there are traditionally several different types of procedures contained in the measurement class:

- Genuine measurement procedures: These procedures aim at reducing the ignorance space as much as possible.

- Exclusion Procedures: This type of procedure aims at excluding a part of the ignorance space, which has been specified beforehand.

- Classification Procedures: This type aims at determining the actual Bernoulli Space from a set of given alternative Bernoulli Spaces.

EXAMPLES

1. **Production process (1)**

   Consider Example 1 of the previous learning unit and assume that the ignorance about the nonconformity probability is too big implying that the predictions are not very useful. In such a case it is necessary to perform a measurement process in order to determine the unknown value p of the nonconformity probability.

   The measurement process may consist of producing a specified number of items and determine how many nonconforming items are among the produced ones.

   The Bernoulli Space for the measurement process is the same as in Example 1 except that there is no ignorance about the number of produced items. The Bernoulli Space enables to develop a stochastic measurement procedure C_D assigning to each observable outcome

a measurement result with respect to the unknown value of the nonconformity probability. Moreover, it is possible to make the procedure as reliable as necessary.

2. **Tensile strength**

   The M5 alloy tubes of Example 2 of the previous learning unit have to stand a certain tensile strength, say t_0, as otherwise it would be dangerous to use them. Therefor the smallest value $\underline{x}(\mu_1, \mu_2)$ is important and must be determined. In fact, if the manufacturing process is very good the specification t_0 should always be met.

   The measurement process consists of checking the tensile strength of a specified number of M5 alloy tubes with the aspect of interest being the minimum tensile strength among them.

   As soon as a Bernoulli Space is available the value $\underline{x}(\mu_1, \mu_2)$ can be measured in order to prove that the specification t_0 is met with certainty.

3. **Production process (2)**

   Consider the case that there is no interest to determine the unknown value of the nonconformity probability, but it shall be shown that the probability does not exceed a given value p_0.

   In this case the aim is to exclude all values larger than p_0 and, therefore, an exclusion procedure is needed. If the exclusion process is modelled by a Bernoulli Space, the exclusion procedure can be developed, which for each observable outcome gives the decision either to exclude or not exclude all values larger than p_0

Unit 2.1.4: The Verification Class

TARGET

This learning unit aims at emphasizing the importance to verify the correctness of a Bernoulli Space. An incorrect Bernoulli-Space does not cover the actually given situation either with respect ignorance or randomness and, therefore, the claimed reliability of the procedures may not be warranted.

CONTENT

Introduction

Before a special Bernoulli Space is acknowledged as part of stochastic science and used as basis for deriving predictions, it should be verified by application and by specially designed verification procedures.

The aim of a verification procedure is to show that the Bernoulli Space meets its purpose to provide β-predictions, i. e., predictions with a probability of occurrence not falling short of β.

Verification Procedures

Each of the components of a Bernoulli Space is the result of an exclusion operation. The aim of an exclusion operation is to exclude as much values as possible, but not to exclude the true ones. The true value of a deterministic variable, as well as the actual range of variability and the uncertainty generated by randomness should be covered by the Bernoulli Space. Thus, the task of a verification procedure is to assure that, the actual values have not been excluded.

If the Bernoulli Space does not cover the entire uncertainty generated by ignorance and randomness, then the required reliability of the corresponding procedures cannot be warranted. Therefore, verifying a Bernoulli Space means to verify the reliability requirements.

Any stochastic procedure derived from a Bernoulli-Space is based on a prediction procedure. Hence, it suffices to check the reliability levels of prediction procedures.

A verification yields new knowledge and, therefore, it must based on a learning process or better said on a verification process. A verification procedure aims at verifying the reliability levels of prediction procedures. If the Bernoulli Space covers the given situation with respect to uncertainty then the actual probabilities of predictions should always be larger then the required reliability level.

Verification Procedures as Measurement Procedures

A verification procedure aims at assessing the probability of a given prediction and comparing it with the claimed reliability level. The probability of a prediction is in any situation a fixed, i. e. it is a determined value, implying that a verification procedure aims at assessing facts. Thus, any verification procedure constitutes a special measurement procedure.

The Bernoulli-Space of a Verification Procedure

A verification procedure can be looked upon as a stochastic measurement procedure which is based on the Bernoulli Space of the verification process and a prediction procedure. The deterministic variable to be measured is the unknown reliability of the prediction procedure. The reliability of a prediction procedure is defined as the probability that the predicted events with respect to the random variable of interest X will actually occur.

Verifying a Bernoulli-Space $\mathbb{B}_{X,D}$ means to determine the probability of a predicted event $A_X(\mathcal{D})$. The verification process consists of a number of independent copies $(X_1|\{d\}, X_2|\{d\}, \ldots)$ of $X|\{d\}$ of the random variable X. The random variable of interest X_v of the verification process is the number of successes, i.e. the number of times the predicted event $A_X(\mathcal{D})$ will actually occur, when performing the verification experiment. The relevant deterministic variable D_v for the verification experiment is given by the number of copies of $X|\{d\}$ and the unknown probability of the predicted event $A_X(\mathcal{D})$.

Preliminary Result

A verification experiment consists of performing a number of experiments and determining how many times a given prediction actually occurs. The aim is to show that the unknown value of the success probability is not smaller than the required reliability of the prediction. The Bernoulli-Space describing this situation is given by the well-known Binomial Bernoulli Space.

EXAMPLE

1. **Manufacturing process**

   Consider a manufacturing process of Uranium pellets for nuclear fuel. The pellets must meet certain specifications as otherwise their safe performance cannot be guaranteed. The conformance of the pellets produced shall be proved by means of a stochastic model, i. e., a Bernoulli Space, for the relevant features of the pellets.

   The random variable refers to the features of the pellets to be produced:

   $$X = \text{ a relevant feature of a pellet to be produced} \tag{5}$$

   The deterministic variable refers to the moments of X. Assume that the probability distribution of X is uni-modal, then according to the principle of minimum information the following deterministic variable is obtained:

   $$D = (E[X], V[X]) \tag{6}$$

   Let the Bernoulli Space for (X, D) be given by:

   $$\mathbb{B}_{X,D} = (\mathcal{D}, \mathcal{X}, \mathcal{P}) \tag{7}$$

   Before, the Bernoulli Space can be used for guaranteeing the conformance of the produced pellets, it must be verified that the Bernoulli Space covers the entire uncertainty generated by ignorance and randomness. In particular, it must be proved that the ignorance space contains the actual, but unknown values (μ, σ^2) of the deterministic variable.

Unit 2.1.5: Specification

The main goal of Learning Unit 2.1.5 is to identify the relevant quality characteristics of stochastic procedures and define specifications for them. Stochastic procedures are products and their quality should therefore be safeguarded by checkable specifications. To this end the meaning of specifications with respect to quality is clarified.
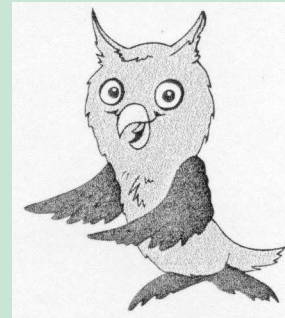
CONTENT

Introduction

A specification is a set of requirements referring to the relevant features of an object, e.g. a product, a system or a service. The specifications are often agreed in by the involved parties, generally the supplier of the product and the purchaser of the product. Meeting the specification should guarantee a certain performance of the product. Any violation of a specification constitutes a shortcoming and gives reason for a warranty claim.

In fact, specifications are necessary for legal security for supplier and purchaser. Specifications are often formulated in form of generally agreed standards which are produced by national and international bodies of standardization.

However, before specifications can be formulated the relevant features have to be identified, where a feature is called relevant, if it has a non-negligible impact on the performance of the product. In industry the relevant features of a product are often called quality characteristics.

Quality Characteristics of Stochastic Procedures

Consider a pair of variables (X, D), a corresponding Bernoulli Space $\mathbb{B}_{X,D}$ and a problem which shall be solved by means of a stochastic procedure being an element of the prediction class, an element of the measurement class or an element of the verification class, respectively. In each case the procedure yields a statement, either about the indeterminate future, or about the determinate past, or, finally, about the given Bernoulli Space.

Thus, a stochastic procedure is a system which produces statements and, therefore, for identifying the relevant quality characteristics for a procedures the corresponding statements, which constitute the performance of the stochastic products, have to be considered.

A procedure which more often yields correct statements than another procedure is called more reliable. Thus, reliability constitutes a major quality characteristic of a stochastic procedure.

Consider two procedures which are equally reliable. In this case the procedure generating statements which are more useful will be considered as better. Therefore, the second quality characteristic of a procedure is its usefulness or precision.

Finally, the procedures of the measurement class and the verification class are based on a process or experiment which afford certain expenses. Consider two procedures with the same reliability and the same precision, then the procedure with the lower expense will be called the better

one. Thus, the expense constitutes a third quality characteristic of stochastic procedures.

Specifications for Stochastic Procedures

A specification for a stochastic procedure must necessarily fix its reliability as an unreliable procedure is worthless no matter whether it is precise or cheap.

The specification may additionally include also the other two quality characteristics. However, one should have in mind that the quality characteristics depend on each other. A higher reliability requirement necessarily leads to a decrease in precision or to a higher expense in the case of the necessity of learning experiments.

The general understanding is that by setting specifications for a system or procedure a certain quality shall be guaranteed in case the specifications are met. Often quality of an object is defined by means of the corresponding specifications. This raises the question of the meaning or definition of the concept "quality," which shall be briefly examined in the following learning unit.

EXAMPLES

1. **Examination**

   Participating in an examination may lead to a success or a failure. Assume that the maximum points which can be achieved during the examination is given by 100 and that you take the test. The random variable is given by:

   $$X = \text{ the number of points you will achieve} \tag{8}$$

   The range of variability of X is given by:

   $$\mathcal{X} = \{0, 1, 2, \ldots, 100\} \tag{9}$$

   You will be successful, if you can meet the specification $\mathcal{S}$ for success, which is a subset of $\mathcal{X}$. For example, the success requirements could be given by by the following specification:

   $$\mathcal{S} = \{60, 61, 62, \ldots, 100\}$$
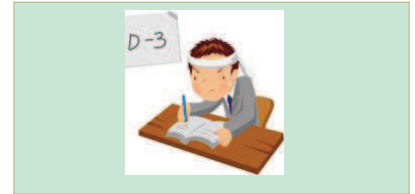
Figure 1: A difficult examination.

2. **Uranium pellet**

   The diameter of the Uranium pellets of Example 1 of the previous learning unit constitutes a relevant feature for a safe use as nuclear fuel. Thus, the random variable is given by:

   $$X = \text{ diameter of a produced Uranium pellet} \tag{10}$$

   The range of variability is limited due to the production conditions:

   $$\mathcal{X} = \{x \mid \underline{x} \le x \le \overline{x}\} \tag{11}$$

   The specification $\mathcal{S}$ is again given by a subset of $\mathcal{X}$:

$$\mathcal{S} \subset \mathcal{X} \tag{12}$$

If for a pellet the diameter falls into $\mathcal{S}$, it meets the corresponding specification and is called "conforming" otherwise it is called "non-conforming".

3. **Standards**

Many technical specifications are developed by standards organizations, especially by the International Organization of Standardization (ISO) which has developed thousands of technical and statistical standards (specifications) which are used worldwide and which enable global trade.

Specifications may be given by a mathematical description, by a software program, data sheet (specification sheet), or a drawing. Specifications are used in engineering, manufacturing and business, and they are of eminent importance for suppliers, purchasers and users of products, materials or services.

Unit 2.1.6: Quality

TARGET

This learning unit aims at clarifying the important concept of quality, which is used in numerous confused and confusing meanings. The learning unit shall illustrate that also the concept quality must be viewed stochastically.

CONTENT

Introduction

The concept "quality" is used in all parts of human societies and, therefore, it should be defined in an unambiguous way. Generally, good and bad quality is distinguished and, therefore, the concept itself is not valuing. Quality is an attribute of an object with regards to certain features of its performance when applied. The object could be a technical system, a human being, a procedure, or almost anything else.

The quality of the object refers to the future result of an application in view of a specified purpose. The result not only depends on the object itself, but also on the environmental conditions, i.e., the quality of an object changes with the circumstance of its application. Note, that the quality of an object generally elapses with the completion of the application, e.g., the quality of a roast beef refers to the state before the meal and not after it.

Thus, it is clear that any definition of the concept of quality must include a pair of quantities, one denoting the object and the other the characteristic of interest of its future performance. Moreover, the environment has to be taken into account.

Preliminary Result

The following implications can be drawn from the above considerations:

1. *Quality refers to a given object and some given features with respect to its application in a given environment.*

2. *Quality makes a statement about the future performance of the object with respect to a given purpose.*

Thus, the question arises how the future performance of an object with respect to the features of interest, which are called "quality characteristics" can be described using mathematical concepts.

As already known, the difficulty of describing a future development is generated by uncertainty.

The future performance is subject of uncertainty and uncertainty is the subject of this e-learning program. In the first course a stochastic model – the Bernoulli Space – for a process has been developed, with focus on describing uncertainty.

Stochastic Model of Quality

The Bernoulli Space is the stochastic model describing uncertainty about a future development. The Bernoulli Space relates to a pair of variable (X, D), where X is a random variable and D a deterministic variable.

The concept quality assumes the following quantities:

- a given object

- some performance characteristics of the object and

- the nature of the object with respect to the performance characteristics.

Considering quality, the two variables (X, D) of a corresponding Bernoulli Space have the following interpretation:

- X: The random variable X stands for the performance characteristics.

- D: The deterministic variable D stands for the nature of the object with respect to the performance characteristics.

The stochastic relation between nature of object/initial conditions and the performance of the object/future development is described by the Bernoulli Space:

$$\mathbb{B}_{X,D} = (\mathcal{D}, \mathcal{X}, \mathcal{P}) \tag{13}$$

As mentioned earlier quality is not valuing, but describes the future performance of an object. The future outcome is, because of uncertainty, not determined by the nature of the object and the initial conditions. However, the future performance is more or less strongly related to them. This stochastic relation is given by the probability distribution as image of the random structure function $\mathcal{P}$. Thus, the following quantitative representation of quality is obtained:

The quality of a given object for specified initial conditions $\{d\}$ with respect to its future performance X is given by the probability distribution $P_{X|\{d\}}$.

The above representation of quality illustrates the fact that it makes no sense to talk about the quality of an object without stating its purpose and the underlying situation.

EXAMPLES

1. **Quality of Cars**

   A Mercedes car is thought to be of a better quality than a Korean Kia car. In fact a Mercedes car is much more expensive than a Kia car, however, this is not the reason for believing that it is of better quality. The reason is different.

   The probability that a Mercedes car will have a good performance with respect to serving well its purpose is believed to be higher than for a Kia car. Thus, the future performance determines the quality of a car or any product. The future performance is adequately described by a probability distribution.

Figure 1: An expensive Mercedes sportscar.

The question whether or not the belief in the Mercedes cars is justified could be answered by determining the probability distributions about the future performance of a Mercedes car and a Kia car, respectively.

2. **Quality assurance**

Nowadays, quality assurance strategies are implemented not only in enterprises, but also in hospitals and even in universities. However, there are at least two problems which should be solved before a quality assurance strategy can be developed for a university. The first one refers to the definition of the quality of a university and the second one to procedures for determining the actual value of the quality. As long as these problems are not solved satisfactory, any quality assurance strategy must be looked at with suspicion.

Unit 2.1.7: Reliability

Target

> The reliability of stochastic procedures
> has already been mentioned in preced-
> ing learning units. This learning unit
> aims at introducing reliability as the
> most important quality characteristic
> particularly for stochastic procedures
> but in general for any object that is
> used for some given purpose.
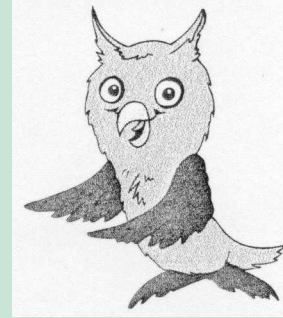
Content

Introduction

The more frequently a procedure yields correct results, the more it will be regarded as being
reliable. The observed frequency of correct results reflects the probability of the event of a
correct result. Hence, the probability of a correct result when applying a stochastic procedure
is called its reliability.

Reliability Specification

Any application of a procedure includes the risk of obtaining a false result. Therefore, the
reliability or at least a lower bound for the reliability must be known before a procedure may
be applied, as otherwise the risk is unknown and the results are more or less meaningless. The
controlled reliability of stochastic procedures distinguish them from unscientific procedures.
Because of the fundamental importance of reliability, it will generally be imposed on the pro-
cedure as a specification, implying that it is not reasonable to compare different procedures, if
they do not meet the same reliability specification.

Reliability Level

A reliability specification of a stochastic procedure is given by the reliability level, the value
of which is generally denoted β. The reliability level β constitutes a lower bound for the
probability of obtaining a correct result when applying the procedure in question.

Thus, any result of a stochastic procedure must necessarily come along with the procedure's
reliability. Note that the reliability refers to the procedure and not to the statement. Even
a highly reliable procedure may eventually produce a wrong result, just as in case of a highly
reliable machine or person.

Reliability in Science

It is a surprising fact that scientific procedures are generally not secured by reliability specifica-
tions. Consequently, it is hardly possible to evaluate scientific procedures. Statistics represents
one exception, as in statistics some procedures are safeguarded by a reliability requirements
given as "confidence level" or "significance level."

In almost any other part of science the reliability of procedures is not stated, maybe because
in the majority of cases the reliability is zero or close to zero. Consider as an example the
well-known "Laws of Nature" derived in physics which constitute procedures of the prediction

class as in general they make a statement about a future development based on some given initial conditions. The procedures are simple mathematical function yielding exactly one value for any given initial conditions. Thus, the precision of the Natural Laws is maximum. However, anybody who has made physical experiments knows that the calculated, i. e., predicted outcome differs with certainty from the observed one. Consequently the reliability of these procedures is zero and the achieved precision worthless.

Note that the reassuring remark that the difference between the predicted outcome and the observed one is not very big, is not helpful in many cases. Sometimes even a small difference may lead to a disaster. And not knowing how much the difference might be, is tantamount of using a procedure and not knowing the risks. Evidently, even in gambling the situation is better.

EXAMPLES

1. **Product Brand**

   A product brand is called more reliable than another if the products delivered operate more often well than the products from the second brand.

   Generally, specifications for certain relevant features of the products are defined. Meeting these specifications shall assure a good performance of the product with high probability.

   If a reliability level is fixed, then the specifications could be selected accordingly.

2. **Measurement Procedure**

   Any measurement procedure can be compared with a production process, where each application of the procedure yields a product in form of the measurement result. If the measurement procedure produces too frequently wrong results, it should not be used. In order to know the risk of getting a wrong result, each measurement procedure should state its reliability level, i. e., a lower bound for the probability of getting a correct result, when applying the procedure.

3. **Prediction Procedure**

   There are many types of prediction procedure. There are people who use procedures based on astrology, others believe in fortune tellers and others rely on scientific predictions. However, as long as no reliability level of the prediction procedure is stated, the risk that the predicted event will not occur is unknown and, therefore, relying on the prediction is hazardous.

Figure 1: Fortune telling by a crystal ball.

Unit 2.1.8: Precision

TARGET

Learning Unit 2.1.8 shall explain that the procedure's quality characteristic "precision" is only meaningful for specified reliability of the procedure. Moreover, it should become clear that precision quantifies the usefulness of a procedure.

CONTENT

Introduction

In the previous learning unit, the reliability of a procedure was introduced as the probability of getting a correct result when applying a given procedure. In the case of a prediction procedure a correct result consists of a predicted event that actually occurs. In the case of a measurement procedure a correct measurement result covers the true value.

A result of a stochastic procedure may be correct, but completely useless, because it is too imprecise and, therefore, cannot serve the underlying purpose sufficiently well. The precision of a procedure is defined as a measure how well the underlying aim is reached by applying the procedure. Different aims need different procedures and different measures of precision.

Procedure's Precision

Explaining quality of something as the degree of meeting its purpose suggests that the precision of a procedure may serve as optimality criterion for procedures meeting the same reliability specification.

If the underlying purpose requires a certain minimum precision, then a corresponding precision specification may be formulated and the procedure is selected accordingly.

Note that similar as in the case of reliability, precision refers to the procedure, too. Requirements for single stochastic statements, i.e., results of applying a procedure, make often not much sense. For example, in the case of a measurement procedure, the results are subject to randomness and, therefore, the achieved precision is generally also subject to randomness, which also shows that the quantification of precision of a stochastic procedure might constitute a difficult task.

Precision and Science

Reliability and precision are two quality characteristics of stochastic procedures, which are to be considered because of the inherent uncertainty about any future development. In classical, deterministic science the inherent uncertainty is neglected. As a consequence neither reliability nor precision are considered for scientific procedures, i.e., *laws of nature*. It is assumed that they represent truth and, therefore, must be correct. The belief in the truth of the *laws of nature* seems to imply that the procedures have a reliability of 1, and, moreover, the results are given by one point and, therefore, most precise.

The fact that the predictions made on the bases of *laws of nature* will never occur and the measurements never yield the true value is explained by stating that observations do not show the truth, but are always biased by an error.

Abandon the belief in truth yields the necessity of regarding the reliability of a procedure and its precision. The imprecision of a given result of a stochastic procedure is given by its size. The smaller the measurement result or prediction, the more precise it is, implying that the precision of a procedure's result is given by the reciprocal value of its size.

EXAMPLES

1. **Measurement**

   Consider to measure the length of an object by means of a ruler, which has a millimeter-scale. Assume that the result of the measurement is stated to be 15.6 cm. Then with almost certainty the result obtained is not correct. It follows that this measurement procedure based on the ruler, has a reliability of almost zero. On the other hand, the wrong measurement result is given by a single value, i. e., has optimum precision.

   However, achieving good precision for a procedure with low reliability is meaningless, as a decision based on a wrong result may lead to a disaster.

   In fact the ruler admits with a relative high reliability only results of a precision of at most 0.5 mm. Therefore, the result 15.6 cm should have been stated in the following form:

   $$\{x \mid 15.55 \leq x \leq 15.65\}$$

   The imprecision of such a measurement procedure would be $15.65 - 15.55 = 0.1$ cm, while the reliability should be determined by a suitable stochastic procedure.

2. **Prediction**

   It is easy to make a correct prediction by predicting an event which covers all the future developments. However, such a prediction would be generally useless. Therefore, the problem is to predict an event, which meets a given reliability specification and is as precise as possible.

   A weather forecast about tomorrow's temperature which covers a large interval is meaningless. On the other hand a forecast of one value will be wrong with certainty. Nevertheless, most weather forecasts consist of one value only.

   In Figure 1 a typical six-day weather forecast is displayed. For each day and each night a certain temperature (in Fahrenheit) is given. It remains unclear how to interpret the given values. For example, for Saturday daytime a temperature of 83° is given, is it the maximum temperature at noon, a mean temperature or anything else?
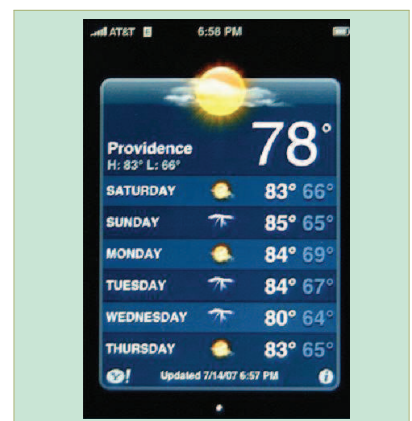

Figure 1: Weather forecast for night and day.

Unit 2.1.9: Expense

TARGET

Generally the expense for applying a procedure must be limited because of the available resources. This learning unit shall show the consequences of limiting the expense on the two other quality characteristics of a stochastic measurement procedure.

CONTENT

Introduction

The expense of a stochastic procedure quantifies the efforts which have to be made in order to execute it. For a procedure of the prediction class the efforts refer to the calculations needed to obtain the desired result based on the Bernoulli Space. If adequate tools are available, the expense should be negligible. In the case of a procedure of the measurement and the verification class, an additional process is needed.

If there are specifications with respect to the reliability and the precision of the procedure, then a minimum expense must be raised for meeting the specifications. In other words, the requirements with respect to the quality of a procedure determine the expense.

Expense and Sample Size

In statistics, procedures are always based on samples and the expense depends on the sample size. For given reliability level (confidence level) a large sample size, i.e., a large expense, leads to a precise result, while a small sample size, i.e., a small expense, yields an imprecise result.

Expense Specification

If the resources for performing the learning process in the frame of a procedure from the measurement or the verification class are limited, then an expense specification has to be imposed. The expense specification together with a reliability specification limits the attainable precision of the procedure.

Expense as Objective Function

The expense, for example the sample size for statistical procedures, may serve as the objective function, with side conditions with respect to reliability and precision. This case would lead to a optimization problem with two side conditions.

EXAMPLE

1. **Measurement Devices**

   Measuring a length by means of a ruler is cheap but rather imprecise. Using a laser based measurement device is much more costly, but the precision obtained is much higher. Note

that generally the reliability of technical measurement procedures are not stated on the measurement devices.

2. **Sample Size**

In stochastics (just as in statistics) samples are generally used as measurement processes. The larger the sample the more reliable and precise is the measurement procedure, because the inherent variability can be reduced for example by considering the arithmetic mean of the random variables comprising the sample. A more reliable procedure will yield more often correct results, and a more precise measurement procedure will yield results that are less imprecise.

Module 2.2: Prediction Procedures

Content and Aim of the Module *Prediction Class*

The ultimate purpose for deriving a Bernoulli Space is to predict the future development. Therefore, prediction procedures based on a Bernoulli Space build the core of Bernoulli Stochastics. The characteristic feature of stochastic prediction procedures consists of the reliability of the predictions made. These will occur with a probability which is not less than a required lower bound.

In this module various types of prediction procedures are introduced, where the type refers to the aim of the prediction and the level of ignorance assumed.

There are prediction procedures probably since mankind came into existence and in fact any decision made by a human being is based explicitly or implicitly on a prediction. Most of the prediction procedures developed by mankind is mysterious and doubtful, especially those which are in the context of science developed and known as *Laws of Nature*. Therefore, a more rational clarification about necessary and desirable properties of a prediction procedure should be elaborated. Moreover, a closer look to so-called scientific predictions as produced, for example in physics, reveals that they are - strictly speaking - meaningless and, therefore, useless.

A successful passing of this module should lead to a better understanding of the stochastic concept of predictions, which, as will be shown in the subsequent modules, are not only useful for disclosing the future, but are also the only way for learning about the past.

Unit 2.2.1: β-Prediction Procedures

TARGET

Learning Unit 2.2.1 aims at introducing a necessary requirement for reliable prediction procedures. This is important as the results of a prediction procedure with unknown reliability are more or less meaningless. Therefore, it is of utmost importance to understand the concept of reliable prediction procedures.

CONTENT

Introduction

The nature of the stochastic relationship between X and D resides in the probability distribution $P_{X|\{d\}}$, i.e., the probability distribution of the random variable X under the initial condition $\{d\} \subset \mathcal{D}$. It is easily appreciated that there are reciprocal relationships between the range of variability $\mathcal{X}(\mathcal{D})$ and the ignorance space $\mathcal{D}$; between subsets of these spaces, and therefore between so-called *predictions* and *measurements*, as defined previously. In Bernoulli Stochastics, these reciprocal relations are used for the development of a unified approach - based exclusively on predictions with respect to the random variable X - towards the development of any stochastic procedure, whether it belongs to the prediction, measurement or verification class.

Reliable Prediction Procedures

A prediction procedure utilizes knowledge about initial conditions contained in $\mathcal{D}$ and the relation between past and future, to generate a prediction for the future outcome in form of a subset of the range of variability $\mathcal{X}(\mathcal{D})$ of the random variable X.

More specifically, a prediction procedure denoted by A_X defines a mapping

$$A_X : \mathcal{T}_D(\mathcal{D}) \rightarrow \mathcal{T}_X\Big(\mathcal{X}(\mathcal{D})\Big) \tag{14}$$

where $\mathcal{T}_D(\mathcal{D})$ and $\mathcal{T}_X\Big(\mathcal{X}(\mathcal{D})\Big)$ are appropriately selected systems of subsets of the ignorance space and the corresponding range of variability, respectively. The elements of $\mathcal{T}_D(\mathcal{D})$ represent appropriately selected levels of ignorance, while the elements of $\mathcal{T}_X\Big(\mathcal{X}(\mathcal{D})\Big)$ represent the events to be predicted.

For a given prediction procedure and a given subset $\mathcal{D}_0 \in \mathcal{T}_D(\mathcal{D})$, a prediction is an element $A_X(\mathcal{D}_0) \in \mathcal{T}_X\Big(\mathcal{X}(\mathcal{D})\Big)$. A prediction error is committed, if the future realization x of X does not fall into the predicted event, i.e., $x \notin A_X(\mathcal{D}_0)$.

To control the reliability of the procedure, a lower bound - denoted by β and called reliability level - for the probability of a correct prediction is selected. As to the range of possible values for β, theoretically it could be the entire unit interval. However, a β value less than or equal to 0.5 does not make much sense. Therefore, β shall be restricted here to the interval $\{\beta \mid 0.5 < \beta \leq 1.0\}$. The value selected for β, will reflect the consequences of an error. If the

consequences are serious, then a value close to 1.0 or even the value 1.0 is indicated; otherwise, a smaller value may be sufficient. The lower bound β constitutes a requirement which must be taken into account when designing a procedure. The result is a general quality concept for procedures, which applies directly to the intended use, and which may be quantified and embedded in the derivation of procedures.

The Procedure's Reliability

A prediction procedure is called β-prediction procedure denoted by $A_X^{(\beta)}$ if its images $A_X^{(\beta)}(\mathcal{D}_0)$ meet the following requirement for any $\mathcal{D}_0 \in \mathcal{T}_D(\mathcal{D})$:

$$P_{X|\{d\}}\left(A_X^{(\beta)}(\mathcal{D}_0)\right) \geq \beta \quad \text{for } d \in \mathcal{D}_0 \tag{15}$$

From (15) it immediately follows that any prediction produced by a β-prediction procedure $A_X^{(\beta)}$ has a reliability of at least β, i.e., will occur with a probability of at least β.

On an average a β-prediction procedure $A_X^{(\beta)}$ will generate predictions, which will occur in at least $100 \cdot \beta$ % of all cases of application.

Graphical Illustration of a β-Prediction Procedure

In Figure 1 below, a β-prediction procedure

$$A_X^{(\beta)} : \mathcal{T}_D(\mathcal{D}) \to \mathcal{T}_X\left(\mathcal{X}(\mathcal{D})\right)$$

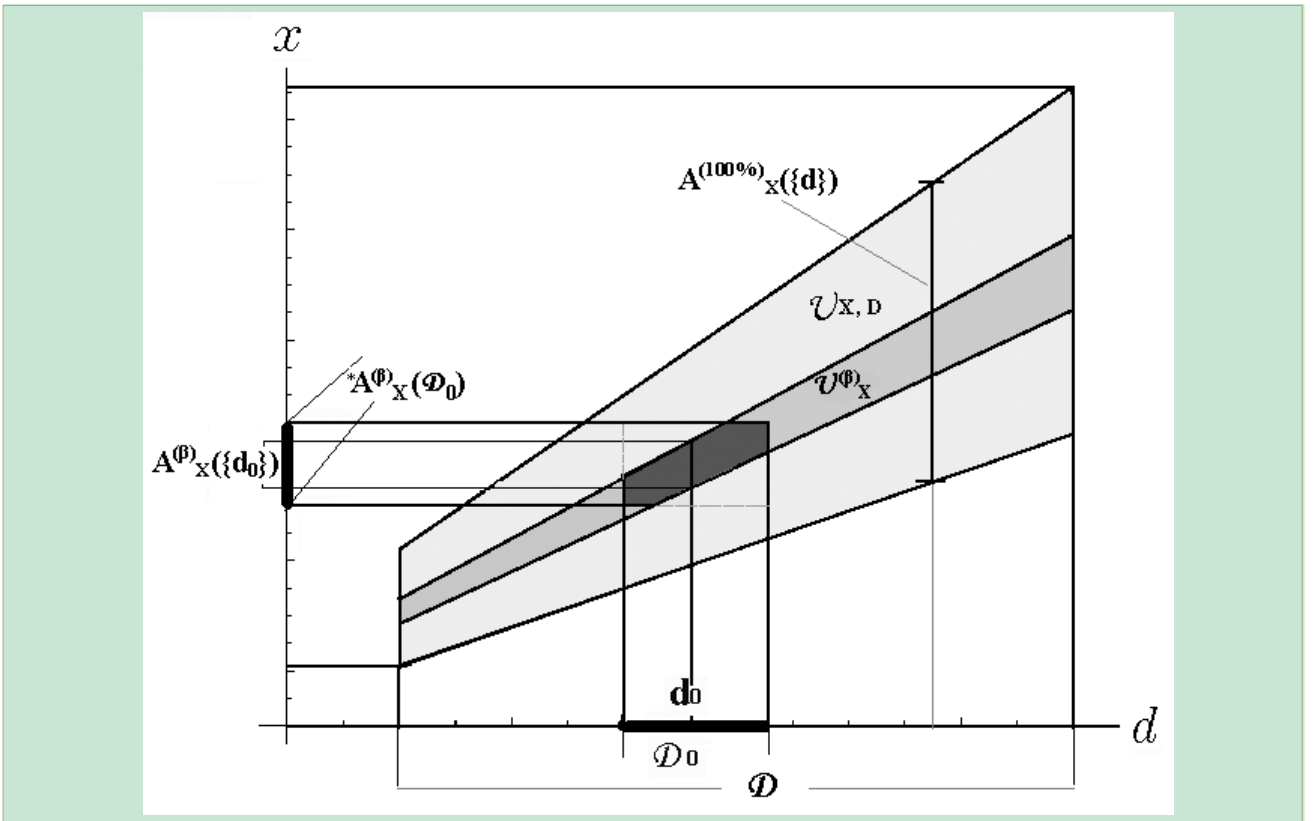is schematically given:

Figure 1: The β-Uncertainty Space $\mathcal{U}_X^{(\beta)}$ and the β-prediction procedure $A_X^{(\beta)}$.

Figure 1 shows two predictions plotted on the x-axis: The first one $A_X^{(\beta)}(\mathcal{D}_0)$ is based on the level of ignorance $\mathcal{D}_0 \in \mathcal{T}_D(\mathcal{D})$ while the second one $A_X^{(\beta)}(\{d_0\})$ assumes complete knowledge.

The two levels of ignorance are plotted on the d-axis, where $d_0 \in \mathcal{D}_0 \subset \mathcal{D}$. The light grey area represents the uncertainty space $\mathcal{U}_{X,D}$, which for $d \in \mathcal{D}$ shows all values $x(d)$ which might be observed. The smaller dark grey area is the so-called β-uncertainty space $\mathcal{U}_X^{(\beta)}$ which is given by the following set:

$$\mathcal{U}_X^{(\beta)} = \left\{ (d, x(d)) \mid d \in \mathcal{D}, x(d) \in A_X^{(\beta)}(\{d\} \right\}$$

The β-uncertainty space $\mathcal{U}_X^{(\beta)}$ shows for $d \in \mathcal{D}$ all values $x(d)$ with $x(d) \in A_X^{(\beta)}(\{d\})$, i.e., which might be observed with probability β. Finally the black area within the β-uncertainty space illustrates how the predictions $A_X^{(\beta)}(\{d_0\})$ and $A_X^{(\beta)}(\mathcal{D}_0)$ ar obtained:

- The prediction $A_X^{(\beta)}(\{d_0\})$ can directly taken from the β-uncertainty space.

- The prediction $A_X^{(\beta)}(\mathcal{D}_0)$ is the result of optimization, which starts with the event:

$$\bigcup_{d \in \mathcal{D}_0} A_X^{(\beta)}(\{d\})$$

with

$$P_{X|\{d\}} \left(\bigcup_{d \in \mathcal{D}_0} A_X^{(\beta)}(\{d\}) \right) > \beta$$

by stepwise reducing the set, the desired prediction $A_X^{(\beta)}(\mathcal{D}_0)$ is obtained.

EXAMPLE

1. **Special subsets of a prediction procedure**

   The simplest system of subsets of $\mathcal{D}$ is the system containing only $\mathcal{D}$ itself. In this case the prediction procedure degenerates to only one single prediction namely $A_X^{(\beta)}(\mathcal{D})$. There are two cases with respect to the amount of ignorance:

   (a) If the actual value d_0 of D is known, each set of the Bernoulli Space has just one element and the uncertainty space $\mathcal{U}_{X,D}$ reduces essentially to the corresponding range of variability $\mathcal{X}(\{d_0\})$ as uncertainty is due to randomness only. Clearly, in this case $\mathcal{T}_D(\mathcal{D})$ is necessarily given as $\{\mathcal{D}\}$ with $\mathcal{D} = \{d_0\}$.

   Let $A_X(\{d_0\})$ denote the prediction when d_0 is known, and let β represent the reliability requirement. The prediction procedure is a β-prediction procedure denoted $A_X^{(\beta)}$ if (16) holds.

   $$P_{X|\{d_0\}} \Big(A_X(\{d_0\}) \Big) \geq \beta \tag{16}$$

   (b) In the general case, i.e., with ignorance), it is only known that the actual value d_0 is an element of a given set $\mathcal{D}$ that contains more than one element. Then uncertainty is not only due to randomness but also due to ignorance.

   In this case a prediction procedure $A_X^{(\beta)}$ meets the reliability requirement given by β if the following condition holds.

   $$P_{X|\{d\}} \Big(A_X(\mathcal{D}) \Big) \geq \beta \quad \text{for } d \in \mathcal{D} \tag{17}$$

Unit 2.2.2: Optimum β-Prediction Procedures

TARGET

In Learning Unit 2.2.2 the precision of a prediction procedure is introduced as optimality criterion. The aim is to show that evaluating the precision of a prediction procedure is only meaningful in combination with the procedure's reliability.

CONTENT

Measuring Sets

Evaluating a stochastic procedure or its results is tantamount to 'measuring' certain subsets of the uncertainty space and, therefore, appropriate measure functions are needed, which have to be derived in accordance with the given situation and with respect to the two variables X and D.

As shown in previous learning units there are three aspects of importance when evaluating a stochastic procedure: *reliability*, *precision* and *expense*. The third aspect, expense, is beyond the scope of this course and it is assumed that the expense is fixed.

The probability measure is appropriate, and has already been adopted, for the evaluation of the reliability aspect, which constitutes one side of uncertainty. However, for assessing the precision aspect of stochastic procedures another measure must be derived.

Measuring Precision

A predictions $A_X(\mathcal{D}_0)$ is a non-empty subset of the range of variability $\mathcal{X}(\mathcal{D})$. The smaller the set, i.e., the predicted event, the more precise is the prediction. The most precise prediction is a set containing exactly one element. The most imprecise prediction consists of the entire range of variability $\mathcal{X}(\mathcal{D})$. Thus, precision of a prediction can be assessed by simply measuring the 'size' of the set that represents a predicted event.

The Size of a Set

There are two types of sets, discrete and continuous ones. The discrete sets contain a countable number of elements. The number may be finite or infinite. For example, $\{1, 2, 3, 4, 5\}$ is a discrete finite set, while the set of integers $\mathbb{N} = \{1, 2, \ldots\}$ is an infinite countable set. A continuous set contains an uncountable number of elements, as for example the set $\{x \in \mathbb{R} \,|\, 0 \leq x \leq 1\}$.

Obviously, for a discrete and finite set an appropriate measure function for evaluating its size is given by the number of elements. This measure is called the counting measure in probability theory.

In case of a continuous set, the so-called Lebesgue[5] measure is appropriate, which in case of an interval, say $\{x \in \mathbb{R} \,|\, a \leq x \leq b\}$ reduces to the interval length $\ell = b - a$.

---

[5]Henri Léon Lebesque (1975 – 1941), French mathematician.

Let $A_X^{(\beta)}(\mathcal{D}_0)$ be a predicted event. The size or precision of $A_X^{(\beta)}(\mathcal{D}_0)$ is denoted by $\left| A_X^{(\beta)}(\mathcal{D}_0) \right|$ and it is given by:

$$
\left| A_X^{(\beta)}(\mathcal{D}_0) \right| = \begin{cases} \text{number of elements of } A_X^{(\beta)}(\mathcal{D}_0) \text{ in the discrete case} \\[2ex] \text{Lebesgue-measure of } A_X^{(\beta)}(\mathcal{D}_0) \text{ in the continuous case} \end{cases} \tag{18}
$$

The Optimality Criterion

Any stochastic procedure belonging to a pair of variables (X, D) has a geometric representation as a subset of the corresponding uncertainty space $\mathcal{U}_{X,D}$. The precision of any given result is assessed by the size of the corresponding set.

A prediction procedure has two quality characteristics namely its reliability and its precision. The reliability of a β-prediction procedure is taken care by the reliability requirement given by the reliability level β. Hence, the only quality criterion left is the procedure's precision.

Optimum β-Prediction Procedures

The quality of a β-prediction procedure is given by its precision. Therefore, a β-prediction procedure is called optimal, if the predicted events are most precise or if their size is minimum.

The precision of a β-prediction $A_X^{(\beta)}(\mathcal{D}_0) \subset \mathcal{X}(\mathcal{D})$, which determines its quality, is given by its size

$$
\left| A_X^{(\beta)}(\mathcal{D}_0) \right| = \begin{cases} \displaystyle\sum_{x \in A_X^{(\beta)}(\mathcal{D}_0)} 1 & \text{for the discrete case} \\[3ex] \displaystyle\int_{A_X^{(\beta)}(\mathcal{D}_0)} dx & \text{for a continuous approximation} \end{cases} \tag{19}
$$

A β-prediction procedure $A_X^{(\beta)}$ is optimal or most precise, if for any $\mathcal{D}_0 \in \mathcal{T}_D(\mathcal{D})$ the size of $A_X^{(\beta)}(\mathcal{D}_0)$ is minimum among all β-prediction procedures.

Note that the continuous case represents always an approximation. The variability range of a realistic random variable is necessarily finite. If nevertheless a continuous approximation shall be used, it must be shown beforehand that the obtained results are useful.

The β-Uncertainty Space of a Prediction Procedure

A β-prediction procedure does not take into account those outcomes of the process in question, which occur with a sufficiently small probability. Neglecting these outcomes means to reduce the considered uncertainty or the resulting uncertainty space.

In the special case

$$
\mathcal{T}_D(\mathcal{D}) = \{\{d\} \mid d \in \mathcal{D}\} \tag{20}
$$

the uncertainty considered by a β-prediction procedure defines the β-uncertainty-space $\mathcal{U}_X^{(\beta)}$ (see preceding learning unit).

In this case the quality of the β-prediction procedure may be evaluated by the volume of the corresponding β-uncertainty-space $\mathcal{U}_X^{(\beta)}$.

$$
\left| \mathcal{U}_X^{(\beta)} \right| \tag{21}
$$

since

$$\mathcal{U}_X^{(\beta)} = \bigcup_{d \in calD} \left\{ (d, x(d)) \mid x(d) \in A_X^{(\beta)}(\{d\}) \right\}$$

is formed by the totality of the β-predictions of $A_X^{(\beta)}$.

EXAMPLE

1. **Determining an optimal prediction**

   Let the probability mass function $f_{X|\{d\}}$ be given as in Figure 1:
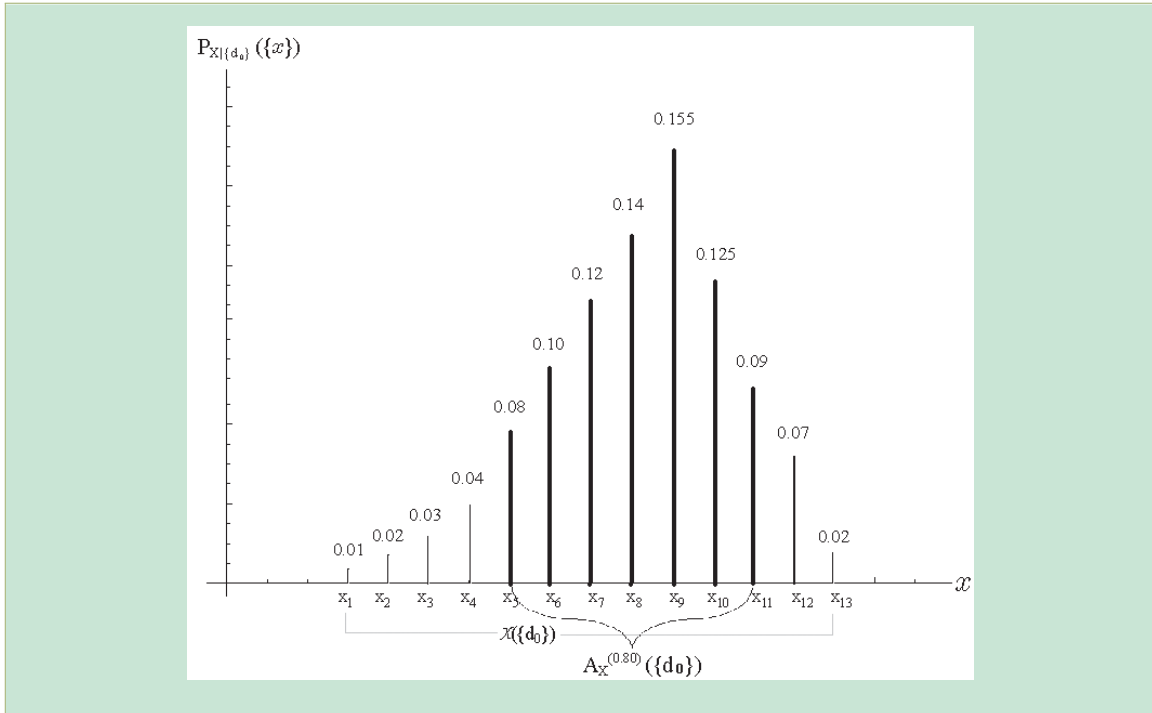

Figure 1: The probability mass function $f_{X|\{d\}}$ for $X|\{d\}$ with range of variability
$\mathcal{X}(\{d\}) = \{x_1, \ldots, x_{13}\}$.

Consider the problem of determining an optimal, i. e., a most precise β-prediction $A_X^{(\beta)}(\{d\})$ for $X|\{d\}$ and reliability level $\beta = 0.80$.

The first elements to be selected for $A_X^{(\beta)}(\{d\})$ is the outcome x_9 which occurs with largest probability, namely with $P_{X|\{d\}}(\{x_9\}) = 0.155$. Next x_8 is selected, because it has the the second largest probability $P_{X|\{d\}}(\{x_8\}) = 0.140$; then the third x_{10} with $P_{X|\{d\}}(\{x_{10}\}) = 0.125$, the forth x_7 with $P_{X|\{d\}}(\{x_7\}) = 0.120$, the fifth x_6 with $P_{X|\{d\}}(\{x_6\}) = 0.100$, the sixth x_{11} with $P_{X|\{d\}}(\{x_{11}\}) = 0.09$, and with the seventh x_5 having a probability of $P_{X|\{d\}}(\{x_5\}) = 0.080$ the sum of probabilities of the selected outcomes exceeds for the first time the required reliability of $\beta = 0.80$.

Taking any of the outcomes out of the predicted event, would lead to a occurrence probability smaller that the required reliability level. Therefore, the event $\{x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}\}$ is a desired optimum β-prediction.

Note that the optimum β-prediction is not unique. If x_5 is replaced by x_{12}, the event $\{x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\}$ also meets the reliability requirement given by $\beta = 0.80$ and is also most precise. However, the actual reliability of the prediction $\{x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}\}$ is larger and, therefore, it makes sense to select it.

Unit 2.2.3: Minimum Prediction Without Ignorance

TARGET

A β-prediction procedure derived from a Bernoulli Space based on complete knowledge constitutes the simplest case. In this learning unit a simple method is proposed for determining an optimum β-prediction procedure, i.e., a prediction procedure which yields the most precise predictions, in a situation without ignorance.

CONTENT

The System of Subsets $\mathcal{T}_D$

Any β-prediction procedure $A_X^{(\beta)}$ depends on the two involved systems of subsets, $\mathcal{T}_D$ and $\mathcal{T}_X$, where the former specifies the level of ignorance to be taken into account and the latter determines the shape of the predictions.

In the case of a β-prediction procedure without ignorance, complete knowledge is assumed for each single predictions. Complete knowledge is represented by an ignorance space containing exactly one element. Therefore, in this case the domain of the prediction function $A_X^{(\beta)}$ is given by the system of singletons:

$$\mathcal{T}_D(\mathcal{D}) = \{\{d\} \,|\, d \in \mathcal{D}\} \tag{22}$$

This case refers to the situation, where the future variability as function of the deterministic variable D is of interest.

The System of Sets $\mathcal{T}_X$

The sets to be predicted build up the system $\mathcal{T}_X(\mathcal{X}(\mathcal{D}))$, which is a system of subsets of $\mathcal{X}(\mathcal{D})$. Generally, the system of sets $\mathcal{T}_X$ is specified according to the underlying aim of the predictions. In such a case the images of a β-prediction procedure must be elements of the given system of sets $\mathcal{T}_X$. In case of a minimum β-prediction procedure, the system of sets $\mathcal{T}_X$ is not specified beforehand, but determined by the requirement to have minimum size, i.e., maximum precsion.

In order that the prediction procedure yields minimum β-predictions, the images $A_X^{(\beta)}(\{d\})$, i.e., the elements of the system of sets $\mathcal{T}_X$, must meet the following two conditions:

- The predictions must have the required reliability, i.e., they must be β-predictions for $d \in \mathcal{D}$.

- The predictions must have minimum size among all β-predictions for $X|\{d\}$ and $d \in \mathcal{D}$.

Determination of $A_X^{(\beta)}(\{d\})$

The above given conditions are met by determining the predictions $A_X^{(\beta)}(\{d\})$ in the following way:

1. For $d \in \mathcal{D}$ let $|\mathcal{X}(\{d\})| = N(d)$; order the elements of the range of variability $\mathcal{X}(\{d\})$ in descending order of the probability of the corresponding singletons yielding the ordered

set $\{x_{(1)}, x_{(2)}, \ldots, x_{N(d)}\}$ with

$$P_{X|\{d\}}(\{x_{(i)}\}) \geq P_{X|\{d\}}(\{x_{(i+1)}\}) \quad \text{for } i = 1, \ldots, N(d) - 1 \tag{23}$$

where the index of an element is called its rank (with respect to the probability of occurrence). If two or more singletons have the same probability the corresponding ranks are assigned appropriately[6].

2. The event to be predicted is filled with the elements $x_{(k)} \in \mathcal{X}(\{d\})$ in the order of their ranks starting with rank-number 1, i.e., $x_{(1)}$.

3. The procedure is stopped as soon as the cumulative probability of all selected outcomes reaches for the first time the reliability level β.

Proceeding like above described yields a prediction $A_X^{(\beta)}(\{d\})$ which on the one hand meets the required reliability level β and has minimum size on the other. The set of all minimum β-predictions builds the system of sets $\mathcal{T}_X$.

The Shape of Minimum β-Predictions

The shape of the elements of $\mathcal{T}_X$ depends on the situation. In the case that X represents a simple and not compound process, the probability distribution is a member of the constant, the monotonic or the uni-modal family. In this case the optimal predictions are intervals, where the lower and the upper bounds, say $x_{(k_1)}$ and $x_{(k_2)}$, are determined by the above optimization algorithm and the elements of the system of sets $\mathcal{T}_X$ have the following form:

$$A_X^{(\beta)}(\{d\}) = \left\{ x \mid x_{(k_1)} \leq x \leq x_{(k_2)}, x_{(k_1)}, x_{(k_2)}, x \in \mathcal{X}(\mathcal{D}) \right\}$$

for any $\{d\}$. The size of a prediction

$$\left| A_X^{(\beta)}(\{d\}) \right| = |k_2 - k_1|$$

is given by the number of its elements which is minimum among all β-predictions for the given Bernoulli Space.

If the probability distribution is a member of a more complex family, then the elements of $\mathcal{T}_X$ might consist of not connected intervals, i.e., if X represents a compound process, then the minimum β-predictions might adopt a variety of different shapes.

EXAMPLES

1. **Number of Female Customers**

   As part of a sales campaign a shop wants to give each female customer of the first 100 customers entering the shop after 1 p.m. a red rose and each male customer a blue forget-me-not. In order not to provide too many or too less roses and forget-me-nots a reliable prediction about the number of female customers among the 100 customers shall be made based on a Bernoulli Space. The reliability level is set to be $\beta = 0.95$.

   - The random variable X is given by:

     $X =$ number of female customers among the first 100 customers

   ---

   [6]'Appropriately' means that the ranks are assigned in a way which yields reasonable predictions.

- The deterministic variable D has two component D_1 and D_2 is given by

$$D_n = \text{Number of customers who are taken into account for getting a rose.}$$
$$D_p = \text{Probability that a customer is female.}$$

- The actual value D_n, i.e., the number of customers to be taken into account, is known to be $n = 100$. The actual value of the probability of a female customer is also assumed to be known, namely $p = 0.65$. Thus the following ignorance space for D_p is obtained:

$$\mathcal{D} = \{(100, 0.65)\}$$

- The variability function $\mathcal{X}$ is known exactly and is given by:

$$\mathcal{X}(\{(100, 0.65)\}) = \{x \,|\, 0 \le x \le 100, x \in \mathbb{N}\}$$

- The random structure function is also known exactly. The probability distribution of $X|\{(100, 0.65)\}$ is given by the binomial distribution $Bi(100, 0.65)$.

Next the elements $x \in \mathcal{X}(\{(100, 0.65)\})$ of the range of variability are ordered according to the probability of occurrence and at the same time accumulated in order to check when the reliability level $\beta = 0.95$ is reached.

x	65	66	64	67	63
$P_{X|\{(100,0.65)\}}$	0.0834047	0.082141	0.0810879	0.0774122	0.0755247
$\sum P_{X|\{(100,0.65)\}}$	0.0834047	0.165546	0.246634	0.324046	0.39957
x	68	62	69	61	70
$P_{X|\{(100,0.65)\}}$	0.0697685	0.0674218	0.0600905	0.0577142	0.0494214
$\sum P_{X|\{(100,0.65)\}}$	0.469339	0.536761	0.596851	0.654566	0.703987
x	60	71	59	72	58
$P_{X|\{(100,0.65)\}}$	0.0473922	0.0387814	0.0373447	0.0290091	0.0282479
$\sum P_{X|\{(100,0.65)\}}$	0.751379	0.79016	0.827505	0.856514	0.884762
x	73	57	56	74	
$P_{X|\{(100,0.65)\}}$	0.020664	0.0205164	0.0143112	0.0140021	
$\sum P_{X|\{(100,0.65)\}}$	0.905426	0.925943	0.940254	0.954256	

The most probable singleton is $\{65\}$, which, therefore, is the first element put into the prediction $A_X^{(0.95)}(\{(100, 0.65)\})$. In order to meet the reliability requirement given by the reliability level $\beta = 0.95$ the prediction is stepwise filled with elements larger and smaller than $x = 65$ yielding the prediction:

$$A_X^{(0.95)}(\{(100, 0.65)\}) = \{x \,|\, 56 \le x \le 74\}$$

for the number of female customers which has a reliability

$$P_{X|\{(100,0.65)\}}\left(A_X^{(0.95)}(\{(100, 0.65)\})\right) = 0.954256 > \beta$$

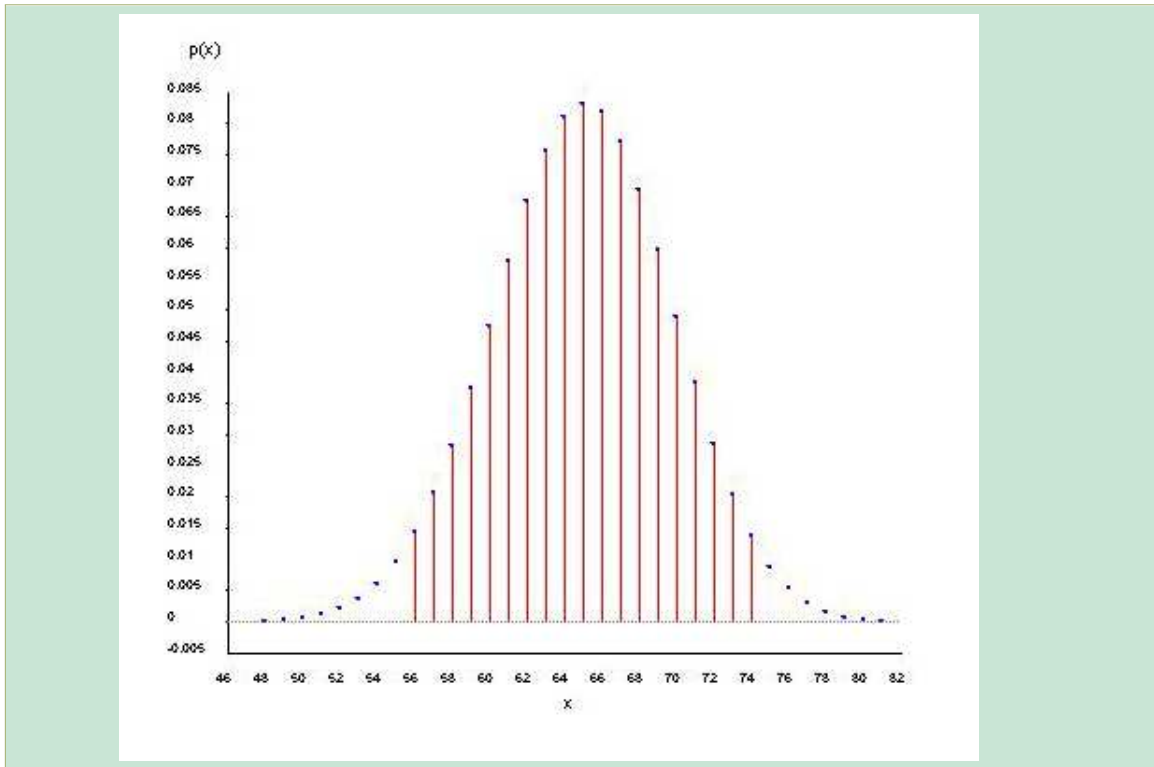which exceeds the required reliability. In Figure 1 the probability mass function $f_{X|\{(100,0.65)\}}$ is displayed.

Fig. 1: The probability mass function of $X|\{(100, 0.65)\}$ and the optimal prediction $A_X^{(0.95)}(\{(100, 0.65)\}) = \{x \,|\, 56 \leq x \leq 74\}$ meeting the reliability requirement given by the reliability level $\beta = 0.95$.

For the campaign the shop orders 74 red roses and 44 forget-me-nots. The risk that there will be more than 74 female customers is smaller than 5%.

2. Final Inspection of Product

A company produces a mass product and ships it in lots of $N = 10000$ items to certain customers all over the world. According to the shipping agreement the number of non-conforming items in a lot should not be larger than 20. The company has investigated the manufacturing process and concludes that the probability of producing a nonconforming item is $p = 0.001$.

In order to decide whether or not to inspect the outgoing lots, a prediction shall be made for the number of nonconforming items in a series of 10000 subsequently produced items.

a) The Bernoulli Space for the situation described above assumes complete knowledge and is given as follows:

– Random Variable:

$$X \;=\; \text{number of nonconforming items among the}$$
$$\text{considered production run}$$

– Deterministic Variable:
$$D = (D_1, D_2)$$

with
$$D_1 = \text{Size of the production run.}$$
$$D_2 = \text{Probability of producing a nonconforming item.}$$

– Ignorance Space:
$$\mathcal{D} = \{(10.000, 0.0001)\}$$

– Variability Function:
$$\mathcal{X}(\mathcal{D}) = \mathcal{X}(\{(10.000, 0.0001)\}) = \{0, 1, 2, \ldots, 10.000\}$$

– Random Structure Function:
$$\mathcal{P}(\{(10.000, 0.0001)\})) = P_{X|\{(10.000, 0.0001)\}}$$

with
$$P_{X|\{(10.000, 0.0001)\}} \sim Bi(10.000, 0.0001)$$

b) The company decides that there will be no costly final inspection if the prediction shows that the number of nonconforming items in a lot will be less than 20 items. Moreover it is decided that the reliability of the prediction should be not less than $\beta = 0.90$. It appeared that a 10% risk could be born, because not every delivered nonconforming item would be detected.

Entering Stochastikon Calculator, the following buttons have to be clicked:
– Binomial Distribution
– *Prediction*
– *Without ignorance*
– *Minimum interval*

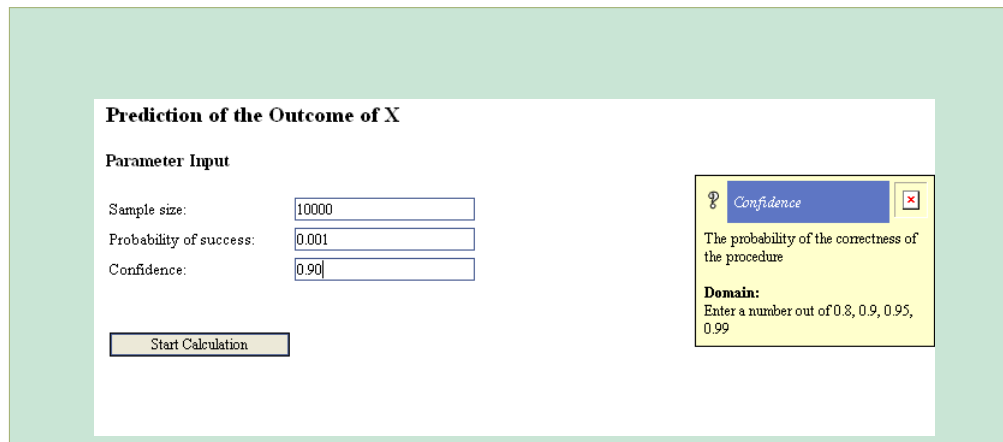then the following input-window opens:

Figure 2: Input-window of Stochastikon Calculator for calculating an optimal prediction in case of the Binomial Distribution.

The input values for the sample size $n = 10.000$, the nonconforming probability $p = 0.001$ and the confidence level $\beta = 0.90$ have to be inserted as shown in the Figure 2. After the *Start Calculation* button has been clicked, it takes some moments to calculate the prediction and produce the report, which contains a description of the situation, i.e., the entire Bernoulli Space, the posed problem and and the solution, i.e., the prediction given as a set. Moreover, the prediction is also illustrated by a graphical representation with relevant part of the probability mass function. The minimum β-prediction for the described case is given by

$$A_X^{(0.90)}(\{(10.000, 0.001)\}) = \{5, 6, \ldots, 15\}$$

Because the predictions does not include the number 20 or higher, it was decided to abandon the costly final inspection of the lots.

In Figure 3, which is part of the report provided by Stohastikon Calculator, the predicted event and the relevant part of the probability mass function are displayed.
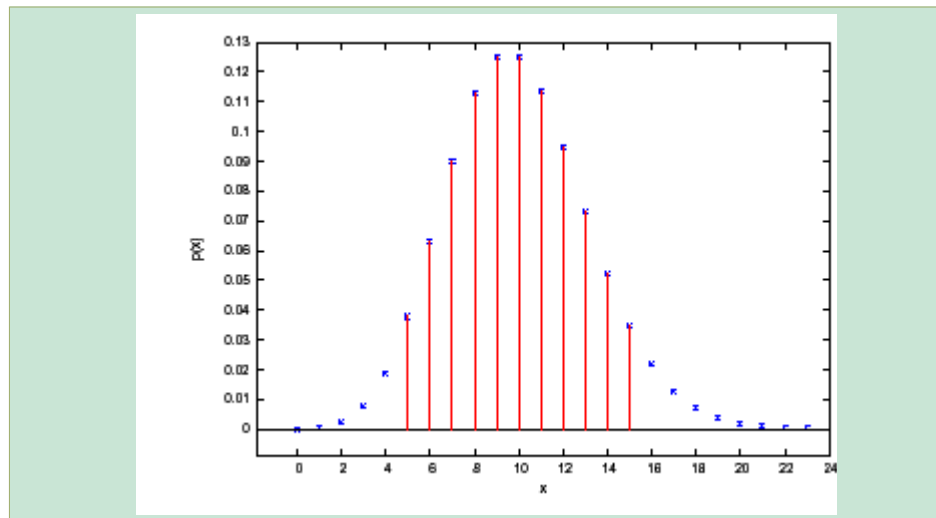

Figure 3: The probability mass function of $X|\{(10.000, 0.001)\}$
and the optimal prediction
$A_X^{(0.90)}(\{(10.000, 0.001)\}) = \{x \,|\, 5 \leq x \leq 15\}$.

Unit 2.2.4: Minimum Upper Bound Prediction Without Ignorance

TARGET

Learning Unit 2.2.4 aims at developing β-predicting procedures in the case that only a maximum is of interest and shall be predicted with given reliability level β. This case is relevant in many safety related situations.

CONTENT

Introduction

In the previous learning unit a method is given for deriving minimum β-prediction procedures assuming complete knowledge, but making no restricting assumption with respect to the random variable X. Thus, the described method is valid irrespective of the dimension of the characteristic of interest X.

Often, the random variable X is one-dimensional and, therefore, the possible outcomes follow the natural order of their size. In such a case the interest about the future outcome of X may be centered on the maximum possible outcome, while smaller outcomes are of less or even no significance. For example, if a decision shall be made upon the height of an embankment of a river, only the maximum height of a future flood is of interest.

The System of Subsets $\mathcal{T}_X$

If only the maximum of the possible values of the future development is of interest, then there is no need to predict simultaneously an upper and a lower bound for the future outcome. It is sufficient to predict an upper bound only. The lower bound is set for each d to the minimum value of the corresponding range of variability $\mathcal{X}(\{d\})$. Thus, the following system $\mathcal{T}_X$ is obtained:

$$\mathcal{T}_X(\mathcal{X}(\{d\})) = \left\{ x \in \mathcal{X}(\{d\}) \mid x \leq u^{(\beta)}_{X|\{d\}} \right\} \tag{24}$$

Clearly, the system of sets $\mathcal{T}_X$ given by (24) means less freedom for the selection of its elements. Each of the elements of $\mathcal{T}_X$ contains for example the smallest possible outcome of X. The only freedom refers to the largest element of the event to be predicted.

Optimum β-Prediction Procedure

Just as before, the optimum β-predictions with shape given by (24) shall have minimum size or, equivalently, a minimum upper bound $u^{(\beta)}_{X|\{d\}}$. The determination of the minimum upper bound is straightforward:

1. Order the elements of the range of variability $\mathcal{X}(\{d\})$ in ascending order of their values.

2. Fill the event to be predicted with elements starting from the smallest and proceeding to the larger ones.

3. Stop as soon as the cumulative probability of all selected outcomes equals or exceeds the required reliability level β for the first time.

Obviously, the above algorithm yields for each d a prediction

$$A_X^{(\beta)}(\{d\}) = \left\{ x \in \mathcal{X}(\{d\}) \mid x \leq u_{X|\{d\}}^{(\beta)} \right\} \tag{25}$$

which meets the reliability requirement and has a minimum upper bound.

EXAMPLE

1. **Maximum Load of a Wind Turbine**

In Example 3 of Learning Unit 1.3.8 the probability distribution of the maximum load for wind turbines has been derived. The random variable of interest was given by:

$$X_i = \text{ maximum load during one load cycle at wind condition } w_i$$

and the corresponding deterministic variable by:

$$D_i = (E[X_i], V[X_i]) \tag{26}$$

Once the Bernoulli Space $\mathbb{B}_{X_i, D_i}$ is available and a reliability level is determined, predictions can be made. In the case of the wind load the only prediction of interest is the maximum load which the turbines will have to stand. Of course, the maximum load should not be unrealistically large, as otherwise the production cost would increase. Therefore, the smallest load shall be predicted which will not be exceeded with a probability of at least the required reliability level, where the reliability level is a very large number close to 1.
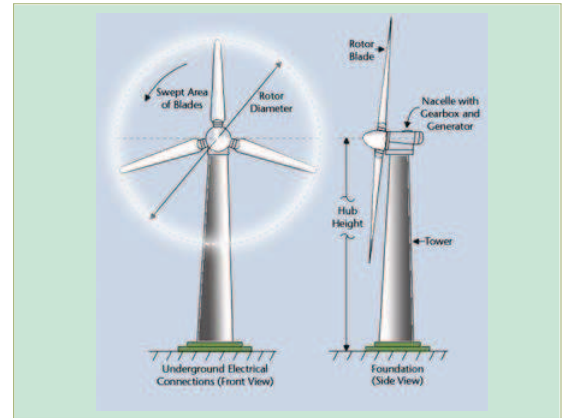

Figure 1: A wind turbine with the points of possible loads.

The smallest maximum load for given reliability level β is obtained by means of the upper quantile function of $X_i|\{d\}$. Thus, the wanted prediction is given as follows:

$$A_{X_i}^{(\beta)}(\{d\}) = \left\{ x \mid x \leq Q_{X_i|\{d\}}^{(u)}(\beta) \right\}$$

This prediction is also the optimum prediction with respect to the precision, as the upper quantile function is defined as:

$$Q_{X_i|\{d\}}^{(u)}(\beta) = \min_{x \in \mathcal{X}_i(\{d\})} \left\{ x \mid F_{X_i|\{d\}}(x) \geq \beta \right\} \tag{27}$$

Unit 2.2.5: Maximum Lower Bound Prediction Without Ignorance

TARGET

Often it is necessary to predict a minimum of an aspect of interest given by a random variable X. For example, the minimum life time of products is important for fixing the warranty time. This learning unit shows how a minimum can be predicted in a reliable way.

CONTENT

Introduction

In the previous learning unit a method is introduced for deriving minimum upper bound β-prediction procedures assuming complete knowledge. Such predictions are needed, if only the maximum outcome of X is of interest.

Besides, there are other cases in which only the minimum outcome is of interest. For example in order to make financial decisions the minimum income to be expected in future is of utmost importance and, hence, a reliable prediction would be very useful.

The System of Subsets $\mathcal{T}_X$

If only the minimum outcome of the future development is of interest, then – just as in the preceding learning unit – there is no need to predict simultaneously an upper and lower bound of the future development. It is sufficient to predict a lower bound only. The upper bound is set for each d to the maximum value of the corresponding range of variability $\mathcal{X}(\{d\})$. Thus, the following system $\mathcal{T}_X$ is obtained:

$$\mathcal{T}_X(\mathcal{X}(\{d\})) = \left\{ x \in \mathcal{X}(\{d\}) \mid \ell_{X|\{d\}}^{(\beta)} \leq x \right\} \tag{28}$$

Clearly, the system of sets $\mathcal{T}_X$ given by (28) means a restriction in the selection of its elements. Each of the elements of $\mathcal{T}_X$ contains for example the largest possible outcome of X. The only freedom refers to the smallest element of the event to be predicted.

Optimum β-Prediction Procedure

Just as before the optimal β-predictions of shape given by (28) have minimum size or, equivalently, maximum lower bound $\ell_{X|\{d\}}^{(\beta)}$. The determination of the maximum lower bound is straightforward:

1. Order the elements of the range of variability $\mathcal{X}(\{d\})$ in descending order of their values.

2. Fill the event to be predicted with elements starting from the largest and proceeding to the smaller ones.

3. Stop as soon as the cumulative probability of all selected outcomes equals or exceeds for the first time the required reliability level β.

Obviously, the above algorithm yields for each d a prediction

$$A_X^{(\beta)}(\{d\}) = \left\{ x \in \mathcal{X}(\{d\}) \, | \, \ell_{X|\{d\}}^{(\beta)} \leq x \right\} \tag{29}$$

which meets the reliability requirement and has a minimum upper bound.

EXAMPLE

1. **Prediction of the State Income**

   Each year an expert committee engaged by the Federal Republic of Germany produces a prediction of the next years state income. The outcome of the prediction is the basis for the expense strategy of the state and, therefore, it has an decisive influence on the state development.

   Generally the prediction has to be amended during the course of the year and in fact the realized income of the next year differs sometimes considerably from the predicted one. The error is not dramatic, if the income is higher than predicted, but if the actual income is much less than the predicted one the consequences may be serious.

   Obviously, what is needed is a prediction of a lower bound of the future income, which is guaranteed by a required and met reliability level β fixed by the government. An early prediction will yield a smaller lower bound than a late prediction, because of the larger uncertainty, unless something extraordinary happens, which changes fundamentally the future course of the economy.

   Let X_i be the ith type of state income and D_i the corresponding deterministic variable, the value d of which specifies the probability distribution $P_{X|\{d\}}$. Moreover, let $\mathbb{B}_{X_i, D_i}$ be a Bernoulli Space for (X_i, D_i). Then for given d the prediction for the state income is obtained as follows:

   $$A_{X_i}^{(\beta)}(\{d\}) = \left\{ x \, | \, x \leq Q_{X_i|\{d\}}^{(\ell)}(\beta) \right\} \tag{30}$$

   i. e., the lower bound for the state income of type i in case of the initial condition d is obtained as the lower quantile $Q_{X_i|\{d\}}^{(\ell)}(\beta)$. This lower bound will not be exceeded with probability β.

Unit 2.2.6: Minimum Prediction With Ignorance

Target

Learning Unit 2.2.6 shall illustrate the way how minimum β-prediction procedures are determined for the case of ignorance with respect to the actual value of the deterministic variable D. Clearly, in most real cases there is ignorance.

Content

Introduction

In case of ignorance, the arguments of a β-prediction procedure $A_X^{(\beta)}$, i.e., the elements of the system $\mathcal{T}_D$ are not singletons anymore, but have more than one element. In this case the derivation of appropriate β-predictions $A_X^{(\beta)}(\mathcal{D}_0)$ for $\mathcal{D}_0 \in \mathcal{T}_D(\mathcal{D})$ becomes more cumbersome than in the case without ignorance.

If a optimal β-prediction procedure $\tilde{A}_X^{(\beta)}$ without ignorance is available, a β-prediction procedure with ignorance is obtained by setting:

$$A_X^{(\beta)}(\mathcal{D}_0) = \bigcup_{d \in \mathcal{D}_0} \tilde{A}_X^{(\beta)}(\{d\}) \tag{31}$$

Clearly, the predictions obtained by (31) are β-predictions, because for $d \in \mathcal{D}_0$ the following relations hold:

$$P_{X|\{d\}}\left(A_X^{(\beta)}(\mathcal{D}_0)\right) \geq P_{X|\{d\}}\left(\tilde{A}_X^{(\beta)}(\{d\})\right) \geq \beta \tag{32}$$

i.e., the resulting prediction procedure (31) meets the reliability requirement given by the reliability level β.

However, in general a prediction given by (31) is not optimal. If the predictions $\tilde{A}_X^{(\beta)}(\{d\})$ are not identical for each $d \in \mathcal{D}_0$, then

$$P_{X|\{d\}}\left(A_X^{(\beta)}(\mathcal{D}_0)\right) > \beta \quad \text{for } d \in \mathcal{D}_0 \tag{33}$$

The reliability is larger than required implying that the size of the predicted event can be reduced without violating the reliability requirement.

Minimum β-Prediction Procedure

If the random variable X is one-dimensional and the predicted event $A_X^{(\beta)}(\mathcal{D}_0)$ is an interval, then a minimum β-prediction is obtained by alternately increasing the lower bound and decreasing the upper bound of the interval until any further reduction would lead to a violation of the reliability requirement given by β.

In the general case the determination of a minimum β-prediction procedure constitutes a technical-mathematical search problem, which will not be discussed here. The result will be a prediction $A_X^{(\beta)}(\mathcal{D}_0)$ with two properties:

- Any further reduction of the size of $A_X^{(\beta)}(\mathcal{D}_0)$ leads to a violation of the reliability requirement.

- Replacing any element of $A_X^{(\beta)}(\mathcal{D}_0)$ by a different element leads does not lead to an increase of the reliability of the prediction.

The second condition considers the fact that β-predictions with minimum size are not necessarily unique. In such a case a minimum prediction with highest reliability should be selected as the optimal one.

EXAMPLE

1. **Prediction of Number of Conforming Products**

   Consider the production run of $n = 100$ of a complex product. Of interest is the number of products produced which are conforming with the specifications. Hence, the random variable if interest is the following:

   $$X = \text{number of conforming items among the 100 produced ones} \tag{34}$$

   The items are produced independent of each other and under the same production conditions. Therefore, the process can be represented by a Bernoulli chain of length $n = 100$ and the deterministic variable D in this case is the probability of producing a conforming item:

   $$D = \text{probability of producing a conforming item} \tag{35}$$

   Then the corresponding Bernoulli Space $\mathbb{B}_{X,D}$ be given by the following components:

   $$\begin{aligned}
   \mathcal{D} &= \{p \mid 0.8 \le p \le 0.9\} \\
   \mathcal{X}(\{p\}) &= \{0, 1, 2, \ldots, 100\} \\
   \mathcal{P}(\{p\}) &= P_{X|\{p\}} \text{ with } X|\{p\} \sim Bi(100, p)
   \end{aligned} \tag{36}$$

   The prediction shall have a reliability of at least $\beta = 0.8$ and shall be most precise, i. e., the following mathematical problem has to be solved:

   Determine a set $A_X^{(0.80)}(\mathcal{D})$ satisfying the following two conditions:

   $$\begin{aligned}
   P_{X|\{d\}}\left(A_X^{(0.80)}(\mathcal{D})\right) &\ge 0.8 \quad \text{for } d \in \mathcal{D} \\
   \left|A_X^{(0.80)}(\mathcal{D})\right| &\overset{!}{=} \text{Minimum}
   \end{aligned} \tag{37}$$

   This problem can easily be solved by means of Stochastikon Calculator. After the input values have been entered, the calculator produces a report, which describes the situation by the corresponding Bernoulli Space and gives the wanted prediction:

   $$A_X^{(0.80)}(\mathcal{D}) = \{77, 78, \ldots, 93\} \tag{38}$$

   i. e., the production run will yield at least 77 conforming items and not more than 93 conforming items. The report also includes a graphical representation of the predicted event, which is given in Figure 1.
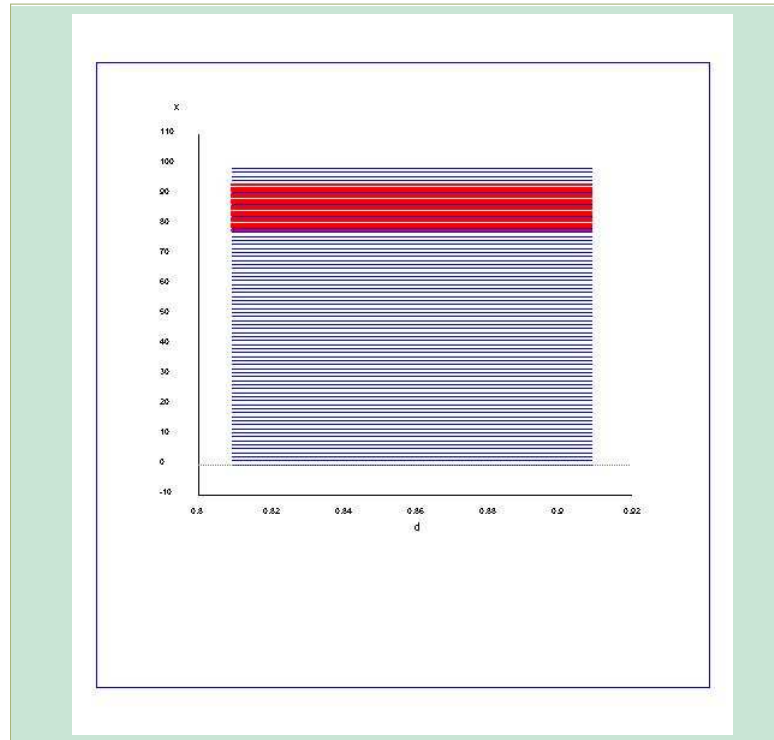
Figure 1: Graphical representation of $A_X^{(0.80)}(\mathcal{D})$

Figure 1 shows the uncertainty space, the ignorance space on the abscissa and the range of variability of the random variable $X|\{d\}$ on the ordinate. Note that in case of the binomial distribution with fixed value of n, the range of variability of $X|\{d\}$ does not depend on d. The prediction itself is highlighted in red.

Unit 2.2.7: Minimum Upper Bound Prediction With Ignorance

TARGET

> The Learning Unit 2.2.7 aims at showing that even in the case of ignorance, a reliable prediction of a maximum of X does not constitute a big problem. This case is of special importance in all safety-relevant situations, which cannot be dealt with appropriately with conventional methods used in engineering science.

CONTENT

Introduction

While the determination of minimum β-predictions in case of ignorance causes some technical troubles, there are no difficulties to determine optimal β-predictions for upper bounds. There are two reasons:

1. A prediction of an upper bound is meaningful only in the case of one-dimensional random variable X.

2. Generally, it is possible to identify exactly one value in $\mathcal{D}_0 \in \mathcal{T}_D(\mathcal{D})$, which is most unfavorable for the upper bound, i.e., if the reliability level is met for this value then it is met for all other values of $\mathcal{D}_0$.

If the second condition holds then the determination of the minimum upper bound prediction with ignorance reduces to the case "without ignorance".

Determination of a β-Prediction in Case of Ignorance

The level of ignorance is given by the set $\mathcal{D}_0 \in \mathcal{T}_D(\mathcal{D})$. The β-prediction for the upper bound $u_X^{(\beta)}(\mathcal{D}_0)$ is obtained in two steps:

- Identify the element $\bar{d} \in \mathcal{D}_0$, which is most unfavorable for an upper bound in the sense that $\bar{d}$ requires the largest upper bound:

$$u_X^{(\beta)}(\{\bar{d}\}) \geq u_X^{(\beta)}(\{d\}) \quad \text{for } d \in \mathcal{D}_0 \tag{39}$$

- Next proceed as in the case without ignorance for determining $u_X^{(\beta)}(\{\bar{d}\})$

The minimum upper bound β-prediction is given by

$$A_X^{(\beta)}(\mathcal{D}_0) = \left\{ x \in \mathcal{X}(\mathcal{D}_0) \,|\, x \leq u_X^{(\beta)}(\{\bar{d}\}) \right\} \tag{40}$$

At the time being, there are only software programs to determine the minimum upper bound β-predictions in the vase of the binomial distribution. IT is, of course, planned to develop programs that cover all types of probability distributions.

EXAMPLE

1. **Number of Successes**

A gambler knows that the success probability in a certain game of chance is between $p = 0.20$ and $p = 0.25$. He plans to play $n = 100$ times and would like to know how many times at most he will win.

The situation is quantified by the following variables and corresponding Bernoulli Space:

Pair of variables:
$$
\begin{array}{rcl}
X & = & \text{number of wins} \\
D & = & \text{success probability}
\end{array}
\tag{41}
$$

Bernoulli Space:
$$
\begin{array}{rcl}
\mathcal{D} & = & \{p \,|\, 0.20 \leq p \leq 0.25\} \\
\mathcal{X}(\{p\}) & = & \{0, 1, 2, \ldots, 100\} \\
\mathcal{P}(\{p\}) & = & P_{X|\{p\}} \text{ with } X_{P|\{p\}} \sim Bi(100, p)
\end{array}
\tag{42}
$$

The problem is to determine a minimum upper bound for the number of future wins, which will be exceeded with a probability not of less than $1 - \beta$, where β denotes the reliability level. This problem can be solved by means of Stochastikon Calculator, which yields the following prediction for $\beta = 0.90$:

$$
A_X^{(0.90)}(\mathcal{D}) = \{0, 1, 2, \ldots, 31\}
\tag{43}
$$

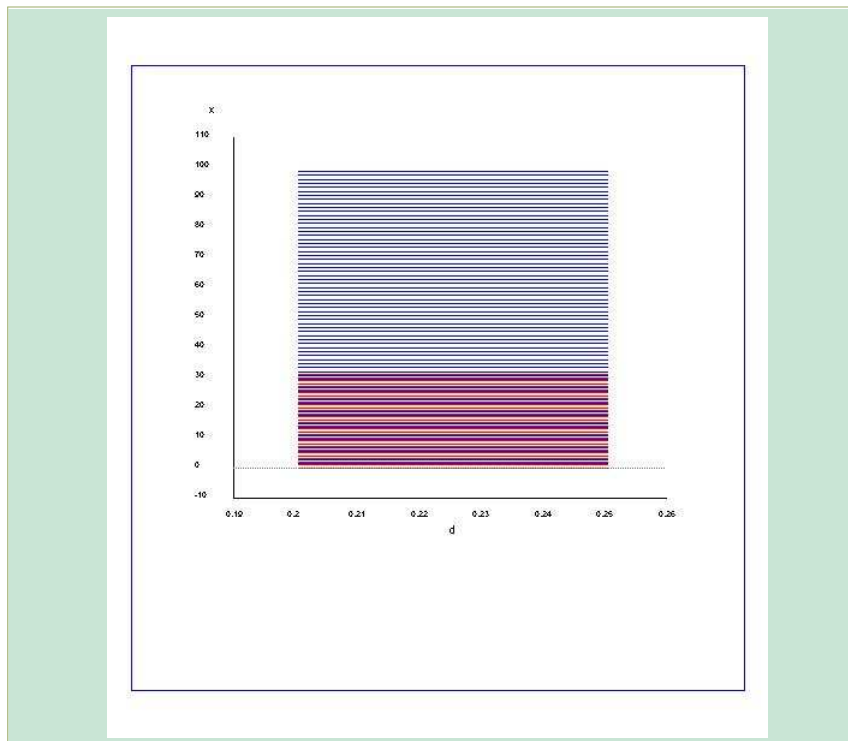The graphical representation of the predicted event is given in Figure 1 below:

Figure 1: Graphical representation of the minimum upper bound prediction $A_X^{(0.80)}(\mathcal{D})$

The graphical representation can be obtained by Stochastikon Calculator or by entering the Graphical Laboratory of Stochastikon Magister.

Unit 2.2.8: Maximum Lower Bound Prediction With Ignorance

TARGET

The Learning Unit 2.2.8 aims at show-
ing how in the case of ignorance given
by $\mathcal{D}_0 \subset \mathcal{D}$, a reliable prediction of
a minimum of X is obtained that will
be undercut only with prescribed small
probability.

CONTENT

Introduction

The determination of β-predictions for a lower bound of a random variable X is carried out
analogously to determining an upper bound. The problem is simple because

1. A prediction of a lower bound is meaningful only in the case of a one-dimensional random
   variable X.

2. Generally, it is possible to identify one value within $\mathcal{D}_0 \in \mathcal{T}_D(\mathcal{D})$ that is most unfavorable
   for the lower bound, i.e., yields the smallest upper bound of all values in $\mathcal{D}_0$.

Determination of a β-Prediction in case of Ignorance

The level of ignorance is given by the set $\mathcal{D}_0 \in \mathcal{T}_D(\mathcal{D})$. The β-prediction for the lower bound
$\ell_X^{(\beta)}(\mathcal{D}_0)$ is obtained in two steps:

- Identify the element $\bar{d} \in \mathcal{D}_0$, which is most unfavorable for a lower bound in the sense
  that $\bar{d}$ requires the smallest lower bound for meeting the reliability requirement:

$$\ell_X^{(\beta)}(\{\bar{d}\}) \leq \ell_X^{(\beta)}(\{d\}) \quad \text{for } d \in \mathcal{D}_0 \tag{44}$$

- Next proceed like in the case without ignorance for determining $\ell_X^{(\beta)}(\{\bar{d}\})$

The required β-prediction is given by

$$A_X^{(\beta)}(\mathcal{D}_0) = \left\{ x \in \mathcal{X}(\mathcal{D}_0) \,|\, \ell_X^{(\beta)}(\{\bar{d}\}) \leq x) \right\} \tag{45}$$

EXAMPLE

1. **Number of Successes**

   Consider the same situation as given in Example 1 of the previous learning unit. The
   gambler plans to play $n = 100$ games and knows that his success probability is between
   $p = 0.20$ and $p = 0.25$.

   He is interested in a lower bound of the number of wins. Of course, the lower bound
   should be as large as possible without violating the required reliability level of $\beta = 0.80$.

The variables as well as the Bernoulli Space are the same as in Example 1 of Learning Unit 2.2.7:

Pair of variables:
$$\begin{array}{rcl} X & = & \text{number of wins} \\ D & = & \text{success probability} \end{array} \tag{46}$$

Bernoulli Space:
$$\begin{array}{rcl} \mathcal{D} & = & \{p \,|\, 0.20 \le p \le 0.25\} \\ \mathcal{X}(\{p\}) & = & \{0, 1, 2, \ldots, 100\} \\ \mathcal{P}(\{p\}) & = & P_{X|\{p\}} \text{ with } X|\{p\} \sim Bi(100, p) \end{array} \tag{47}$$

In contrast to the previous learning unit, in this case a lower bound shall be predicted with reliability level $\beta = 0.80$. Again this problem is solved by means of Stochastikon Calculator, which for the case of the binomial distribution contains all types of prediction procedures.

After entering the values of the input parameters the following prediction with lower bound $x = 17$ is obtained:

$$A_X^{(0.80)}(\mathcal{D}) = \{17, 18, \ldots, 100\} \tag{48}$$

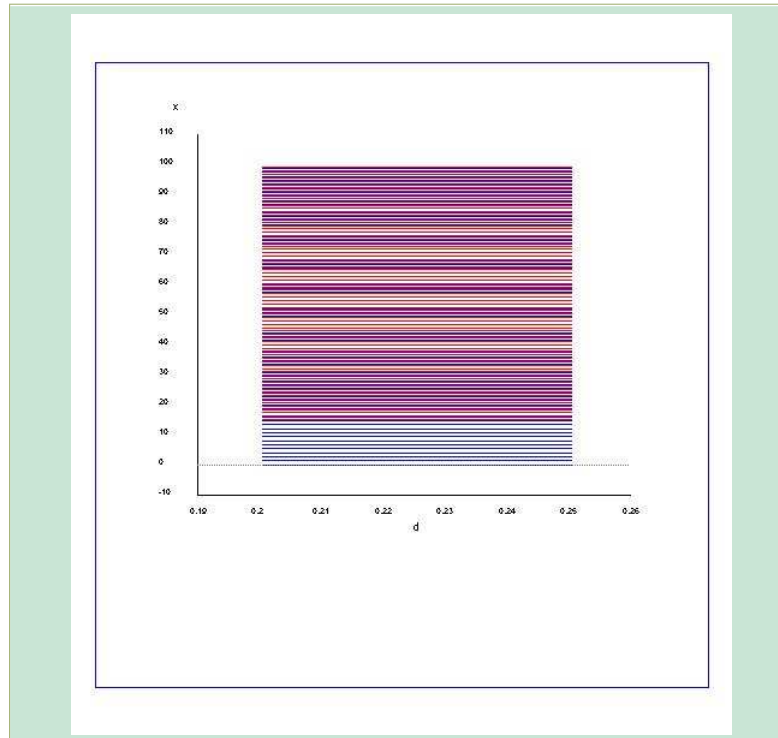The following graphical representation is obtained by means of the Graphical Laboratory:

Figure 1: Graphical representation of the maximum
lower bound prediction $A_X^{(0.80)}(\mathcal{D})$

Module 2.3: Measurement Procedures

Content and Aim of the Module Measurement Procedures

When Jakob Bernoulli introduced the concept of probability, he noted in his masterpiece Ars conjectandi that "To predict something is to measure its probability." Bernoulli wanted to say that a prediction is meaningless, if the probability that it actually will occur, is unknown. The probability is fixed by the initial conditions and determining a probability means to measure it. Measurement procedures are therefore the complement of prediction procedures.

The aim of this module is to develop and introduce a stochastic metrology, i. e., a stochastic science of measurement. This stochastic metrology could replace the conventional metrology which emerged as a branch of physics and therefore is not able to consider appropriately the inherent uncertainty in measurements. The quality of a measurement procedure depends essentially on the way that measurement uncertainty is taken into account. This module aims at developing general rules how to consider and model and finally how to compensate measurement uncertainty.

This module makes it possible to develop better measurement procedures, i.e., measurement procedures that meet a prescribed reliability requirement and are optimal with respect to measurement precision.

This module wants also to lay the foundation for a scientific handling of measurement uncertainty which could replace the already mentioned ISO *Guide to the Expression of Uncertainty in Measurement.* This guide was issued in 1993 and since then the the disputes about its appropriateness have not stopped. Since measurements are one of the most important issues for human civilizations, measurement uncertainty should be taken into account adequately. The fact that the GUM contains many different expressions for measurement uncertainty should be sufficient to abandon it. Moreover, even fundamental concepts used in the guide are not explained in a way that a user could understand them. For example, the concept "probability" is defined with the following words:
probability: a real number in the scale 0 to 1 attached to a random event.

Figure 1: Cover of the GUM as reprinted by the JCGM[7].

---

[7]JCGM = Joint Committee for Guides in Metrology.

Unit 2.3.1: β-Measurement Procedures

TARGET

Learning Unit 2.3.1 aims at introducing the necessary requirements for developing reliable and precise measurement procedures. Any learning process aims at revealing unknown facts, which is possible only by by means of a measurement procedure. However, any measurement procedure is necessarily based on an appropriate prediction procedure showing that without Bernoulli Stochastics it is not possible to develop a rational metrology.

CONTENT

Introduction

The ignorance space $\mathcal{D}$ represents the existing ignorance about the initial conditions given by d_0. If the ignorance space is too large, any β-prediction procedure will yield more or less useless predictions. In order to obtain more precise predictions, the existing ignorance about the actual value d_0 must be reduced based on a measurement process and a suitably selected observable random variable $X|\{d_0\}$.

A procedure for determining the actual value d_0 of D by reducing the ignorance space $\mathcal{D}$ is called measurement procedure. It is based on $X|\{d_0\}$ in the sense that those values of $\mathcal{D}$ are excluded which are not consistent with the observed outcome x of $X|\{d_0\}$, while those values which are consistent with the outcome constitute the measurement result and form the new reduced ignorance space.

Thus, a measurement procedure is a function denoted C_D with

$$C_D : \mathcal{T}_X(\mathcal{X}(\mathcal{D})) \to \mathcal{T}_D(\mathcal{D}) \tag{49}$$

A measurement procedure assigns to each observation given by an element of $\mathcal{T}_X(\mathcal{X}(\mathcal{D}))$ a measurement result given by an element of $\mathcal{T}_D(\mathcal{D})$.

The system of subsets $\mathcal{T}_X$ consists of the observations made with respect to the random variable X. If the realizations of X are observable, the observations are the singletons of $\mathcal{X}(\mathcal{D})$. Because this case is generally encountered in praxis, it shall be considered here, i.e.:

$$\mathcal{T}_X\left(\mathcal{X}(\mathcal{D})\right) = \{\{x\} \,|\, x \in \mathcal{X}(\mathcal{D})\}$$

Reliability of Measurement Procedures

A measurement procedure should yield measurement results $C_D(\{x\})$ which contain the unknown actual value d_0. The more frequently a measurement result obtained by a measurement procedure contains d_0, the more reliable is the measurement procedure.

Let $d_0 \in \mathcal{D}$ be the true, but unknown value of D. Then the reliability of a measurement procedure C_D is defined by the probability of obtaining a correct measurement result, i.e., a result $C_D(\{x\})$, which contains d_0. This probability is given by:

$$P_{X|\{d_0\}}\left(\{x \,|\, d_0 \in C_D(\{x\})\}\right) \tag{50}$$

β-Measurement Procedures

In an analogous manner as in the case of a prediction procedure, a reliability specification for measurement procedures is introduced by the *reliability level β*. The reliability level β specifies a lower bound for the procedure's reliability. A measurement procedure meeting the specification given by β, i.e.:

$$P_{X|\{d\}}\left(\{x \mid d \in C_D(\{x\})\}\right) \geq \beta \qquad \text{for } d \in \mathcal{D} \tag{51}$$

is called β-measurement procedure denoted $C_D^{(\beta)}$.

A β-measurement procedures yields correct results $C_D^{(\beta)}(\{x\})$ with a probability not less than β, where a correct measurement result is defined as one that contains the actual value of the deterministic variable.

β-Measurement Procedures and β-Prediction Procedure

Any measurement procedure can be looked upon as a method for excluding elements of $\mathcal{D}$ which are not consistent with the observation $\{x\}$. Not consistent means that those values d of D can be excluded which would not lead to a prediction of the observed outcome $\{x\}$. Therefore, for each $d \in \mathcal{D}$ a prediction of the future outcome is needed for deciding which $d \in \mathcal{D}$ may be excluded.

For predicting the future development a prediction procedure is necessary. Requiring a reliability level of β for the measurement procedure means that a β-prediction procedure $A_X^{(\beta)}$ must be used.

Assuming that the outcome $\{x\} \in \mathcal{T}_X(\mathcal{X}(\mathcal{D}))$ of X has been observed, then the measurement $C_D^{(\beta)}(\{x\})$ consists of all those $d \in \mathcal{D}$ for which the observed outcome $\{x\}$ had been predicted:

$$C_D^{(\beta)}(\{x\}) = \left\{ d \in \mathcal{D} \,\middle|\, x \in A_X^{(\beta)}(\{d\}) \right\} \tag{52}$$

Next, it is shown that applying a measurement procedure as defined by (52) will, in fact, yield a correct result with probability of at least β. Let $d_0 \in \mathcal{D}$ be the actual but unknown value of D and $A_X^{(\beta)}(\{d_0\})$ the corresponding β-prediction. Then, d_0 is an element of $C_D^{(\beta)}(\{x\})$ if and only if $x \in A_X^{(\beta)}(\{d_0\})$ implying that the occurrence of $A_X^{(\beta)}(\{d_0\})$ is equivalent with the occurrence of a correct measurement for d_0. It follows:

$$
\begin{aligned}
P_{X|\{d_0\}}\left(\left\{x \mid d_0 \in C_D^{(\beta)}(\{x\})\right\}\right) &= P_{X|\{d_0\}}\left(\left\{x \mid x \in A_X^{(\beta)}(\{d_0\})\right\}\right) \\
&= P_{X|\{d_0\}}\left(A_X^{(\beta)}(\{d_0\})\right) \\
&\geq \beta
\end{aligned}
$$

where the latter inequality holds by definition of $A_X^{(\beta)}$.

From (52) is is seen that any measurement procedure is based on a prediction procedure. Using a β-prediction procedure for defining the measurement procedure yields a β-measurement procedure.

Completeness of a β-Measurement Procedure

Clearly, it is desirable that any possible observation $\{x\}$ yields a non-empty measurement result $C_D^{(\beta)}(\{x\})$. A measurement procedure which meets this requirement is called a complete measurement procedure. In order to guarantee the completeness property each possible observation $\{x\} \in \mathcal{T}_X(\mathcal{X}(\mathcal{D}))$ must be contained in at least one of the predictions $A_X^{(\beta)}(\{d\})$ obtained by

the prediction procedure used for defining $C_D^{(\beta)}$. Therefore, the defining prediction procedure $A_X^{(\beta)}$ must meet the following completeness requirement:

$$\bigcup_{d \in \mathcal{D}} A_X^{(\beta)}(\{d\}) = \mathcal{X}(\mathcal{D}) \tag{53}$$

β-Uncertainty Space for D

Consider the following subset of the uncertainty-space $\mathcal{U}_{X,D}$:

$$\mathcal{U}_D^{(\beta)} = \left\{ (x, d(x)) \mid x \in \mathcal{X}(\mathcal{D}), d(x) \in C_D^{(\beta)}(\{x\}) \right\} \tag{54}$$

$\mathcal{U}_D^{(\beta)}$ shows the region of consistent pairs $(x, d(x))$ within the uncertainty-space. The set of values $d(x)$ belonging to the observation $\{x\}$ cannot be excluded and, therefore, represent the ignorance after the learning or measurement process has been performed. The set $\mathcal{U}_D^{(\beta)}$ is referred to as the *β-uncertainty-space of D*.

Graphical Illustration of a β-Measurement Procedure

In Figure 1 below, a β-measurement procedure

$$C_D^{(\beta)} : \mathcal{T}_X \left(\mathcal{X}(\mathcal{D}) \to \mathcal{T}_D(\mathcal{D}) \right)$$
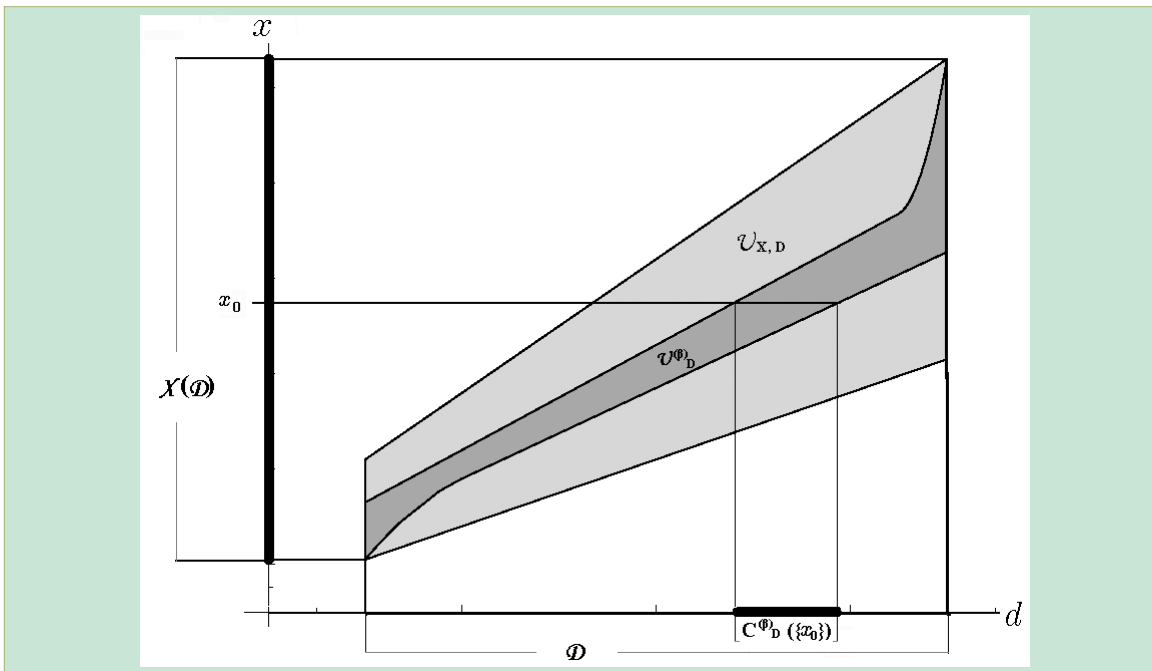
is schematically given:

Figure 1: β-Uncertainty Space $\mathcal{U}_D^{(\beta)}$ and β-measurement procedure $C_D^{(\beta)}$.

Note that the β-Uncertainty Space $\mathcal{U}_D^{(\beta)}$ could be called the calibration region for the measurement procedure $C_D^{(\beta)}$, where $A_X^{(\beta)}(\{d\})$ constitutes the predicted response for each $d \in \mathcal{D}$. If there is no reliability requirement, i.e., $\beta = 0$, then the calibration region shrinks to a calibration curve as assumed in conventional metrology. This last fact illustrates the serious shortcomings in conventional metrology and the necessity to replace it by a more rational stochastic metrology.

1. **Success Probability**

A frequently arising problem is to determine the success probability for a given experiment or process. Assume that one wishes to know the success probability of a certain medical treatment for more or less comparable cases. The measurement process consists of a clinical study with $n = 80$ patients, who undergo the identical treatment. In order to quantify the situation a pair of variables has to be defined and subsequently a Bernoulli Space has to be developed.

Figure 2: Clinical study with $n = 80$ patients.

Pair of variables:

$$
\begin{aligned}
X &= \text{number of treatment successes among the } n = 80 \text{ patients} \\
D &= \text{success probability}
\end{aligned}
\tag{55}
$$

Bernoulli Space:

$$
\begin{aligned}
\mathcal{D} &= \{p \,|\, 0 \leq p \leq 1.0\} \\
\mathcal{X}(\{p\}) &= \{0, 1, 2, \ldots, 80\} \\
\mathcal{P}(\{p\}) &= P_{X|\{p\}} \text{ with } X|\{p\} \sim Bi(80, p)
\end{aligned}
\tag{56}
$$

The ignorance space is selected assuming that there is no knowledge at all about the efficiency of the treatment. This will rarely be the case, as in general, preliminary studies have already indicated its effectiveness.

The stochastic measurement procedure $C_D^{(\beta)}$ is a mapping of the observed result onto a subset of the ignorance space. Of course, any measurement procedure should be derived before the measurement experiment is performed. There are two important conditions to be considered when deriving a measurement procedure. The first refers to meeting the reliability level and the second to assure that each possible observation yields a meaningful measurement result.

The reliability level is met by $C_D^{(\beta)}$, if the following holds:

$$
C_D^{(\beta)}(\{x\}) = \left\{ p \,|\, x \in A_X^{(\beta)}(\{p\}) \right)
\tag{57}
$$

The second requirement is met, if the predictions used for the measurement procedure meet the following condition:

$$
\bigcup_{p \in \mathcal{D}} A_X^{(\beta)}(\{p\}) = \{0, 1, 2, \ldots, 80\}
\tag{58}
$$

Unit 2.3.2: Optimum β-Measurement Procedures

Target

Any measurement procedure is characterized by three properties: reliability, precision and expense. The specified reliability level β provides the necessary reliability, while the precision or more precisely said the imprecision serves as an optimality criterion for given expense of the procedure. This learning unit derives a mathematical expression for the mean imprecision of a β-measurement procedure $C_D^{(\beta)}$.

Content

Introduction

Any β-measurement procedure

$$C_D^{(\beta)} : \left\{ \{x\} \,\middle|\, x \in \mathcal{X}(\mathcal{D}) \right\} \rightarrow \mathcal{T}_D(\mathcal{D}) \tag{59}$$

is completely determined by the set of measurement results:

$$\left\{ C_D^{(\beta)}(\{x\}) \,\middle|\, x \in \mathcal{X}(\mathcal{D}) \right\} \tag{60}$$

where the measurement results are obtained by means of the predictions $A_X^{(\beta)}(\{d\})$ provided by a β-prediction procedure $A_X^{(\beta)}$.

Measurement precision is besides the reliability of a measurement procedure the most important feature. Because of reasons which will become obvious below, not the precision, but the imprecision of a measurement is used for optimization. The imprecision of a measurement result $C_D^{(\beta)}(\{x\})$ is simply defined by its size:

$$\left| C_D^{(\beta)}(\{x\}) \right| = \begin{cases} \text{number of elements of } C_D^{(\beta)}(\{x\}) \\ \text{Lebesgue measure of } C_D^{(\beta)}(\{x\}) \end{cases} \tag{61}$$

The larger the size of $C_D^{(\beta)}(\{x\})$, the less precise is the measurement result. The most precise measurement result is given by a singleton.

Imprecision of a β-Measurement Procedure

Defining the imprecision of a measurement procedure constitutes a problem, because the measurement result that will be obtained is subjected to randomness as it depends on the future outcome $\{x\}$ of the random variable $X|\{d_0\}$ and in general the size of the different measurement results depends on the observed event $\{x\}$.

Let the random size of a measurement result be given by the random variable S:

$$S = \left| C_D^{(\beta)}(\{X\}) \right| \tag{62}$$

and let d_0 be the actual but unknown value of D. Then we have

$$S|\{d_0\} = \left| C_D^{(\beta)}(\{X|\{d_0\}\}) \right| \tag{63}$$

and the mean imprecision of the β-measurement procedure is given by the corresponding value of the first moment of $S|\{d_0\}$:

$$E\left[S|\{d_0\}\right] = \sum_{x \in \mathcal{X}(\{d_0\})} \left| C_D^{(\beta)}(\{x\}) \right| P_{X|\{d_0\}}(\{x\}) \tag{64}$$

Since the actual value is not known and each of the elements of the ignorance space must be considered equally when applying the β-measurement procedure $C_D^{(\beta)}$, the mean imprecision of a measurement procedure is obtained straightforwardly:

$$\frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} E\left[S|\{d\}\right] = \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \sum_{x \in \mathcal{X}(\{d\})} \left| C_D^{(\beta)}(\{x\}) \right| P_{X|\{d\}}(\{x\}) \tag{65}$$

Optimum β-Measurement Procedure

An optimum β-measurement procedure $^*C_D^{(\beta)}$ is defined by a β-measurement procedure with minimum mean imprecision, which is equivalent with maximum mean measurement precision. Let $\mathcal{C}_D^{(\beta)}$ be the set of all β-measurement procedures for the actual value of the deterministic variable D. Then the optimum β-measurement procedure $^*C_D^{(\beta)}$ meets the following relation:

$$\sum_{d \in \mathcal{D}} \sum_{x \in \mathcal{X}(\{d\})} \left| ^*C_D^{(\beta)}(\{x\}) \right| P_{X|\{d\}}(\{x\}) \leq \sum_{d \in \mathcal{D}} \sum_{x \in \mathcal{X}(\{d\})} \left| C_D^{(\beta)}(\{x\}) \right| P_{X|\{d\}}(\{x\})$$

$$\text{for } C_D^{(\beta)} \in \mathcal{C}_D^{(\beta)}$$

An optimum β-measurement procedure $^*C_D^{(\beta)}$ guarantees a reliability of β and simultaneously a minimum mean imprecision or equivalently a maximum mean precision of the measurement procedure.

EXAMPLE

1. **Kidney Cancer**

Kidney cancer appears as Type 1 and Type 2 where the latter is accompanied by an abnormal rise of the interleukin-6 level (IL-6 is a certain protein). From preliminary past studies it is known that the probability of Type 2 cancer is not smaller than 0.25 and not larger than 0.90.

Figure 1: A human kidney with cancer.

A study[8] is performed to improve the knowledge about the probability of Type 2 cancer. A group of 23 kidney cancer patients are randomly selected and checked for their interleukin-6 level. The probability of Type 2 cancer shall be determined by means of a stochastic measurement procedure with required reliability level $\beta = 90\%$.

The Bernoulli Space in this example refers to the variables:

---

[8]The study was performed by the Urological University Clinic of Tübingen.

- Random variable X: number of Type 2 cancer.
- Deterministic variable D: probability of Type 2 cancer.

The preliminary Bernoulli Space $\mathbb{B}_{X,D} = (\mathcal{D}, \mathcal{X}, \mathcal{P})$ that shall be improved by the study is given as follows:

- Ignorance space: $\mathcal{D} = \{p \mid 0.25 \leq p \leq 0.90\}$

- Variability function: $\mathcal{X}(\{p\}) = \{0, 1, 2, ..., 23\}$ for $p \in \mathcal{D}$

- Random structure function: $\mathcal{P}(\{p\}) = \{P_{X|\{d\}}\}$ where $P_{X|\{p\}}$ is the Binomial distribution $Bi(23; p)$ with:

$$P_{X|\{p\}}(\{x\}) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x \in \mathcal{X}(\{p\})$$

The study resulted in $x = 10$ cases of Type 2 kidney cancer in the group of $n = 23$ cancer patients. Applying an optimum 0.90-measurement procedure yields the following result for the probability of Type 2 cancer:

$$C_D^{(0.90)}(\{10\}) = \{p \mid 0.34 \leq p \leq 0.61\}$$

which constitutes a rather large improvement of the so far available knowledge.

The complete stochastic 0.90-measurement procedure applied in the cancer study is displayed in Figure 1 below:
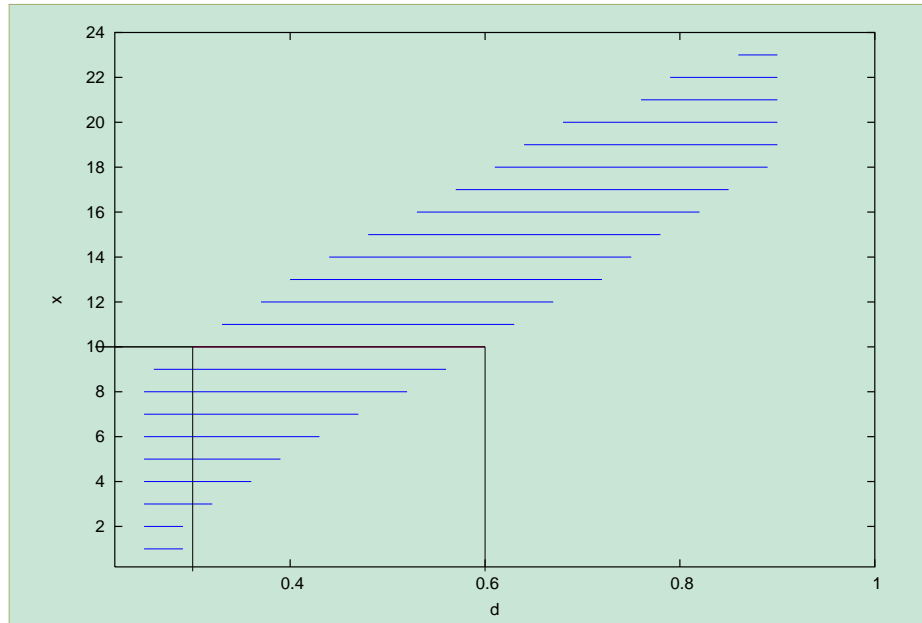

Figure 2: Graphical representation of the optimum
β-measurement procedure for the example with kidney cancer.

Note that the measurement procedure is available independently whether or not the measurement experiment is performed. Performing a measurement process and only subsequently develop a measurement procedure, which yields the expected or desired result, is at least a doubtful proceeding.

Unit 2.3.3: Minimum Measurement Without Ignorance

TARGET

The value of the mean imprecision of a measurement procedure depends heavily on the subset $\mathcal{T}_D(\mathcal{D})$ containing the measurement results. This learning unit introduces the minimum β-measurement procedure, which have the smallest mean imprecision among all optimum β-measurement procedure.

Of course, the mean imprecision also depends on the amount of ignorance about those components of the deterministic variable, which are not subject of the measurement. In this learning unit it is assumed that there is ignorance only with respect to the component of the deterministic variable which shall be determined, while there is complete knowledge with respect to all other components of D.

CONTENT

Introduction

From the previous learning unit it is known that any β-measurement procedure is determined by a β-prediction procedure which is the solution of an optimization problem under side conditions.

The side-conditions refer to the required procedure's reliability and the necessary completeness of the procedure, i. e. the procedure shall yield a non-empty measurement for any possible observation. The objective function is given by the procedure's mean imprecision.

A prediction procedure is given by a function

$$A_X : \mathcal{T}_D(\mathcal{D}) \to \mathcal{T}_X(\mathcal{X}(\mathcal{D}))$$

where the two subsets essentially determine the properties of the procedure.

Formulation of the Problem

In case that there is no ignorance about other components of the deterministic variable, the set of arguments of the prediction procedure is given by the singletons $\{d\}$:

$$\mathcal{T}_D(\mathcal{D}) = \left\{ \{d\} \,\middle|\, d \in \mathcal{D} \right\}$$

Because a minimum β-measurement procedure shall be derived, the elements of the image set $\mathcal{T}_X(\mathcal{X}(\mathcal{D}))$ of the prediction procedure must not be subjected to any restrictions. Thus, we obtain:

$$\mathcal{T}_X(\mathcal{X}(\mathcal{D})) = \mathfrak{P}(\mathcal{X}(\mathcal{D}))$$

where $\mathfrak{P}$ denotes the power set, i. e., the system of all subsets of a given set.

The problem may be formulated as follows: Determine a prediction procedure $^*A_X^{(\beta)}$ with fol-

lowing properties:

$$P_{X|\{d\}}\left(\,^*A_X^{(\beta)}(\{d\})\right) \geq \beta \qquad \text{for } d \in \mathcal{D}$$

$$\bigcup_{d \in \mathcal{D}} \,^*A_X^{(\beta)}(\{d\}) = \mathcal{X}(\mathcal{D}) \tag{66}$$

$$\sum_{d \in \mathcal{D}} \sum_{x \in \mathcal{X}(\{d\})} \left| \left\{ d \,|\, x \in \,^*A_X^{(\beta)}(\{d\}) \right\} \right| P_{X|\{d\}}(\{x\}) = \min$$

with no restrictions whatsoever being made with respect to the form of the predictions $^*A_X^{(\beta)}(\{d\})$ and, hence, of the form of the measurement results $^*C_D^{(\beta)}(\{x\})$. Both are determined by the optimization procedure.

In the one-dimensional case the predictions as well as the measurement results are in general intervals (at least for simple, i.e., not compound measurement processes):

$$^*A_X^{(\beta)}(\{d\}) = \left\{ x \in \mathcal{X}(\mathcal{D}) \,|\, \underline{x}(d) \leq x \leq \overline{x}(d) \right\} \tag{67}$$

$$^*C_D^{(\beta)}(\{x\}) = \left\{ d \in \mathcal{D} \,|\, \underline{d}(x) \leq d \leq \overline{d}(x) \right\} \tag{68}$$

Note that in this case a lower and an upper bound for the unknown value d_0 of the deterministic variable have to be determined for each observable event $\{x\}$.

The set of all measurements $^*C_D^{(\beta)}(\{x\})$ represents the entire measurement procedure.

Determination of Minimum β-Measurement Procedures

Minimum β-measurement procedures will be made available by *Stochastikon Calculator*, which presently is under construction. However, it already covers the case of Bernoulli Spaces generated by binomial random variables $X \sim Bi(n,p)$. Therefore, the proceedings for determining a measurement result will be demonstrated here by means of a binomial random variable.

Consider a production process, which produces under more or less identical conditions. In order to determine the nonconformace probability, $n = 1000$ items shall be produced, where the aspect of interest X is the number of nonconforming items among them. This number has a binomial probability distribution $Bi(1000, p)$ and, therefore, Stochastikon Calculator can be used for determining the nonconformance probability. The reliability level is set to be $\beta = 0.90$ and it can be excluded by past experience that the nonconformance probability exceeds the value 0.1. Thus, the ignorance space for the nonconformance probability is given by

$$\mathcal{D}_p = \{p \,|\, 0 \leq p \leq 0.1\}$$

The nonconformance probability shall be determined as precisely as possible and, therefore, a minimum β-measurement procedure is needed on the basis of the observed outcome $x = 3$. After the values characterizing the situation have been entered, Stochastikon Calculator determines the optimum β-measurement procedure and the measurement result $C_{D_p}^{(0.90)}(\{3\})$ and produces a report which summarizes the Bernoulli Space and specifies the measurement result:

$$C_{D_p}^{(0.90)}(\{3\}) = \{p \,|\, 0.0012 \leq p \leq 0.0085\}$$

Moreover, the following graphical representation of the entire optimum 0.90-measurement procedure is displayed in the report.
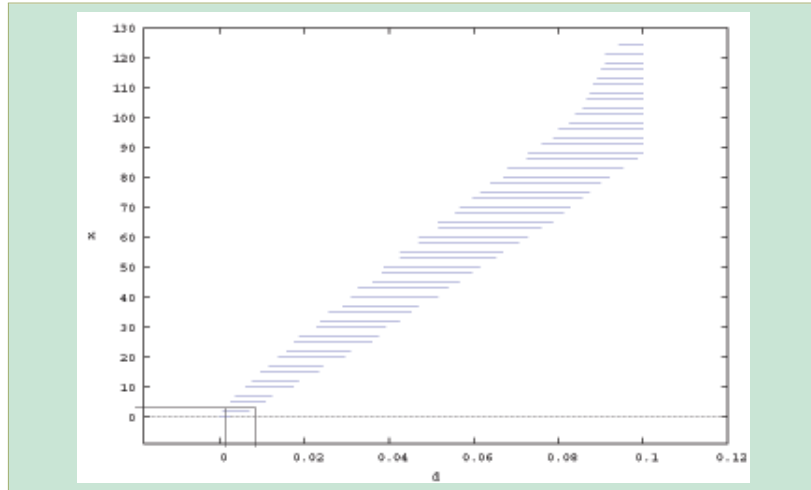
Figure 1: Graphical representation of the optimum
(0.90)-measurement procedure.

EXAMPLE

1. **Success Probability**

Example 1 in Learning Unit 2.3.1 treated the success probability of a medical treatment.
The problem was to develop a measurement procedure for the success probability. The
situation was quantified in the following way:

Pair of variables:

$$
\begin{aligned}
X &= \text{number of treatment successes among the 80 patients} \\
D &= \text{success probability}
\end{aligned}
\tag{69}
$$

Bernoulli Space:

$$
\begin{aligned}
\mathcal{D} &= \{p \,|\, 0 \leq p \leq 1.0\} \\
\mathcal{X}(\{p\}) &= \{0, 1, 2, \ldots, 80\} \\
\mathcal{P}(\{p\}) &= P_{X|\{p\}} \text{ with } X|\{p\} \sim Bi(80, p)
\end{aligned}
\tag{70}
$$

The measurement procedure shall have a reliability level of $\beta = 0.85$.

The clinical study is performed and for $x = 66$ patients a significant improvement was
observed. The measurement procedure $C_D^{(0.85)}$ and the corresponding measurement result
are displayed in Figure 1. Note that for any possible observation there is a measure-
ment result. The reliability level $\beta = 0.85$ implies that applying the given measurement
procedure will yield with probability of at least 0.85 a correct measurement result, i. e.,
the actual, but unknown value of the success probability is covered by the measurement
result.

The measurement yields the following result for the success probability of the given treat-
ment:

$$
C_D^{(0.85)}(\{66\}) = \{p \,|\, 0.76 \leq p \leq 87\}
\tag{71}
$$

60

A graphical representation of the entire measurement procedure including the obtained measurement result is readily obtained by means of the Graphical Laboratory.
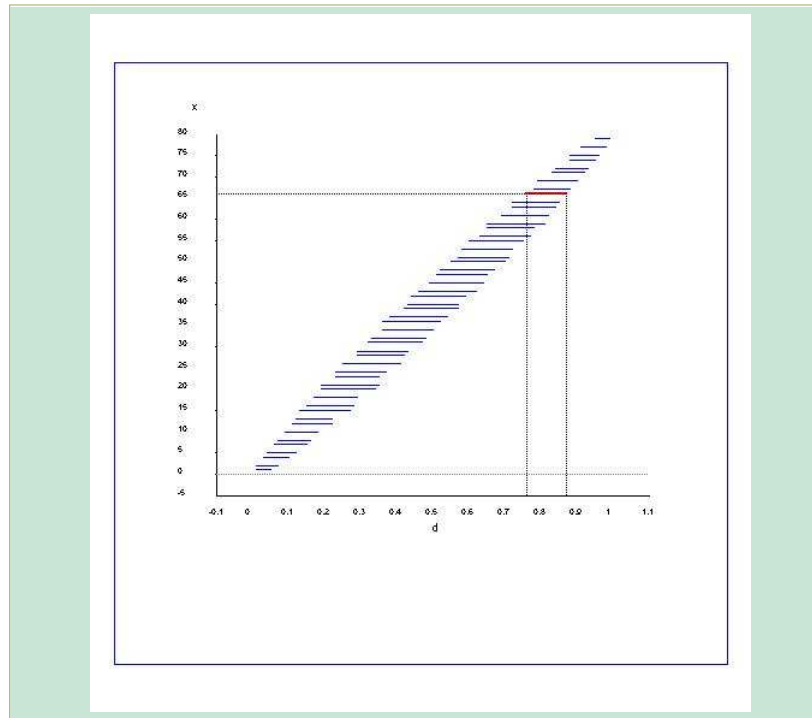

Figure 1: Graphical representation of the stochastic measurement $C_D^{(0.85)}(\{66\})$

Unit 2.3.4: Minimum Upper Bound Measurement Without Ignorance

TARGET

This case refers to the one-dimensional case, i. e., there is only one component of the deterministic variable with ignorance. Moreover, one is not interested how small the unknown value d_0 is, but only how large it can be.

This learning unit answers the question of how large the unknown value d_0 of a deterministic variable is by determining an upper bound for d_0.

CONTENT

Introduction

In this learning unit it is assumed that there is complete knowledge about the components of the deterministic variable except for one univariate component, say D, with unknown value d_0 and ignorance space $\mathcal{D}$.

The question shall be answered how large d_0 is, while there is no interest in the question how small d_0 might be.

For answering the question an upper bound for the unknown value d_0 must be determined implying that the measurement results have the following form:

$$C_D^{(\beta)}(\{x\}) = \{d \in \mathcal{D} \mid d \le \bar{d}(x)\}$$

Note the difference between the minimum β-measurement procedure and the minimum upper bound β-measurement procedure. In the former case an lower and upper bound for d_0 must be determined, while in the latter only an upper bound is needed.

Analogously to the case of a minimum β-measurement procedure, an appropriate prediction procedure has to be derived.

$$A_X : \mathcal{T}_D(\mathcal{D}) \to \mathcal{T}_X(\mathcal{X}(\mathcal{D}))$$

where the domain of $A_X^{(\beta)}$ is again given by the set of singletons $\{d\} \subset \mathcal{D}$. The co-domain of $A_X^{(\beta)}$, however, is different, because only an upper bound $\bar{d}(x)$ must be determined. In this case the needed predictions are of the following form:

$$A_X^{(\beta)}(\{d\}) = \{x \in \mathcal{X}(\mathcal{D}) \mid \underline{x}(d) \le x\} \tag{72}$$

i. e., the co-domain of the prediction procedure consists of intervals with lower bounds.

Assume that a prediction procedure $A_X^{(\beta)}$ of the form given by (72) is available, then the corresponding measurement results are given as follows:

$$
\begin{aligned}
C_D^{(\beta)}(\{x\}) &= \{d \in \mathcal{D} \mid x \in A_X^{(\beta)}(\{d\})\} \\
&= \{d \in \mathcal{D} \mid d \le \bar{d}(x)\} \tag{73}
\end{aligned}
$$

where $\bar{d}(x)$ is the largest value $d \in \mathcal{D}$ for which the observed event $\{x\}$ has been predicted. Note that the problem of determining a stochastic procedure can always be reduced to determining

a corresponding measurement procedure. This fact makes the development of a unified theory of uncertainty possible.

Formulation of the Problem

A minimum upper bound β-measurement procedure is based on predictions which have a lower bound, but no upper one. Thus the problem may be formulated as follows:

Determine a prediction procedure ${}^{*}A_X^{(\beta)}$ of the form given by (72) with following properties:

$$\text{reliability:} \qquad P_{X|\{d\}}\left({}^{*}A_X^{(\beta)}(\{d\})\right) \geq \beta \quad \text{for } d \in \mathcal{D}_0$$

$$\text{completeness:} \quad \min_{d \in \mathcal{D}_0}\{\underline{x}(d)\} = \min \mathcal{X}(\mathcal{D}_0) \tag{74}$$

$$\text{precision:} \qquad \sum_{d \in \mathcal{D}_0}\sum_{x \in \mathcal{X}(\{d\})} |\{d \in \mathcal{D}_0 \,|\, \underline{x}(d) \leq x\}|P_{X|\{d\}}(\{x\}) = \min$$

Note that the problem of determining a minimum upper bound β-measurement procedure is much simpler compared with that of determining a minimum β-measurement procedure as described in the previous learning unit.

EXAMPLE

1. **Failure Probability**

   Consider a case that a new medical treatment might lead to failure. In order to judge the risk an upper bound for the probability of a failure of the treatment shall be determined. To this end $n = 35$ treatments shall be performed with the number of failures being the quantity of interest. Moreover, it is decided to set the reliability level to be $\beta = 0.80$.

   Quantification leads to the following Bernoulli Space:

   Pair of variables:

   $$\begin{aligned} X &= \text{number of treatment failures among the 35 patients} \\ D &= \text{failure probability} \end{aligned} \tag{75}$$

   Bernoulli Space:

   $$\begin{aligned} \mathcal{D} &= \{p \,|\, 0 \leq p \leq 0.50\} \\ \mathcal{X}(\{p\}) &= \{0, 1, 2, \ldots, 20\} \\ \mathcal{P}(\{p\}) &= P_{X|\{p\}} \text{ with } X|\{p\} \sim Bi(20, p) \end{aligned} \tag{76}$$

   In this case, it is known that values exceeding 0.50 for the failure probability p can be excluded. The measurement experiment is performed and $x = 3$ failures are observed.

   The measurement result is again obtained by means of the Stochastikon Calculator along with a report and a graphical illustration of the entire measurement procedure $C_D^{(0.80)}$, which is given in Figure 1.
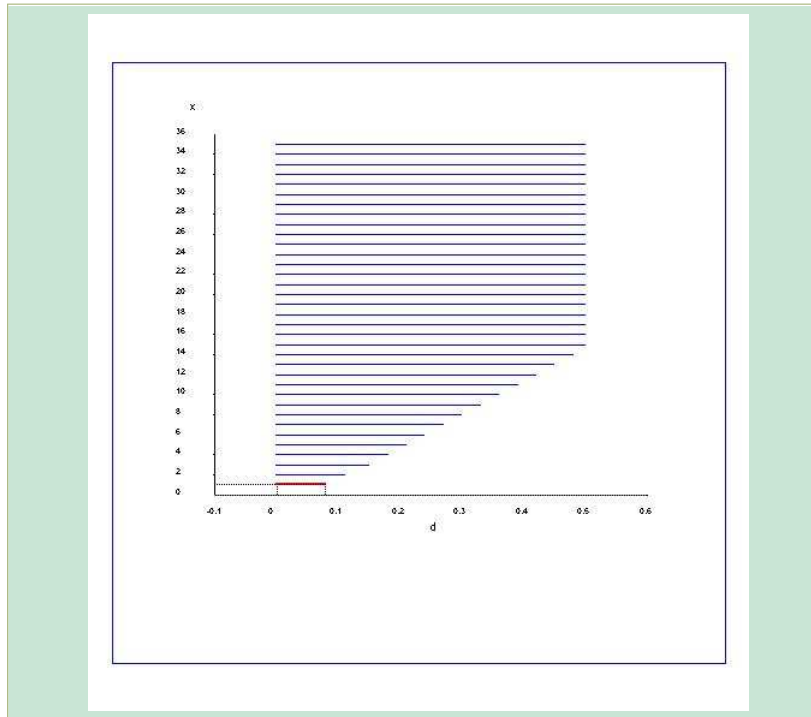
Figure 1: Graphical representation of the stochastic
measurement procedure $C_D^{(0.80)}$.

From Figure 1, the following measurement result $C_D^{(0.80)}(\{3\})$ is obtained:

$$C_D^{(0.80)}(\{3\}) = \{p \,|\, 0 \leq p \leq 0.15\} \tag{77}$$

This measurement result means that the risk of a treatment failure is not larger than 15%.

Unit 2.3.5: Maximum Lower Bound Measurement Without Ignorance

Target

This learning unit refers to the same situation as the previous one, i.e., there is only one component of the deterministic variable with ignorance, while the values of any other component is known. In contrast to the previous learning unit, however, it is not of interest how large the unknown value d_0 might be, but how small it is.

Content

Introduction

The situation considered in this learning unit is similar as that in the previous one. There is complete knowledge about the deterministic variable except for one univariate component, say D with unknown value d_0 and ignorance space $\mathcal{D}$.

The question shall be answered how small d_0 is, while there is no interest in the question how large d_0 might be.

For answering the question a lower bound for the unknown value d_0 must be determined implying that the measurement results have the following form:

$$C_D^{(\beta)}(\{x\}) = \{d \in \mathcal{D} \,|\, \underline{d}(x) \leq d\}$$

Note the difference between the minimum β-measurement procedure and the maximum lower bound β-measurement procedure. In the former case an lower and upper bound for d_0 must be determined, while in the latter only a lower bound is needed.

Analogously to the previous cases, an appropriate prediction procedure has to be derived.

$$A_X : \mathcal{T}_D(\mathcal{D}_0) \rightarrow \mathcal{T}_X(\mathcal{X}(\mathcal{D}))$$

where the domain of $A_X^{(\beta)}$ is again given by the singletons $\{d\} \subset \mathcal{D}$, because there is no ignorance about possibly further components of the deterministic variable. The co-domain of $A_X^{(\beta)}$, however, is different, because only a lower bound $\underline{d}(x)$ must be determined. In this case the needed predictions are of the following form:

$$A_X^{(\beta)}(\{d\}) = \{x \in \mathcal{X}(\mathcal{D}) \,|\, x \leq \overline{x}(d)\} \tag{78}$$

i.e., the co-domain of the prediction procedure consists of intervals with upper bounds.

Assume that a prediction procedure $A_X^{(\beta)}$ with predictions of the form given by (78) is available, then the corresponding measurement results are given as follows:

$$
\begin{aligned}
C_D^{(\beta)}(\{x\}) &= \left\{d \in \mathcal{D} \,|\, x \in A_X^{(\beta)}(\{d\})\right\} \\
&= \{d \in \mathcal{D} \,|\, \underline{d}(x) \leq d\} \tag{79}
\end{aligned}
$$

where $\underline{d}(x)$ is the largest value of the deterministic variable for which the observed event $\{x\}$ had been predicted.

Formulation of the Problem

A maximum lower bound β-measurement procedure is based on predictions which have an upper bound, but no lower one.

Thus the problem may be formulated as follows: Determine a prediction procedure ${}^{*}A_X^{(\beta)}$ of the form given by (78) with following properties:

$$
\begin{aligned}
&\text{reliability:} &&P_{X|\{d\}}\left({}^{*}A_X^{(\beta)}(\{d\})\right) \geq \beta &&\text{for } d \in \mathcal{D}_0 \\[2mm]
&\text{completness:} &&\max_{d\in\mathcal{D}_0}\{\overline{x}(d)\} = \max \mathcal{X}(\mathcal{D}_0) &&(80) \\[2mm]
&\text{precision:} &&\sum_{d\in\mathcal{D}_0}\sum_{x\in\mathcal{X}(\{d\})}\left|C_{\mathcal{D}_0}^{(\beta)}(\{x\})\right|P_{X|\{d\}}(\{x\}) = \min
\end{aligned}
$$

Note that the problem of determining a maximum lower bound β-measurement procedure is similar as in the case of a minimum upper bound β-measurement procedure and much simpler than that of determining a minimum β-measurement procedure.

EXAMPLE

1. **Success Probability**

   Example 1 in Learning Unit 2.3.1 deals with the success probability of a medical treatment. The problem was to develop a measurement procedure for the success probability. The situation was quantified in the following way:

   Pair of variables:

   $$
   \begin{aligned}
   X &= \text{number of treatment successes among the 80 patients} \\
   D &= \text{success probability}
   \end{aligned}
   \qquad (81)
   $$

   Bernoulli Space:

   $$
   \begin{aligned}
   \mathcal{D} &= \{p \,|\, 0 \leq p \leq 1.0\} \\
   \mathcal{X}(\{p\}) &= \{0, 1, 2, \ldots, 80\} \\
   \mathcal{P}(\{p\}) &= P_{X|\{p\}} \text{ with } X|\{p\} \sim Bi(80, p)
   \end{aligned}
   \qquad (82)
   $$

   The measurement procedure shall have a reliability level of $\beta = 0.85$ and shall yield a guaranteed success probability, which is obtained as a maximum lower bound for the true success probability. A graphical representation of the corresponding measurement procedure is readily obtained by means of the Graphical Laboratory. The clinical study is performed and for $x = 66$ patients a significant improvement was observed. The measurement procedure $C_D^{(0.85)}$ and the corresponding measurement result are displayed in Figure 1 below.
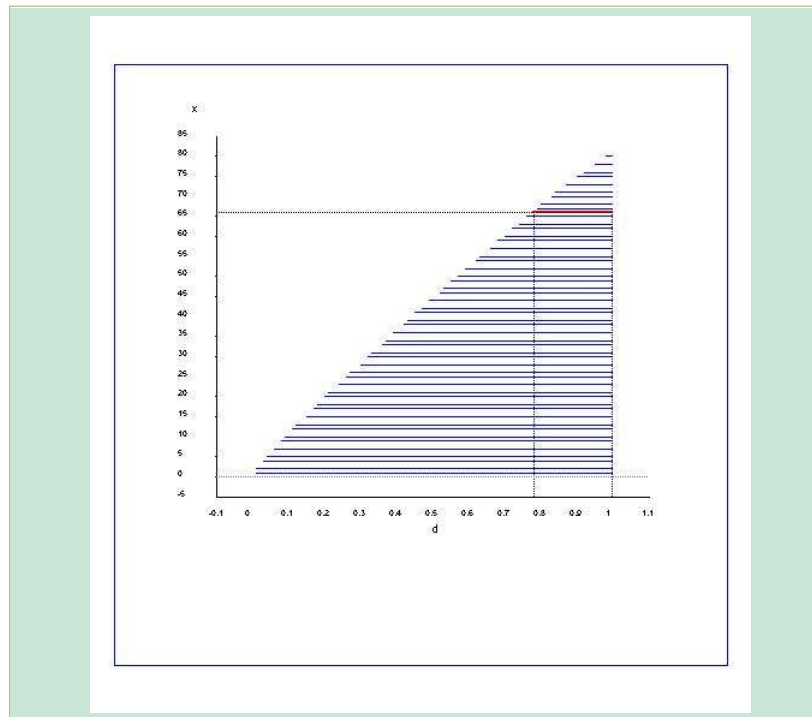
Figure 1: Graphical representation of the stochastic measurement $C_D^{(0.85)}$ with the highlighted measurement result $C_D^{(0.85)}(\{66\})$.

The measurement yields a maximum lower bound of $p = 0.78$ for the success probability of the given treatment. Note that the probability statement of obtaining a correct result hold only for the procedure, but not hold for the given measurement result $C_D^{(0.85)}(\{66\})$, as the result is either correct or not.

Unit 2.3.6: Minimum Measurement With Ignorance

TARGET

The mean imprecision of a measurement procedure depends also on the amount of ignorance about those components of the deterministic variable, which shall not be measured. In this learning unit it is assumed that there is ignorance not only with respect to the component of the deterministic variable which shall be determined, but also to some of the other components of the deterministic variable.

CONTENT

Introduction

From the previous learning units it is known that any β-measurement procedure is determined by a β-prediction procedure which is the solution of an optimization under side conditions. By the way, this is true for any stochastic procedure. Every stochastic procedure can be reduced to certain prediction procedures.

Just as in the case without ignorance, the side-conditions here refer to the required procedure's reliability and the necessary completeness of the procedure. The objective function is given by the procedure's mean imprecision.

A prediction procedure is given by a function

$$A_X : \mathcal{T}_D(\mathcal{D}) \to \mathcal{T}_X(\mathcal{X}(\mathcal{D}))$$

where the two systems of subsets $\mathcal{T}_D(\mathcal{D})$ and $\mathcal{T}_X(\mathcal{X}(\mathcal{D}))$, respectively, determine essential the properties of the procedure.

Formulation of the Problem

In the case of ignorance about several components of the deterministic variable, the set of arguments of the prediction procedure is not anymore given by singletons, but by larger sets.

Consider a deterministic variable with k components, i.e.

$$D = (D_1, \ldots, D_{j_0}, \ldots, D_k) \tag{83}$$

with D_{j_0} being the component of interest with actual value $d_\ell^{(j_0)}$, which shall be determined by a β-measurement procedure.

Let the ignorance space $\mathcal{D}$ have N elements:

$$\mathcal{D} = \{d_1, \ldots, d_{j_0}, \ldots, d_N\} \qquad \text{with} \qquad d_i = (d_{i,1}, \ldots, d_{i,j_0}, \ldots, d_{i,k}) \tag{84}$$

for $i = 1, \ldots, N$. Moreover, consider the situation that the there are n potential values with respect to the component D_{j_0} of interest given by the set:

$$\mathcal{D}_{j_0} = \{d_1^{(j_0)}, \ldots, d_\ell^{(j_0)}, \ldots, d_n^{(j_0)}\} \tag{85}$$

which includs the actual, but unknown value $d_\ell^{(j_0)}$.

As shown in the previous learning units, the measurement procedure with respect to D_{j_0} is developed based on a prediction procedure $A_X^{(\beta)}$. The domain of the prediction procedure $A_X^{(\beta)}$ is given by the following system of n subsets of $\mathcal{D}$, where each of the n subsets relates to one of the possible values of D_{j_0}:

$$
\begin{aligned}
\mathcal{T}_{D_{j_0}}(\mathcal{D}) &= \left\{ \{ d_i \in \mathcal{D} \mid d_{i,j_0} = d_h^{(j_0)} \} \,\Big|\, 1 \le h \le n \right\} \\
&= \left\{ \mathcal{D}_{j_0,h} \,\Big|\, 1 \le h \le n \right\}
\end{aligned}
$$

with the set $\mathcal{D}_{j_0,h}$ representing the hth possible value ($h = 1, 2, \ldots, n$) of the component of interest D_{j_0}. The set $\mathcal{D}_{j_0,h}$ consists of all those elements of the overall ignorance space $\mathcal{D}$ which have the value $d_h^{(j_0)}$ for the component of interest D_{j_0}.

Because a minimum β-measurement procedure shall be derived, the elements of the image set $\mathcal{T}_X(\mathcal{X}(\mathcal{D}))$ of the prediction procedure $A_X^{(\beta)}$ must not be subjected to any restrictions. Thus, we select

$$
\mathcal{T}_X(\mathcal{X}(\mathcal{D})) = \mathfrak{P}(\mathcal{X}(\mathcal{D}))
$$

where $\mathfrak{P}$ denotes the power set, i. e., the system of all subsets of a given set.

The problem to be solved may be formulated as follows:
Determine a prediction procedure ${}^*A_X^{(\beta)}$ on $\mathcal{T}_{D_{j_0}}(\mathcal{D})$ with following properties:

reliability: $\qquad P_{X|\{d\}}\left({}^*A_X^{(\beta)}(\mathcal{D}_{j_0,h}) \right) \ge \beta \qquad$ for $d \in \mathcal{D}_{j_0,h}$ and $h = 1, \ldots, n$

completeness: $\qquad \bigcup_{h=1}^{n} {}^*A_X^{(\beta)}(\mathcal{D}_{j_0,h}) = \mathcal{X}(\mathcal{D})$

precision: $\qquad \displaystyle\sum_{h=1}^{n} \frac{1}{|\mathcal{D}_{j_0,h}|} \sum_{d \in \mathcal{D}_{j_0,h}} \sum_{x \in \mathcal{X}(\{d\})} \left| \left\{ d_k^{(j_0)} \mid x \in {}^*A_X^{(\beta)}(\mathcal{D}_{j_0,k}) \right\} \right| P_{X|\{d\}}(\{x\}) = \min$

No restrictions are made with respect to the form of the predictions ${}^*A_X^{(\beta)}(\mathcal{D}_{j_0,h})$, which is determined by the optimization procedure.

Minimum β-Measurement Procedure

Once the prediction procedure ${}^*A_X^{(\beta)}$ is available, the minimum β-measurement procedure for D_{j_0} is given by the measurement results:

$$
{}^*C_{D_{j_0}}^{(\beta)}(\{x\}) = \left\{ d_k^{j_0} \mid x \in {}^* A_X^{(\beta)}(\mathcal{D}_{j_0,k}) \right\} \tag{86}
$$

In the one-dimensional case for X and for D_{j_0}, respectively, the predictions as well as the measurement results are in general intervals:

$$
{}^*A_X^{(\beta)}(\mathcal{D}_{j_0,h}) = \left\{ x \in \mathcal{X}(\mathcal{D}) \mid \underline{x}(\mathcal{D}_{j_0,h}) \le x \le \overline{x}(\mathcal{D}_{j_0,h}) \right\} \tag{87}
$$

$$
{}^*C_{D_{j_0}}^{(\beta)}(\{x\}) = \left\{ d_k^{(j_0)} \in \mathcal{D}_{j_0} \mid \underline{d}^{(j_0)}(x) \le d_k^{(j_0)} \le \overline{d}^{(j_0)}(x) \right\} \tag{88}
$$

In this case a lower and an upper bound for the unknown value $d_\ell^{(j_0)}$ have to be determined using (86).

Example

1. **Success Probability**

Consider the Example 1 of Learning Unit 2.3.3 which assumed that the number of patients were exactly known to be $n = 80$. Consider now the case that the exact number of treatments has been lost. It is only known, that the number of treatments is between $n = 80$ and $n = 86$. In this case there is also ignorance about the number of Bernoulli trials and, therefore, the deterministic variable has to cover not only the success probability, but also the number of treatments.

Pair of variables:

$$
\begin{aligned}
X &= \text{number of successes among the treated patients} \\
D &= (D_1, D_2) \text{ with} \\
&\quad D_1 = \text{number of treatments} \\
&\quad D_2 = \text{success probability}
\end{aligned}
\tag{89}
$$

Bernoulli Space:

$$
\begin{aligned}
\mathcal{D} &= \{(n, p) \,|\, 80 \leq n \leq 86, 0 \leq p \leq 1.0\} \\
\mathcal{X}(\{(n, p)\}) &= \{0, 1, 2, \ldots, n\} \\
\mathcal{P}(\{(n, p)\}) &= P_{X|\{(n,p)\}} \text{ with } X|\{(n, p)\} \sim Bi(n, p)
\end{aligned}
\tag{90}
$$

Note that in this case, the range of variability $\mathcal{X}(\{(n, p)\})$ depends on the argument $\{(n, p)\}$. The range of variability of $X|\{(n, p)\}$ increases with increasing value of n.

Just as in the case in Learning Unit 2.3.3, the measurement procedure $C_{D_2}^{(\beta)}$ for the success probability shall have a reliability level of $\beta = 0.85$. The clinical study is performed and for $x = 66$ patients a significant improvement was observed.

The measurement procedure $C_{D_2}^{(0.85)}$ and the corresponding measurement result are obtained by means of Stochastikon Calculator. Because of the additional ignorance with respect to the number of patients, it is to be expected that the measurement results have a larger size.

The measurement result for the success probability of the treatment is obtained by Stochastikon Calculator and given as follows:

$$
C_{D_2}^{(0.85)}(\{66\}) = \{p \,|\, 0.71 \leq p \leq 0.87\}
\tag{91}
$$

Comparing the result above with that obtained with exact knowledge about the number of trials shows how the measurement is enlarged due to the additional ignorance.

A graphical representation of the entire measurement procedure including the obtained result is either obtained by Stochastikon Calculator or by using the Graphical Laboratory. The graphical representation is given by Figure 1:
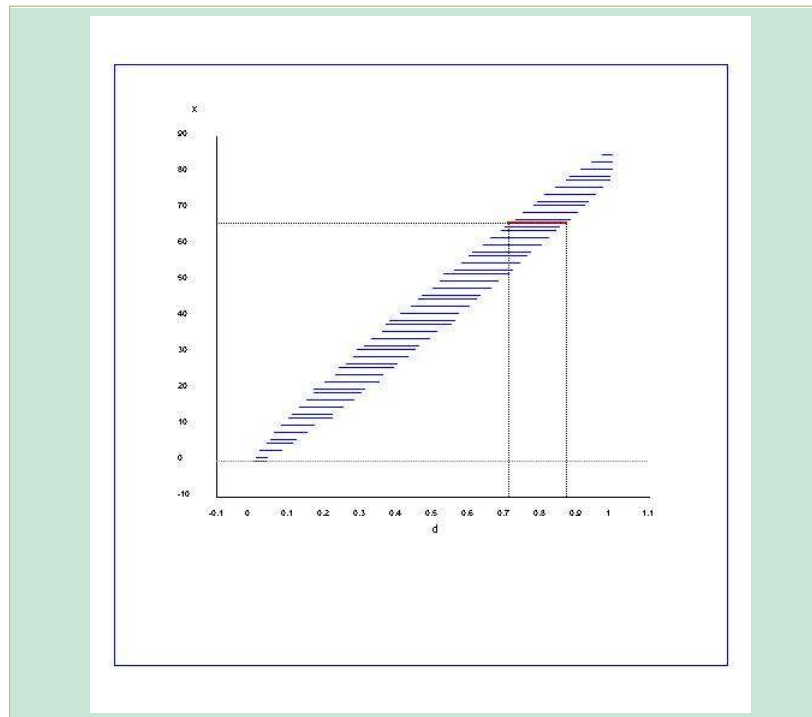
Figure 1: Graphical representation of the stochastic measurement $C_{D_2}^{(0.85)}(\{66\})$ in the case of ignorance with respect to the number of treatments.

Unit 2.3.7: Minimum Upper Bound Measurement With Ignorance

TARGET

This learning unit investigates a similar situation as that of Learning Unit 2.3.4 with the difference that now there is ignorance also about those components which are of no interest with respect to the measurement procedure. Similar as in Unit 2.3.4 one is not interested how small the unknown value of interest is, but only how large it is.

CONTENT

Introduction

In this learning unit the situation is considered that not only the value of the component of interest is unknown, but there is also ignorance about the values of some other components of the deterministic variable.

The situation is identical to that of Learning Unit 2.3.6. It is assumed that the deterministic variable has k components, i. e.

$$D = (D_1, \ldots, D_k) \tag{92}$$

The ignorance space $\mathcal{D}$ has N elements:

$$\mathcal{D} = \{d_1, \ldots, d_N\} \qquad \text{with} \qquad d_i = (d_{i,1}, \ldots, d_{i,k}), i = 1, \ldots, N \tag{93}$$

The component of interest is D_{j_0} with actual value $d_\ell^{(j_0)}$, for which an upper bound shall be determined by a β-measurement procedure $C_{D_{j_0}}^{(\beta)}$.

Let the ignorance space with respect to the component D_{j_0} of interest contain n elements, given by

$$\mathcal{D}_{j_0} = \{d_1^{(j_0)}, \ldots, d_n^{(j_0)}\} \tag{94}$$

including the actual value $d_\ell^{(j_0)}$. The question shall be answered how large $d_\ell^{(j_0)}$ is, while there is no interest in the question how small $d_\ell^{(j_0)}$ is.

For answering the question an upper bound for the unknown value $d_\ell^{(j_0)}$ must be determined implying that the measurement results have the following form:

$$C_{D_{j_0}}^{(\beta)}(\{x\}) = \{d_h^{(j_0)} \in \mathcal{D}_{j_0} \mid d_h^{(j_0)} \leq \overline{d}^{j_0}(x)\}$$

Note the difference between the minimum β-measurement procedure and the minimum upper bound β-measurement procedure. In the former case an lower and upper bound for $d_\ell^{(j_0)}$ must be determined, while in the latter only an upper bound is needed.

Analogously to the case of a minimum β-measurement procedure, an appropriate prediction procedure has to be derived.

$$A_X : \mathcal{T}_{D_{j_0}}(\mathcal{D}) \to \mathcal{T}_X(\mathcal{X}(\mathcal{D}))$$

where the domain of $A_X^{(\beta)}$ is the same as in Unit 2.3.6 given by the sets $\mathcal{D}_{j_0,h}$, $h = 1, \ldots, n$. The co-domain of $A_X^{(\beta)}$, however, is different, because only an upper bound $\overline{d}(x)$ must be determined. In this case the needed predictions are of the following form:

$$A_X^{(\beta)}(\mathcal{D}_{j_0,h}) = \{x \in \mathcal{X}(\mathcal{D}_0) \mid \underline{x}(\mathcal{D}_{j_0,h}) \leq x\} \tag{95}$$

i. e., the co-domain of the prediction procedure consists of intervals with only lower bounds.

Assume that a prediction procedure $A_X^{(\beta)}$ of the form given by (95) is available, then the corresponding measurement results are given as follows:

$$\begin{aligned}
C_{D_{j_0}}^{(\beta)}(\{x\}) &= \{d_k^{(j_0)} \in \mathcal{D}_{j_0} \mid x \in A_X^{(\beta)}(\mathcal{D}_{j_0,k})\} \\
&= \{d_k^{(j_0)} \in \mathcal{D}_{j_0} \mid d_k^{(j_0)} \leq \overline{d}^{(j_0)}(x)\}
\end{aligned} \tag{96}$$

where $\overline{d}^{(j_0)}(x)$ is the largest value in $\mathcal{D}_{j_0}$ foe which the observed event $\{x\}$ had been predicted.

Formulation of the Problem

A minimum upper bound β-measurement procedure is based on predictions which have a lower bound, but not an upper one.

Thus the problem may be formulated as follows: Determine a prediction procedure $^*A_X^{(\beta)}$ of the form given by (95) with following properties:

reliability: $\qquad P_{X|\{d\}}\left(A_X^{(\beta)}(\mathcal{D}_{j_0,h})\right) \geq \beta$ for $d \in \mathcal{D}_{j_0,h}$ and $h = 1, \ldots, n$

completeness: $\qquad \bigcup_{h=1}^{n} A_X^{(\beta)}(\mathcal{D}_{j_0,h}) = \mathcal{X}(\mathcal{D})$

precision: $\qquad \sum_{h=1}^{n} \frac{1}{|\mathcal{D}_{j_0,h}|} \sum_{d \in \mathcal{D}_{j_0,h}} \sum_{x \in \mathcal{X}(\{d\})} \left|\{d_k^{(j_0)} \mid x \in A_X^{(\beta)}(\mathcal{D}_{j_0,k})\}\right| P_{X|\{d\}}(\{x\}) = \min$

The properties guarantee that the procedure's reliability is at least equal to the required reliability level β, that the resulting measurement procedure is complete, and that it has minimum mean imprecision.

EXAMPLE

1. **Failure Probability**

   Consider Example 1 of Learning Unit 2.3.4, where knowledge about the number of treatments was assumed. Now we consider the case that there is ignorance about the number of trials and, therefore, the deterministic variable has to be extended. We obtain:

   Pair of variables:
   $$\begin{aligned}
   X &= \text{ number of failures among the treatments} \\
   D &= (D_1, D_2) \text{ with} \\
   &\quad D_1 = \text{number of treatments} \\
   &\quad D_2 = \text{failure probability}
   \end{aligned} \tag{97}$$

   Bernoulli Space:
   $$\begin{aligned}
   \mathcal{D} &= \{(n, p) \mid 18 \leq n \leq 22, 0 \leq p \leq 0.50\} \\
   \mathcal{X}(\{(n, p)\}) &= \{0, 1, 2, \ldots, n\} \\
   \mathcal{P}(\{(n, p)\}) &= P_{X|\{(n,p)\}} \text{ with } X|\{(n,p)\} \sim Bi(n, p)
   \end{aligned} \tag{98}$$

The measurement experiment is performed and $x = 3$ failures are observed. The corresponding measurement result $C_{D_2}^{(0.80)}(\{3\})$ is obtained by means of the Stochastikon Calculator:

$$C_D^{(0.80)}(\{3\}) = \{p \,|\, 0 \le p \le 0.28\} \tag{99}$$

A graphical representation of the measurement procedure is given in Figure 1. It is contained in the report of Stochastikon Calculator or can be obtained by means of the Graphical Laboratory.
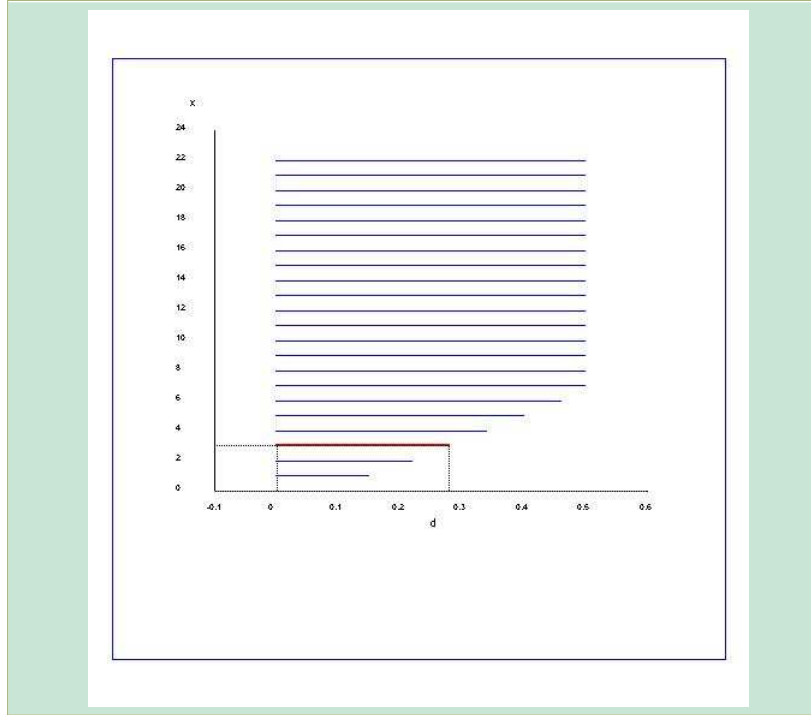

Figure 1: Graphical representation of the stochastic measurement procedure $C_D^{(0.80)}$ with the highlighted measurement result $C_D^{(0.80)}(\{3\})$

Unit 2.3.8: Maximum Lower Bound Measurement With Ignorance

TARGET

This learning unit refers just as the previous one to a one-dimensional variable of interest and considers the same situation as that in Learning Unit 2.3.7 and Learning Unit 2.3.6. Here the question is answered, how small the unknown value of a component of interest is by determining a lower bound.

CONTENT

Introduction

In this learning unit the same situation is considered as in the Learning Units 2.3.6 and 2.3.7. The deterministic variable has k components, i. e.

$$D = (D_1, \ldots, D_k) \tag{100}$$

with ignorance space $\mathcal{D}$ having N elements:

$$\mathcal{D} = \{d_1, \ldots, d_N\} \qquad \text{with} \qquad d_i = (d_{i,1}, \ldots, d_{i,k}), i = 1, \ldots, N \tag{101}$$

Let D_{j_0} be the component of interest with actual value $d_\ell^{(j_0)}$. It is required to determine a lower bound for $d_\ell^{(j_0)}$ by means of a β-measurement procedure.

The ignorance space with respect to the component D_{j_0} of interest contains n elements and is given by

$$\mathcal{D}_{j_0} = \{d_1^{(j_0)}, \ldots, d_n^{(j_0)}\} \tag{102}$$

with $d_\ell^{(j_0)} \in \mathcal{D}_{j_0}$.

For determining a lower bound for the unknown value $d_\ell^{(j_0)}$ a measurement procedure is needed with measurement results of the following form:

$$C_{D_{j_0}}^{(\beta)}(\{x\}) = \{d_h^{(j_0)} \in \mathcal{D}_{j_0} \mid \underline{d}^{j_0}(x) \le d_h^{(j_0)}\}$$

Note the difference between the minimum β-measurement procedure and the maximum lower bound β-measurement procedure. In the former case a lower and an upper bound for $d_\ell^{(j_0)}$ must be determined, while in the latter only an lower bound is needed.

Analogously to the case of a minimum upper bound β-measurement procedure, an appropriate prediction procedure has to be derived.

$$A_X : \mathcal{T}_{D_{j_0}}(\mathcal{D}) \to \mathcal{T}_X(\mathcal{X}(\mathcal{D}))$$

where the domain of $A_X^{(\beta)}$ is again given by the sets $\mathcal{D}_{j_0,h}$, $h = 1, \ldots, n$. The co-domain of $A_X^{(\beta)}$, however, is not the power set of $\mathcal{X}(\mathcal{D})$. Since only a lower bound for the unknown value $d_\ell^{(j_0)}$ must be determined, the predictions are of the following form:

$$A_X^{(\beta)}(\mathcal{D}_{j_0,h}) = \{x \in \mathcal{X}(\mathcal{D}_{j_0,h}) \mid x \le \bar{x}(\mathcal{D}_{j_0,h})\} \tag{103}$$

i. e., the co-domain of the prediction procedure consists of intervals which are bounded above.

Assume that a prediction procedure $A_X^{(\beta)}$ of the form given by (103) is available, then the corresponding measurement results are given as follows:

$$
\begin{aligned}
C_{D_{j_0}}^{(\beta)}(\{x\}) &= \{d_h^{(j_0)} \in \mathcal{D}_{j_0} \mid x \in A_X^{(\beta)}(\mathcal{D}_{j_0,h})\} \\
&= \{d_h^{(j_0)} \in \mathcal{D}_{j_0} \mid d_h^{(j_0)} \geq \underline{d}^{(j_0)}(x)\}
\end{aligned}
\tag{104}
$$

where $\underline{d}^{(j_0)}(x)$ is the smallest value in $\mathcal{D}_{j_0}$ for which the observed event $\{x\}$ had been predicted.

Formulation of the Problem

A maximum lower bound β-measurement procedure is based on a prediction procedure $A_X^{(\beta)}$ with predictions that are bounded above. The problem to be solves may be formulated as follows:

Determine a prediction procedure $^*A_X^{(\beta)}$ of the form given by (103) with the following properties:

reliability: $\qquad P_{X|\{d\}}\left(A_X^{(\beta)}(\mathcal{D}_{j_0,h})\right) \geq \beta$ for $d \in \mathcal{D}_{j_0,h}$ and $h = 1, \ldots, n$

completeness: $\qquad \bigcup_{h=1}^{n} A_X^{(\beta)}(\mathcal{D}_{j_0,h}) = \mathcal{X}(\mathcal{D})$

precision: $\qquad \sum_{h=1}^{n} \frac{1}{|\mathcal{D}_{j_0,h}|} \sum_{d \in \mathcal{D}_{j_0,h}} \sum_{x \in \mathcal{X}(\{d\})} \left| \left\{ d_k^{(j_0)} \mid x \in A_X^{(\beta)}(\mathcal{D}_{j_0,k}) \right\} \right| P_{X|\{d\}}(\{x\}) = \min$

The properties guarantee that the procedure's reliability is at least equal to the required reliability level β, that the resulting measurement procedure is complete, and that it has minimum mean imprecision.

EXAMPLE

1. **Success Probability**

   Again Example 1 of Learning Unit 2.3.1, which was also considered Learning Unit 2.3.5 shall be investigated, however, now there is ignorance also about the number of trials. The situation is quantified in the following way:

   Pair of variables:

   $$
   \begin{aligned}
   X &= \text{number of successes among the treated patients} \\
   D &= (D_1, D_2) \text{ with} \\
   &\quad D_1 = \text{number of treatments} \\
   &\quad D_2 = \text{success probability}
   \end{aligned}
   \tag{105}
   $$

   Bernoulli Space:

   $$
   \begin{aligned}
   \mathcal{D} &= \{(n,p) \mid 70 \leq n \leq 80, 0 \leq p \leq 1.0\} \\
   \mathcal{X}(\{(n,p)\}) &= \{0, 1, 2, \ldots, n\} \\
   \mathcal{P}(\{(n,p)\}) &= P_{X|\{p\}} \text{ with } X|\{(n,p)\} \sim Bi(n,p)
   \end{aligned}
   \tag{106}
   $$

   A measurement procedure with reliability level $\beta = 0.85$ is required. The clinical study is performed and for $x = 66$ patients a significant improvement was observed.

The measurement procedure $C_D^{(0.85)}(\{66\})$ can be obtained by means of Stochastikon Calculator:

$$C_{D_2}^{(\beta)}(\{66\}) = \{p \,|\, 0.78 \leq p\}$$

In Figure 1 a illustration of the entire measurement procedure is given. Such an illustration can be obtained by the Graphical Laboratory.
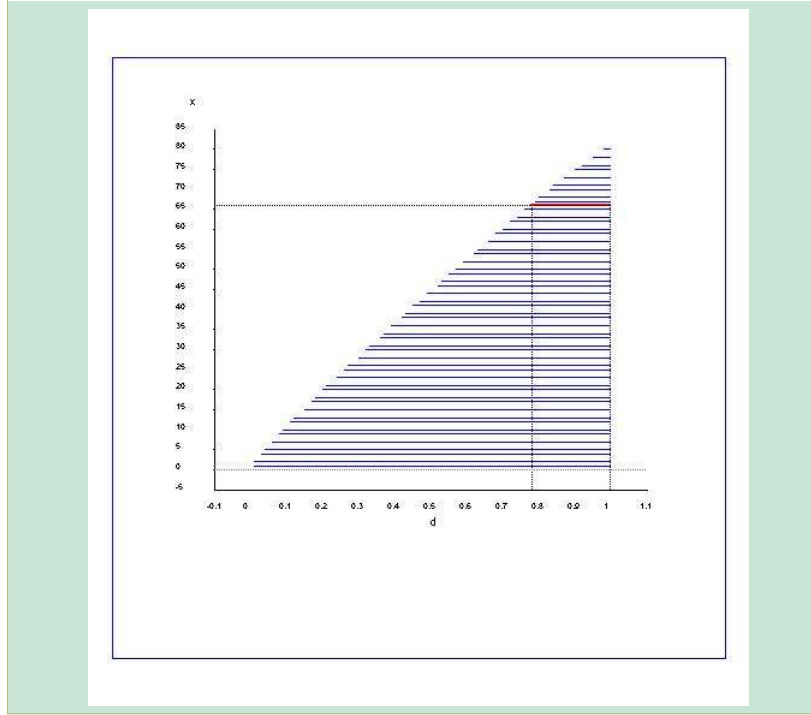

Figure 1: Graphical representation of the stochastic measurement $C_D^{(0.85)}$ with highlighted measurement result $C_D^{(0.85)}(\{66\})$.

The measurement yields a maximum lower bound of $p = 0.78$ for the success probability of the given treatment, which is the same as in Example 1 of Learning Unit 2.3.5. The reason that the additional ignorance does not change the measurement result is the fact that the upper bound $n = 80$, which was assumed in Learning Unit 2.3.5, represents the worst case with respect to the measurement result.

Unit 2.3.9: Measurement of the First Moment

TARGET

The central problem in stochastic mod-
eling is the determination of the vari-
ability function and the random struc-
ture function for a given pair of vari-
ables (X, D). In many cases this can
be achieved in good approximation by
identifying a probability distribution
which agrees with the true distribution
in the range of variability, the value of
the first moment and the value of the
variance.

In this learning unit a stochastic procedure for measuring the value of the first moment of
a random variable X is introduced. Actually, this procedure is well-known in statistics, but
it will be derived here, as a stochastic β-measurement procedure for $E[X]$ in the framework
given by the Bernoulli Space.

CONTENT

Introduction

The first moment $E[X]$ of a random variable X is, besides its range of variability, the most im-
portant deterministic quantity, and the development of an appropriate stochastic measurement
procedure for $E[X]$ has therefore a high priority. This is particularly true, because procedures
for determining the range of variability are based on the known values of the first moments
(see Learning Units 2.3.11 and 2.3.12). Therefore, a measurement procedure for $E[X]$ must be
developed, which does not assume that the corresponding range of variability is known.

In fact, since the beginnings of statistics, the problem of determining the actual value μ of $E[X]$
built the center of the joint efforts. When in 1908 William S. Gosset succeeded to derive the
so-called t-distribution, the problem was solved in an approximate, but rather general way.

Measurement Experiment

Let a pair of variables (X, D) be given, and assume that the first moment of X is one of
the components of D. The problem is to determine the unknown value μ of the first moment
$E[X|\{d\}]$ of the random variable $X|\{d\}$, where d is the actual value of the deterministic variable
D in the given situation.

The measurement process consists of a number of independent copies of $X|\{d\}$, which are called
random sample of size n of $X|\{d\}$:

$$(X_1|\{d\}, X_2|\{d\}, \ldots, X_n|\{d\})$$

The following approximation is based on the normal distribution, i.e., on the Central Limit
Theorem, which refers to a suitably normalized sum of random variables. If the probability
distribution $P_{X|\{d\}}$ is of constant or uni-modal type, then already a small number of copies is
sufficient for using the normal distribution as an approximation, since the probability distri-
bution of the normalized sum of $X_i|\{d\}$, $i = 1, \ldots, n$, converges very rapidly to the normal

distribution. If the probability distribution is of any other type, the number of copies must generally be much larger, because the convergence is slower.

The Normal Distribution

The normal or Gauss distribution is the most frequently assumed probability distribution in statistics, since a majority of statistical methods are based on the normal distribution. The normal distribution is a continuous approximation $\hat{P}_{X|\{d\}}$ of a probability distribution $P_{X|\{d\}}$ of a random variable $X|\{d\}$. Originally, it has been derived as a limit for the probabilities of a binomial distribution by Abraham Moivre. The continuous distribution function is given below:

$$\hat{F}_{X|\{d\}}(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \, dy \quad \text{for } x \in \mathbb{R} \tag{107}$$

The integral is analytical not solvable, however, there are tables and software programs providing the solutions. The value μ stands for the first moment and the value σ^2 for the variance of $\hat{P}_{X|\{d\}}$. Note that the true range of variability is not approximated but just replaced by the set of real numbers.

The Observational Variable

For the measurement process $(X_1|\{d\}, X_2|\{d\}, \ldots, X_n|\{d\})$ an observational (random) variable must be defined. The observed value of this variable will be used for determining the measurement result with respect to the unknown value μ of $E[X|\{d\}]$.

The observational variable is defined on the random sample and represents the result of the measurement experiment, which is the input value for the stochastic measurement procedure to be developed.

- Random observational variable:

$$T(\mu)|\{(n,d)\} = \frac{\overline{X}|\{(n,d)\} - \mu}{S|\{(n,d)\}} \sqrt{n}$$

with

$$\overline{X}|\{(n,d)\} = \frac{1}{n} \sum_{i=1}^{n} X_i|\{d\}$$

$$S^2|\{(n,d)\} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i|\{d\} - \overline{X}|\{(n,d)\})^2$$

For each realization $(x_1, \ldots, x_n)$ one obtains the following quantity $t(\mu)$ as realization of $T(\mu)|\{(n,d)\}$:

$$t(\mu) = \frac{\overline{x} - \mu}{s} \sqrt{n} \qquad \text{for } \mu \in \mathbb{R}$$

where

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{and} \quad s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

It follows that the range of variability of $T(\mu)|\{(n,d)\}$ consists of decreasing straight lines $t(\mu)$ in μ:

$$\mathcal{X}_{T(\mu)} = \{t(\mu) \mid \mu \in \mathbb{R}\}$$

As to the random structure of $T(\mu)|\{(n, d)\}$, William Gosset and later Ronald A. Fisher showed that a generally good approximation of the true probability distribution is given by the Student t-distribution, which is independent of the actual value d of D, which is therefore abandoned in the notation of $T(\mu)$:

$$T(\mu)|\{n\} \sim t_{n-1}$$

where t_{n-1} denotes the t-distribution with $(n-1)$ degrees of freedom.

The degree of freedom is fixed by the sample size, which completely determines the corresponding t-distribution. Thus, the deterministic variable N denoting the sample size with actual value n is selected as deterministic variable with respect to the random observational variable $T(\mu)$.

The density function of the t-distribution resembles very much the density function of the standardized normal distribution. The difference is that density function of the t-distribution is slightly flatter than that of the standardized normal distribution.
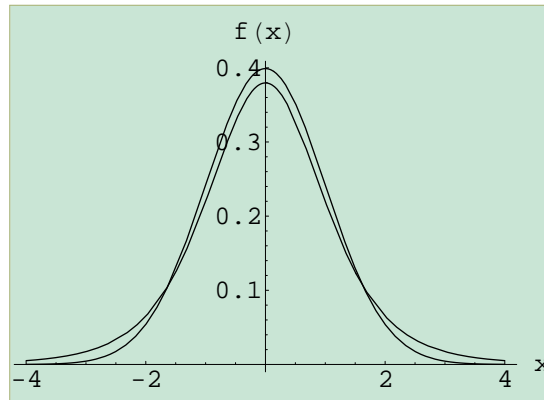

Figure 1: Density functions of the
t-distribution t_5 and the standardized
normal distribution $N(0, 1)$

The two first moments of $T(\mu)|\{n\}$ are given as follows:

$$\begin{aligned} E[T(\mu)|\{n\}] &= 0 & \text{for} \quad n > 2 \\ V[T(\mu)|\{n\}] &= \tfrac{n-1}{n-3} & \text{for} \quad n > 3 \end{aligned}$$

Next, a relation between the value n of the deterministic variable N, which constitutes the initial condition, and the future outcome of $T(\mu)|\{n\}$ has to be established by means of prediction procedure $A_{T(\mu)}^{(\beta)}$.

Lemma 1: An prediction procedure $A_{T(\mu)}^{(\beta)}$ determining a minimum β-measurement procedure for $E[X]$ is given as follows:

$$A_{T(\mu)}^{(\beta)}(\{n\}) = \left\{ t(\mu) \,\Big|\, Q_{t(n-1)}^{(\ell)}\left(\frac{1+\beta}{2}\right) \leq t(\mu) \leq Q_{t(n-1)}^{(u)}\left(\frac{1+\beta}{2}\right) \right\}$$

80

where $Q^{(\ell)}_{t(n-1)}$ denotes the lower quantile function and $Q^{(u)}_{t(n-1)}$ the upper quantile function of the t-distribution with $(n-1)$ degrees of freedom.

Proof: The t distribution is symmetric, i. e.:

$$f_{T(\mu)|\{n\}}(-x) = f_{T(\mu)|\{n\}}(x) \quad \text{for } x \in \mathbb{R}$$

It follows that the probability of exceeding the upper bound of the prediction and the probability of falling short the lower bound must be equal for an optimal, i. e., minimum β-measurement procedure.

The overall error probability must not exceed $1 - \beta$, as otherwise the reliability requirement is not met. Thus, the probability of exceeding the upper bound must not go beyond $\frac{1-\beta}{2}$ and the same holds for the probability for falling short the lower bound.

Hence, the upper bound is given by the upper quantile of order $1 - \frac{1-\beta}{2} = \frac{1+\beta}{2}$, while the lower bound is given by the lower quantile of the same order.

A measurement procedure is complete, if for any possible observation there is at least one prediction having a non-empty intersection with the observation. The observations are given by the decreasing straight lines $t(\mu)$ for $\mu \in \mathbb{R}$, while the predictions $A^{(\beta)}_{T(\mu)}(\{n\})$ for $\mu \in \mathbb{R}$ are given as two horizontal parallels. Obviously, each straight line $t(\mu)$ intersects a number of predictions, which completes the proof.

Note that the completeness requirement is not met, if the value μ is restricted to a bounded set. However, since is was assumed that there is complete ignorance about μ, the completeness requirement is met.   ●

Figure 2, below, displays the predictions $A^{(\beta)}_{T(\mu)}(\{n\})$ for $5 \leq \mu \leq 10$.
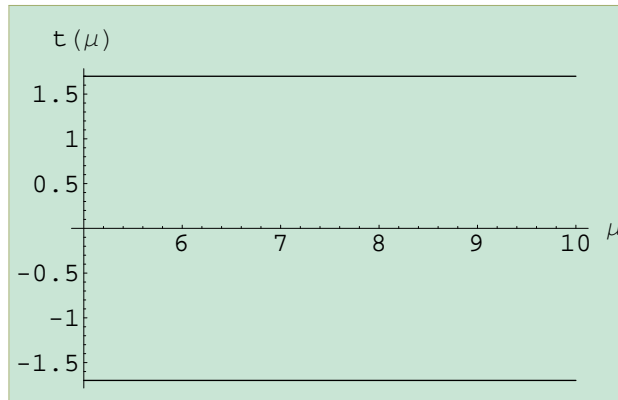

Figure 2: The predictions $A^{(\beta)}_{T(\mu)}(\{n\})$
depending on μ given as two parallels for
$n = 30$ and $\beta = 0.90$

The Bernoulli Space $\mathbb{B}_{T(\mu),N}$

For illustration the Bernoulli Space for the measurement experiment given by the variables $(T(\mu), N)$ shall be stated here.

- Ignorance space:

$$\mathcal{D}_N = \{n\}$$

i. e., there is complete knowledge about the initial conditions.

- Variability function:

$$\mathcal{X}_{T(\mu)}(\{n\}) = \{t(\mu) \,|\, \mu \in \mathbb{R}\}$$

The elements of the range of variability are decreasing straight lines.

- Random structure function:

$$\mathcal{P}_{T(\mu)}(\{n\}) = P_{T(\mu)|\{\mu\}} \text{ with } T(\mu)|\{n\} \sim t_{n-1} \quad \text{for } \mu \in \mathbb{R}$$

The elements of the images of the random structure function are approximated by the corresponding t-distribution with $(n-1)$ degrees of freedom.

The t-distribution has the following density function, i.e., the derivative of the distribution function:

$$f_{T(\mu)|\{n\}}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \qquad \text{for } x \in \mathbb{R}$$

By means of the predictions, which are available for each possible value μ of $E[X|\{d\}]$, it becomes possible to derive the measurement procedure $C_{E[X]}^{(\beta)}$ by the corresponding measurement results.

Measurement Procedure for $E[X]$

As soon as the predictions are available, the problem of deriving a measurement procedure $C_{E[X]}^{(\beta)}$ for given reliability level β for the unknown value μ of $E[X]$ is solved by setting:

$$C_{E[X]}^{(\beta)}(\{t(\mu)\}) = \left\{\mu \,|\, t(\mu) \in A_{T(\mu)}^{(\beta)}(\{n\})\right\} \tag{108}$$

where

$$P_{T(\mu)|\{n\}}\left(A_{T(\mu)}^{(\beta)}(\{n\})\right) \geq \beta \quad \text{for } \mu \in \mathbb{R}$$

Since the prediction procedure meets the completeness requirement, there is for any observation $t(\mu)$ a non-empty measurement result.

Replacing $t(\mu)$ by $\frac{\bar{x}-\mu}{s}\sqrt{n}$ and by inserting the predictions $A_{T(\mu)}^{(\beta)}(\{n\})$ the following measurement results are obtained:

$$
\begin{aligned}
C_{E[X]}^{(\beta)}(\{t(\mu)\}) &= C_{E[X]}^{(\beta)}\left(\left\{\frac{\bar{x}-\mu}{s}\sqrt{n}\right\}\right) \\
&= \left\{\mu \,|\, \bar{x} + Q_{T(\mu)|\{n\}}^{(\ell)}\left(\frac{1+\beta}{2}\right)\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + Q_{T(\mu)|\{n\}}^{(u)}\left(\frac{1+\beta}{2}\right)\frac{s}{\sqrt{n}}\right\}
\end{aligned}
$$

Graphical Representation of $C_{E[X]}^{(\beta)}$

Let $n = 30$ and $\beta = 0.95$, then the following prediction is obtained:

$$A_{T(\mu)}^{(0.95)}(\{30\}) = \{t(\mu) \,|\, -1.99 \leq t(\mu) \leq 1.99\}$$

Consider a realization $(x_1, x_2, \ldots, x_{30})$ yielding $\bar{x} = 7$ and $s^2 = 0.5$.

Figure 3 displays the observation $t(\mu)$ as function of μ, which is given by the straight line running from top left to bottom right, and the prediction $A_T^{(\beta)}(\{30\})$ given by the two parallels.
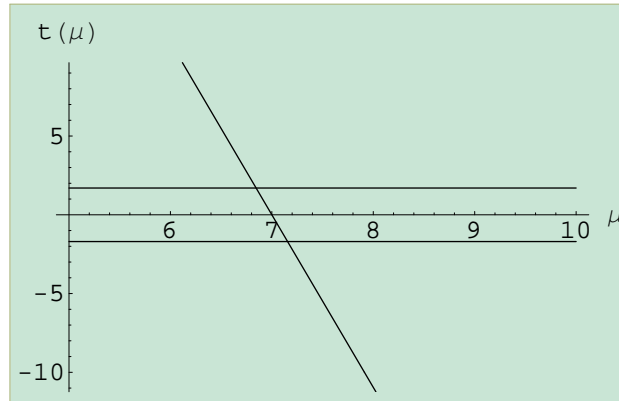

Figure 3: Graphical representation of the measurement result of the procedure $C_{E[X]}^{(0.95)}$ for $t(\mu) = \frac{7-\mu}{0.5} \sqrt{30}$.

The measurement result contains all those values of $\mu \in \mathbb{R}$, for which $t(\mu)$ is within the prediction given by the parallels in Figure 1.

The measurement result for the above given realization of the measurement experiment is as follows:

$$C_{E[X]}^{(0.95)} \left(\frac{7-\mu}{0.5} \sqrt{30} \right) = \{\mu \,|\, 6.743 \leq \mu \leq 7.257\}$$

EXAMPLE

1. **First Moment of a Wind Load**

Figure 4: Offshore wind turbines have to stand high load values.

The design of wind turbines is mainly determined by the maximum loads the turbines have to withstand. In order to determine the maximum loads, the probability distributions of different load types have to be determined. Let the load X be of Type M_x, which represents the momentum in x-direction. It is known that for given wind conditions the random variable X has a uni-modal probability distribution.

The first moment $E[X]$ is the most important deterministic variable with respect to a given probability distribution. The determination of the actual value μ_0 of the wind load $X|\{d\}$ for given wind conditions is based on a sample $(X_1|\{d\}, X_2|\{d\}, \ldots, X_n|\{d\})$ and

the two sample functions:

$$\overline{X}|\{(n,d)\} \;=\; \tfrac{1}{n}\sum_{i=1}^{n} X_i|\{d\}$$

$$S^2|\{(n,d)\} \;=\; \tfrac{1}{n-1}\sum_{i=1}^{n}(X_i|\{d\} - \overline{X}|\{(n,d)\})^2$$

Because an erroneously too low value of the first moment could lead to safety problems, the reliability requirement is set to be $\beta = 0.995$ and the sample size to $n = 250$. The problem is to determine an upper bound for the unknown value μ_0 of $E[X|\{d\}]$.

For determining an upper bound, the observational variable is given by $T(\mu)|\{n\}$ with the predictions procedure given by the predictions of the following form:

$$A_{T(\mu)}^{(0.995)}(\{\mu\}) = \left\{ t(\mu) \,|\, t(\mu) \geq Q_{T(\mu)|\{250\}}^{(\ell)}(0.995) \right\}$$

where $T(\mu)|\{n\} = \dfrac{\overline{X}|\{(n,d)\} - \mu}{S|\{(n,d)\}}\sqrt{250}$ is distributed according to the t-distribution with 249 degrees of freedom, and $Q_{T(\mu)|\{250\}}^{(\ell)}$ is the lower quantile function of $T(\mu)|\{250\}$.
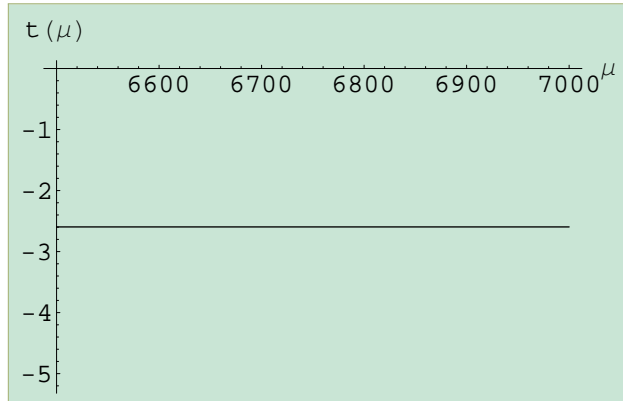

Figure 5: The predictions $A_{T(\mu)}^{(0.995)}(\{250\})$
defining the measurement procedure
$C_{E[X]}^{(0.995)}(\{t(\mu)\})$

The measurement experiment is performed with $n = 250$ and yields the following values $\bar{x} = 6784$ and $s^2 = 423.5$ by means of which the observation $t(\mu)$ is obtained.

In Figure 6 the predictions, which do not depend on μ, are displayed by the line parallel to the μ-axes through $(0, -2.596)$, which represents the lower bound of $A_{T(\mu)}^{(0.995)}(\{250\})$. The realization $t(\mu)$ of $T(\mu)|\{250\}$ is represented by the decreasing straight line.
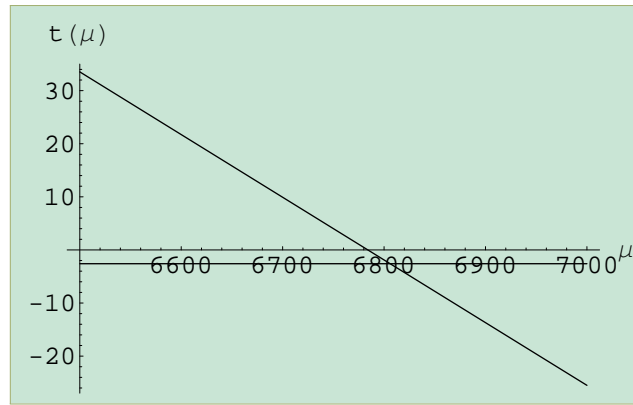
Figure 6: The predictions $A_{T(\mu)}^{(0.995)}(\{250\}) =$
$\{t(\mu) \,|\, t(\mu) \geq Q_{T(\mu)|\{250\}}^{(\ell)}(0.995)\}$ and the
realization of the measurement variable $t(\mu)$

The value of μ where $t(\mu)$ leaves the predicted area is the upper bound of the measurement result:

$$C_{E[X]}^{(0.995)}(\{t(\mu)\}) = \{\mu \,|\, \mu \leq 6805.99\}$$

This result means that the value of the first moment of the considered wind load does not exceed the value 6805.99.

Unit 2.3.10: Measurement of the Variance

Target

Besides the range of variability and the first moment, the variance of a random variable is extremely important, as it reflects size of the variability and, hence, the underlying uncertainty. Therefore, a β-measurement procedure for $V[X]$ is necessary. Similar as in the case of the first moment, a measurement procedure is needed, which does not assume knowledge about the actual probability distribution.

In this learning unit a measurement procedure is introduced based on the χ^2-distribution, which was proposed in 1876 by the German geodesist and mathematician Friedrich Robert Helmert[9].

Content

Introduction

According to the minimum information principle, for deriving an appropriate approximation for a uni-modal probability distribution, the value of the corresponding variance must be known. As many random variables X have a uni-modal probability distribution, the development of an appropriate stochastic measurement procedure for $V[X]$ is extremely important.

This is also important, because the procedure for determining the range of variability of a random variable assumes that the value σ^2 of the variance is known.

Similar as in the case of a measurement procedure for the first moment $E[X]$, the problem to determine the actual value σ^2 of the variance $V[X]$ is one of the oldest problems dealt with in statistics. The German geodesist and mathematician Friedrich Robert Helmert derived the χ^2-distribution as the probability distribution for a sum of squares of random variables each having a standardized normal probability distribution[10]. In the following the χ^2-distribution is used for deriving a generally applicable measurement procedure for determining the actual value of the variance of a random variable.

Measurement Process

Analogously as in the preceding learning unit, let a pair of variables (X, D) be given, and assume that the variance of X is one of the components of D. The problem is to determine the unknown value σ of the variance $V[X|\{d\}]$ of the random variable $X|\{d\}$, where d is the actual value of the deterministic variable D in the given situation.

The measurement process consists of n independent copies of $X|\{d\}$, i.e., of a random sample of size n of $X|\{d\}$:

$$(X_1|\{d\}, X_2|\{d\}, \ldots, X_n|\{d\})$$

---

[9]Friedrich Robert Helmert born 1843 in Freiberg (Sachsen) and death in 1917 in Potsdam). Helmert is considered the founder of the mathematical theory of geodesy.

[10]A normal distribution with $\mu = 0$ and $\sigma^2 = 1$ is called "standardized normal distribution".

Similar as the approximation used in the preceding learning unit, also the following approximation is based on the normal distribution, i. e., on the Central Limit Theorem. Therefore, again, if the probability distribution $P_{X|\{d\}}$ of $X|\{d\}$ is of constant or uni-modal type, already a small number of copies is sufficient for obtaining a sufficiently good approximation, because the probability distribution of a suitably normalized sum of the random variables will tend rapidly to the normal distribution. If the probability distribution is of any other type, the number of copies must be larger, because the convergence is slower.

For the measurement process $(X_1|\{d\}, X_2|\{d\}, \ldots, X_n|\{d\})$ an observational (random) variable must be defined. The observed value of this variable will be used for determining the measurement result with respect to the unknown value of $V[X|\{d\}]$.

The Observational Variable

The observational variable is defined on the sample and represents the result of the measurement process, which is the input value for the stochastic measurement procedure to be developed.

- Random observational variable:

$$R(\sigma^2)|\{(n,d)\} = (n-1)\frac{S^2|\{(n,d)\}}{\sigma^2}$$

    with

$$S^2|\{(n,d)\} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i|\{d\} - \overline{X}|\{(n,d)\})^2$$

For each realization $(x_1, \ldots, x_n)$, the quantity $r(\sigma^2)$ is obtained as realization of $R(\sigma^2)|\{(n,d)\}$:

$$r(\sigma^2) = (n-1)\frac{s^2}{\sigma^2} \qquad \text{for } \sigma^2 \in \mathbb{R}^+$$

where

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

It follows that each element $r(\sigma^2)$ of the range of variability of $R(\sigma^2)|\{(n,d)\}$ is a decreasing and concave curve in σ^2:

$$\mathcal{X}_{R(\sigma^2)} = \left\{ r(\sigma^2) \,|\, \sigma^2 \in \mathbb{R}^+ \right\}$$

The following limits are obtained for the function $r(\sigma^2)$:

$$\lim_{\sigma^2 \to 0} r(\sigma^2) = \infty$$
$$\lim_{\sigma^2 \to \infty} r(\sigma^2) = 0$$

As to the random structure of $R(\sigma^2)|\{(n,d)\}$, a generally very good approximation of the true probability distribution, is given by the χ^2-distribution which is independent of the actual value d of D. Hence we set:

$$R(\sigma^2)|\{n)\} \sim \chi^2_{n-1}$$

where χ^2_{n-1} denotes the χ^2-distribution with $(n-1)$ degrees of freedom.

The degree of freedom determined by the sample size n is the only deterministic variable of the χ^2, which completely determines the χ^2-distribution. Thus, the deterministic variable N denoting the sample size with actual value n is selected as deterministic variable for the random observational variable $R(\sigma^2)$.

The density function of the χ^2-distribution resembles already for moderate large values of n much the density function of the corresponding standardized normal distribution. The difference is that density function of the χ^2-distribution is not symmetric like the standardized normal distribution and slightly moved to the left with range of variability $\mathbb{R}^+$.
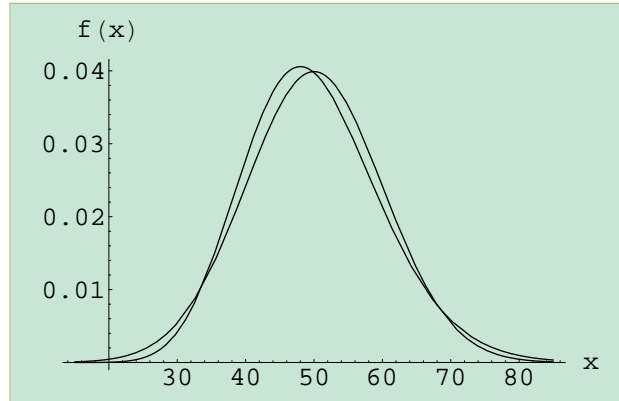

Figure 1: Density functions of the
χ^2-distribution χ^2_{50} and the standardized
normal distribution $N(50, 10)$

The two first moments of $R(\sigma^2)|\{n\}$ are given as follows:

$$E[R(\sigma^2)|\{n\}] = n - 1 \qquad \text{for} \quad n > 1$$
$$V[R(\sigma^2)|\{n\}] = 2(n - 1) \quad \text{for} \quad n > 1$$

Next, a relation between the value n of the deterministic variable N, which constitutes the initial conditions, and the future outcome of $R(\sigma^2)|\{n\}$ has to be established by means of a prediction procedure $A_{R(\sigma^2)}^{(\beta)}$.

Lemma 1: For any y with $\beta \leq y \leq 1$, a β-prediction procedure $A_{R(\sigma^2)}^{(\beta)}$ is given by the following predictions:

$$A_{R(\sigma^2)}^{(\beta)}(\{n\}) = \left\{ r(\sigma^2) \,\middle|\, Q_{\chi^2(n-1)}^{(\ell)}(1 - y + \beta) \leq r(\sigma^2) \leq Q_{\chi^2_{n-1}}^{(u)}(y) \right\}$$

where $Q_{\chi^2_{n-1}}^{(\ell)}$ denotes the lower quantile function and $Q_{\chi^2_{n-1}}^{(u)}$ the upper quantile function of the χ^2-distribution with $(n-1)$ degrees of freedom.

Proof:
The probability that the upper bound of the prediction $A_{R(\sigma^2)}^{(\beta)}(\{n\})$ is exceeded is given by the value $1 - y$, while the probability that the lower bound is not exceeded is given by the value $1 - (1 - y + \beta)$. Hence the overall error probability is obtained as:

$$(1 - y) + 1 - (1 - y + \beta) \leq 1 - \beta$$

Thus, the given predictions meet the reliability requirement β.

The measurement procedure is complete, if for any possible observation there is at least one prediction having a non-empty intersection with the observation. The observations are given by the decreasing concave curves $r(\sigma^2)$ for $\sigma^2 \in \mathbb{R}^+$, while the predictions $A^{(\beta)}_{R(\sigma^2)}(\{n\})$ for $\mu \in \mathbb{R}$ are given as two parallels. Obviously, each of the curves $r(\sigma^2)$ intersects a prediction which completes the proof.

Note that, similar as in the case of $E[X]$, the completeness requirement is not met, if the possible values σ^2 are restricted to a bounded set.   ●

Figure 2, below, displays the predictions $A^{(\beta)}_{R(\sigma^2)}(\{n\})$ and the observation $r(\sigma^2)$.
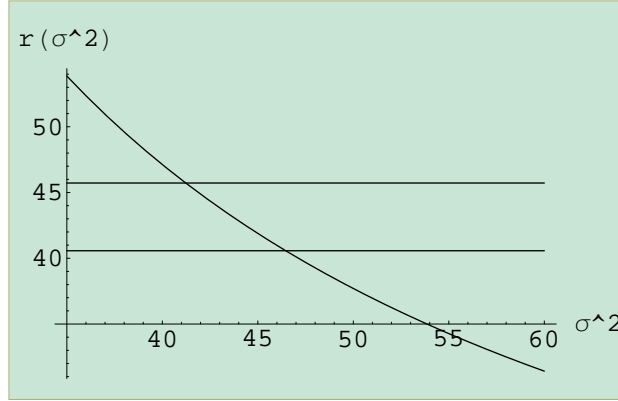

Figure 2: The predictions $A^{(0.90)}_{R(\sigma^2)}(\{30\})$ for
$y = 0.975$ and the observation $r(\sigma^2)$ for
$s^2 = 65$.

The Optimization Problem

By Lemma 1, any β-measurement procedure $C^{(\beta)}_{V[X]}$ for $V[X]$ is given as follows:

$$C^{(\beta)}_{V[X]}(\{r(\sigma^2)\}) = \left\{\sigma^2 \,|\, r(\sigma^2) \in A^{(\beta)}_{R(\sigma^2)}(\{n\})\right\} \tag{109}$$

An equivalent formulation based on the random variable S^2 instead of $R(\sigma^2)$ is given by:

$$
\begin{aligned}
C^{(\beta)}_{V[X]}(\{s^2\}) &= \left\{\sigma^2 \,|\, s^2 \in A^{(\beta)}_{S^2}(\{(n, \sigma^2)\})\right\} \\
&= \left\{\sigma^2 \,\Big|\, \frac{Q^{(\ell)}_{\chi^2(n-1)}(1 - y + \beta)}{n-1}\sigma^2 \le s^2 \le \frac{Q^{(u)}_{\chi^2_{n-1}}(y)}{n-1}\sigma^2\right\} \\
&= \left\{\sigma^2 \,\Big|\, \frac{n-1}{Q^{(u)}_{\chi^2_{n-1}}(y)}s^2 \le \sigma^2 \le \frac{n-1}{Q^{(\ell)}_{\chi^2(n-1)}(1 - y + \beta)}s^2\right\} \tag{110}
\end{aligned}
$$

With the above representation of $C^{(\beta)}_{V[X]}$ as function of the observations s^2 of the random variable $S^2|\{n, \sigma^2\}$, the following result with respect to an optimal β-measurement procedure is obtained:

Lemma 2: The β-measurement procedure

$$
^*C^{(\beta)}_{V[X]}(\{s^2\}) = \left\{\sigma^2 \,\Big|\, \frac{n-1}{Q^{(u)}_{\chi^2(n-1)}(y^*)}s^2 \le \sigma^2 \le \frac{n-1}{Q^{(\ell)}_{\chi^2_{n-1}}(1 - y^* + \beta)}s^2\right\}
$$

for the variance $V[X|\{d\}]$ of a random variable $X|\{d\}$ is optimal with respect to the accuracy, if y^* is a solution of the following optimization problem:

$$y^* = \arg \min_{\beta \leq y \leq 1} \left(\frac{1}{Q^{(\ell)}_{\chi^2(n-1)}(1 - y + \beta)} - \frac{1}{Q^{(u)}_{\chi^2_{n-1}}(y)} \right) \tag{111}$$

Proof: The proof of Lemma 2 is straightforward. The β-measurement procedure is optimal, if the mean length of the measurement results is minimum. Here, however, (see (110)), it is possible to minimize the results simultaneously, which can be seen by the following representation:

$$
\begin{aligned}
|C^{(\beta)}_{V[X]}(\{s^2\})| &= \frac{n-1}{Q^{(\ell)}_{\chi^2_{n-1}}(1 - y^* + \beta)} s^2 - \frac{n-1}{Q^{(u)}_{\chi^2(n-1)}(y^*)} s^2 \\
&= \left(\frac{1}{Q^{(\ell)}_{\chi^2(n-1)}(1 - y + \beta)} - \frac{1}{Q^{(u)}_{\chi^2_{n-1}}(y)} \right) (n-1)s^2
\end{aligned}
\tag{112}
$$

Thus, the length of the measurement results for any observation s^2 gets minimum, if y^* given by (111) is used.   ●.

Comparison of Measurement Procedures for $V[X]$

1. Traditional Statistical Method

   Here a brief numerical comparison between the traditional statistical method denoted by $^{tr}C^{(\beta)}_{V[X]}$, a β-measurement procedure based on the shortest predictions denoted by $^{sh}C^{(\beta)}_{V[X]}$ and the optimal β-measurement procedure $^*C^{(\beta)}_{V[X]}$ obtained according to Lemma 2 shall illustrate the differences.

   The measurement procedure for $V[X]$ in traditional statistics is defined by symmetric error probabilities as follows:

$$^{tr}C^{(\beta)}_{V[X]}(\{s^2\}) = \left\{ \sigma^2 \,\middle|\, \frac{n-1}{Q^{(u)}_{\chi^2_{n-1}}\left(\frac{1+\beta}{2}\right)} s^2 \leq \sigma^2 \leq \frac{n-1}{Q^{(\ell)}_{\chi^2(n-1)}\left(\frac{1+\beta}{2}\right)} s^2 \right\} \tag{113}$$

   Consider the special case $n = 40$ and $\beta = 0.95$ and assume that the measurement process yields the observation $s^2 = 0.2$, then the following measurement result is obtained by means of the traditional statistical method:

$$^{tr}C^{(\beta)}_{V[X]}(\{s^2\}) = \{\sigma^2 \,|\, 0.134 \leq \sigma^2 \leq 0.330\} \tag{114}$$

   with length:

$$\left| {}^{tr}C^{(\beta)}_{V[X]}(\{s^2\}) \right| = 0.196$$

2. Measurement Procedure Based on Shortest Predictions

   In statistical textbooks it is sometimes mentioned that the measurement procedure (113) is not optimal, because the χ^2-distribution is not symmetric and, therefore, fixing symmetric

error probabilities do not yield a shortest prediction interval. However, it is also mentioned that the differences are small and may be neglected.

In a first step the value y in (109) is determined, which minimizes the length of the prediction interval $A_{R(\sigma^2)}^{(\beta)}(\{n\})$ for given values of n and β.

The obtained value of y is used for calculating the measurement results according to (110). Consider the same example of $n = 40$ and $\beta = 0.95$ as in the case of the traditional method.

$$^{sh}C_{V[X]}^{(\beta)}(\{s^2\}) = \left\{\sigma^2 \,|\, 0.138 \leq \sigma^2 \leq 0.346\right\} \tag{115}$$

with length:

$$\left|\,^{sh}C_{V[X]}^{(\beta)}(\{s^2\})\,\right| = 0.208$$

Surprisingly, the shortest predictions do not yield an improvement over the traditional method, but contrary a change to the worse. This clearly shows that optimizing a prediction procedure is a different problem than optimizing a measurement procedure.

3. Optimal Measurement Procedure

By solving the optimization problem (111), the optimal value y^* is obtained yielding with (110) optimal, i. e., shortest measurement results.

Taking again the example of $n = 40$ and $\beta = 0.95$ as in the preceding cases, the optimal β-measurement procedure yields:

$$^*C_{V[X]}^{(\beta)}(\{s^2\}) = \left\{\sigma^2 \,|\, 0.125 \leq \sigma^2 \leq 0.312\right\} \tag{116}$$

with length:

$$\left|\,^*C_{V[X]}^{(\beta)}(\{s^2\})\,\right| = 0.187$$

The improvement of the optimal measurement result when compared with the result of the traditional method is with almost 5% astonishing large. If the optimal result is compared with the result based on shortest predictions, the improvement is even larger than 11%. Thus, the use of the optimal β-measurement procedures leads generally to non-negligible advantages.

Bernoulli Space $\mathbb{B}_{R(\sigma^2),N}$

For completeness, the Bernoulli Space $\mathbb{B}_{R(\sigma^2),N}$ generated by the pair of variables $(R(\sigma^2), N)$ is given below:

- Ignorance space:
$$\mathcal{D}_N = \{n\}$$
Just as in the case of the measurement process for $E[X]$, there is complete knowledge for the approximation.

- Variability function:
$$\mathcal{X}_{R(\sigma^2)}(\{\sigma^2\})) = \{r(\sigma^2) \,|\, \sigma^2 \in \mathbb{R}^+\}$$

- Random structure function:

$$\mathcal{P}_{R(\sigma^2)}(\{n\}) = P_{R(\sigma^2)|\{n\}} \text{ with } R(\sigma^2)|\{n\} \sim \chi^2_{n-1}$$

The actual probability distributions are approximated by the χ^2-distribution with $(n-1)$ degrees of freedom. The density function of the χ^2-distribution with $(n-1)$ degrees of freedom is given below:

$$f_{R(\sigma^2)|\{n\}}(x) = \frac{x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)} \qquad \text{for } x \in \mathbb{R}^+$$

Graphical Representations of $C^{(\beta)}_{V[X]}$

Let $n = 30$ and $\beta = 0.90$, then we obtain the following predictions with the traditional (symmetric) approach and the optimal stochastic approach

- Traditional approach:

$$^{tr}A^{(0.95)}_{R(\sigma^2)}(\{\sigma^2)\}) = \left\{r(\sigma^2)\,|\,17.708 \leq r(\sigma^2) \leq 42.557\right\}$$

which is equivalent to:

$$^{tr}A^{(0.95)}_{S^2}(\{\sigma^2)\}) = \left\{s^2\,|\,0.610634\sigma^2 \leq s^2 \leq 1.46748\sigma^2\right\}$$

and

- Stochastic approach:

$$^*A^{(0.95)}_{R(\sigma^2)}(\{\sigma^2)\}) = \left\{r(\sigma^2)\,|\,18.3635 \leq r(\sigma^2) \leq 44.022\right\}$$

which is equivalent to:

$$^*A^{(0.95)}_{S^2}(\{\sigma^2)\}) = \left\{s^2\,|\,0.633224\sigma^2 \leq s^2 \leq 1.518\sigma^2\right\}$$

Consider a realization $(x_1, x_2, \ldots, x_{30})$ yielding $s^2 = 2.1$. Then Figure 3 displays the observation $r(\sigma^2)$ as function of σ^2, which is given by the curve from top left to bottom right, and the prediction $^{tr}A^{(\beta)}_R(\{\sigma^2\})$ given by the two parallels.
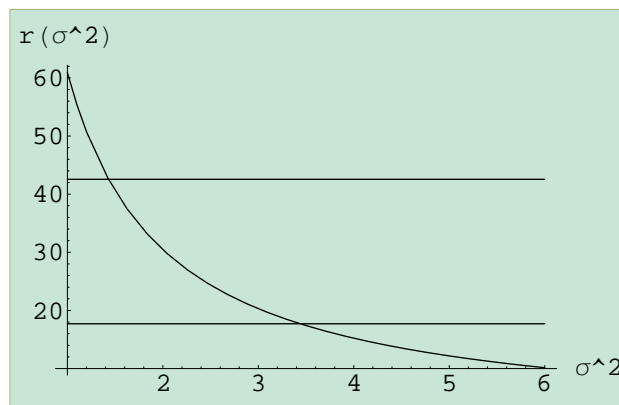

Figure 3: Graphical representation of the
traditional measurement procedure for $V[X]$
in case of the observation $r(\sigma^2) = 29\,\frac{2.1}{\sigma^2}$.

In Figure 4, the same situation as in Figure 3 is shown, however, by means of the random variable $S^2|\{n, \sigma^2\}$ with the realization given by $s^2 = 2.1$.
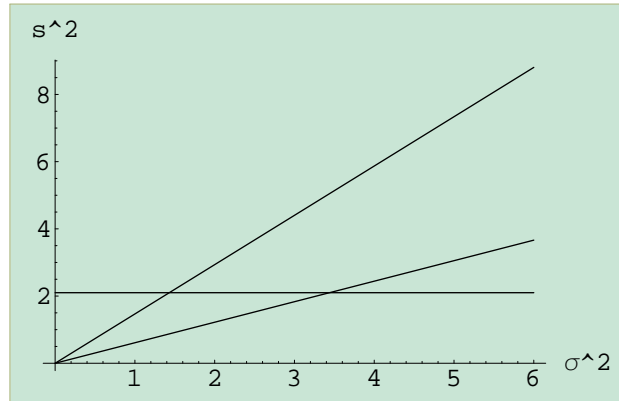

Figure 4: Graphical representation of the traditional measurement procedure for $V[X]$ in case of the observation $s^2 = 2.1$.

The measurement result contains all those values of σ^2, for which $r(\sigma^2)$ is within the prediction $^{tr}A_{R(\sigma^2)}^{(0.95)}(\{\sigma^2)\})$ given by the parallels in Figure 3.

The measurement result for the above given realization of the measurement experiment is as follows:

$$^{tr}C_{V[X]}^{(0.95)}\left(29\,\frac{2.1}{\sigma^2}\right) = \left\{\sigma^2\,|\,1.43 \leq \sigma^2 \leq 3.44\right\}$$

with length:

$$\left|^{tr}C_{V[X]}^{(0.95)}\left(29\,\frac{2.1}{\sigma^2}\right)\right| = 2.008$$

The measurement result using the optimal measurement procedure $^*C_{V[X]}^{(0.95)}$ is as follows:

$$^*C_{V[X]}^{(0.95)}\left(29\,\frac{2.1}{\sigma^2}\right) = \left\{\sigma^2\,|\,1.3824 \leq \sigma^2 \leq 3.31636\right\}$$

with length:

$$\left|^*C_{V[X]}^{(0.95)}\left(29\,\frac{2.1}{\sigma^2}\right)\right| = 1.933$$

Also in this example, the improvement is not at all negligible.

Concluding Remarks

The result of this learning unit shows the advantages of the stochastic approach, when compared with the statistical one. Just by applying the principles of the stochastic approach rigorously, a statistical method could be improved. This is in particular surprising, because the method, which is improved here, refers to one of the central results in statistics.

Besides the possibility to improve conventional procedures, the stochastic approach also leads to new procedures which have been not considered in statistics so far. This includes procedures which solve new problems and procedures which solve old problems, however, based on different more realistic conditions.

EXAMPLE

1. **Variance of a Wind Load**

Consider the example of the previous learning unit, which deals with a wind load X of Type M_x referring to a wind turbine. In order to specify a Bernoulli Space for X, not only the value of the first moment, but also the value of the variance $V[X]$ has to be determined based on a sample $(X_1|\{d\}, X_2|\{d\}, \dots, X_n|\{d\})$ and the sample function:

$$
\begin{aligned}
R(\sigma^2)|\{n\} &= (n-1)\frac{S^2|\{(n,d)\}}{\sigma^2} \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(X_i|\{d\} - \overline{X}|\{(n,d)\})^2
\end{aligned}
$$

If only an upper bound for the unknown value σ_0 of the variance $V[X]$ is required, the predictions procedure $A_{R(\sigma^2)}^{(\beta)}$ is of the following form:

$$
A_{R(\sigma^2)}^{(\beta)}(\{n\}) = \left\{ r(\sigma^2) \mid r(\sigma^2) \geq Q_{R(\sigma^2)|\{n\}}^{(\ell)}(\beta) \right\}
$$

The required reliability level is set to $\beta = 0.95$. Moreover, it is decided to have a sample size of $n = 250$. Next the predictions are determined by means of the χ^2-distribution with 249 degrees of freedom. In Figure 5 the predictions, which do not depend on σ^2, are displayed by the line parallel to the σ^2-axes through $(0, 213.465)$, which represents the lower bound of $A_{R(\sigma^2)}^{(0.95)}(\{\sigma^2\})$.
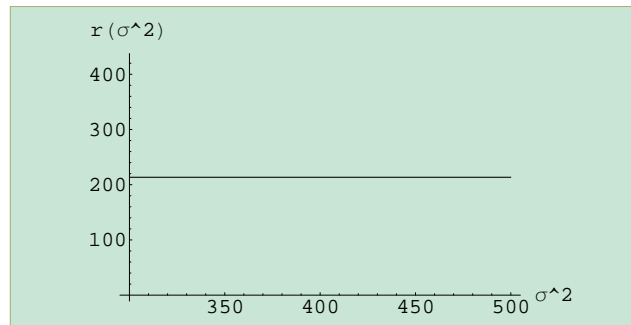

Figure 5: The lower bound $Q_{R(\sigma^2)|\{250\}}^{(\ell)}(0.95)$
of the predictions $A_{R(\sigma^2)}^{(0.95)}(\{\sigma^2\})$.

The measurement process is performed with sample size $n = 250$ and results in $s^2 = 342.98$. This result yields the observation $r(\sigma^2)$. In Figure 6 the predictions, which do not depend on σ^2, are displayed by the line parallel to the σ^2-axes through $(0, 213.465)$, which represents the lower bound of $A_{R(\sigma^2)}^{(0.95)}(\{\sigma^2\})$. Moreover, the realization $r(\sigma^2)$ of $R(\sigma^2)|\{249\}$ is represented by the decreasing curve from the left top to the right bottom.
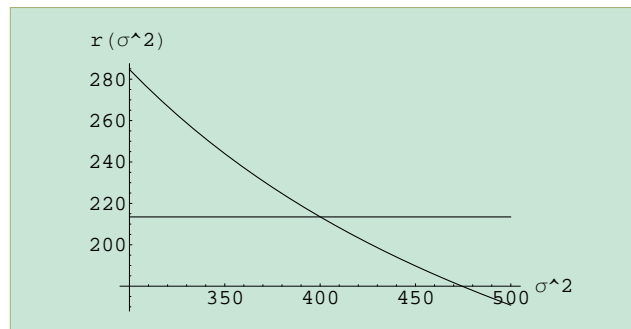
Figure 6: The predictions $A_{R(\sigma^2)}^{(0.95)}(\{\sigma^2\})$ and
the observation $r(\sigma^2)$.

The value of σ^2 for which $r(\sigma^2)$ crosses the lower bound $Q_{R(\sigma^2)|\{250\}}^{(\ell)}(0.95)$ is the upper bound of the measurement result:

$$C_{V[X]}^{(0.95)}(\{r(\sigma^2)\}) = \{\sigma^2 \,|\, \sigma^2 \le 400.075\}$$

Unit 2.3.11: Measurement of the Smallest Value

One of the most important problems concerning uncertainty refers to the extreme values which might occur in future. An extreme value is often related to a danger and only if the extreme value is known, an appropriate prevention strategy can be implemented. In this learning unit a method is introduced to measure the smallest value of a random variable X, i.e., the smallest value of its range of variability.

CONTENT

Introduction

The problem of determining the extreme values of the range of variability are not considered so far in statistical literature. One reason for this surprising fact might be that in statistics generally an unbounded range of variability is assumed. This assumption is necessary, because a majority of statistical methods has been developed based on probability distribution with unbounded support.

Besides the moments $E[X^n]$, $n = 1, 2, \ldots$, of a random variable X, the range of variability $\mathcal{X}$ is of special importance and, therefore, measurement procedures for the smallest value $\min \mathcal{X}$ and the largest value $\max \mathcal{X}$ of the range of variability of X should be derived and made available. Actually, in many cases only these extreme values are of interest.

In the previous learning units measurement procedures for $E[X]$ and $V[X]$ were introduced, which can be extended to other moments, if necessary. These measurement procedures are based on the normal approximation for sums of random variables and, therefore, can be applied without knowing the true probability distribution of X. Therefore, we assume in the following that not only the family of the unknown probability distribution is known, but also the values of the moments to be considered.

As to the range of variability $\mathcal{X}$ of the random variable X, it is assumed that it is known except for its smallest value, which shall be determined by a stochastic β-measurement procedure.

Measurement Process

Any process refers to a pair of variables (X, D), where X represents the future indeterminate outcome and D the determinate initial conditions. In a given situation the actual value d of the deterministic variable D determines the range of variability $\mathcal{X}(\{d\})$ and the probability distribution $\mathcal{P}(\{d\}) = P_{X|\{d\}}$ of the random variable $X|\{d\}$.

The measurement process for determining $\min \mathcal{X}(\{d\})$ is the same as in the case of determining the moments and is given by a random sample, i.e., by a number of independent copies of $X|\{d\}$:

$$(X_1|\{d\}, X_2|\{d\}, \ldots, X_n|\{d\})$$

where n is the actual value of the sample size, represented by the deterministic variable N.

Next, the observational random variable X must be defined on the measurement experiment $(X_1|\{d\}, X_2|\{d\}, \ldots, X_n|\{d\})$.

The Observational Variable

The sample moments or transformations of them emerged as natural observational variables for the moments, which according to the law of large numbers tend for $n \to \infty$ to the true values. Analogously, the sample minimum is the natural observational variable for the deterministic variable $\min \mathcal{X}(\{d\})$, since it tends to the true value for $n \to \infty$.

As stated in the introduction, it is assumed that the probability distribution of $X|\{d\}$ is known except for $\min \mathcal{X}(\{d\})$. Thus, we obtain the following pair of variables representing the measurement process for determining the value of $\min \mathcal{X}$:

- Random observational variable:

$$X_{min}|\{(n, d)\} = \min(X_1|\{d\}, X_2|\{d\}, \ldots, X_n|\{d\})$$

- Deterministic variable:

$$D = \min \mathcal{X}$$

Since values of all other distributional parameters (except for $\min \mathcal{X}(\{d\})$) are assumed to be known, there is no need to include them as further components in the deterministic variable D.

The Bernoulli Space $\mathbb{B}_{X_{min}, \min \mathcal{X}}$

It is always possible to specify a lower bound a_{low} for $\min \mathcal{X}$, while the known value $\mu(d)$ of the first moment $E[X|\{d\}]$ represents a global upper bound for $\min \mathcal{X}$. However, this global upper bound can often be improved considerably by utilizing the close relations between the bounded range of variability, the moments and the type of the probability distribution of X. Let a_{max} be the largest value of $\min \mathcal{X}$ preserving the given type of the probability distribution of X. Then an upper bound for $\min \mathcal{X}$ is given by $\min\{\lfloor \mu(d) \rfloor, a_{max}\}$, where $\lfloor \mu(d) \rfloor$ is the largest value in $\mathcal{X}(\{d\})$ smaller than $\mu(d)$. Hence, the following Bernoulli Space $\mathbb{B}_{X_{min}, \min \mathcal{X}}$ is obtained:

- Ignorance space:

$$\mathcal{D} = \{d \,|\, a_{low} \le d \le \min\{\lfloor \mu(d) \rfloor, a_{max}\}$$

- Variability function:

$$\mathcal{X}(\{d\}) = \{x \,|\, d \le x \le b\} \quad \text{for } d \in \mathcal{D}$$

- Random structure function:

$$\mathcal{P}(\{d\}) = P_{X_{min}|\{d\}}$$

  with

$$P_{X_{min}|\{d\}}\left(\{x \,|\, x \ge y\}\right) = \left(P_{X|\{d\}}\left(\{x \,|\, x \ge y\}\right)\right)^n \quad \text{for } y \in \mathcal{X}(\{d\})$$

  where the probability distribution $P_{X|\{d\}}$ is known.

Measurement Procedure

The problem is to derive a β-measurement procedure $C_{\min \mathcal{X}(\{d\})}^{(\beta)}$ for the unknown value of $\min \mathcal{X}(\{d\})$, where the measurement procedure is defined by an appropriate prediction procedure $A_{X_{min}}^{(\beta)}$.

Lemma 1: The measurement procedure $C_{\min \mathcal{X}}^{(\beta)}$ is defined by its measurement results as follows:

$$C_{\min \mathcal{X}}^{(\beta)}(\{x\}) = \left\{ d \,|\, x \in A_{X_{min}}^{(\beta)}(\{d\}) \right\} \tag{117}$$

where

$$A_{X_{min}}^{(\beta)}(\{d\}) = \begin{cases} \left\{ x \,|\, a_{low} \le d \le x \le Q_{X_{min}|\{d\}}^{(u)}(\beta) \right\} & \text{for } d < \overline{d} \\[2mm] \left\{ x \,|\, \overline{d} \le x \le b \right\} & \text{for } d = \overline{d} \end{cases} \tag{118}$$

with $\overline{d} = \min\{\lfloor \mu(d) \rfloor, a_{max}\}$ and $Q_{X_{min}|\{d\}}^{(u)}$ is the upper quantile function of $X_{min}|\{d\}$.

Proof: We have to show that $C_{\min \mathcal{X}}^{(\beta)}$ fulfills the reliability requirement β, that it is complete and that it is optimal.

- Reliability:
  The measurement procedure is a β-measurement procedure, if the defining predictions meet the reliability requirement given by β. This is trivially fulfilled, because by definition of the upper quantile function we have for $a_{low} \le d < \overline{d}$:

$$\begin{aligned} P_{X_{min}|\{d\}}\left(A_{X_{min}}^{(\beta)}(\{d\}) \right) &= P_{X_{min}|\{d\}}\left(\left\{ x \,|\, x \le Q_{X_{min}|\{d\}}^{(u)}(\beta) \right\} \right) \\ &\ge \beta \end{aligned}$$

  The prediction for $d = \overline{d}$ occurs with certainty implying that the reliability requirement is met in this case, too.

- Completeness: The completeness requirement means that for any observation $x \in \mathcal{X}(\mathcal{D})$, i.e., $x \in \{x \,|\, a_{low} \le x \le b\}$ there is at least one prediction $A_{\min X}^{(\beta)}(\{d\})$ with $x \in A_{\min X}^{(\beta)}(\{d\})$.

  We have to show that the union of all predictions cover the entire range of variability $\mathcal{X}(\mathcal{D})$. For any realistic reliability level β, i.e., a not too small value, we have $Q_{X_{min}|\{d\}}^{(u)}(\beta) \ge \mu \ge \min\{\lfloor \mu(d) \rfloor, a_{max}\} = \overline{d}$. Thus, from (118) we obtain:

$$\bigcup_{d \in \mathcal{D}} A_{\min X}^{(\beta)}(\{d\}) = \{x \,|\, a_{low} \le x \le b\} = \mathcal{X}(\mathcal{D})$$

- Optimality: The measurement procedure (121) is optimal, because it is a lower bound measurement procedure and any optimal lower bound measurement procedure is defined by predictions given by the upper quantile function.   ●

With (118) the following measurement procedure for $\min \mathcal{X}$ is obtained:

$$C_{\min \mathcal{X}}^{(\beta)}(\{x\}) = \begin{cases} \left\{ d \,|\, d(x) \le d \le \min(\overline{d}, x) \right\} & \text{for } \quad d(x) < \min(\overline{d}, x) \\[2mm] \{\overline{d}\} & \text{for } \quad d(x) \ge \min(\overline{d}, x) \end{cases}$$

where $d(x)$ is obtained as the solution of the following equation:

$$Q_{X_{min}|\{d(x)\}}^{(u)}(\beta) = x$$

Concluding Remarks

The here derived procedure for determining a bound of the range of variability $\mathcal{X}(\{d\})$ of a given random variable $X|\{d\}$ represents a new class of measurement procedures with a great potential for application.

The bounds of the range of variability are the extreme values, which can be adopted by the random variable. The extreme values play an important role in many areas of application. Thus, the here outlined approach could replace the so-called extreme value theory, which generally starts with the unrealistic assumption of an unbounded range of variability of the considered random variable.

Similar as in the case of the measurement procedure for the variance of a random variable, the result concerning the smallest value of a random variable illustrates the benefits of the unified stochastic approach, when compared with the statistical approach.

EXAMPLE

1. **Constant Probability Distribution**

   The constant (or uniform) probability distributions represent the simplest case with respect to the random structure. Let X be a random variable with constant probability distribution over its range of variability $\mathcal{X}$. The maximum value of $\mathcal{X}$ is known with $\max \mathcal{X} = 100$. The minimum value $\min \mathcal{X} = a$ is unknown except for the fact that $0 \leq a \leq 50$ holds. The problem is to determine the smallest value that X may adopt, i.e., $\min \mathcal{X}$, by means of a stochastic measurement procedure.

   The reliability specification is set to $\beta = 0.99$ and the measurement process (= random sample) $(X_1|\{a\}, X_2|\{a\}, \ldots, X_n|\{a\})$ has a sample size of $n = 50$.

   The observational random variable and the corresponding deterministic variable are as follows:

   $$\begin{aligned} X_{min} &= \min(X_1, X_2, \ldots, X_{50}) \\ D &= \min \mathcal{X} \end{aligned}$$

   with the following Bernoulli Space $\mathbb{B}_{X_{min}, \min \mathcal{X}}$:

   $$\begin{aligned} \mathcal{D} &= \{a \mid 0 \leq a \leq 50\} \\ \mathcal{X}(\{a\}) &= \{x \mid a \leq x \leq 100\} \\ \mathcal{P}(\{a\}) &= \{P_{X_{min}|\{a\}}\} \end{aligned}$$

   Determination of $P_{X_{min}|\{a\}}$

   The survival function of $X|\{a\}$ is given as follows:

   $$\overline{F}_{X|\{a\}}(x) = \frac{100 - x}{100 - a} \quad \text{for } x \in \mathcal{X}(\{a\})$$

   implying that the survival function of $X_{min}|\{a\}$, which determines the images $P_{X_{min}|\{a\}}$ of the random structure function, is given by:

$$\overline{F}_{X_{min}|\{a\}}(x) = \left(\frac{100-x}{100-a}\right)^{50} \quad \text{for } x \in \mathcal{X}(\{a\})$$

which gives the following distribution function:

$$F_{X_{min}|\{a\}}(x) = 1 - \left(\frac{100-x}{100-a}\right)^{50} \quad \text{for } x \in \mathcal{X}(\{a\})$$

with the upper quantile function:

$$Q^{(u)}_{X_{min}|\{a\}}(p) = 100 - (100-a)(1-p)^{\frac{1}{50}} \quad \text{for } 0 < p \le 1$$

The Prediction Procedure $A^{(0.99)}_{X_{min}}$

The upper quantiles define the prediction procedure $A^{(0.99)}_{X_{min}}$, which are used to derive the measurement procedure $C^{(0.99)}_{\min \mathcal{X}}$:

$$A^{(0.99)}_{X_{min}}(\{a\}) = \left\{ x \mid x \le 100 - (100-a)(1-0.99)^{\frac{1}{50}} \right\}$$

The Measurement Procedure $C^{(0.99)}_{\min \mathcal{X}}$

The prediction procedure defines the measurement procedure. For $x \le 50$ the measurement results are as follows:

$$
\begin{aligned}
C^{(0.99)}_{\min \mathcal{X}}(\{x\}) &= \left\{ a \mid x \in A^{(0.99)}_{X_{min}}(\{a\}) \right\} \\[2mm]
&= \left\{ a \mid 100 - \frac{100-x}{(1-0.99)^{\frac{1}{50}}} \le a \le x \right\} \\[2mm]
&= \left\{ a \mid -9.64782 + 0.912011x \le a \le x \right\}
\end{aligned}
$$

Figure 1 shows a graphical representation of the measurement procedure $C^{(0.99)}_{\min \mathcal{X}}$, which is given by the two straight lines. For given observation x, the measurement interval is obtained by the points of intersection of the horizontal line through $(0, x)$ and the two straight lines in Figure 1.
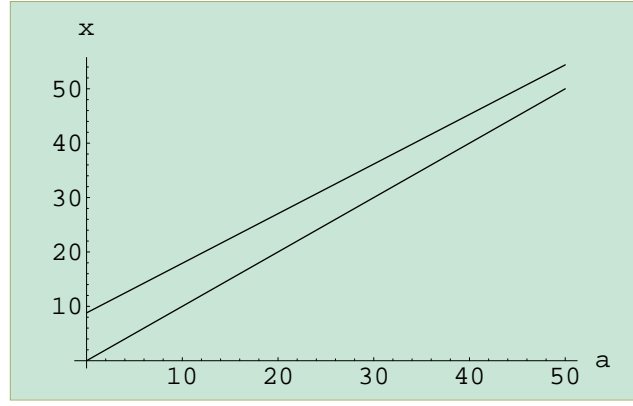
Figure 1: The measurement procedure
$C_{\min \mathcal{X}}^{(0.99)}$ for the smallest value.

For example, assume that the measurement experiment results in the observation $x = 23$, then the following measurement result with respect to $\min \mathcal{X}$ is obtained:

$$C_{\min \mathcal{X}}^{(0.99)}(\{23\}) = \{a \,|\, 15.57 \leq a \leq 23\} \tag{119}$$

By means of (119) it is shown that the smallest value that can be adopted by X is not less than 15.57.

2. Monotonic Probability Distribution

Let X be a random variable with a monotonic increasing probability distribution. Moreover, it is assumed that the range of variability together with the values of the first moment and of the variance determine completely the probability distribution. Let the actual values of the first moment and the variance be given as:

$$\begin{aligned} \mu &= 700 \\ \sigma^2 &= 250^2 \end{aligned}$$

Moreover, it is known that X can adopt only non-negative integers smaller than or equal to $b = 1000$.

The only unknown value characterizing the initial conditions refers to the deterministic variable $D = \min \mathcal{X}$, the actual value of which shall be measured by a β-measurement procedure.

In a first step the ignorance space $\mathcal{D}$ of the deterministic variable D is determined as sharp as possible. A lower bound of D is given by 0, an upper bound by $\mu = 700$. However, the type of distribution function with given values μ and $\sigma^2 = 250^2$ depends on the range of variability. It can be shown that for $d \geq 95$ the type of $P_{X|\{d\}}$ changes from a monotonic increasing probability distribution to a U-shaped probability distribution and, finally, if d is further increased to a monotonic decreasing probability distribution.

Thus, the initial ignorance space with respect to $D = \min \mathcal{X}$, which shall be reduced by the β-measurement procedure $C_{\min \mathcal{X}}^{(\beta)}$, is given by:

$$\mathcal{D} = \{d \,|\, 0 \leq d \leq 94\}$$

Consider a measurement process with sample size $n = 250$ and a reliability level of $\beta = 0.95$. Then according to (118) the prediction procedure that defines the measurement procedure $C^{(0.95)}_{\min \mathcal{X}}$ is given as follows:

$$
A^{(0.95)}_{X_{min}}(\{d\}) =
\begin{cases}
\left\{ x \,|\, d \leq x \leq Q^{(u)}_{X_{min}|\{d\}}(0.95) \right\} & \text{for } 0 \leq d \leq 93 \\[2ex]
\{ x \,|\, 94 \leq x \leq 1000 \} & \text{for } d = 94
\end{cases}
\tag{120}
$$

Once the predictions are available, the measurement procedure is obtained by the measurement results:

$$
C^{(\beta)}_{\min \mathcal{X}}(\{x\}) =
\begin{cases}
\{ d \,|\, d(x) \leq d \leq \min(94, x) \} & \text{for} \quad d(x) < 94 \\[2ex]
\{94\} & \text{for} \quad d(x) \geq 94
\end{cases}
$$

where $d(x)$ is defined as solution of the following equation:

$$
Q^{(u)}_{X_{min}|\{d(x)\}}(0.95) = x
$$

In order to derive the prediction procedure $A^{(0.95)}_{X_{min}}$, the probability distributions $P_{X|\{d\}}$ and $P_{X_{min}|\{d\}}$ have to be determined. Subsequently, the corresponding value $Q^{(u)}_{X_{min}|\{d\}}(0.95)$ of the upper quantile function of $X_{min}|\{d\}$ can be calculated.

For each value $d \in \mathcal{D}$ and given values of the first moment and the variance of $X|\{d\}$, the probability distribution can be determined by solving the system of equations:

$$
\begin{aligned}
E[X^0|\{d\}] &= 1 \\
E[X^1|\{d\}] &= \mu \\
E\left[(X^2|\{d\} - \mu)^2\right] &= \sigma^2
\end{aligned}
$$

For example, for $d = 0$ we obtain:

$$
P_{X|\{0\}}(\{x\}) = \frac{0.014119211280766233}{250\sqrt{2\pi}} e^{0.0299247376665377777 \frac{(x+2195.6998165057234)^2}{250^2}}
$$

In Figure 2 the monotonic increasing probability mass function of $X|\{0\}$ is displayed.
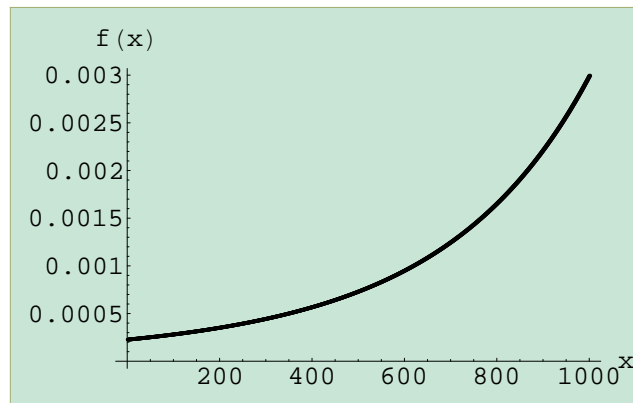

Figure 2: Monotonic increasing probability mass function $f_{X|\{0\}}$ of the random variable $X|\{0\}$

In contrast to the monotonic increasing probability distribution of the given random variable $X|\{0\}$, the random observational variable $X_{min}|\{(250,0)\}$ has for a sufficiently large sample size n, a monotonic decreasing probability distribution, i.e., the smallest value $\min \mathcal{X}$ occurs with largest probability.

In Figure 3, the probability distribution of $X_{min}|\{(250,0)\}$ is displayed.
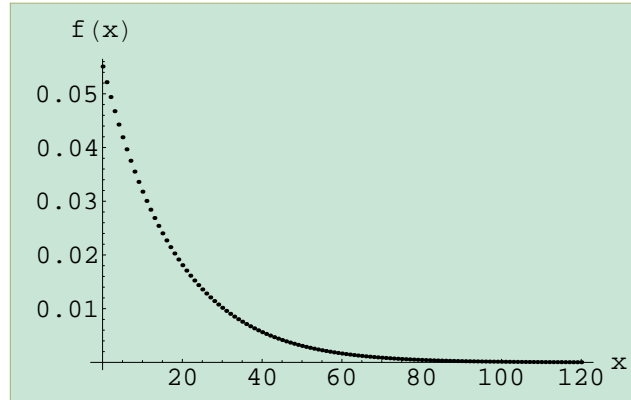

Figure 3: Monotonic decreasing probability mass function $f_{X_{min}|\{(250,0)\}}$ of the random variable $X_{min}|\{(250,0)\}$

The predictions are given by the upper quantiles of order $\beta = 0.95$. In other words, for each value of $d \in \mathcal{D}$ and the given values of n, μ and σ^2, the probability mass function $f_{X_{min}|\{(n,d)\}}$ has to be determined. Subsequently the upper quantiles of order $\beta = 0.95$ are calculated, by means of which the measurement results are obatined.

In the case of $d = 0$ the following value is obtained:

$$Q^{(u)}_{X_{min}|\{(250,0)\}}(0.95) = 49$$

Thus, if the observed value x does not exceed the limit 49 then the value $d = 0$ cannot be excluded from the set of possible values. However, if the observation exceeds 49, it is concluded that the lower bound of $\mathcal{X}$ is larger than 0.

The measurement result $C^{(0.95)}_{\min \mathcal{X}}(\{x\})$ contains all those values of $d \in \mathcal{D}$, for which the observation x is an element of the prediction $A^{(0.95)}_{X_{min}}(\{d\})$.

The entire measurement procedure $C^{(0.95)}_{\min \mathcal{X}}$ is given in Table 1. For example, if the measurement experiment yields the result $x = 56$, then the following measurement result is obtained:

$$C^{(0.95)}_{\min \mathcal{X}}(\{56\}) = \{d|10 \leq d \leq 56\}$$

The obtained measurement result represents the new ignorance space for the deterministic variable $\min \mathcal{X}$.

Note extreme values are not subject to randomness but are fixed by the given initial conditions. Therefore, extreme values must be represented by deterministic variables and not by random variables as done in the extreme value theory. Modeling extreme values by random variables means not to regard the actual "extreme value", but something different which is not really explained.

Table 1: Complete measurement procedure $C_{\min \mathcal{X}}^{(0.95)}$ for $\min \mathcal{X}$

x	$C_{\min \mathcal{X}}^{(0.95)}(\{x\})$	x	$C_{\min \mathcal{X}}^{(0.95)}(\{x\})$	x	$C_{\min \mathcal{X}}^{(0.95)}(\{x\})$
$x \leq 49$	$\{d \mid 0 \leq d \leq x\}$	73	$\{d \mid 35 \leq d \leq 73\}74$	97	$\{d \mid 67 \leq d \leq 94\}$
50	$\{d \mid 1 \leq d \leq 50\}$	74	$\{d \mid 36 \leq d \leq 74\}$	98	$\{d \mid 68 \leq d \leq 94\}$
51	$\{d \mid 2 \leq d \leq 51\}$	75	$\{d \mid 37 \leq d \leq 75\}$	99	$\{d \mid 70 \leq d \leq 94\}$
52	$\{d \mid 4 \leq d \leq 52\}$	76	$\{d \mid 39 \leq d \leq 76\}$	100	$\{d \mid 71 \leq d \leq 94\}$
53	$\{d \mid 5 \leq d \leq 53\}$	77	$\{d \mid 40 \leq d \leq 77\}7$	101	$\{d \mid 72 \leq d \leq 94\}$
54	$\{d \mid 7 \leq d \leq 54\}$	8	$\{d \mid 42 \leq d \leq 78\}$	102	$\{d \mid 74 \leq d \leq 94\}$
55	$\{d \mid 8 \leq d \leq 55\}$	79	$\{d \mid 43 \leq d \leq 79\}$	103	$\{d \mid 75 \leq d \leq 94\}$
56	$\{d \mid 10 \leq d \leq 56\}$	80	$\{d \mid 44 \leq d \leq 80\}$	104	$\{d \mid 76 \leq d \leq 94\}$
57	$\{d \mid 11 \leq d \leq 57\}$	81	$\{d \mid 46 \leq d \leq 81\}$	105	$\{d \mid 77 \leq d \leq 94\}$
58	$\{d \mid 13 \leq d \leq 58\}$	82	$\{d \mid 47 \leq d \leq 82\}$	106	$\{d \mid 79 \leq d \leq 94\}$
59	$\{d \mid 14 \leq d \leq 59\}$	83	$\{d \mid 48 \leq d \leq 83\}$	107	$\{d \mid 80 \leq d \leq 94\}$
60	$\{d \mid 16 \leq d \leq 60\}$	84	$\{d \mid 50 \leq d \leq 84\}$	108	$\{d \mid 81 \leq d \leq 94\}$
61	$\{d \mid 17 \leq d \leq 61\}$	85	$\{d \mid 51 \leq d \leq 85\}$	109	$\{d \mid 83 \leq d \leq 94\}$
62	$\{d \mid 19 \leq d \leq 62\}$	86	$\{d \mid 52 \leq d \leq 86\}$	110	$\{d \mid 84 \leq d \leq 94\}$
63	$\{d \mid 20 \leq d \leq 63\}$	87	$\{d \mid 54 \leq d \leq 87\}$	111	$\{d \mid 85 \leq d \leq 94\}$
64	$\{d \mid 22 \leq d \leq 64\}$	88	$\{d \mid 55 \leq d \leq 88\}$	112	$\{d \mid 86 \leq d \leq 94\}$
65	$\{d \mid 23 \leq d \leq 65\}$	89	$\{d \mid 56 \leq d \leq 89\}$	113	$\{d \mid 88 \leq d \leq 94\}$
66	$\{d \mid 25 \leq d \leq 66\}$	90	$\{d \mid 58 \leq d \leq 90\}$	114	$\{d \mid 89 \leq d \leq 94\}$
67	$\{d \mid 26 \leq d \leq 67\}$	91	$\{d \mid 59 \leq d \leq 91\}$	115	$\{d \mid 90 \leq d \leq 94\}$
68	$\{d \mid 27 \leq d \leq 68\}$	92	$\{d \mid 60 \leq d \leq 92\}$	116	$\{d \mid 92 \leq d \leq 94\}$
69	$\{d \mid 28 \leq d \leq 69\}$	93	$\{d \mid 62 \leq d \leq 93\}$	117	$\{d \mid 93 \leq d \leq 94\}$
70	$\{d \mid 30 \leq d \leq 70\}$	94	$\{d \mid 63 \leq d \leq 94\}$	$x \geq 118$	$\{94\}$
71	$\{d \mid 32 \leq d \leq 71\}$	95	$\{d \mid 64 \leq d \leq 94\}$		
72	$\{d \mid 33 \leq d \leq 72\}$	96	$\{d \mid 66 \leq d \leq 94\}$		

Unit 2.3.12: Measurement of the Largest Value

Target

This learning unit extends the result obtained in the previous learning unit to the case that the largest value $\max \mathcal{X}$ of the range of variability $\mathcal{X}$ of a random variable X shall be determined. The measurement procedure $C^{(\beta)}_{\max \mathcal{X}}$ is derived analogously to $C^{(\beta)}_{\min \mathcal{X}}$. Therefore, this learning unit can be taken as a repetition and exercise of the previous one.

Content

Introduction

In many situations the maximum value, which a random variable X may adopt, is of utmost importance. For example in order to decide about the height of a river embankment, knowledge about the largest value of future floods would be very valuable. Especially in the area of safety, there are many situations in which the maximum value is of great significance.

Analogously as in the case of a measurement procedure for $\min \mathcal{X}$, it is assumed that the probability distribution of X is known except for the upper bound of the range of variability.

Measurement Process

The measurement process is the same as in the case of $C^{(\beta)}_{\min \mathcal{X}}$ and is given by a random sample of size n, i. e., by n independent copies of $X|\{d\}$:

$$(X_1|\{d\}, X_2|\{d\}, \ldots, X_n|\{d\})$$

For the given number n of copies, a random observational variable must be defined having a probability distribution, which is essentially related to the unknown value of $\max \mathcal{X}$.

The Observational Variable

Clearly, the sample maximum constitutes a natural observational variable for determining the actual value of the deterministic variable $\max \mathcal{X}$. Thus, we obtain the following pair of variables representing the measurement experiment for the value of $\max \mathcal{X}$:

- Random observational variable:

$$X_{max} = \max(X_1, X_2, \ldots, X_n$$

- Deterministic variable:

$$D = \max \mathcal{X}(\{d\})$$

The values of all other distributional parameters (except for $\max \mathcal{X}(\{d\})$) are assumed to be known. Hence, there is no need to include them as further components in the deterministic variable.

The Bernoulli Space $\mathbb{B}_{X_{max},\max \mathcal{X}}$

It is always possible to specify an upper bound b_{up} for $\max \mathcal{X}(\{d\})$, while the value $\mu(d)$ of the first moment $E[X|\{d\}]$ represents a global lower bound for $\max \mathcal{X}$. However, the global lower bound can often be improved considerably by utilizing the close relations between the range of variability, the moments and the type of the probability distribution of X. Let b_{max} be the smallest value of $\max \mathcal{X}$ preserving the given type of the probability distribution of X. Then a lower bound for $\max \mathcal{X}$ is given by $\max\{\lceil \mu(d) \rceil, b_{min}\}$, where $\lceil \mu \rceil$ denotes the smallest element of $\mathcal{X}$ larger than or equal $\mu(d)$. Hence, the following Bernoulli Space $\mathbb{B}_{X_{max},\max \mathcal{X}}$ is obtained:

- Ignorance space:
$$\mathcal{D} = \left\{ d \,\middle|\, \max\{\lceil \mu \rceil, b_{min}\} \le d \le b_{up}\} \right\}$$

- Variability function:
$$\mathcal{X}\Big(\{d\}\Big) = \{x \,|\, a \le x \le d\} \quad \text{for } d \in \mathcal{D}$$

- Random structure function:
$$\mathcal{P}\Big(\{d\}\Big) = P_{X_{max}|\{d\}}$$

  with

$$P_{X_{max}|\{d\}}\left(\{x \,|\, x \le y\}\right) = \Big(P_{X|\{d\}}\left(\{x \,|\, x \le y\}\right)\Big)^n$$

  where the probability distribution $P_{X|\{d\}}$ is known.

Measurement Procedure

The problem is to derive a β-measurement procedure $C^{(\beta)}_{\max \mathcal{X}}$ for the unknown value of $\max \mathcal{X}(\{d\})$, where the measurement procedure is defined by an appropriate prediction procedure $A^{(\beta)}_{X_{max}}$.

Lemma 1: The measurement procedure $C^{(\beta)}_{\max \mathcal{X}}$ is defined by its measurement results as follows:

$$C^{(\beta)}_{\max \mathcal{X}}(\{x\}) = \left\{ d \,\middle|\, x \in A^{(\beta)}_{X_{max}}(\{d\}) \right\} \tag{121}$$

where for $d \in \mathcal{D}$:

$$A^{(\beta)}_{X_{max}}(\{d\}) = \begin{cases} \left\{ x \,\middle|\, Q^{(\ell)}_{X_{max}|\{d\}}(\beta) \le x \le d \right\} & \text{for} \quad \underline{d} < d \le b_{up} \\[2mm] \{x \,|\, a \le x \le \underline{d}\} & \text{for} \quad d = \underline{d} \end{cases} \tag{122}$$

with $\underline{d} = \max\{\lceil \mu \rceil, b_{min}\}$ and $Q^{(\ell)}_{X_{max}|\{d\}}$ is the lower quantile function of $X_{max}|\{d\}$.

Proof: We have to show that $C^{(\beta)}_{\max \mathcal{X}}$ fulfills the reliability requirement β, that it is complete and that it is optimal.

- Reliability:
  The measurement procedure is a β-measurement procedure, if the defining predictions meet the reliability requirement given by β. This is trivially fulfilled, because by definition of the lower quantile function we have for $\max\{\lceil \mu \rceil, b_{min}\} < d \le b_{up}$:

$$P_{X_{max}|\{d\}}\left(A_{X_{max}}^{(\beta)}(\{d\})\right) \;=\; P_{X_{max}|\{d\}}\left(\left\{x \,|\, Q_{X_{max}|\{d\}}^{(\ell)}(\beta) \leq x\right\}\right)$$
$$\geq\; \beta$$

The prediction for $d = \max\{\lceil\mu\rceil, b_{min}\}$ occurs with certainty and, thus, the reliability requirement is fulfilled.

- Completeness: The completeness requirement is proved, if we show that for any observation $x \in \mathcal{X}(\mathcal{D}) = \{x \,|\, a \leq x \leq b_{up}\}$ there is at least one prediction $A_{\max X}^{(\beta)}(\{d\})$ with $x \in A_{\max X}^{(\beta)}(\{d\})$. This is achieved, if we show that the union of all predictions cover the entire range of variability $\mathcal{X}(\mathcal{D})$. For any realistic reliability level β, i.e., a not too small value, we have:
$$Q_{X_{max}|\{d\}}^{(l)}(\beta) \leq \mu \leq \max\{\lceil\mu(d)\rceil, b_{min}\} = \underline{d}$$

and thus, from (122) we obtain:

$$\bigcup_{d\in\mathcal{D}} A_{\max X}^{(\beta)}(\{d\}) = \{x \,|\, a \leq x \leq b_{up}\} = \mathcal{X}(\mathcal{D})$$

- Optimality: The measurement procedure (121) is optimal, because it is a upper bound measurement procedure and any optimal upper bound measurement procedure is defined by predictions given by the lower quantile function. •

With (122) the following measurement procedure for $\max \mathcal{X}$ is obtained:

$$C_{\max\mathcal{X}}^{(\beta)}(\{x\}) = \begin{cases} \{d \,|\, \max(\underline{d}, x) \leq d \leq d(x)\} & \text{for} \quad d(x) > \max(\underline{d}, x) \\ \{\underline{d}\} & \text{for} \quad d(x) \leq \max(\underline{d}, x) \end{cases}$$

where $d(x)$ is obtained as the solution of the following equation:

$$Q_{X_{max}|\{d(x)\}\}}^{(\ell)}(\beta) = x$$

Remark on Extreme Value Theory

In order to cope with the problem to determine the extreme values (smallest and largest values) a random variable may adopt, the so-called *extreme value theory* has been developed. However, in contrast to the here proposed method to measure an extreme value by an upper or lower bound directly by means of a stochastic measurement procedure, extreme value theory looks at the limiting distributions of extreme values within a sample. The main result in extreme value theory is due to the Russian mathematician Boris V. Gnedenko[11], who in 1943 showed that the limiting distribution, if it exists, has one of three possible forms. The here proposed method gets by without asymptotics and avoids the difficulty to judge how much the limit departs from reality. Therefore, stochastic measurement procedures for determining the smallest or largest values of a range of variability of a random variable should at least in some instances replace the methods based on extreme value theory. Particularly, for safety-related situations, where a violation of a limit might lead to a catastrophe, one should not rely on asymptotic developments, but on a model that is derived using only the available knowledge about the situation and which guarantees the required reliability level. Moreover, the here proposed procedures are adaptable to new experience and the results may thus improved with each new observation.

---

[11]Gnedenko, B.V. (1943): Les extreêmes des distributions statistiques. *Ann. Inst. H. Poincaré* 5, 115–158.

1. **Constant Probability Distribution**

Let X be a random variable with a constant probability distribution and known minimum $a = 0$ of the range variability and unknown value $\max \mathcal{X}$ of the maximum. As to the maximum, a global upper bound is known and given by $b_{up} = 150$ and a global lower bound is given by $b_{low} = 30$, i.e., it is known that $30 \leq \max \mathcal{X} \leq 150$. The measurement procedure shall meet the reliability specification given by $\beta = 0.999$. Finally, the measurement experiment shall consist of $n = 100$ independent copies of X.

The initial ignorance space with respect to $D = \max \mathcal{X}$ is given by:

$$\mathcal{D} = \{d \mid 30 \leq d \leq 150\}$$

it shall be reduced by the measurement procedure $C_{\max \mathcal{X}}^{(\beta)}$.

The β-measurement procedure $C_{\max \mathcal{X}}^{(0.999)}$ is defined by a prediction procedure given by the following predictions:

$$A_{X_{max}}^{(0.999)}(\{d\}) = \begin{cases} \left\{ x \mid Q_{X_{max}|\{d\}}^{(\ell)}(0.999) \leq x \leq d \right\} & \text{for} \quad 30 < d \leq 150 \\ \{x \mid 0 \leq x \leq 30\} & \text{for} \quad d = 30 \end{cases}$$

In order to derive the prediction procedure $A_{X_{max}}^{(0.999)}$, the lower quantile function of $X_{max}|\{d\}$ has to be known, which is no problem here, since the distribution function of $X|\{d\}$ is given by:

$$F_{X|\{d\}}(x) = \frac{x}{d} \quad \text{for } 0 \leq x \leq d$$

With $F_{X|\{d\}}$, the probability distribution of $P_{X_{max}|\{d\}}$ is immediately obtained. Subsequently, the corresponding value $Q_{X_{max}|\{d\}}^{(\ell)}(0.999)$ of the lower quantile function of $X_{max}|\{d\}$ is calculated by means of the survival function of $X_{max}|\{d\}$ given by:

$$\overline{F}_{X_{max}|\{d\}}(x) = 1 - \left(\frac{x}{d}\right)^{100} \tag{123}$$

The survival function yields the following lower quantile function:

$$Q_{X_{max}|\{d\}}^{(\ell)}(p) = d(1-p)^{\frac{1}{n}}$$

Hence, for $30 < d \leq 150$ we obtain:

$$\begin{aligned} A_{X_{max}}^{(0.999)}(\{d\}) &= \left\{ x \mid Q_{X_{max}|\{d\}}^{(\ell)}(0.999) \leq x \leq d \right\} \\ &= \left\{ x \mid 0.001^{\frac{1}{100}} d \leq x \leq d \right\} \end{aligned}$$

Note that the predictions for obtaining an upper bound measurement procedure for $\max \mathcal{X}$ are lower bound predictions, while the predictions for determining a lower bound measurement procedure for $\min \mathcal{X}$ were upper bound predictions.
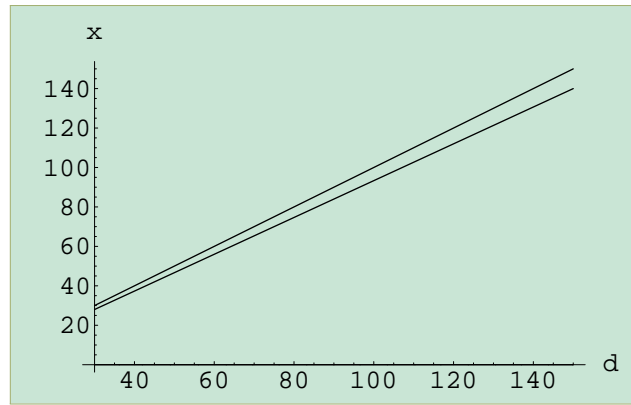
Figure 1: Predictions $A_{X_{max}}^{(0.999)}(\{d\})$ for
$30 < d \leq 150$

From the predictions, the measurement procedure is immediately obtained:

$$C_{\max \mathcal{X}}^{(0.999)}(\{x\}) = \left\{ d \,|\, x \leq d \leq \frac{x}{0.001^{\frac{1}{100}}} \right\}$$

For example the observation $x = 67$ yields the following measurement interval for the upper bound $\max \mathcal{X}$ of the range of variability of the considered random variable X:

$$C_{\max \mathcal{X}}^{(0.999)}(\{67\}) = \{d \,|\, 67.0 \leq b \leq 71.8\}$$

2. Binomial Distribution

This example shall not illustrate the procedure introduced in Learning Unit 2.3.12, but show that the problem of measuring $\max \mathcal{X}$ may occur in situations, which can hardly be handled by extreme value theory. However, even then it is possible to solve the problems by applying the same principles given in this learning unit.

Consider n identical units, which operate during one time unit independently of each other. Each of them fails during the time with a certain probability. Assume that the number n of units is unknown, except for the fact that it is larger than 400 and smaller than 500.

The number of units shall be determined by means of a stochastic measurement procedure. The measurement experiment is to have the units operate one time unit and observe the number of failed units X, which is binomial distributed with parameters n and p. Assume that it is known that the value of the failure probability of a unit is at least 0.05 and at most 0.06.

The Pair of Variables (X, D)

The observational random variable X is defined as follows:

$$X = \text{number of failed units}$$

The deterministic variable D determines the distribution of X, which is binomial. Thus, the following two-dimensional deterministic variable is obtained:

$$D = (D_1, D_2)$$

with

$$
\begin{aligned}
D_1 &= \text{ number of operating units} \\
D_2 &= \text{ probability of failure}
\end{aligned}
$$

Bernoulli Space $\mathbb{B}_{X,D}$

The three components of the Bernoulli Space are as follows:

$$
\begin{aligned}
\mathcal{D} &= \{(n,p) \,|\, 400 \le n \le 500 \,, 0.05 \le p \le 0.06\} \\
\mathcal{X}(\{(n,p)\}) &= \{0, 1, \ldots, n\} \\
\mathcal{P}(\{(n,p)\}) &= P_{X|\{(n,p)\}} \quad \text{with } X|\{(n,p)\} \sim Bi(n,p)
\end{aligned}
$$

Measurement Procedure

The measurement procedure is based on a prediction procedure $A_X^{(\beta)}$, which assigns to each element of an appropriate system $\mathcal{T}_D(\mathcal{D})$ of subsets of $\mathcal{D}$ an event, which will occur with a probability not falling short of β.

The system of subsets $\mathcal{T}_D(\mathcal{D})$ is given by the following elements:

$$\mathcal{D}_n = \{(n,p) \,|\, 0.05 \le p \le 0.6\} \quad \text{for } n = 400, 401, \ldots, 500$$

Because the upper bound of $\mathcal{X}$ shall be determined, the predictions have the following form:

$$A_X^{(\beta)}(\mathcal{D}_n) = \left\{ x \,|\, q_n^{(\beta)} \le x \right\}$$

where $q_n^{(\beta)}$ is determined so that the corresponding prediction meets the reliability condition:

$$P_{X|\{d\}} \left(A_X^{(\beta)}(\mathcal{D}_n) \right) \ge \beta \quad \text{for } d(n,p) \in \mathcal{D}_n$$

Once the predictions are available, the measurement procedure is given by:

$$C_{D_1}^{(\beta)}(\{x\}) = \left\{ n \,|\, x \in A_X^{(\beta)}(\mathcal{D}_n) \right\}$$

Prediction Procedure $A_X^{(\beta)}$

The prediction procedure $A_X^{(\beta)}$ assigns to each possible value of n a prediction. The derivation of the prediction procedure is difficult, because of the ignorance with respect to the failure probability p. However, the problem can be solved by means of the computer algebra system *Stochastikon Calculator*. The desired predictions calculated by *Stochastikon Calculator* are displayed in Table 1 below.

$\mathcal{D}_n$	$A_X^{(0.9)}(\mathcal{D}_n)$
$\{(n,p) \mid n = 400, 0.05 \le p \le 0.06\}$	$\{x \mid 0 \le x\}$
$\{(n,p) \mid n, 0.05 \le p \le 0.06\}$ for $401 \le n \le 422$	$\{x \mid 15 \le x\}$
$\{(n,p) \mid n, 0.05 \le p \le 0.06\}$ for $423 \le n \le 445$	$\{x \mid 16 \le x\}$
$\{(n,p) \mid n, 0.05 \le p \le 0.06\}$ for $446 \le n \le 470$	$\{x \mid 17 \le x\}$
$\{(n,p) \mid n, 0.05 \le p \le 0.06\}$ for $471 \le n \le 491$	$\{x \mid 18 \le x\}$
$\{(n,p) \mid n, 0.05 \le p \le 0.06\}$ for $492 \le n \le 500$	$\{x \mid 19 \le x\}$

Table 1: Predictions $A_X^{(0.9)}(\mathcal{D}_n)$ for $n = 400, 401, \ldots, 500$

With Table 1, the measurement procedure for an upper bound of the number of units is immediately obtained:

x	$C_D^{(0.9)}(\{x\})$
$x \le 15$	$C_D^{(0.9)}(\{x\}) = \{400\}$
15	$C_D^{(0.9)}(\{15\}) = \{n \mid 400 \le n \le 422\}$
16	$C_D^{(0.9)}(\{16\}) = \{n \mid 400 \le n \le 445\}$
17	$C_D^{(0.9)}(\{17\}) = \{n \mid 400 \le n \le 470\}$
18	$C_D^{(0.9)}(\{18\}) = \{n \mid 400 \le n \le 491\}$
19	$C_D^{(0.9)}(\{19\}) = \{n \mid 400 \le n \le 500\}$
$x \ge 20$	$C_D^{(0.9)}(\{x\}) = \{500\}$

Table 2: Measurements $C_D^{(0.9)}(\{x\})$ for $x = 0, 1, \ldots, 500$

If, for example, $x = 17$ failures are observed, then the measurement result is given by

$$C_D^{(0.9)}(\{17\}) = \{n \mid 400 \le n \le 470\}$$

and we conclude that there were not more than 470 units.

Module 2.4: Exclusion Procedures

Content and Aim of the Module *Exclusion Procedures*

The significance test constitutes the most frequently used statistical method. This fact must surprise as significance tests are rather inefficient as they use only partially the information provided by a random process. Moreover, the result of an significance test is often misunderstood and misinterpreted.

The following entry can be found in the well-known *Encyclopedia of Statistical Sciences*:

> *A satisfactory significance test must (1) stipulate a suitable hypothesis of chance, (2) find a test statistic to rank possible experimental outcomes, (3) determine the level of significance of the experimental outcome from the test statistics probability distribution, and (4) reject or fail to reject the hypothesis of chance. The history of significance tests may be viewed as a concerted effort to construct tests that satisfy these criteria. Much controversy remains today over the degree to which tests can and do satisfy these four criteria.*

This citation indicates some fundamental difficulties with significance tests. One of the reasons for these difficulties is the fact that not the aim of a significance test is stated, but only four vague properties. Accordingly a test is based on a hypothesis, however, what is the meaning of a *hypothesis of chance* and what *justifies the ordering provided by the test statistics?*[12]

This learning unit introduces the stochastic analogue of a (statistical) significance test, which is called *exclusion procedure*. An exclusion procedure is imbedded into a Bernoulli Space, and starts with formulating its purpose allowing to deduce necessary and desirable properties.

The aim of this module is threefold:

- The module introduces and explains a stochastic procedure for reducing the ignorance space by excluding a certain specified part of it.

- The second aim of the module is to enable a better understanding of statistical significance tests.

- Passing this module should enable to successfully apply the exclusion procedures provided by *Stochastikon Calculator* for solving related problems in an optimal way.

In contrast to significance tests, which are mainly restricted to null-hypotheses given as singletons, exclusion procedures can be applied to any subset of the ignorance space. Moreover, the result of an exclusion procedure is given in a way that any misinterpretation is hardly possible.

This module provides not only a clarification of the role of significance tests and their weaknesses, but also offers an alternative which is developed on a clear and sound basis.

References

[1] Kotz, S. and Johnson N.L. (Eds.) (1988): *Encyclopedia of Statistical Sciences*, Vol. 8, p. 466.

---

[12]Question posed by Jerzy Neyman cited according to the Encyclopedia of Statistical Sciences, Vol. 8, p. 469.

Unit 2.4.1: Exclusion Procedure and Measurement Procedure

Target

<div style="background-color:#d9e9d9; padding:1em;">

In this learning unit the difference between a stochastic measurement procedure and a stochastic exclusion procedure is shown. Both procedures have in common that they aim at reducing the ignorance space. However, a measurement procedure yields the largest reduction, while in case of an exclusion procedure the desired reduction is specified beforehand.

</div>

Content

Introduction

There are situations, when the actual value d_0 of a deterministic variable D is not of primary interest, but one wants to prove that the actual value d_0 is not an element of a specified subset $\mathcal{D}_0$ of the ignorance space $\mathcal{D}$.

In such a situation an exclusion procedure may be applied instead of a measurement procedure. An exclusion procedures answers only the question whether the subset $\mathcal{D}_0$ can be excluded or not. It does not at all aim at determining the unknown value d of D as precisely as possible.

An exclusion procedure consists of an random process which yields an observation and a stochastic procedure which answers the question whether or not $\mathcal{D}_0$ may be excluded.

The process to be performed is related to the pair of variables (X, D), with X being the random variable which will be observed, and D the deterministic variable. The exclusion procedure aims at showing that the actual value of D is not an element of $\mathcal{D}_0$.

The stochastic model of the process represented by (X, D) is given by the corresponding Bernoulli Space $\mathbb{B}_{X,D}$. If the evaluation of the observed value x results in an exclusion of $\mathcal{D}_0$ and, hence, in an reduction of the ignorance space, then the procedure was applied successful, otherwise it was a failure.

In the following section it is shown, how a given measurement procedure can be utilized for excluding $\mathcal{D}_0$.

Measurement Procedure as Exclusion Procedure

A measurement procedure aims at reducing as much as possible the given ignorance space, i. e., the measurement process shall be utilized in an optimal way to learn about the unknown value of the deterministic variable.

Any β-measurement procedure $C_D^{(\beta)}$ is based on a β-prediction procedure $A_X^{(\beta)}$ defined for each of the potential values d of the deterministic variable D. After having performed the measurement experiment resulting in an observation x, each value d of D is abandoned from the ignorance space, for which the respective prediction $A_X^{(\beta)}(\{d\})$ does not contain the observed value x. The β-prediction procedure is selected so that:

- the reliability requirement for the measurement procedure is met,

- the completeness of the measurement procedure is given, and

- the measurement results are on an average most precise.

Let the set to be excluded denoted by $\mathcal{D}_0$ and let the measurement result be given by $C_D^{(\beta)}(\{x\})$. If

$$\mathcal{D}_0 \cap C_D^{(\beta)}(\{x\}) = \emptyset$$

holds, then none of the predictions $A_X^{(\beta)}(\{d\})$ for $d \in \mathcal{D}_0$ contains the observation x and, therefore, the set $\mathcal{D}_0$ can be excluded. Exactly this is the aim of an exclusion procedure. If the subset $\mathcal{D}_0$ can be excluded, one obtains the reduced ignorance space $\mathcal{D} \setminus \mathcal{D}_0$.

If $\mathcal{D}_0$ can be excluded on the basis of the measurement result, the measurement procedure allows to exclude additionally the following part of the ignorance space:

$$\overline{C}_D^{(\beta)}(\{x\} \setminus \mathcal{D}_0 \tag{124}$$

where $\overline{C}_D^{(\beta)}(\{x\}) = \mathcal{D} \setminus C_D^{(\beta)}(\{x\})$ is the complement of the measurement result. The measurement result $C_D^{(\beta)}(\{x\})$ constitutes the new ignorance space, which is generally much smaller than $\mathcal{D} \setminus \mathcal{D}_0$.

Note that even if $\mathcal{D}_0$ cannot be excluded on the basis of the measurement result, a measurement procedures yields in general a reduction of the ignorance space. Therefore, a measurement procedure is more efficient for reducing the ignorance space than an exclusion procedure.

Exclusion Procedure

An exclusion procedure is derived very similar as a measurement procedure. Let $\mathcal{D}$ be the ignorance space and $\mathcal{D}_0 \subset \mathcal{D}$ the subset to be excluded on the result of a random process represented by the random variable X. The exclusion procedure is based on a prediction procedure and the prediction $A_X(\mathcal{D}_0)$. The hypothesis $\mathcal{D}_0$ is excluded, if the observed event $\{x\}$ of the random process has not been predicted, otherwise it cannot be excluded.

Reasons for Applying an Exclusion Procedure

The reason for applying an exclusion procedure instead of a measurement procedure has already been indicated in the introduction. Whenever one does not want to disclose the true value of a deterministic variable, but only show that it is not an element of a given set, then an exclusion procedure is indicated.

Such situations occur often in the pharmaceutical industry, when a new medicine shall be introduced. Then the aim is to prove that the new medicine is more effective than the old one, but often not to disclose the true value of the effectiveness.

EXAMPLE

1. **Marketing Campaign**

   A company that makes Brand B laundry detergent knows that the proportion of house-wives favoring Brand B over all other detergents is $p = 0.20$. For increasing the proportion the marketing department launched an intensive and costly advertising campaign. In order to show the success of the campaign, it was decided to apply a β-exclusion procedure with $\beta = 0.90$ for excluding the set $\mathcal{D}_0 = \{p \mid p \leq 0.2\}$.

By random, $n = 300$ housewives are to be selected and asked about their favorite laundry detergent. The variable of interest is the number of housewives in favor of Brand B.

The process represented by the questionnaire is modelled by a Bernoulli Space with random structure given by the binomial distribution and with the deterministic variable given by the success probability, which equals the proportion of housewives favoring Brand B. A suitable exclusion procedure is offered by *Stochastikon Calculator*.

The questionaire was performed and results in $x = 72$ housewives favoring Brand B. Inserting the input-values into the corresponding exclusion procedure of *Stochastikon Calculator* leads to an exclusion of $\mathcal{D}_0 = \{p \,|\, p \leq 0.2\}$ and, thus, proving the success of the advertisement campaign.

Figure 1: Housewife filling a questionnaire.

The graphical representation of the corresponding exclusion procedure is readily obtained by the Graphical Laboratory. In Figure 1 the prediction $A_X(\mathcal{D}_0)$ in case $\mathcal{D}_0$ is highlighted in dark blue. As can be seen the observation represented by the red line does not cross the prediction and, therefore, the set $\mathcal{D}_0$ can be excluded.
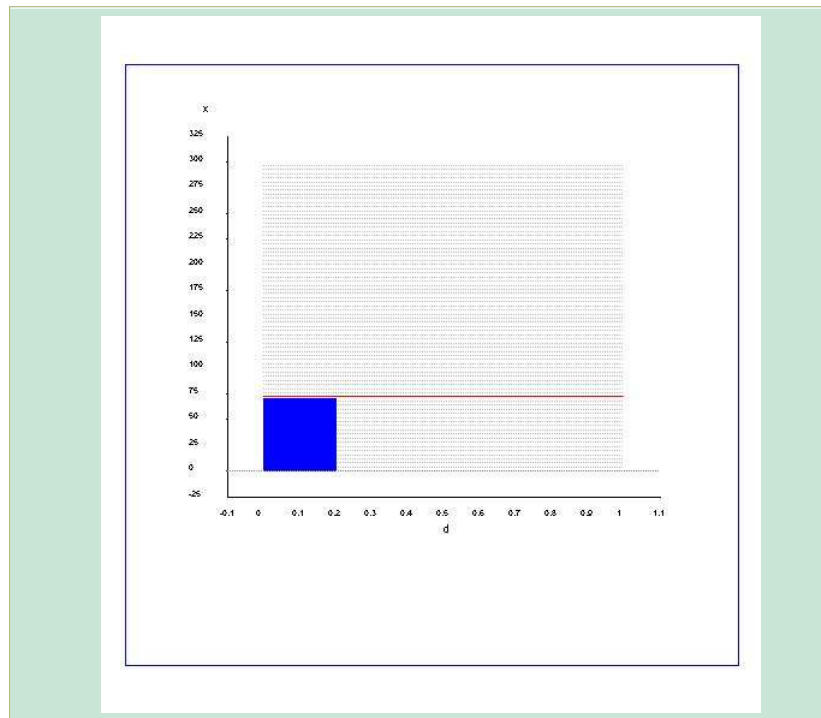
Figure 1: Graphical representation of the stochastic exclusion procedure for the set $\mathcal{D}_0 = \{p \,|\, p \leq 0.2\}$

The company's management decides to get more detailed information about the new situation by applying a 0.90-measurement procedure, which is also offered by *Stochastikon Calculator*. The 0.90-measurement procedure yields the following result:

$$C_D^{(0.90)}(\{72\}) = \{p \,|\, 0.21 \leq p \leq 0.28\}$$

Based on this result management can judge the effects of the advertisement better, since the result not only shows that the marketing campaign was successful, but it also quantifies the success. Thus, the measurement result provides a sound basis for any decisions with respect to Brand B.

The graphical representation of measurement procedure used for getting the above result can easily be generated by the Graphical Laboratory:
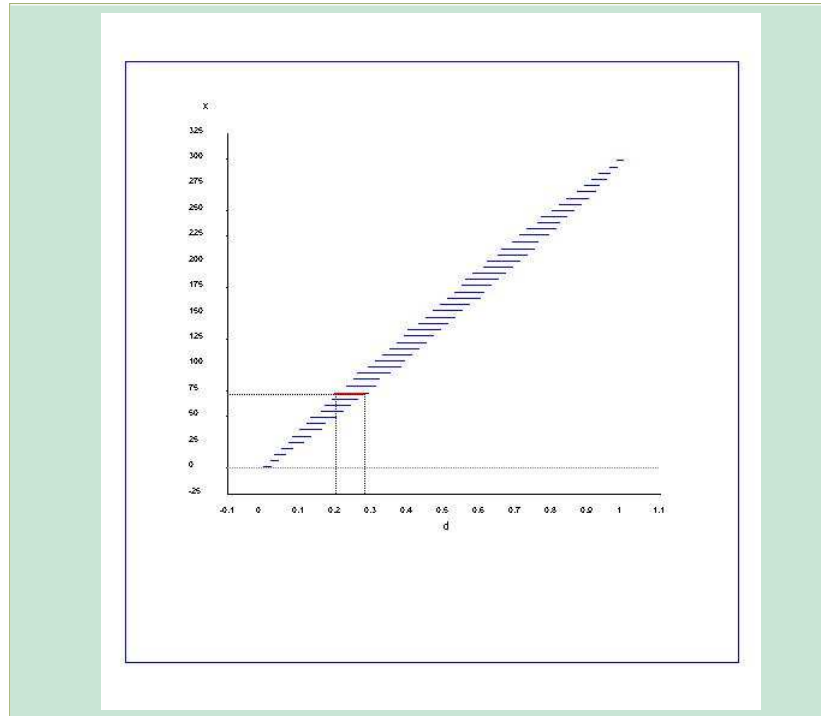

Figure 2: Graphical representation of the stochastic measurement $C_D^{(0.90)}$ and the measurement result $C_D^{(0.90)}(\{72\})$

Unit 2.4.2: β-Exclusion Procedure

TARGET

The aim of a exclusion procedure is to reduce the ignorance space $\mathcal{D}$ by showing that a given subset $\mathcal{D}_0 \in \mathcal{T}_D(\mathcal{D})$ can be excluded, thus reducing the ignorance space to $\mathcal{D} \setminus \mathcal{D}_0$. In this learning unit the reliability level β for an exclusion procedure is introduced and it is shown how the corresponding procedures are derived.

CONTENT

Introduction

A β-exclusion procedure on the basis of an exclusion process may be used for excluding a specified subset $\mathcal{D}_0$ of the ignorance space $\mathcal{D}$ of a certain deterministic variable D. An exclusion procedure is given by a decision procedure φ with two possible outcomes, namely to exclude or not exclude the given subset $\mathcal{D}_0 \subset \mathcal{D}$. If the decision function adopts the value 1, the set $\mathcal{D}_0$ is excluded, if the value 0 is adopted, $\mathcal{D}_0$ is not excluded.

An exclusion procedure defined on a set of subsets $\mathcal{T}_D(\mathcal{D})$ of the ignorance space is called β-exclusion procedure denoted $\varphi_{\mathcal{D}_0}^{(\beta)}$, if the probability of a correct decision when applied is at least β.

The decision to exclude a considered subset $\mathcal{D}_0 \in \mathcal{T}_D(\mathcal{D})$ means to reduce the ignorance space from $\mathcal{D}$ to $\mathcal{D} \setminus \mathcal{D}_0$ and, clearly, the decision is wrong, if the true value of D is an element of $\mathcal{D}_0$.

The decision not to exclude $\mathcal{D}_0$ means that no reduction of the ignorance space $\mathcal{D}$ is made and this decision is, of course, always correct. However, not to exclude $\mathcal{D}_0$ represents a failure of the attempt to exclude $\mathcal{D}_0$. Consequently, a false decision can happen only in case of an exclusion, i. e., in case the decision function $\varphi_{\mathcal{D}_0}^{(\beta)}$ adopts the value 1.

The exclusion process is modelled by a Bernoulli Space $\mathbb{B}_{X,D}$, where X represents the variable to be observed and D the deterministic variable related to the set $\mathcal{D}_0$ to be excluded. The Bernoulli Space enables the development of prediction procedures, which are necessary for deriving the required exclusion procedure $\varphi_{\mathcal{D}_0}^{(\beta)}$.

Formal Representation

An exclusion procedure is given by a decision function $\varphi_{\mathcal{D}_0}$:

$$\varphi_{\mathcal{D}_0} : \left\{ \{x\} \,|\, x \in \mathcal{X}(\mathcal{D}) \right\} \to \{0,1\} \tag{125}$$

with the following meaning:

$$\begin{aligned} \varphi_{\mathcal{D}_0}(\{x\}) = 1 &\Rightarrow \mathcal{D}_0 \text{ can be excluded and } \mathcal{D} \setminus \mathcal{D}_0 \text{ is accepted} \\ \varphi_{\mathcal{D}_0}(\{x\}) = 0 &\Rightarrow \mathcal{D}_0 \text{ cannot be excluded and } \mathcal{D} \text{ is maintained} \end{aligned} \tag{126}$$

Thus, a false decision occurs if:

$$\varphi_{\mathcal{D}_0}(\{x\}) = 1 \qquad \text{for } d \in \mathcal{D}_0$$

117

In order to have the probability of a correct decision, when applying the exclusion procedure, sufficiently high, a reliability specification is set by means of the reliability level β. An exclusion procedure meets the reliability level β, if the probability for a correct decision is at least equal to β.

An exclusion procedure meeting the reliability specification is called β-exclusion procedure denoted $\varphi_{\mathcal{D}_0}^{(\beta)}$. The probability of a wrong decision is smaller then $1 - \beta$:

- $P_{X|\{d\}}\left(\{x \,|\, \varphi_{\mathcal{D}_0}(\{x\}) = 1\}\right) < 1 - \beta$ for $d \in \mathcal{D}_0$

From the above we obtain the following probabilities of correct decisions:

- Case 1: $d \in \mathcal{D}_0$

  A correct decision is made if $\varphi_{\mathcal{D}_0}(\{x\}) = 0$ with probability:

$$P_{X|\{d\}}\left(\{x \,|\, \varphi_{\mathcal{D}_0}(\{x\}) = 0\}\right) \geq \beta$$

- Case 2: $d \in \mathcal{D} \setminus \mathcal{D}_0$

  In this case either of the two possible decisions is correct. Thus:

$$P_{X|\{d\}}\left(\{x \,|\, \varphi_{\mathcal{D}_0}(\{x\}) \in \{0, 1\}\}\right) = 1$$

β-Exclusion Procedure for $\mathcal{D}_0$

A β-exclusion procedure refers to the past which is represented by the deterministic variable D. Hence, exclusion procedures are members of the measurement class of procedures.

Analogous to β-measurement procedures, it is based on a β-prediction procedure with respect to the outcome of the exclusion process X.

$$A_X^{(\beta)} : \mathcal{T}_D(\mathcal{D}) \to \mathcal{T}_X(\mathcal{X}(\mathcal{D}) \tag{127}$$

where the system of subsets $\mathcal{T}_X(\mathcal{X}(\mathcal{D}))$ is not the system of singletons as in the case of a measurement procedure, but the system containing only one element namely $\mathcal{D}_0$. Exactly this is the characteristic difference between a measurement procedure and an exclusion procedure. A measurement procedure has the highest possible resolution expressed by the domain of the corresponding prediction procedure, consisting of all the singletons of $\mathcal{D}$. The domain of the prediction procedure of a exclusion procedure for $\mathcal{D}_0$ is the system of subsets which contains only one element, namely $\mathcal{D}_0$, i.e.:

$$\mathcal{T}_D(\mathcal{D}) = \{\mathcal{D}_0\} \text{ in case of an exclusion procedure fro } \mathcal{D}_0$$

The β-prediction procedure for a β-exclusion procedures is therefore completely specified by one prediction:

$$A_X^{(\beta)}(\mathcal{D}_0)$$

Once a β-prediction $A_X^{(\beta)}(\mathcal{D}_0)$ is available, the decision function $\varphi_{\mathcal{D}_0}^{(\beta)}$ is obtained in a straightforward manner:

$$\varphi_{\mathcal{D}_0}(\{x\}) = \begin{cases} 0 & \text{for} \quad x \in A_X^{(\beta)}(\mathcal{D}_0) \\ 1 & \text{for} \quad x \notin A_X^{(\beta)}(\mathcal{D}_0) \end{cases} \tag{128}$$

The set $\mathcal{D}_0$ is excluded, if the observed event $\{x\}$ is not a subset of the predicted event $A_X^{(\beta)}(\mathcal{D}_0)$. If the observation falls into the prediction, the set $\mathcal{D}_0$ can, of course, not be excluded.

A Note on Significance Tests

In many significance tests the set to be excluded is a singleton. However, generally a continuous approximation for the deterministic variable D is used, and excluding one single point and admitting all adjacent neighbors makes therefore no sense at all.

This is the reason that *Stochastikon Calculator* does not accept such a situation (continuous deterministic variable D, singleton to be excluded) and will record an error.

EXAMPLE

1. **Pharmaceutics**

   Exclusion procedures (as given by significance tests – see Learning Unit 2.4.5) – are the most frequently used stochastic methods in the pharmaceutics. Whenever a new drug shall be put on market, it has to be proved that it is more effective than an already existing one.

   Let the efficacy of the drugs been given by the probability of being effective with the following ignorance spaces for the old and the new drug:

   $$\begin{aligned} \text{old drug:} \quad & \mathcal{D}_0 = \{p \,|\, p \le p_0\} \\ \text{new drug:} \quad & \mathcal{D} = \{p \,|\, p \le p_1\} \end{aligned}$$

   Because of the experience with the old drug, the value p_0 is generally known, while the value p_1 is only anticipated to be larger than p_0. In order to prove that the new drug is actually more efficient than the old one, it is necessary to show that $\mathcal{D}_0$ can be excluded from the ignorance space $\mathcal{D}$ with respect to the new drug.

   The exclusion experiment consists of administering the new drug to a set of n more or less equal patients. Thus, the exclusion experiment represents a Bernoulli-chain of length n. The number of successes X_n among the n trials has a binomial probability distribution with a two-dimensional deterministic variable $D = (D_1, D_2)$ with:

   $$\begin{aligned} D_1 = & \ \text{number of trials} \\ D_2 = & \ \text{success probability} \end{aligned}$$

   The resulting Bernoulli Space is the so-called Binomial Bernoulli Space, which is covered by *Stochastikon Calculator*. Assume here the following case:

   Number of trials: $\quad n = 50$
   Ignorance space with resepect to the success probability of new drug:
   $\quad \mathcal{D} = \{p \,|\, p \le 0.85\}$
   Ignorance space with respect to the success probability of old drug:
   $\quad \mathcal{D}_0 = \{p \,|\, p \le 0.55\}$
   Required reliability level: $\quad \beta = 0.95$

   Entering the above input values into Stochastikon Calculator yields the following β-prediction for $\mathcal{D}_0$:
   $$A_{X_n}^{(0.95)}(\mathcal{D}_0) = \{x \,|\, x \le 33\}$$

and, hence, the following decision function with respect to $\mathcal{D}_0$ is obtained:

$$\varphi_{\mathcal{D}_0}^{(\beta)}(\{x\}) = \begin{cases} 1 & \text{if} \quad x \geq 34 \\ 0 & \text{if} \quad x \leq 33 \end{cases}$$

The exclusion process is performed and yields $x = 45$ successes.

Inserting the number of successes into $\varphi_{\mathcal{D}_0}^{(\beta)}(\{x\})$, yields the following decision:

$$\varphi_{\mathcal{D}_0}^{(\beta)}(\{45\}) = 1$$

i. e., the set $\mathcal{D}_0$ may be excluded showing that the new drug is superior to the old one. The graphical representation of the above derived exclusion procedure is given in Figure 1. The prediction area for $\mathcal{D}_0$ is highlighted in dark blue, while the observation is represented by the red line. Because the line does not cross the prediction area, the set $\mathcal{D}_0$ is excluded.
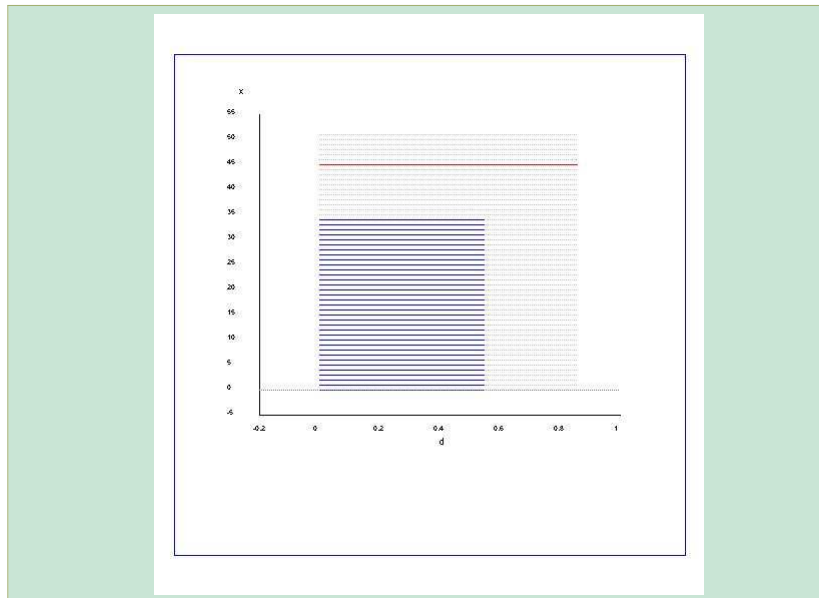
Figure 1: Graphical representation of the exclusion procedure $\varphi_{\mathcal{D}_0}^{(0.95)}$.

Unit 2.4.3: Optimum β-Exclusion Procedure

TARGET

In the previous learning unit the reliability specification for exclusion procedures has been introduced, which represents a necessary side condition.

In this learning unit a natural objective function for β-exclusion procedures is derived and optimal β-exclusion procedures are defined meeting the reliability requirement.

CONTENT

The Mean Exclusion Probability

A β-exclusion procedure $\varphi_{\mathcal{D}_0}^{(\beta)}$ aims at excluding a subset $\mathcal{D}_0$ of the ignorance space $\mathcal{D}$ on condition that the probability of a correct decision is not less than the reliability level β. There are two possibilities of a correct decision:

1. The exclusion fails and the given ignorance space is maintained.

2. The set $\mathcal{D}_0$ is excluded from the ignorance space for $d \in \mathcal{D} \setminus \mathcal{D}_0$.

The search for an optimal exclusion procedure is restricted on the set of β-exclusion procedures, i. e., those procedures which meet the reliability side condition. The procedure's reliability is taken care of by the side condition and, thus, it needs not be considered by the objective function.

Any β-exclusion procedure which leads more frequently to the desired exclusion than another β-exclusion procedure will be judged as being the better one.

In other words, a β-exclusion procedure with a higher probability of exclusion is better than one with a lower exclusion probability. The problem is that the exclusion probability of any set $\mathcal{D}_0 \in \mathcal{T}_D(\mathcal{D})$ depends on the unknown value $d \in \mathcal{D}$ of the deterministic variable D. Therefore, not the exclusion probability for a given d can be selected as objective function, but the mean exclusion probability, with the mean taken over all the possible values of $d \in \mathcal{D}$.

Let the ignorance space consists of N elements, i. e., $\mathcal{D} = \{d_1, \ldots, d_N\}$. Let the actual value of the deterministic variable D be given by d_i. Then, the exclusion probability for $\mathcal{D}_0$ is equal to the probability that an observation x will be made, which is not an element of the corresponding prediction $A_X^{(\beta)}(\mathcal{D}_0)$. The exclusion probability may also be represented by the first moment of the corresponding decision function $\varphi_{\mathcal{D}_0}^{(\beta)}$:

$$E_{X|\{d_i\}}\left[\varphi_{\mathcal{D}_0}^{(\beta)}\right] \;=\; P_{X|\{d_i\}}\Big(\{x \in \mathcal{X}(\mathcal{D}) \,|\, x \notin A_X^{(\beta)}(\mathcal{D}_0)\}\Big) \tag{129}$$

$$=\; P_{X|\{d_i\}}\Big(\{x \in \mathcal{X}(\mathcal{D}) \,|\, \varphi_{\mathcal{D}_0}^{(\beta)}(\{x\}) = 1\}\Big) \tag{130}$$

In case the actual value d_i is not known, the mean exclusion probability for the entire ignorance

space denoted by $E_{X|\mathcal{D}}\left[\varphi_{\mathcal{D}_0}^{(\beta)}\right]$ must be considered. It is given as follows:

$$
\begin{aligned}
E_{X|\mathcal{D}}\left[\varphi_{\mathcal{D}_0}^{(\beta)}\right] &= \frac{1}{|\mathcal{D}|}\sum_{d_i \in \mathcal{D}} P_{X|\{d_i\}}\left(\{x \in \mathcal{X}(\mathcal{D}) \,|\, x \notin A_X^{(\beta)}(\mathcal{D}_0)\}\right) \\
&= \frac{1}{N}\sum_{i=1}^{N} P_{X|\{d_i\}}\left(\{x \in \mathcal{X}(\mathcal{D}) \,|\, \varphi_{\mathcal{D}_0}^{(\beta)}(\{x\}) = 1\}\right)
\end{aligned}
\tag{131}
$$

If a continuous approximation is used, the sum is replaced by the corresponding integral.

The Optimality Criterion

The aim of a β-exclusion procedure is to exclude a specified element of $\mathcal{T}_D(\mathcal{D})$ based on an exclusion experiment. Thus, the quality of a β-exclusion procedure is determined by its success probability. Because the actual value of the deterministic variable and hence the actual probability distribution is unknown, the mean success probability is taken as optimality criterion.

A β-exclusion procedure $^*\varphi_{\mathcal{D}_0}^{(\beta)}$ is called optimal, if

$$
E_{X|\mathcal{D}}\left[^*\varphi_{\mathcal{D}_0}^{(\beta)}\right] = \max_{\varphi_{\mathcal{D}_0}^{(\beta)}} E_{X|\mathcal{D}}\left[\varphi_{\mathcal{D}_0}^{(\beta)}\right]
\tag{132}
$$

Remarks:

- Note that the optimality criteria directly reflects the objective of an exclusion procedure taking into account that the actual value of the deterministic variable D is unknown. Thus, whatever situation with respect to d is the true one, the procedure will perform well.

- Note also that the optimality criterion does not take into account the correctness of decisions, since this is done by the reliability requirement, which is specified by the user and which represents an admissible risk of $1 - \beta$ for making a wrong decision.

Optimality Criteria of Significance Tests

A significance test refers to a so-called null hypothesis, which sometimes goes along with a so-called alternative hypothesis. The aim is often not mentioned clearly. However, safeguarding by means of the significance level α refers only to the case of a wrongly rejected null hypothesis. Therefore, the null hypothesis may be regarded as the set $\mathcal{D}_0$ within an $(1 - \alpha)$-exclusion procedure aiming at excluding $\mathcal{D}_0$ on the condition of the reliability level $\beta = 1 - \alpha$. By the way, erroneously rejecting the null hypothesis is called the Type I Error in statistics.

The generally used optimality criterion for significance tests is its power, which is based on the so-called Type II Error. This error is defined as the probability of not excluding the null hypothesis on the condition that $d \notin \mathcal{D}_0$.

Evidently, the Type II Error is not a decision error, because the decision not to exclude $\mathcal{D}_0$ only means to maintain the given ignorance space $\mathcal{D}$ which is correct by assumption.

The Type II Error refers to the failure of the procedure due to a too small sample size n. The power of a test is given by the probability of not committing a Type II Error and, therefore, depends heavily on the unknown actual value d of the deterministic variable.

Note that the correctness of a decision and a failure of the procedure are two very different things, which should not be confused. Putting the Type I Error side by side to the Type II Error must almost necessarily lead to misunderstandings and misinterpretations. Particularly, an exclusion procedure just as a significance test is not able to classify a given situation to one of two or more alternatives, which would make a Type II Error meaningful.

The frequent misuse of significance tests is illustrated by the following citation[13]:

> *Because SSTs (statistical significance tests) have been so frequently misapplied, some reflective researchers (e.g. Carver, 1978; Meehl, 1978; Schmidt, 1996; Shulmann, 1970) have recommended that SSTs be completely abandoned as a method for evaluating statistical results. In fact Carver (1993) not only recommended abandoning statistical significance testing, but referred to it as a "corrupt form of the scientific method" (p. 288).*

EXAMPLE

1. **Gambling**

   The exclusion procedures used in the examples of the previous learning units were already optimum exclusion procedures. Here the concept shall be illustrated by a series of games of chance (e.g. throwing a coin), where each of the two participants A and B win with probability $p = 0.5$.

   A and B play a number of games, but at the end this number has not be documented; however, player A knows exactly that he has won $x = 45$ times. B insists that they have played at least 120 time but not more than 140 times. A believes that they played less than 120 times, but at least 100 times. In order to answer the question, A decides to use a stochastic exclusion procedure in order to exclude the claim $\mathcal{D}_0 = \{n \,|\, 120 \leq n \leq 140\}$. The situation is quantified in the following way:

   Pair of variables:
   $$\begin{aligned} X &= \text{number of successes of } A \\ D &= \text{number of games} \end{aligned} \tag{133}$$

   Bernoulli Space:
   $$\begin{aligned} \mathcal{D} &= \{n \,|\, 100 \leq n \leq 140\} \\ \mathcal{X}(\{n\}) &= \{0, 1, 2, \ldots, n\} \\ \mathcal{P}(\{n\}) &= P_{X|\{n\}} \text{ with } X|\{n\} \sim Bi(n, 0.50) \end{aligned} \tag{134}$$

   The aim is to exclude the set $\mathcal{D}_0 = \{n \,|\, 120 \leq n \leq 140\}$ by means of an exclusion procedure $\varphi_{\mathcal{D}_0}^{(0.90)}$. The result of the exclusion process is $x = 45$. The optimal decision function is obtained by the Stochastikon Calculator:

   $$\varphi_{\mathcal{D}_0}^{(0.90)}(\{45\}) = 1 \tag{135}$$

   i.e., the set $\mathcal{D}_0$ is excluded from the ignorance space.

   By means of the Graphical Laboratory the following representation of the exclusion procedure is obtained:

---

[13]Larry G. Daniel (1998): Statistical Significance Testing: A Historical Overview of Misuse and Misinterpretations with Implications for the Editorial Policies of Educational Journals. *Research in Schools* 5, 23-32.
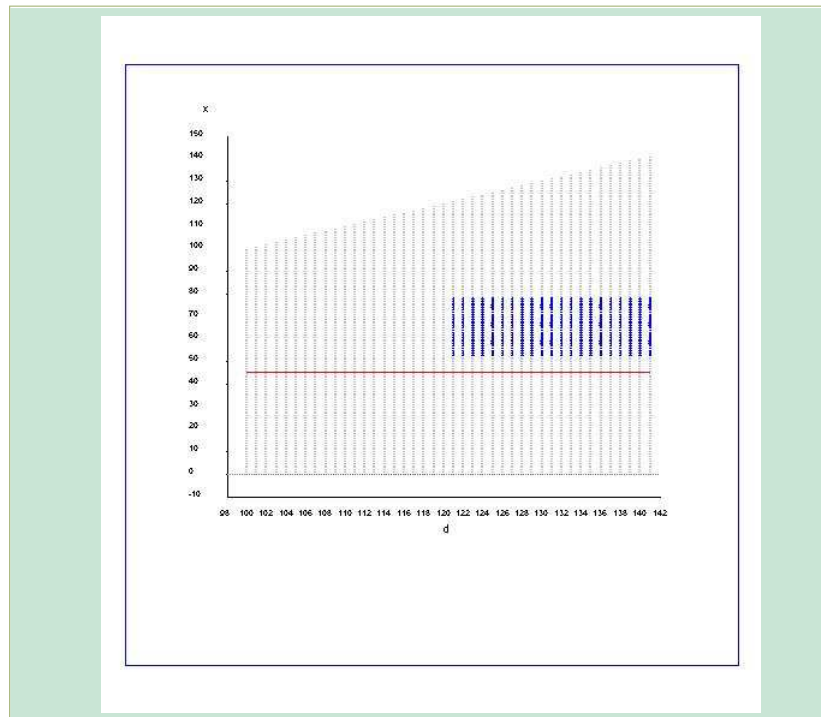
Figure 1: Graphical representation of the stochastic exclusion procedure $\varphi_{\mathcal{D}_0}^{(0.90)}$ for $\mathcal{D}_0 = \{n \in \mathbb{N} \mid 120 \leq n \leq 140\}$

Unit 2.4.4: Graphical Representation

TARGET

A β-exclusion procedure is to some extent comparable with a significance test on significance level $\alpha = 1 - \beta$. This learning unit aims at illustrating aim and mode of operation of a β-exclusion procedure by means of a graphical representation, which also can be used for illustrating significance test and their results.

CONTENT

Introduction

Actually, a β-exclusion procedure is one of the simplest stochastic procedures, which can be seen from the fact that the domain of $\varphi_{\mathcal{D}_0}^{(\beta)}$ consists of one element $\{\mathcal{D}_0\}$ and the codomain of only two element namely $\{0, 1\}$.

Nevertheless, it seems to be indicated to explain an exclusion procedure by means of a graphical representation, because of the frequent malpractice of significance tests in all fields of applications and the fact that statistical significance tests and stochastic exclusion procedures are relatively similar.

The graphical representation will be restricted to the case that the random structure is given by a binomial probability distribution.

The Binomial Case

The general problem is to exclude a certain part $\mathcal{D}_0$ of a given ignorance space $\mathcal{D}$ with respect to a deterministic variable D based on an exclusion process described by the random variable X.

Let the deterministic variable D be defined as the probability of a certain event. In this case the exclusion process consists of n independent repetitions of the process in question with X being the number of times the specified event will occur.

Let $n = 130$ be the number of repetitions, and consider the case that from past experience it is known that the probability of occurrence of the certain event is positive but does not exceed the value 0.5. Then the pair of variables is given as follows:

$$X = \text{number of successes}$$
$$D = (D_n, D_p)$$

with

$$D_n = \text{number of repetitions}$$
$$D_p = \text{probability of the specified event}$$

The Bernoulli Space $\mathbb{B}_{X,D}$ for (X, D) is given by

$$\mathcal{D} = \{130, p)\,|\, 0 < p \leq 0.5\}$$
$$\mathcal{X}(\{(130, p)\})\,|\, = \{0, 1, \ldots, 130\} \quad \text{for } (130, p) \in \mathcal{D}$$
$$\mathcal{P}(\{(130, p)\}) = P_{X|\{(130,p)\}} \text{ with } X|\{(130, p)\} \sim Bi(130, p)$$

The problem is to exclude the set $\mathcal{D}_0 = \{p\,|\, 0 < p \leq 0.2\}$ by a β-exclusion procedure with $\beta = 0.95$.

The optimal β-exclusion procedure is given by a prediction $A^{(\beta)}_{X|\mathcal{D}_0}$, which maximizes the mean exclusion probability.

Graphical Representation

The graphical representation should refer to the involved sets and points. Thus, it should include the ignorance space $\mathcal{D}_p$ with respect to p, the subset to be excluded $\mathcal{D}_0$, the prediction $A^{(\beta)}_{X|\mathcal{D}_0}$ and finally the realization x indicating whether $\mathcal{D}_0$ can be excluded or not. Below, the graphical representation for the special case given above is displayed.

The abscissa refers to the deterministic variable D. The elements of the ignorance space $(0 < p \leq 0.5)$ are marked in light blue. The ordinate refers to the outcome x of the exclusion experiment and runs from 0 to 130.

The uncertainty space is given by the light blue area in the (p, x)-plane:

$$\mathcal{U}_{X,D_p} = \{(p, x)\,|\, 0 < p \leq 0.5, x \in \mathbb{N}, 0 \leq x \leq 130\}$$

The set $\mathcal{D}_0 = \{p\,|\, 0 < p \leq 0.2\}$ to be excluded is marked dark blue on the abscissa. While the optimal prediction $A^{(\beta)}_X(\mathcal{D}_0) = \{x\,|\, 0 \leq x \leq 34\}$ is indicated by the dark blue area $\mathcal{D}_0 \times A^{(\beta)}_X(\mathcal{D}_0)$.
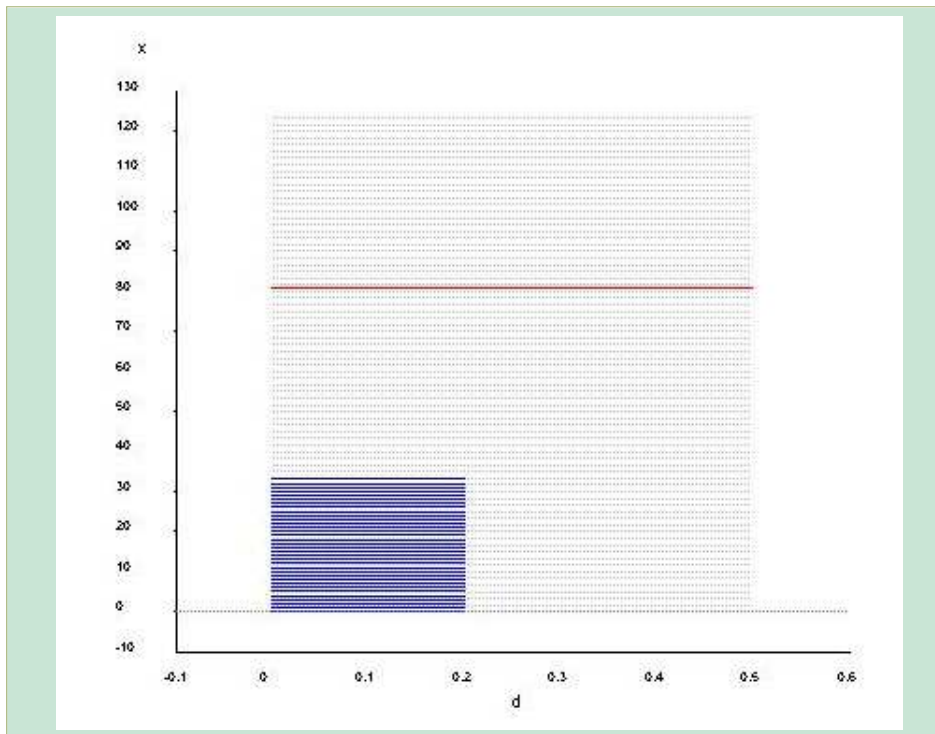
Figure 1: The set $\mathcal{D}_0 \times A^{(\beta)}_X(\mathcal{D}_0)$ within the uncertainty space $\mathcal{U}_{X,D_p}$.

If the line $\{(p, x_0) \,|\, 0 < p \leq 0.2\}$ where x_0 denotes observation of the exclusion process has no common element with the above direct set product, then $\mathcal{D}_0$ can be excluded.

The observation $x_0 = 81$ is indicated by the corresponding horizontal line in the uncertainty space. If this line would cross the dark blue area, then it would not be possible to exclude $\mathcal{D}_0$, otherwise it can be excluded.

In the example the observation lies far outside of the prediction $A_X^{(\beta)}(\mathcal{D}_0)$ and, therefore, the set $\mathcal{D}_0$ can be excluded.

EXAMPLE

1. **Uncertainty Space and Prediction**

   An exclusion procedure is given by an decision function $\varphi_{\mathcal{D}_0}^{(\beta)}$, where $\mathcal{D}_0$ constitutes the subset of the ignorance space, which shall be excluded, while β represents the reliability level of the stochastic procedure.

   The decision is based on a prediction for the set of initial conditions $\mathcal{D}_0$. If the predicted event actually occurs, the exclusion procedures results in a failure and $\mathcal{D}_0$ cannot be excluded.

   However, if the predicted event does not occur, then $\mathcal{D}_0$ can be excluded and the procedure results in a success.

   The graphical representations shows the uncertainty space with the horizontal axes reserved for the deterministic variable and the vertical axes for the random variable. The set $\mathcal{D}_0$ and the predicted event $A_X^{(\beta)}(\mathcal{D}_0)$ form a rectangle in the uncertainty space. The observation of the exclusion experiment is represented by a red line; if it crosses the rectangle representing the prediction, the set $\mathcal{D}_0$ cannot be excluded, otherwise it can.

   In the following some exclusion procedures are illustrated by their graphical representations:

   (a) Pair of variables:

   $$\begin{aligned} X &= \text{number of successes among } n = 65 \text{ trials} \\ D &= \text{success probability} \end{aligned} \tag{136}$$

   Bernoulli Space:

   $$\begin{aligned} \mathcal{D} &= \{p \,|\, 0.50 \leq p \leq 0.80\} \\ \mathcal{X}(\{p\}) &= \{0, 1, 2, \ldots, 65\} \\ \mathcal{P}(\{p\}) &= P_{X|\{p\}} \text{ with } X|\{p\} \sim Bi(65, p) \end{aligned} \tag{137}$$

   Problem:
   The subset $\mathcal{D}_0 = \{p \,|\, 0.60 \leq p \leq 0.70\}$ shall be excluded from the ignorance space $\mathcal{D}$ by means of an exclusion procedure with reliability level $\beta = 0.80$.

   Graphical Representation:
   Let the exclusion process yield the observation $x = 25$, then Figure 2 is produced by the Graphical Laboratory, it displays the following quantities:

   - The ignorance space $\mathcal{D}$ is displayed on the horizontal axes.

- The range of variability $\mathcal{X}(\mathcal{D})$ is displayed on the vertical axes.
- The entire uncertainty space is highlighted in light blue.
- The rectangle $\mathcal{D}_0 \times A_X^{(\beta)}(\mathcal{D}_0)$ is highlighted in dark blue.
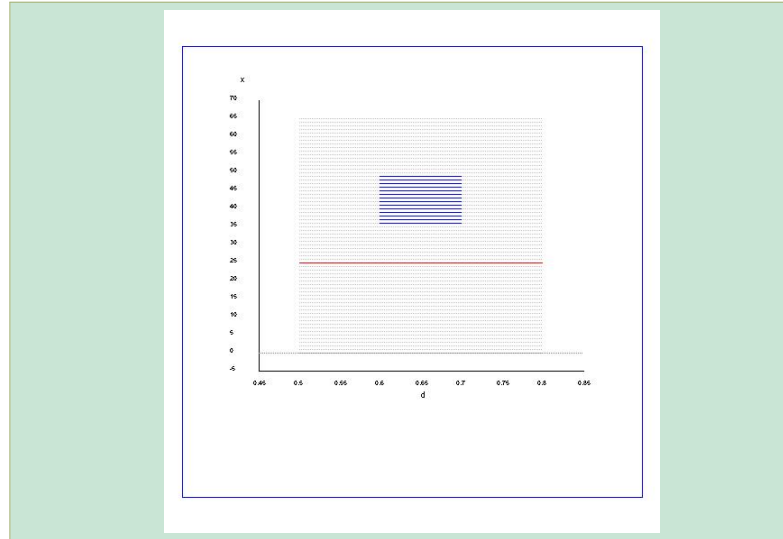- The observation $x = 25$ is represented by the red line.

Figure 2: Graphical representation of an exclusion
procedure and the observation $x = 25$.

The set $\mathcal{D}_0$ is excluded, because the red line does not cross the dark blue area, i. e., the exclusion function adopts the value 1:

$$\varphi_{\mathcal{D}_0}^{(0.80)}(\{25\}) = 1 \tag{138}$$

(b) Let the exclusion experiment yield the observation $x = 40$, then the following graph is produced by the Graphical Laboratory:
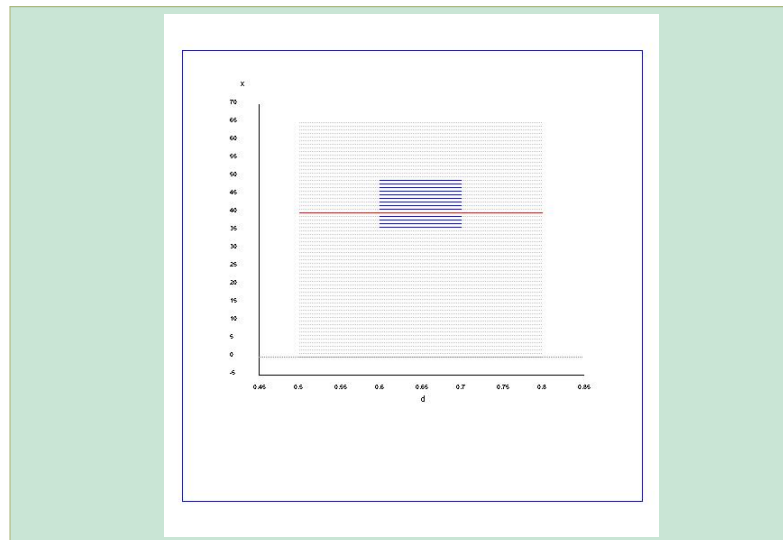
Figure 3: Graphical representation of an exclusion
procedure and the observation $x = 40$.

In this case it is not possible to exclude $\mathcal{D}_0$, because the red line crosses the dark blue area, i. e., the prediction. In other words the prediction $A_X^{(0.80)}(\mathcal{D}_0)$ actually occurred and, thus, $\mathcal{D}_0$ cannot be excluded:

$$\varphi_{\mathcal{D}_0}^{(0.80)}(\{40\}) = 0 \tag{139}$$

Unit 2.4.5: Exclusion Procedure and Significance Test

TARGET

A significance test with significance level $\alpha = 1 - \beta$ and null hypothesis H_0 can be regarded as a special β-exclusion procedure for $H_0 : \mathcal{D}_0$. This learning unit aims at showing the differences between a statistical significance test and a stochastic exclusion procedure. The comparison shall help to understand the two concepts and facilitate their correct application.

CONTENT

Significance Test

The introduction of a statistical significance test in textbooks does in general nor start with stating the problem to be solved, but with lists of the components of the test (null hypothesis, applied statistic, significance level, p-value) and a directive what to do in order to perform the test. The consequence is hat the user often apply the test for solving wrong problems. For example, in many cases s significance test is used to answer the question, which one of two specified alternatives with respect to the initial conditions is the true one. Unfortunately, a significance test i not suitable in this case as the risk of a wrong decision is not controlled and may be extremely large.

The many misunderstandings of users about significance tests reflect the vagueness of the professional description of the essentials of a significance test. In order to clarify the situation the four "criteria" stated in the citation of the *Encyclopedia of Statistics* given in Learning Unit 2.4.1 shall be looked at from the viewpoint of a stochastic exclusion procedure:

(1) The mentioned "suitable hypothesis of chance" represents a subset $\mathcal{D}_0 \subset \mathcal{D}$, with $\mathcal{D}$ of the ignorance space. Calling the set $\mathcal{D}_0$ a "suitable hypothesis of chance" is indeed misleading as there is no indication of the meaning of "suitable." Moreover, the meaning of "chance" is unclear! The set $\mathcal{D}_0$ or any other "hypothesis" is fixed, i.e., not at all subject to randomness or chance. It is also not selected by chance as otherwise it would hardly be called "suitable".

(2) The "test statistic" stands for the observational random variable X, which describes the outcomes of the exclusion process. Why and how X should rank the possible experimental outcomes, is a disputed matter, as any ranking includes some arbitrariness.

(3) The "level of significance" is the equivalent to $(1 - \beta)$, i.e., the complement of the reliability level and, therefore, an upper bound for making a wrong decision to exclude $\mathcal{D}_0$. Unfortunately, the *Encyclopedia of Statistical Sciences* does not explain that a significance test must not be used in order to confirm $\mathcal{D}_0$ as it is often done in practice.

(4) From (4) it can be seen that a significance test is similar to an exclusion procedure, as the results of a significance test are either to reject (exclude) the null hypothesis, i.e., $\mathcal{D}_0$, or not to reject the null hypothesis, i.e., not to exclude $\mathcal{D}_0$.

The only safeguarded aim, which can be reached with a significance test, is to reject the hypothesis, i.e., to exclude a given subset $\mathcal{D}_0$ of the ignorance space $\mathcal{D}$. However, this aim is

hardly mentioned in any statistical textbooks. Therefore, users apply significance tests having in mind different aims and do not know that the obtained results are not safe at all.

Exclusion Procedure Versus Significance Test

Despite the similarity, there are some principle differences between a statistical significance test and a stochastic exclusion procedure.

- One of the striking differences is the fact that nothing is explained clearly with respect to a significance test. The vagueness starts with its name "test," which has hundreds of different meanings. While a β-exclusion procedure is clearly defined by a function meeting certain side conditions and being the result of optimization with respect to the aim, a significance test is explained by means of some construction details.

- In a significance test it is often demanded to formulate two hypotheses, the null hypothesis to be rejected and an alternative hypothesis. This suggest a kind of symmetry between the two hypotheses, which actually does not exist. In contrast, in a stochastic exclusion procedure only the set or hypothesis to be excluded (rejected) as an element of a system of subsets of $\mathcal{D}$ has to be specified, which makes misinterpretations more difficult.

- Significance tests are more or less restricted to hypotheses, which uniquely determine the probability distribution implying that the hypotheses are singletons or can be treated by means of singletons (one-sided case). Stochastic exclusion procedures are applied for any hypothesis. Note that excluding only one single value and admitting the values in any neighborhood of the excluded one makes clearly not much sense.

- Significance tests are generally introduced by means of examples distinguishing between the small sample case and the large sample case. In contrast, there is a general definition of stochastic exclusion procedures which covers all distribution functions and all samples sizes.

- In statistics, there are many approaches for constructing a significance test and defining good or optimal tests. In contrast, there is only one approach for constructing a stochastic exclusion procedure, namely by deriving an appropriate prediction procedure.

- While the optimality criteria used for defining "optimal significance tests" are obscure and hardly connected with the not clearly stated purpose of a significance test, the optimality criterion for a stochastic β-exclusion procedure is given by the probability of a successful application of the test, i. e., of the exclusion of $\mathcal{D}_0$.

When to Use an Exclusion Procedure

Exclusion procedures aim at reducing the ignorance space. Any measurement procedure has the same aim. Comparing exclusion procedure and measurement procedure by means of the degree of reduction of the ignorance space shows that there is always a measurement procedure, which performs at least as well as the exclusion procedure. In other words, using an exclusion procedure (or a significance test) instead of a measurement procedure is tantamount of abandoning information being obtained by the experiment.

The reason for the above is simply the better resolution of a measurement procedure with respect to the set $\mathcal{T}_D(\mathcal{D})$ which is maximum for a measurement procedure. The better resolution yields more precise results and, therefore, a measurement procedure is always better than a exclusion procedure with respect to the achieved precision.

An exclusion procedure might be of advantage, if it is not intended to reveal the true value of the deterministic variable D, but only wants to show that it is not an element of a "critical" set of values.

EXAMPLES

The following example of a significance test is taken from Statistics Tutorial on Significance tests which available in internet (`http://stattrek.com/Lesson5/HypothesisTesting.aspx`). The example shall illustrate the methodology as well as the the terminology used in statistical hypothesis testing. Subsequently, the same example is formulated and solved by means of Bernoulli Stochastics.

1. **Statistical Significance Test**

   The CEO of a large electric utility claims that at least 80 percent of the company's 1,000,000 customers are very satisfied. 100 customers are surveyed using simple random sampling. The result: 73 percent are very satisfied. Based on these results, should we accept or reject the CEO's hypothesis? Assume a significance level of 0.05.

   Solution: *The solution to this problem takes four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results. We work through those steps below:*

   - *State the hypotheses. The first step is to state the null hypothesis and an alternative hypothesis.*

Null hypothesis:	$P \geq 0.80$
Alternative hypothesis:	$P < 0.80$

     Note that these hypotheses constitute a one-tailed test[14]. The null hypothesis will be rejected only if the sample proportion is too small.

   - *Formulate an analysis plan. For this analysis, the significance level is 0.05. The test method, shown, is a one-sample z-test[15].*

   - *Analyze sample data. Using sample data, we calculate the standard deviation (σ) and compute the z-score test statistic (z).*

   $$\sigma = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{0.8 \cdot 0.2}{100}} = \sqrt{0.0016} = 0.04$$

   $$z = \frac{p-P}{\sigma} = \frac{0.73 - 0.80}{0.04} = -1.75$$

   where P is the hypothesized value of population proportion in the null hypothesis, p is the sample proportion, and n is the sample size.

   Since we have a one-tailed test, the P-value is the probability that the z-score is less than -1.75. We use the Normal Distribution Calculator to find $P(z < -1.75) = 0.04$. Thus, the P-value = 0.04.

---

[14]The "one-tailed" situation is deliberately selected here, because the "two-tailed" situation leads to a singleton as null hypothesis, which makes hardly any sense.

[15]The z-test is explained as follows: "A one-sample z-test is used to test whether a population parameter is significantly different from some hypothesized value. The test statistic is a z-score (z) defined by the following equation.

$$z = (\bar{x} - M)/[\sigma/sqrt(n)]$$

where $\bar{x}$ is the observed sample mean, M is the hypothesized population mean (from the null hypothesis), and sσ is the standard deviation of the population."

- *Interpret results. Since the P-value[16] (0.04) is less than the significance level (0.05), we cannot accept the null hypothesis.*

2. **Stochastic Exclusion Procedure**

In order to illustrate the difference between a stochastic exclusion procedure and a statistical significance test, the same question as stated above shall be answered by applying an adequate stochastic exclusion procedure.

The exclusion process consists of drawing by random $n = 100$ customers from the set of $N = 1000000$ customers. The random observational variable is given by:

$$X = \text{number of very satisfied customers in the smaple.}$$

The deterministic variable D is given by:

$$D = \text{proportion of very satisfied customers among all the customers.}$$

Next the Bernoulli Space for the pair of variables (X, D) must be derived. Consider the case that no sure information about the true value of the D is available, except for the fact that there are very satisfied customers as well as there are unsatisfied customers. In this case we obtain:

$$
\begin{aligned}
\mathcal{D} &= \{p \,|\, 0 < p < 1\} \\
\mathcal{X}(\{p\}) &= \{x \in \mathbb{N}_0 \,|\, 0 \leq x \leq n\} \\
\mathcal{P}(\{p\}) &= \{P_{X|\{p\}}\} \text{ with } X|\{p\} \sim H(1000000, p, 100)
\end{aligned}
$$

Since the number of customers N is by far larger than the sample size n, the binomial distribution is a very good approximation of the hypergeometric distribution and we may replace the random structure function given above by:

$$\mathcal{P}(\{p\}) = \{P_{X|\{p\}}\} \text{ with } X|\{p\} \sim Bi(p, 100)$$

Since the CEO aims at proving that more than 80% of the customers are very satisfied, the complement must be excluded. Consequently the subset $\mathcal{D}_0 \subset \mathcal{D}$ of the ignorance space to be excluded is given by:

$$\mathcal{D}_0 = \{p \,|\, p < 0.8\}$$

The reliability level is given by $\beta = 1 - \alpha$. Since the significance level in the example is assumed to be 5%, we obtain $\beta = 0.95$.

After entering the Stochastikon Calculator, case binomial distribution, case exclusion of p, an input mask opens for inserting the corresponding values. Subsequently the Calculator provides a report with the exclusion result, which also contains a graphical illustration of the procedure and the result which is given in Figure 1.

---

[16]The p-value is explained as follows: "A P-value measures the strength of evidence in support of a null hypothesis. Suppose the test statistic in a hypothesis test is equal to S. The P-value is the probability of observing a test statistic as extreme as S, assuming the null hypothesis is true. If the P-value is less than the significance level, we reject the null hypothesis."
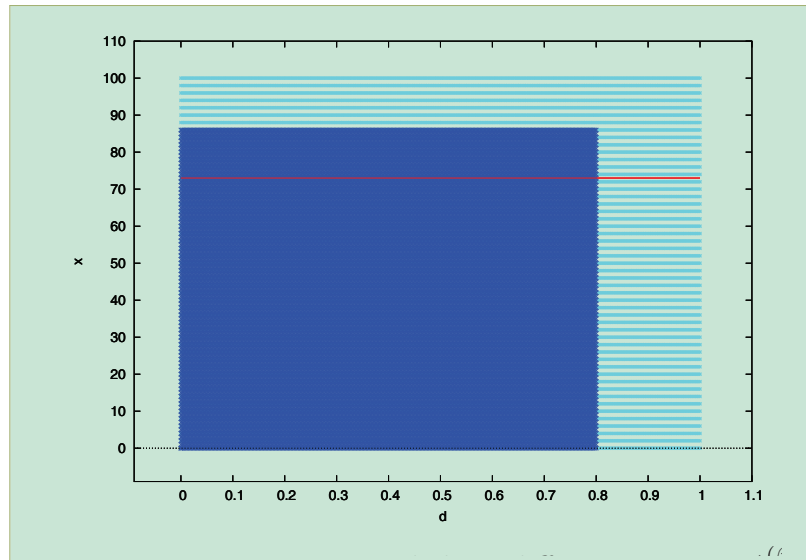
Figure 1: Graphical representation of the exclusion procedure with the dark blue area representing the prediction $A_X^{(0.95)}(\mathcal{D}_0)$ and the red line representing the process outcome.

From Figure 1 it is seen that the observation had been predicted under the condition of $\mathcal{D}_0$ and therefore it can not be excluded and the claim of the CEO has not been confirmed by the sample result.

In the above example, the significance test and the exclusion procedure yield the same result, however, in order to understand the approximations and concepts made in course of the formulation and execution of the statistical significance test, one would need several learning units. Moreover, consider the formulation of the test problem and the test result. Accordingly, not the rejection of the null hypothesis is the aim, but the acceptance of H_0. Although an rejection of the alternative hypothesis is equivalent with an acceptance of H_0, formulating the problem in this way must be confusing since in general the rejection of H_0 is the aim of a significance test.

Module 2.5: Classification Procedures

Content and Aim of the Module Classification Procedures

Consider an operational system subject to a number of different faults. Each fault results in a certain operational state and often the problem is to identify the actual state of the system. The above outlined problem is solved by a stochastic classification procedure allowing to assign or classify a given situation to one of m specified alternatives each modelled by a Bernoulli Space.

The analogous statistical procedure would be an alternative test, which is sometimes called binary classification procedure. In contrast to a stochastic classification procedure an alternative test is restricted to two alternatives.

Classification problems occur for instance

- in medicine, where a patient can have one of several diseases,

- in quality control, where different faults may lead to different quality states of a system.

Classification problems are often dealt with in artificial intelligence aiming at developing so-called *expert system* or *knowledge based systems* which are realized often by decision trees, Bayesian networks, and neural networks. In general these solution methods do not take into account sufficiently well the inherent random variation and are based on subjective belief, rather than on objective reasoning. Consequently, it is generally impossible to make any founded statement on the reliability of the considered procedure.

This module has, similar as the previous modules, two aims. The first one is to introduce and make available the stochastic classification procedure. The second one is to compare a stochastic classification procedure with a statistical alternative tests in order to make both concepts better understood.

Classification problems are of great importance in medical diagnosis, where the concepts *sensitivity* and *specificity* are often used for evaluating the performance of a medical test. These concepts are closely related to the so-called errors of Type I and Type II considered in the framework of significance test. Stochastic classification can be applied not only in the case of binary classification, but also in the often occurring case of several alternative. Moreover, stochastic classification procedures are designed in order to meet specified reliability requirements with respect to the possible decisions, which make stochastic classification procedures superior to the analogous statistical methods.

Unit 2.5.1: Alternatives and Classification

TARGET

In real life there are many situations in which one has to classify an object, a system or a human being according to a number of given alternatives. This learning unit aims at introducing a stochastic procedure called classification procedure which aims at assigning the state of a given situation to one of several alternatives by means of a classification process.

CONTENT

Alternatives

Consider the problem of making a diagnosis about the disease of a patient. Although the possible alternatives are known, they cannot be observed directly.

The actual state can be detected only by means of tests subject to randomness. Thus, in order to determine the actual disease the random process has to be modelled by a Bernoulli Space. Each of the distinct diseases (or alternatives) is represented by different initial conditions and, therefore, each of the states yields a different Bernoulli Space. Assume that m different alternatives with respect to the actual state are possible, then m different Bernoulli Spaces must be considered for the diagnosis test.

Note that the different alternatives refer to a given deterministic variable D and each alternative is represented by a separate ignorance space $\mathcal{D}_i$. In order to make a unique classification meaningful, the sets of values representing the m different alternatives must be disjoint. Otherwise a given situation could belong to several alternatives and a decision for one would be meaningless.

$$\mathcal{D}_i \cap \mathcal{D}_j = \emptyset \quad \text{for } i \neq j$$

The deterministic variable D of a Bernoulli Space is generally represented by the corresponding distributional parameters of P_X. It follows that the deterministic variable D is the same for each of the alternatives, if the random variable X and, hence, the classification process, is the same for each of the alternatives. However, the situations themselves might well be represented by different situation-related deterministic variables.

Bernoulli Space

There is only one random variable X to be observed when performing the classification test in order to decide about the actual alternative. Moreover, let the deterministic variable D be given by the corresponding distributional parameters. Hence the classification process refers independently of the actual alternative to the same pair of variables (X, D).

The m alternatives have different ignorance spaces, i.e. the Bernoulli Space corresponding to the ith alternative, $i = 1, \ldots, m$ is characterized by the ignorance space $\mathcal{D}_i$ and it is denoted by:

$\mathbb{B}^{(i)}_{X,D}$, $i = 1, \ldots, m$. It is given by:

$$\mathbb{B}^{(i)}_{X,D} = (\mathcal{D}_i, \mathcal{X}, \mathcal{P}) \quad \text{for } i = 1, \ldots, m$$

While the ignorance spaces are different for each of the considered alternatives, the variability function $\mathcal{X}$ and the random structure function $\mathcal{P}$ may be the same for each partial Bernoulli Space $\mathbb{B}^{(i)}_{X,D}$.

A classification means to decide in favor of one of the m alternatives, i.e., in determining, which of the m Bernoulli Spaces is the actual one.

The Decision Alternatives

The problem is to select the actual ignorance space out of the m alternatives that are represented by m disjoint sets $\mathcal{D}_i$. Each alternative $\mathcal{D}_i$ is a non-empty and connected set resulting in an overall ignorance space

$$\mathcal{D} = \bigcup_{i=1}^{m} \mathcal{D}_i \tag{140}$$

The classification procedure as considered in this learning unit may lead to one of the following decisions:

- Decision in favor of $\mathcal{D}_1$
- Decision in favor of $\mathcal{D}_2$
- $\cdots$
- Decision in favor of $\mathcal{D}_m$
- Decision in favor of $\mathcal{D}$

The last result means that no decision in favor of one of the alternatives is made. Deciding for $\mathcal{D}_0$ is, of course, always correct, but constitutes a classification failure. A classification failure occurs if the variability of the random variable X is too large in order to decide in a reliable way for one of the alternatives. Note that a classification failure does not constitute a wrong decision and, hence, should not be called an *error*.

Decision Function

Similarly to the case of an exclusion procedure, a classification procedure is given by a decision function φ assigning to each of the observed events $\{x\} \in \mathcal{X}(\mathcal{D})$ an integer between 0 and m. Formally the decision function is defined as follows:

$$\varphi_{\mathcal{D}_1 \mathcal{D}_2 \cdots \mathcal{D}_m} : \left\{ \{x\} \,\middle|\, x \in \bigcup_{i=1}^{m} \mathcal{X}(\mathcal{D}_i) \right\} \rightarrow \{0, 1, 2, \ldots, m\} \tag{141}$$

with the following meaning:

If $\varphi_{\mathcal{D}_1 \mathcal{D}_2 \cdots \mathcal{D}_m}(\{x\}) = i$ such that $i \neq 0$, then the decision is made in favour of $\mathcal{D}_i$, i.e. in favor of the ith alternative, $i = 1, 2, \ldots, m$.

If $\varphi_{\mathcal{D}_1 \mathcal{D}_2 \cdots \mathcal{D}_m}(\{x\}) = 0$, then the decision is made in favour of $\mathcal{D}$.

The aim of a stochastic classification procedure is to assign the given situation to one of the m alternatives. This aim is reached, if the decision function yields an element of $\{1, 2, \ldots, m\}$.

Note the following terminology:

- The application of a classification procedure is successful and correct, if D_i is true, and $\varphi_{\mathcal{D}_1 \mathcal{D}_2 \cdots \mathcal{D}_m}(\{x\}) = i$.

- The application of a classification procedure is successful but wrong, if D_i is true, but $\varphi_{\mathcal{D}_1 \mathcal{D}_2 \cdots \mathcal{D}_m}(\{x\}) = j \notin \{i, 0\}$

- The application of a classification procedure is a failure, if $\varphi_{\mathcal{D}_1 \mathcal{D}_2 \cdots \mathcal{D}_m}(\{x\}) = 0$

A failure of a stochastic classification procedure is not to be looked at as an error. It simply indicates that the uncertainty is too large in order to decide about the true alternative by means of the given classification process.

Classification Procedure and Measurement Procedure

Consider the case that a piece of literature is detected and the circumstance allow only three known writers to be the authors of the text. In this case there are three distinct authors. Let the writing style of the authors be characterized by a random variable X, which might, for instance, represent the number of occurrences of a specific word within one page.

Clearly, the number of occurrences of the specified word within one written page is a random variable for each of the three authors. However, the random pattern, i.e. the probability distribution of using the word are different for the three authors. It follows that the values of the distributional parameter D are different for the three writers.

By means of some works of the different authors and a β measurement procedure the potential values of the deterministic variable D may be determined, leading to three different ignorance spaces and, thus, to three different Bernoulli Spaces representing the three considered authors.

In order to identify the writer of the given text, n pages of the text are checked with respect to the number of aforesaid occurrences and the writer is identified by means of a stochastic classification procedure.

EXAMPLES

1. **Medicine**

   For deciding whether or not a person has a specified disease medical tests are used. The variable of observation X is represented by a random variable, while the initial condition represented by the deterministic variable D are either characterized by the absence or the presence of the disease, i.e., the deterministic variable is given by an indicator variable.

   The resulting Bernoulli Spaces are

$$\mathbb{B}_{X,D}^{(1)} = (\mathcal{D}_1, \mathcal{X}, \mathcal{P})$$
$$\mathbb{B}_{X,D}^{(2)} = (\mathcal{D}_2, \mathcal{X}, \mathcal{P})$$

   where the upper index (1) refers to the absence and (2) to the presence of the considered disease.

2. **Process Control**

   Processes are often monitored in order to detect and subsequently to remove disturbances. Any process disturbance affects some process characteristic, which, therefore, are selected

as observational variable. The disturbance alters process mechanics and, thus, the initial conditions. Similar as in the above example of medical tests two Bernoulli Spaces are obtained, the one representing a fault-free process and the other the process in the presence of a disturbance.

$$\mathbb{B}_{X,D}^{(1)} = (\mathcal{D}_1, \mathcal{X}, \mathcal{P})$$
$$\mathbb{B}_{X,D}^{(2)} = (\mathcal{D}_2, \mathcal{X}, \mathcal{P})$$

3. **Archeology**

Consider an archaeological artifact with unknown origin. Assume that four different possible origins were identified. Each origin leads to a slightly different chemical composition of the material, which is known by past experience. The variable of observation X reflects the chemical composition and the deterministic variable is representative for the origin of the artifact. In this case the result of stochastic modeling are four different Bernoulli Spaces:

$$\mathbb{B}_{X,D}^{(1)} = (\mathcal{D}_1, \mathcal{X}, \mathcal{P})$$
$$\mathbb{B}_{X,D}^{(2)} = (\mathcal{D}_2, \mathcal{X}, \mathcal{P})$$
$$\mathbb{B}_{X,D}^{(3)} = (\mathcal{D}_3, \mathcal{X}, \mathcal{P})$$
$$\mathbb{B}_{X,D}^{(4)} = (\mathcal{D}_4, \mathcal{X}, \mathcal{P})$$

Unit 2.5.2: $\beta_1, \ldots, \beta_m$-Classification Procedure

TARGET

Reliability is a characteristic property of any stochastic procedure. The reliability of a stochastic procedure is given by the probability of obtaining a correct result, when applying the procedure. This learning unit is devoted to the reliability of classification procedures.

CONTENT

Introduction

Consider a situation as introduced in the previous learning unit. There are m different alternatives giving rise to m Bernoulli Spaces, whose ignorance spaces are pairwise disjoint.

$$
\begin{aligned}
\mathbb{B}_{X,D}^{(1)} &= (\mathcal{D}_1, \mathcal{X}, \mathcal{P}) \\
\mathbb{B}_{X,D}^{(2)} &= (\mathcal{D}_2, \mathcal{X}, \mathcal{P}) \\
&\ldots \\
\mathbb{B}_{X,D}^{(m)} &= (\mathcal{D}_m, \mathcal{X}, \mathcal{P})
\end{aligned}
$$

with $\mathcal{D}_i \cap \mathcal{D}_j = \emptyset$ for $i \neq j$ and let

$$
\bigcup_{i=1}^{m} \mathcal{D}_i = \mathcal{D}
$$

A decision in favor of $\mathcal{D}$ is correct with certainty, because the true alternative is always included in $\mathcal{D}$. Deciding for a specific alternative may be incorrect. The reliability of the respective decisions is controlled by specifying for each alternative a required reliability level. Employing the same notation as in the case of exclusion procedures, and for the same reason, the required values of the reliability levels are denoted by $\beta_1, \ldots, \beta_m$. Thus, the value β_i denotes the required lower bound for the probability of a correct decision in the situation when the actual value d is an element of $\mathcal{D}_i$.

Reliability Requirement

Let $\mathcal{D}_i$ be true, then a correct decision means either to decide for $\mathcal{D}_i$ or to decide for $\mathcal{D}$, which is always correct. Thus, the decision function $\varphi_{\mathcal{D}_1 \cdots \mathcal{D}_m}$ meets the reliability requirement given by β_i, if for $i = 1, 2, \ldots, m$, the following holds:

$$
P_{X|\{d\}} \left(\left\{ x \mid \varphi_{\mathcal{D}_1 \cdots \mathcal{D}_m}(\{x\}) \in \{0, i\} \right\} \right) \geq \beta_i \quad \text{for } d \in \mathcal{D}_i \tag{142}
$$

A decision function $\varphi_{\mathcal{D}_1 \cdots \mathcal{D}_m}$ meeting all the requirements given by the reliability levels $(\beta_1, \ldots, \beta_m)$ is denoted by

$$
\varphi_{\mathcal{D}_1 \mathcal{D}_2 \cdots \mathcal{D}_m}^{(\beta_1, \ldots, \beta_m)} \tag{143}
$$

The lower bound β_i for the probability of a correct decision, when the ith alternative is true, is called *reliability level of the ith alternative.*

Let $\mathcal{D}_i$ be the true alternative. Then, the probability that the application of a classification procedure is successful and correct, is given by:

$$P_{X|\{d\}} \left(\left\{ x \mid \varphi_{\mathcal{D}_1 \cdots \mathcal{D}_m}^{(\beta_1, \ldots, \beta_m)}(\{x\}) = i \right\} \right) \tag{144}$$

No general statement about the actual value of the probability (144) can be made. In contrast, the probability that the application of a classification procedure is successful, but wrong, meets the following condition:

$$P_{X|\{d\}} \left(\left\{ x \mid \varphi_{\mathcal{D}_1 \cdots \mathcal{D}_m}^{(\beta_1, \ldots, \beta_m)}(\{x\}) = j \notin \{0, i\} \right\} \right) < 1 - \beta_i \tag{145}$$

The condition (145) means that all risks when applying a stochastic classification procedure $\varphi_{\mathcal{D}_1 \mathcal{D}_2 \cdots \mathcal{D}_m}^{(\beta_1, \ldots, \beta_m)}$ are explicitly known.

Derivation of a Stochastic Classification Procedure

A $(\beta_1, \ldots, \beta_m)$-classification procedure Table 1: The most probable values of $X|\{(100, 0.65)$ are derived by means of a prediction procedure for the outcome of the classification process. In contrast to the prediction procedures introduced in the previous learning units, there is not one reliability level for all predictions, but different reliability levels for different predictions:

$$A_X^{(\beta_1, \ldots, \beta_m)} : \{\mathcal{D}_i \mid i = 1, \ldots, m\} \to \mathcal{T}_X \left(\mathcal{X}(\mathcal{D}) \right) \tag{146}$$

where $\mathcal{T}_X \left(\mathcal{X}(\mathcal{D}) \right)$ denotes an appropriate system of subsets of $\mathcal{X}(\mathcal{D})$.

The prediction procedures $A_X^{(\beta_1, \ldots, \beta_m)}$ yields for each alternative $\mathcal{D}_i$, $i = 1, 2, \ldots, m$, a prediction $A_X^{(\beta_1, \ldots, \beta_m)}(\mathcal{D}_i)$ with

$$P_{X|\{d\}} \left(A_X^{(\beta_1, \ldots, \beta_m)}(\mathcal{D}_i) \right) \geq \beta_i \quad \text{for } d \in \mathcal{D}_i \tag{147}$$

Classification Success and Classification Failure

The procedure is applied successfully, only if it yields one of the non-trivial alternatives $\mathcal{D}_i$, $i = 1, \ldots, m$. As already mentioned, a decision in favor of $\mathcal{D}$, though correct, constitutes a classification failure.

Similar as in the case of a β-exclusion procedure, a necessary condition for a decision in favour of $\mathcal{D}_i$ is that the outcome x of X lies in a β_i-prediction $A_X^{(\beta_1, \ldots, \beta_m)}(\mathcal{D}_i)$. However, if the observed outcome x is an element of two or more of the β_k-predictions, $k = 1, \ldots, m$, then a decision in favour of one of the alternatives meeting the reliability requirements is impossible and the decision must be in favour of $\mathcal{D}$ constituting a classification failure.

Thus, the following two conditions are necessary and sufficient for a decision in favour of $\mathcal{D}_i$:

$$\begin{array}{l} \bullet \quad x \in A_X^{(\beta_1, \ldots, \beta_m)}(\mathcal{D}_i) \\ \bullet \quad x \notin A_X^{(\beta_1, \ldots, \beta_m)}(\mathcal{D}_j) \text{ for } j \neq i \end{array} \tag{148}$$

A classifi-ction success is obtained, if for one $i \in \{1, \ldots, m\}$ the conditions (148) are fulfilled.

A classification failure occurs, if the observed value x is an element of at least two predictions, i. e. there are i and j with:

$$x \in A_X^{(\beta_1, \ldots, \beta_m)}(\mathcal{D}_i) \cap A_X^{(\beta_1, \ldots, \beta_m)}(\mathcal{D}_j)$$

Successful and Correct Classification

Let $A_X^{(\beta_1,\ldots,\beta_m)}(\mathcal{D}_i)$, $i = 1, \ldots, m$, be the β_i-predictions generated for a classification procedure. Note that these predictions need not be disjoint, but may overlap. There are two cases to be considered with respect to the β_i-predictions of a classification procedure $A_X^{(\beta_1,\ldots,\beta_m)}$.

- **Case 1**

In Case 1 the union of all β_i predictions covers the entire range of variability:

$$\bigcup_{i=1}^{m} A_X^{(\beta_1,\ldots,\beta_m)}(\mathcal{D}_i) = \mathcal{X}(\mathcal{D}) \tag{149}$$

Define for $i = 1, \cdots, m$ the following m disjoint sets:

$$\mathfrak{A}_X^{(\beta_1,\ldots,\beta_m)}(\mathcal{D}_i) = A_X^{(\beta_1,\ldots,\beta_m)}(\mathcal{D}_i) \setminus \left(\bigcup_{\substack{j=1 \\ j \neq i}}^{m} A_X^{(\beta_1,\ldots,\beta_m)}(\mathcal{D}_j) \right) \tag{150}$$

Clearly, any $x \in \mathfrak{A}_X^{(\beta_1,\ldots,\beta_m)}(\mathcal{D}_i)$, $i = 1, \ldots, m$, meets the sufficient conditions (148) for a decision in favour of $\mathcal{D}_i$. Therefore, $\mathfrak{A}_X^{(\beta_1,\ldots,\beta_m)}(\mathcal{D}_i)$ is called β_i-*acceptance region for* $\mathcal{D}_i$.

Those outcomes of X which are not contained in one of the acceptance regions are combined to a set denoted $\mathfrak{A}_X(\mathcal{D})$

$$\mathfrak{A}_X(\mathcal{D}) = \mathcal{X}(\mathcal{D}) \setminus \left(\bigcup_{i=1}^{m} \mathfrak{A}_X^{(\beta_1,\ldots,\beta_m)}(\mathcal{D}_i) \right) \tag{151}$$

As any outcome $x \in \mathfrak{A}_X(\mathcal{D})$ leads to a classification failure, this set is called *indifference region*.

Then a $(\beta_1, \ldots, \beta_m)$-classification procedure is defined by the following decision function:

$$\varphi_{\mathcal{D}_1,\mathcal{D}_2\ldots\mathcal{D}_m}^{(\beta_1,\ldots,\beta_m)}(\{x\}) = \begin{cases} i & \text{if } x \in \mathfrak{A}_X^{(\beta_1,\ldots,\beta_m)}(\mathcal{D}_i), \quad i = 1, \cdots, m \\ 0 & \text{if } x \in \mathfrak{A}_X(\mathcal{D}) \end{cases} \tag{152}$$

- **Case 2**

In Case 2, the union of all β_i predictions does not cover the entire variability space:

$$\begin{aligned} \bigcup_{i=1}^{m} A_X^{(\beta_1,\ldots,\beta_m)}(\mathcal{D}_i) &\subset \mathcal{X}(\mathcal{D}) \\ \bigcup_{i=1}^{m} A_X^{(\beta_1,\ldots,\beta_m)}(\mathcal{D}_i) &\neq \mathcal{X}(\mathcal{D}) \end{aligned} \tag{153}$$

which equivalently means

$$\mathcal{X}(\mathcal{D}) \setminus \bigcup_{i=1}^{m} A_X^{(\beta_1,\ldots,\beta_m)}(\mathcal{D}_i) \neq \emptyset \tag{154}$$

In this case the non-empty set (154) may be used for two purposes:

1. The elements of (154) are used for reducing a possible overlap of the β_i-predictions and, thus, for reducing the probability of a classification failure.

2. The elements of (154) are used for increasing the reliability of some of the predictions and, thus, for reducing the probability of wrong decisions.

The first aim is reached by appropriately replacing elements of an overlap by elements of the non-empty set (154). The reliability of β_i-predictions is increased by simply adding elements of (154) to the prediction $A_X^{(\beta_1,\ldots,\beta_m)}(\mathcal{D}_i)$.

Optimal Classification Procedure $A_X^{(\beta_1, \ldots, \beta_m)}$

The aim of a stochastic classification procedure is to assign in a reliable way a given situation to one of several alternatives. The reliability of the procedure is given by the corresponding reliability levels, which constitute side-conditions or specifications. The objective function is given by the probability of reaching the aim, i. e., the probability for a successful classification. The probability depends on the actual initial condition $d \in \mathcal{D}$ and, therefore, the mean success probability is selected as objective function:

$$P_{X|\mathcal{D}} \left(\bigcup_{i=1}^{m} \mathfrak{A}_X^{(\beta_1, \ldots, \beta_m)}(\mathcal{D}_i) \right) = \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} P_{X|\{d\}} \left(\bigcup_{i=1}^{m} \mathfrak{A}_X^{(\beta_1, \ldots, \beta_m)}(\mathcal{D}_i) \right) \qquad (155)$$

A stochastic classification procedure $*A_X^{(\beta_1, \ldots, \beta_m)}$ is called optimal, if for any $A_X^{(\beta_1, \ldots, \beta_m)}$ the following holds:

$$P_{X|\mathcal{D}} \left(\bigcup_{i=1}^{m} *\mathfrak{A}_X^{(\beta_1, \ldots, \beta_m)}(\mathcal{D}_i) \right) \geq P_{X|\mathcal{D}} \left(\bigcup_{i=1}^{m} \mathfrak{A}_X^{(\beta_1, \ldots, \beta_m)}(\mathcal{D}_i) \right) \qquad (156)$$

Clearly, if the β_i-predictions $A_X^{(\beta_1, \ldots, \beta_m)}(\mathcal{D}_i)$, $i = 1, \ldots, m$, are disjoint, then any resulting stochastic $(\beta_1, \ldots, \beta_m)$-classification procedure is optimal.

EXAMPLE

1. **Process Control**

   A production process shall be monitored which produces a certain product with given product specifications $\mathcal{S}$. The monitoring process consists of taking a number of subsequently produced items from the line and checking the specifications. Let $I_\mathcal{S}$ denote the indicator variable for the set $\mathcal{S}$ representing the specifications. Then the monitoring variable is the sum of the indicator variables of the tested items:

   $$X = \text{number of conforming items among the tested items}$$

   The monitoring process is represented by a so-called Bernoulli-chain consisting of a number of independent and identical Bernoulli-Experiments implying that the number X of conforming items is binomially distributed. From this we immediately obtain the deterministic variable D:

   $$\begin{aligned} D \quad &= \quad (D_1, D_2) \\ \text{with} \\ D_1 \quad &= \quad \text{number of tested items} \\ D_2 \quad &= \quad \text{probability } P_{I_\mathcal{S}}(\{1\}) \end{aligned}$$

   The process is monitored in order to detect the occurrence of disturbances which decrease the conforming probability.

   It is known that the conforming probability is larger than $p_\ell = 0.995$, if no disturbance has been occurred. If a disturbance occurs the conforming probability drops at least to the value of $p_u = 0.800$. The two situations are described by the two Bernoulli Spaces:

   $$\begin{aligned} \mathbb{B}_{X,D}^{(1)} &= (\mathcal{D}_1, \mathcal{X}, \mathcal{P}) \\ \mathbb{B}_{X,D}^{(2)} &= (\mathcal{D}_2, \mathcal{X}, \mathcal{P}) \end{aligned}$$

Let the number of items to be tested be known and given by n. Then we get:

$$\mathcal{D}_1 = \{(n,p) \,|\, 0.995 \le p < 1.0\}$$
$$\mathcal{D}_2 = \{(n,p) \,|\, 0 < p \le 0.8\}$$
$$\mathcal{X}(\{(n,p)\}) = \{0,1,\ldots,n\}$$
$$\mathcal{P}(\{(n,p)\}) = P_{X|\{(n,p)\}} \quad \text{with } P_{X|\{(n,p)\}}(\{k\}) = \binom{n}{k}p^k(1-p)^{n-k}$$

There are two alternatives and, therefore, two reliability levels have to be fixed with respect to the two substantial decisions. Assuming that a false alarm is extremely expensive leads to the decision to require the high reliability level of $\beta_1 = 0.99$, while the reliability level for the second alternative is set $\beta_2 = 0.90$ because not immediately detecting a disturbance is considered as less serious compared with a false alarm.

If the number n of tested items is too small, then it is not always possible to obtain a classification success. For instance, selecting $n = 15$ leads to following acceptance regions:

$$\mathfrak{A}_X^{(0.99,0.90)}(\mathcal{D}_1) = \{15\}$$
$$\mathfrak{A}_X^{(0.99,0.90)}(\mathcal{D}_2) = \{0,1,\ldots,13\}$$

where in case of $x = 14$ no decision can be made.

In order to have the probability of a failure to be zero, the number of item has to be increased. The minimum n for which the failure probability becomes zero is $n = 18$ with the following acceptance regions:

$$\mathfrak{A}_X^{(0.99,0.90)}(\mathcal{D}_1) = \{18\}$$
$$\mathfrak{A}_X^{(0.99,0.90)}(\mathcal{D}_2) = \{0,1,\ldots,17\}$$

Thus, the stochastic classification procedure for monitoring the process is given by the following decision function:

$$\varphi_{\mathcal{D}_1,\mathcal{D}_2}^{(0.99,0.90)}(\{x\}) = \begin{cases} 1 & \text{if } x = 18 \\ 2 & \text{if } x \le 17 \end{cases}$$

The acceptance regions can be obtained by means of Stochastikon Calculator.

Unit 2.5.3: Optimum Classification Procedure

> In the preceding learning unit optimum classification procedures were briefly outlined. In this learning unit the objective function for determining optimum $\beta_1, \ldots, \beta_m$-classification procedure is introduced in more detail.

Content

Introduction

The reliability requirements of a $(\beta_1, \ldots, \beta_m)$-classification procedure given by the reliability levels $(\beta_1, \ldots, \beta_m)$ represent side conditions, which have to be met necessarily, as otherwise the result would be more or less meaningless.

Applying a $(\beta_1, \ldots, \beta_m)$-classification procedure is successful, if the decision is made in favour of one of the given alternatives $\mathcal{D}_i$. A failure occurs, if the decision is made in favor of the entire ignorance space $\mathcal{D}$.

Therefore, the smaller the failure probability, the better the performance of a $(\beta_1, \ldots, \beta_m)$-classification procedure.

Objective Function

The procedure's reliability (specified by the reliability levels $\beta_1, \cdots, \beta_m$) constitutes a necessary side condition, which has to be taken into account by the optimization procedure. The only quality characteristic left is the success probability of the $(\beta_1, \ldots, \beta_m)$-classification procedure, or equivalently, its failure probability. Unfortunately, the failure probability depends on the actual value d of the deterministic variable D, which is unknown. Therefore, the mean failure probability is selected as objective function, where the mean is taken over all elements of the entire ignorance space:

$$\mathcal{D} = \bigcup_{i=1}^{m} \mathcal{D}_i \tag{157}$$

Thus, the objective function is given by:

$$P_{X|\mathcal{D}}\left(\left\{x \mid \varphi_{\mathcal{D}_1,\ldots,\mathcal{D}_m}^{(\beta_1,\ldots,\beta_m)}(\{x\}) = 0\right\}\right) =$$
$$\frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} P_{X|\{d\}}\left(\left\{x \mid \varphi_{\mathcal{D}_1,\ldots,\mathcal{D}_m}^{(\beta_1,\ldots,\beta_m)}(\{x\}) = 0\right\}\right) \tag{158}$$

Optimum $(\beta_1, \ldots, \beta_m)$-Classification Procedure

A $(\beta_1, \cdots, \beta_m)$-classification procedure $^*\varphi_{\mathcal{D}_1,\ldots,\mathcal{D}_m}^{(\beta_1,\ldots,\beta_m)}$ given by the acceptance regions

$$\left(^*\mathfrak{A}_X^{(\beta_1)}(\mathcal{D}_1), \ldots, ^*\mathfrak{A}_X^{(\beta_m)}(\mathcal{D}_m), ^*\mathfrak{A}_X(\mathcal{D})\right) \tag{159}$$

is called an optimum $(\beta_1, \ldots, \beta_m)$-classification procedure, if

$$\bullet \quad P_X\left(\left\{x \mid {}^*\varphi_{\mathcal{D}_1,\ldots,\mathcal{D}_m}^{(\beta_1,\ldots,\beta_m)}(\{x\}) = 0\right\}\right) =$$
$$\min_{\varphi_{\mathcal{D}_1,\ldots,\mathcal{D}_m}^{(\beta_1,\ldots,\beta_m)}} P_X\left(\left\{x \mid \varphi_{\mathcal{D}_1,\ldots,\mathcal{D}_m}^{(\beta_1,\ldots,\beta_m)}(\{x\}) = 0\right\}\right) \tag{160}$$

where the minimum is taken over all $(\beta_1, \ldots, \beta_m)$-classification procedure $\varphi_{\mathcal{D}_1,\ldots,\mathcal{D}_m}^{(\beta_1,\ldots,\beta_m)}$ for the alternatives $\mathcal{D}_1, \cdots, \mathcal{D}_m$.

In terms of the prediction procedures an optimum $(\beta_1, \ldots, \beta_m)$-classification procedure is obtained, if the predictions meet the following conditions:

$$\text{completeness:} \quad \bigcup_{i=1}^m A_X^{(\beta_i)}(\mathcal{D}_i) = \mathcal{X}(\mathcal{D}) \tag{161}$$

$$\text{reliability:} \quad P_{X|\{d\}}\left(A_X^{(\beta_i)}(\mathcal{D}_i)\right) \geq \beta_i \quad \text{for } i = 1, \ldots, m \tag{162}$$

$$\text{precision:} \quad \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} P_{X|\{d\}}\left(\bigcup_{i \neq j}\left(A_X^{(\beta_i)}(\mathcal{D}_i) \cap A_X^{(\beta_j)}(\mathcal{D}_j)\right)\right) \overset{!}{=} \min \tag{163}$$

Clearly, any $(\beta_1, \cdots, \beta_m)$-classification procedure $\varphi_{\mathcal{D}_1,\ldots,\mathcal{D}_m}^{(\beta_1,\ldots,\beta_m)}$ with

$$\bigcup_{i \neq j}\left(A_X^{(\beta_i)}(\mathcal{D}_i) \cap A_X^{(\beta_j)}(\mathcal{D}_j)\right) = \emptyset \tag{164}$$

is an optimum one. Moreover, an optimum $(\beta_1, \cdots, \beta_m)$-classification procedure meeting (164) will with certainty lead to a positive decision, because the failure probability is zero. Therefore, an optimum $(\beta_1, \cdots, \beta_m)$-classification procedure meeting (164) may be called an *ideal classification procedure*.

EXAMPLE

1. **Process Control**

   Consider the example of the previous learning unit with:

   $$X = \text{number of conforming items among the tested items}$$

   and

   $$\begin{aligned} D &= (D_1, D_2) \\ &\text{with} \\ D_1 &= \text{number of tested items} \\ D_2 &= \text{probability } P_{I_S}(\{1\}) \end{aligned}$$

   The classification procedure shall distinguish the following two Bernoulli Spaces given by the respective ignorance spaces:

   $$\begin{aligned} \mathcal{D}_1 &= \{(n,p) \mid 0.95 \leq p < 1.0\} \\ \mathcal{D}_2 &= \{(n,p) \mid 0 < p \leq 0.8\} \end{aligned}$$

The reliability levels for the first alternative is set to $\beta_1 = 0.99$ and for the second alternative to $\beta_2 = 0.90$. The classification process consists of $n = 60$ trials. The number of observed successes is given by $x = 52$. The optimum (β_1, β_2)-classification procedure is graphically illustrated in Figure 1.
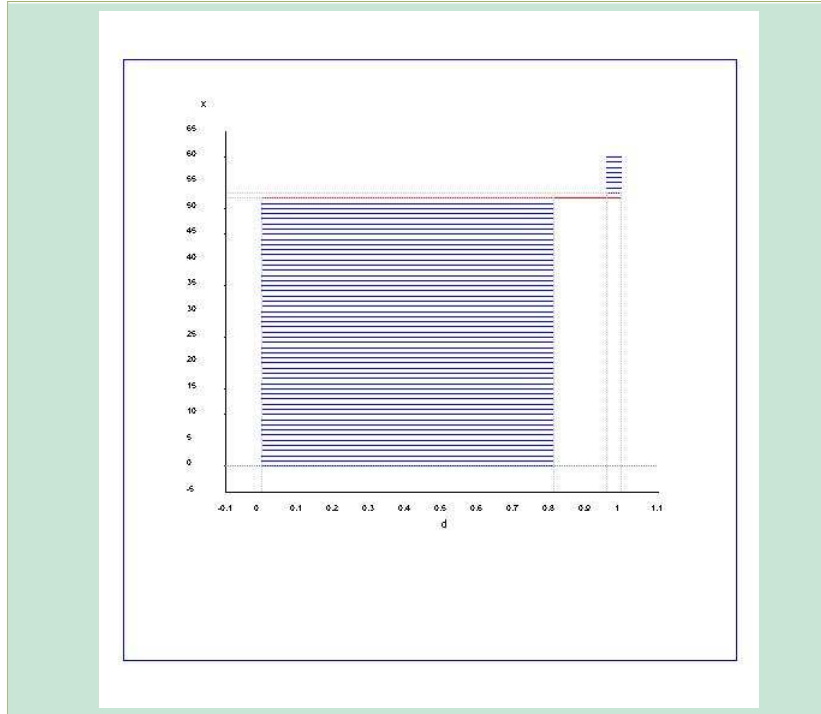
Figure 1: Optimum (β_1, β_2)-classification procedure with
the result $\varphi_{\mathcal{D}_1,\mathcal{D}_2}^{(0.99,0.90)}(\{52\}) = 2$

As can be seen from the graph, the optimum classification procedure is also an ideal one, because a positive decision is made with certainty. The graph shows the two acceptance regions for the first and the second Bernoulli Space and the union of these two regions cover the entire range of variability.

Unit 2.5.4: Two Alternatives

TARGET

In this learning unit the simplest classification procedure is discussed, where the simplicity refers to the number of alternatives and the dimension of the deterministic classification variable. The simplest case is characterized by two alternatives for a univariate variable D.

CONTENT

Introduction

In the case of two alternatives, there are two Bernoulli Spaces for a pair of variables (X, D) differing only with respect to the ignorance space.

$$
\begin{aligned}
\mathbb{B}_{X,D}^{(1)} &= (\mathcal{D}_1, \mathcal{X}, \mathcal{P}) \\
\mathbb{B}_{X,D}^{(2)} &= (\mathcal{D}_2, \mathcal{X}, \mathcal{P})
\end{aligned}
\tag{165}
$$

A stochastic (β_1, β_2)-classification procedure is based on an classification process $(X_1, \ldots, X_n)$ yielding the random variable $X^{(n)} = h(X_1, \ldots, X_n)$ with h selected appropriately.

The deterministic variable is assumed to be one-dimensional and, therefore, the two ignorance spaces can be arranged as follows:

$$
\begin{aligned}
\mathcal{D}_1 &= \{d \,|\, \underline{d}_1 \leq d \leq \overline{d}_1\} \\
\mathcal{D}_2 &= \{d \,|\, \underline{d}_2 \leq d \leq \overline{d}_2\}
\end{aligned}
\tag{166}
$$

with $\overline{d}_1 < \underline{d}_2$.

Predictions

Any stochastic procedure is either a prediction procedure or it is based on a prediction procedure. In case of a classification problem with two alternatives, two prediction procedures are needed, which yield two predictions, one for the first alternative and the other for the second alternative.

In view of the objective function the two predictions should overlap as less as possible. If they are disjoint, then the indifference region is the empty set and, hence, the corresponding (β_1, β_2)-classification procedure is ideal.

The following prediction procedures guarantee obviously the smallest overlap of the two predictions:

$$
A_{X^{(n)}}^{(\beta_1)} : \left\{ \mathcal{D}_1 \right\} \rightarrow \left\{ \{x \,|\, \min \mathcal{X}(\mathcal{D}_1) \leq x \leq \overline{x}_1 \} \right\}
\tag{167}
$$

$$
A_{X^{(n)}}^{(\beta_2)} : \left\{ \mathcal{D}_2 \right\} \rightarrow \left\{ \{x \,|\, \underline{x}_2 \leq x \leq \max \mathcal{X}(\mathcal{D}_2) \} \right\}
\tag{168}
$$

where the non-trivial interval bounds are given by the respective values of the quantile functions:

$$\overline{x}_1 = Q^{(u)}_{X|\{\overline{d}_1\}}(\beta_1)$$

$$\underline{x}_2 = Q^{(\ell)}_{X|\{\underline{d}_2\}}(\beta_2)$$

If $\overline{x}_1 < \underline{x}_2$ the two predictions are disjoint and, hence, it is possible to derive an ideal (β_1, β_2)-classification procedure by selecting a point x_0 with $\overline{x}_1 \leq x_0 \leq \underline{x}_2$ and taking the following acceptance regions:

$$\mathfrak{A}^{(\beta_1)}_{X^{(n)}}(\mathcal{D}_1) = \{x \,|\, \min \mathcal{X}(\mathcal{D}_1) \leq x \leq x_0\} \tag{169}$$

$$\mathfrak{A}^{(\beta_2)}_{X^{(n)}}(\mathcal{D}_2) = \{x \,|\, x_0 < x \leq \max \mathcal{X}(\mathcal{D}_2)\} \tag{170}$$

Optimum (β_1, β_2)-Classification Procedure

If $\overline{x}_1 \geq \underline{x}_2$, then there is no ideal classification procedure. In this case the optimum (β_1, β_2)-classification procedure $^*\varphi^{(\beta_1,\beta_2)}_{\mathcal{D}_1,\mathcal{D}_2}$ for two alternatives is given as follows:

$$^*\varphi^{(\beta_1,\beta_2)}_{\mathcal{D}_1,\mathcal{D}_2}(\{x\}) = \begin{cases} 0 & \text{for} \quad \underline{x}_2 \leq x \leq \overline{x}_1 \\[2mm] 1 & \text{for} \quad x < \underline{x}_2 \\[2mm] 2 & \text{for} \quad x > \overline{x}_1 \end{cases} \tag{171}$$

The (β_1, β_2)-classification procedure given by (171) may lead to a failure with mean failure probability given by:

$$\frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} P_{X^{(n)}|\{d\}} \left(A^{(\beta_1)}_{X^{(n)}}(\mathcal{D}_1) \cap A^{(\beta_2)}_{X^{(n)}}(\mathcal{D}_2) \right) \tag{172}$$

EXAMPLE

1. **Bernoulli Chain**

   Consider the example of a Bernoulli Chain with

   $$X = \text{number of successes among the performed trials}$$

   and

   $$\begin{aligned} D &= (D_1, D_2) \\ \text{with} & \\ D_1 &= \text{number of trials} \\ D_2 &= \text{success probability} \end{aligned}$$

   Consider the case of $n = 50$ trials implying that there is ignorance only with respect to D_2. The classification procedure shall distinguish the following two Bernoulli Spaces given by the corresponding ignorance spaces:

   $$\mathcal{D}_1 = \{p \,|\, 0.2 \leq p < 0.3\}$$
   $$\mathcal{D}_2 = \{p \,|\, 0.4 < p \leq 0.6\}$$

   The classification procedure with reliability levels $\beta_1 = 0.95$ and $\beta_2 = 0.90$ is illustrated in Figure 1.
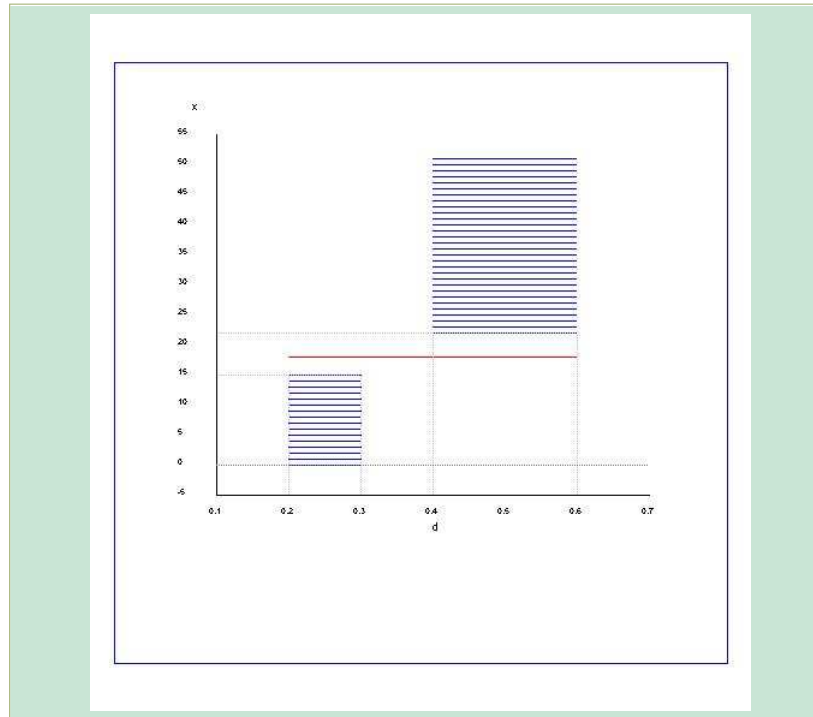
Figure 1: Graphical representation of the two acceptance
regions of a classification procedure with two alternatives
and the classification process result $x = 18$.

Figure 1 may be obtained by the Graphical Laboratory. The graph shows that a positive
decision cannot be made between the two alternatives without violating the reliability
requirements given by the reliability levels. Therefore, no decision is made and applying
the classification procedure was a failure.

Unit 2.5.5: Three Alternatives

Target

This learning unit investigates the case of three alternatives leading to some problems, which are not existing for the case of two alternatives as studied in the previous learning unit.

Content

Introduction

In the case of three alternatives, there are three Bernoulli Spaces for the pair of variables (X, D) differing only with respect to the ignorance space. Let the three Bernoulli Spaces be given by:

$$
\begin{aligned}
\mathbb{B}_{X,D}^{(1)} &= (\mathcal{D}_1, \mathcal{X}, \mathcal{P}) \\
\mathbb{B}_{X,D}^{(2)} &= (\mathcal{D}_2, \mathcal{X}, \mathcal{P}) \\
\mathbb{B}_{X,D}^{(3)} &= (\mathcal{D}_3, \mathcal{X}, \mathcal{P})
\end{aligned}
\tag{173}
$$

The deterministic variable is assumed to be one-dimensional and, therefore, the three ignorance spaces can be arranged as follows:

$$
\begin{aligned}
\mathcal{D}_1 &= \{d \,|\, \underline{d}_1 \le d \le \overline{d}_1\} \\
\mathcal{D}_2 &= \{d \,|\, \underline{d}_2 \le d \le \overline{d}_2\} \\
\mathcal{D}_3 &= \{d \,|\, \underline{d}_3 \le d \le \overline{d}_3\}
\end{aligned}
\tag{174}
$$

with $\overline{d}_1 < \underline{d}_2 \le \overline{d}_2 < \underline{d}_3$.

The stochastic $(\beta_1, \beta_2, \beta_3)$-classification procedure is based on an classification process $(X_1, \ldots, X_n)$ yielding a random variable $X^{(n)} = h(X_1, \ldots, X_n)$ with h selected appropriately.

The Predictions

In case of a classification problem with three alternatives, three prediction procedures with respect to $X^{(n)}$ are needed, yielding three predictions, one for each of the three alternatives.

In view of the objective function the three predictions should overlap as less as possible. If they are disjoint, then the indifference region is the empty set and, hence, there is an ideal $(\beta_1, \beta_2, \beta_3)$-classification procedure.

For deriving an optimum or if possible an ideal $(\beta_1, \beta_2, \beta_3)$-classification procedure the following prediction procedures must be used:

$$
A_{X^{(n)}}^{(\beta_1)} : \left\{ \mathcal{D}_1 \right\} \to \left\{ \{x \,|\, \min \mathcal{X}(\mathcal{D}_1) \le x \le \overline{x}_1 \} \right\}
\tag{175}
$$

$$
A_{X^{(n)}}^{(\beta_3)} : \left\{ \mathcal{D}_3 \right\} \to \left\{ \{x \,|\, \underline{x}_3 \le x \le \max \mathcal{X}(\mathcal{D}_3) \} \right\}
\tag{176}
$$

where

$$\overline{x}_1 = Q^{(u)}_{X^{(n)}|\{\overline{d}_1\}}(\beta_1)$$

$$\underline{x}_3 = Q^{(\ell)}_{X^{(n)}|\{\underline{d}_3\}}(\beta_3)$$

Ideal and Optimum $(\beta_1, \beta_2, \beta_3)$-Classification Procedure

If the following condition is met, then an ideal $(\beta_1, \beta_2, \beta_3)$-classification procedure exists.

- $P_{X^{(n)}|\{d\}}\left(\mathcal{X}(\mathcal{D}) \setminus \left(A^{(\beta_1)}_{X^{(n)}}(\mathcal{D}_1) \cup A^{(\beta_2)}_{X^{(n)}}(\mathcal{D}_3)\right)\right) \geq \beta_2$ (177)
  for any $d \in \mathcal{D}_2$

Condition (177) implies that

$$\overline{x}_1 < \underline{x}_3 \tag{178}$$

and that

$$P_{X^{(n)}|\{d\}}\left(\{x \,|\, \overline{x}_1 < x < \underline{x}_3\}\right) \geq \beta_2 \quad \text{for any } d \in \mathcal{D}_2 \tag{179}$$

An optimum, but not ideal, $(\beta_1, \beta_2, \beta_3)$-classification procedure is obtained, if (177) is not met, by solving the following optimization problem:

Determine

$$A^{(\beta_2)}_{X^{(n)}}(\mathcal{D}_2) = \{x \,|\, \underline{x}_2 \leq x \leq \overline{x}_2\}$$

that

$$P_{X^{(n)}|\{d\}}\left(A^{(\beta_2)}_{X^{(n)}}(\mathcal{D}_2)\right) \geq \beta_2 \quad \text{for } d \in \mathcal{D}_2$$
$$\frac{1}{|\mathcal{D}|}\sum_{d \in \mathcal{D}} P_{X^{(n)}|\{d\}}\left(\{x \,|\, \underline{x}_2 \leq x \leq \overline{x}_1\} \cup \{x \,|\, \underline{x}_3 \leq x \leq \overline{x}_2\}\right) \overset{!}{=} \min \tag{180}$$

EXAMPLE

1. **Bernoulli Chain**

   Consider the example of a Bernoulli Chain with

   $$X = \text{number of successes among the performed trials}$$

   and

   $$\begin{aligned} D \quad &= \quad (D_1, D_2) \\ \text{with} \\ D_1 \quad &= \quad \text{number of trials} \\ D_2 \quad &= \quad \text{success probability} \end{aligned}$$

   The number of trials is given by $n = 100$ and, therefore, there is ignorance only with respect to D_2. The classification procedure shall distinguish the following three Bernoulli Spaces given by the respective ignorance spaces:

   $$\begin{aligned} \mathcal{D}_1 &= \{p \,|\, 0.201 \leq p < 0.251\} \\ \mathcal{D}_2 &= \{p \,|\, 0.4 < p \leq 0.6\} \\ \mathcal{D}_3 &= \{p \,|\, 0.809 < p \leq 0.999\} \end{aligned}$$

   The confidence levels shall be the same for each of the given alternatives with $\beta = 0.90$. The classification procedure and the result for $x = 52$ is illustrated in Figure 1.
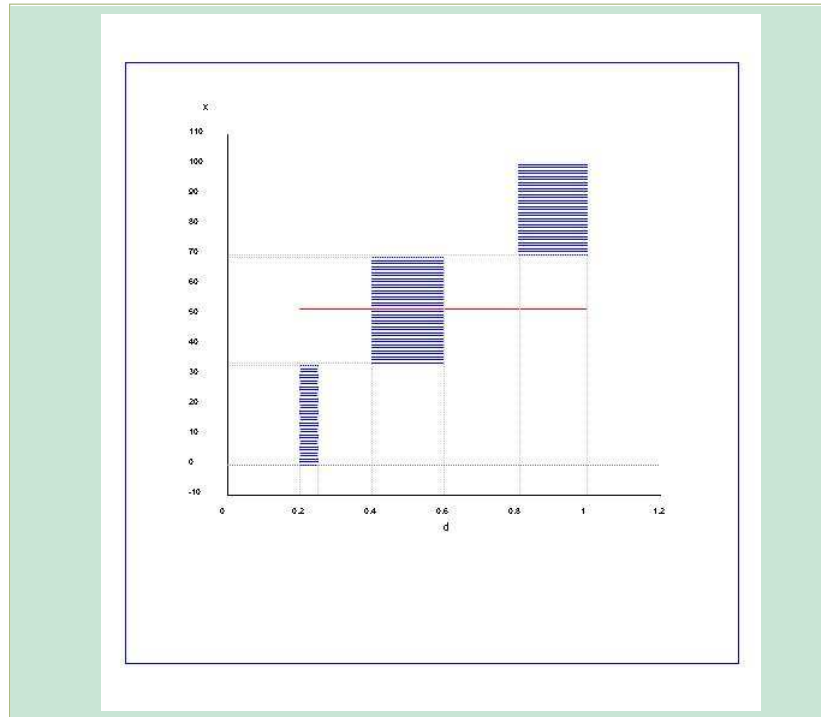
Figure 1: Graphical representation of the classification
procedure with three alternatives

Figure 1 shows the ignorance space of the three alternatives and the corresponding dark blue acceptance regions. The observed result $x = 52$ is represented by the red line which passes through the acceptance regions of the second alternative which therefore is accepted.

Unit 2.5.6: Graphical Representation

TARGET

This learning unit shall support the understanding of a $(\beta_1, \ldots, \beta_m)$-classification procedure by displaying and discussing the corresponding graphical representations.

CONTENT

Introduction

Any stochastic procedure can be reduced to certain predictions, i. e., sets representing future events. Therefore, any stochastic procedure can be represented graphically.

The graphics show the states of ignorance on the one hand and the corresponding predictions on the other. As for classification procedures, the states of ignorance are given by the alternative ignorance spaces $\mathcal{D}_i$, $i = 1, \ldots, m$, while the future is displayed by the corresponding acceptance regions.

Two Alternatives

Consider a similar case than in Learning Unit 2.4.4, where the pair of (X, D) generates a Bernoulli Space based on the binomial distribution, i. e.:

$$
\begin{aligned}
X &= \text{ number of successes} \\
D &= (D_n, D_p) \quad \text{with} \\
&\quad D_n = \text{number of Bernoulli Experiments} \\
&\quad D_p = \text{success probability}
\end{aligned}
$$

with the Bernoulli Spaces $\mathbb{B}_{X,D}^{(1)}$ and $\mathbb{B}_{X,D}^{(2)}$:

$$
\begin{aligned}
\mathbb{B}_{X,D}^{(1)} &= (\mathcal{D}_1, \mathcal{X}, \mathcal{P}) \quad \text{with} \\
&\quad \mathcal{D}_1 = \{(80, p) \,|\, 0.29 \leq p \leq 0.39\} \\
\mathbb{B}_{X,D}^{(2)} &= (\mathcal{D}_2, \mathcal{X}, \mathcal{P}) \quad \text{with} \\
&\quad \mathcal{D}_2 = \{(80, p) \,|\, 0.50 \leq p \leq 0.60\}
\end{aligned}
$$

The graphical representation of the optimum (0.90,0.92)-classification procedure is given in Figure 1.

The figure shows that the procedure is not ideal, because the indifference region is not empty. On the abscissa the two ignorance spaces with respect to the success probability are displayed. Above each ignorance space the corresponding acceptance region is shown. The gap between the two acceptance regions is the indifference region.
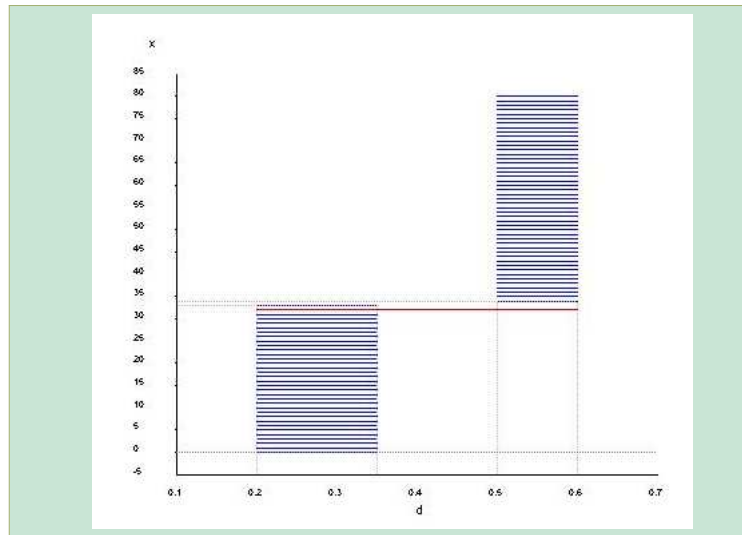
Figure 1: Graphical representation of the optimum
(0.90,0.92)-classification procedure.

The observation $x = 31$ yields a decision in favor of the first alternative since the red line that represent the observation passes through the acceptance region of the first alternative.

Three Alternatives

We consider a similar case as above, namely a Bernoulli Space modeling an experiment consisting of $n = 100$ independent and identical Bernoulli experiments with the following ignorance spaces:

$$\mathcal{D}_1 = \{(100, p) \mid 0.20 \leq p \leq 0.30\}$$
$$\mathcal{D}_2 = \{(100, p) \mid 0.50 \leq p \leq 0.60\}$$
$$\mathcal{D}_6 = \{(100, p) \mid 0.75 \leq p \leq 0.80\}$$

A graphical representation of the optimum (0.90,0.92,0.80)-classification procedure is given by Figure 2.
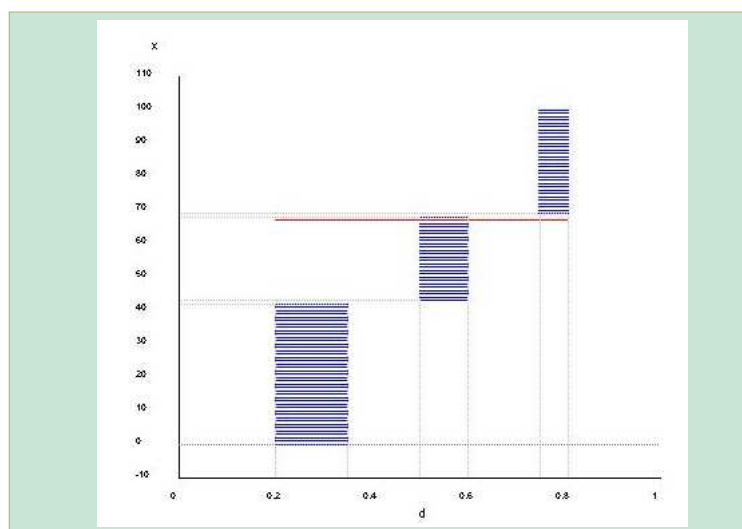
Figure 2: Graphical representation of the optimum
(0.90,0.92,0.80)-classification procedure.

The observation $x = 67$ yields a decision in favor of the second alternative since the red line that represent the observation passes through the acceptance region of the second alternative.

Example

1. **Uncertainty Space and Prediction**

A classification procedure is given by an decision function $\varphi_{\mathcal{D}_1,\ldots,\mathcal{D}_m}^{(\beta_1,\ldots,\beta_m)}$, where each $\mathcal{D}_i$ constitutes an alternative for the ignorance space. The decision function shall decide on the basis of predictions and the observation, which of the alternatives is the true one. The graphical representations shows the overall uncertainty space with the horizontal axes reserved for the deterministic variable and the vertical axes for the random variable. The sets $\mathcal{D}_i$ and the corresponding acceptance events form rectangles in the uncertainty space. The observation of the classification process is represented by a red line; if it passes through one rectangle representing a certain alternative, then this alternative is chosen to be the actual one.

Below, two classification procedures are illustrated by their graphical representations:

(a) Pair of variables:

$$\begin{aligned} X &= \text{number of successes among the 65 trials} \\ D &= \text{success probability} \end{aligned} \tag{181}$$

Two alternatives:

$$\begin{aligned} \mathcal{D}_1 &= \{p \,|\, 0.50 \leq p \leq 0.70\} \\ \mathcal{D}_2 &= \{p \,|\, 0.75 \leq p \leq 0.90\} \end{aligned} \tag{182}$$

Problem:
The actual alternative shall be determined by requiring the reliability levels $\beta_1 = \beta_2 = 0.85$. The number of trials to be performed is given by $n = 45$.

Graphical Representation:
Let the classification process yield the observation $x = 25$, then the following graph is produced by the Graphical Laboratory:
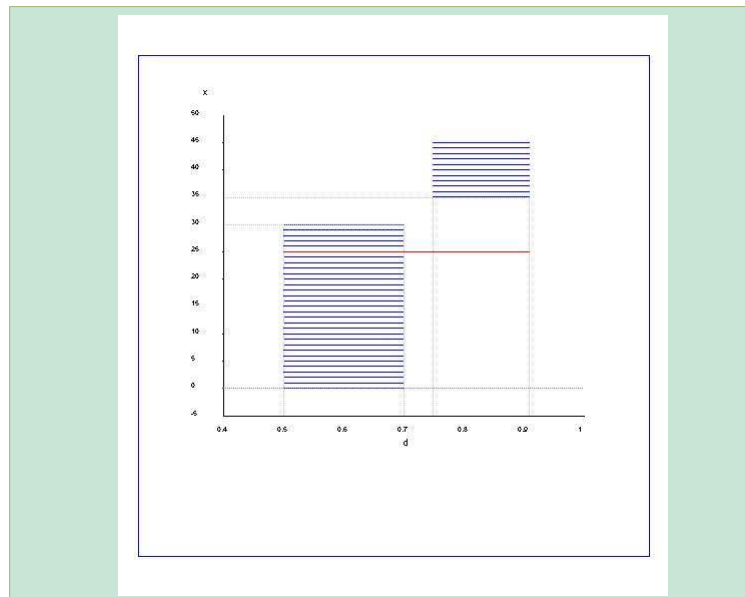
Figure 3: Optimum $(0.85, 0.85)$-classification
procedure for $\mathcal{D}_1$ and $\mathcal{D}_2$.

Figure 3 displays the following quantities:

- The alternative ignorance spaces are displayed on the horizontal axes.
- The range of variability is displayed on the vertical axes.
- The different acceptance regions are highlighted in dark blue.
- The observation $x = 25$ is represented by a red line.

Classification Result:
The red line representing the observation passes through the acceptance region for $\mathcal{D}_1$ and, therefore, the decision is made in favor of the first alternative.

$$\varphi_{\mathcal{D}_1,\mathcal{D}_2}^{(0.80,0.80)}(\{25\}) = 1 \tag{183}$$

(b) For the same example, let the classification process yield now the observation $x = 33$, then the following graph is produced by the Graphical Laboratory:
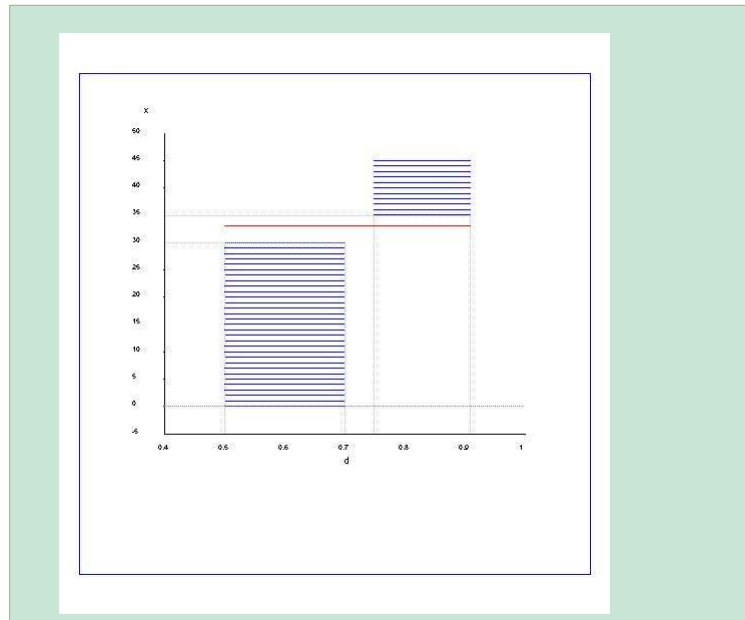
Figure 4: The $(0.85, 0.85)$-classification procedure with $x = 33$.

Classification Result:
In this case the red line does not cross any of the acceptance regions and, therefore, no positive decision can be made:

$$\varphi_{\mathcal{D}_1,\mathcal{D}_2}^{(0.80,0.80)}(\{33\}) = 0 \tag{184}$$

Module 2.6: Verification Procedures

Content and Aim of the Module Verification Procedures

So far the modules aimed at utilizing or improving the Bernoulli Space. This module is devoted to verify the usefulness of a given Bernoulli Space, by comparing the predictions and the corresponding reliabilities made on the basis of the Bernoulli Space with the actually occurring events.

A Bernoulli Space might be wrong, i.e., it does not cover the actual conditions or it might be a too bad approximation. In the first case the predictions will be too often wrong, in the second one the predictions will be of no great use. The verification procedure aims at disclosing the quality of a Bernoulli Space.

Unfortunately, no verification procedures has been completely developed so far. Therefore, this module is only announced here. Hopefully, Module 2.6 will be contained in the next version of Stochastikon Magister.

Bibliography

[1] Bartosyzyński and Niewiadomska-Bugaj, M. (1996): *Probability and Statistical Inference.* Wiley, New York.

[2] Bernoulli, J. (1713): *Ars conjectandi.* Basel.

[3] Bernoulli, J. (2006): *Ars conjectandi together with Letter to a Friend on Sets in Court Tennis.* Translated with an introduction and notes by Edith Dudley Sylla. The John Hopkins University Press, Baltimore.

[4] Boole, G. (1854): *The Laws of Thought.* Macmillan, London.

[5] Box, G. (1997): Scientific Method: The Generation of Knowledge and Quality. *Quality Progress*, January 1997, 47–50.

[6] Byrne, E.F. (1968): *Probability and Opinion. A Study in the Medieval Presuppositions of Post-Medieval Theories of Probability.* Martinus Nijhoff, The Hague.

[7] Cartwright, N. (1983): *How the Laws of Physics Lie.* Claredon Press, New York.

[8] Cartwright, N. (1999): *The Dappled World – A Study of the Boundaries of Science.* University Press, Cambridge.

[9] Carver, R.P. (1978): The case against statistical significance testing. *Harvard Educational Review* 48, 378 – 299.

[10] Collani, E.v. and Dräger, K. (2001): *Binomial Distribution Handbook for Applied Scientists and Engineers.* Birkh¨äuser, Boston.

[11] Collani, E.v. (2002): Theoretical Stochastics. In *Defining the Science of Stochastics*, Ed. E.v.Collani. Heldermann Verlag, Lemgo, 147 – 174.

[12] Collani, E.v. (2002): Empirical Stochastics. In *Defining the Science of Stochastics*, Ed. E.v.Collani. Heldermann Verlag, Lemgo, 175 – 213.

[13] Collani, E.v. and Zhai, X. (2005): *Stochastics.* (In Chinese), Beijing Publisher Group, Beijing.

[14] Collani, E.v. and Baur, K. (2007): *Was zum Teufel ist Qualität?* Heldermann Verlag, Lemgo.

[15] Collani, E.v. (2010): A Note on the XIX IMEKO World Congress Fundamental and Applied Metrology. *EQC* 24, 287 – 307.

[16] Daniel, L.G. (1998): Statistical Significance Testing: A Historical Overview of Misuse and Misinterpretations with Implications for the Editorial Policies of Educational Journals. *Research in Schools* 5, 23-32.

[17] Davies, P. (1992): *The Mind of God: The Scientific Basis for a Rational World.* Simon & Schuster, New York.

[18] Feyerabend, P. (1988): *Against Method: Outline of an Anarchystic Theory of Knowledge.* Verso, London.

[19] Feynman, R. (1965): *The Character of Physical Laws.* MIT Press, Cambridge.

[20] Gnedenko, B.V. (1943): Les extreêmes des distributions statistiques. *Ann. Inst. H. Poincaré* 5, 115–158.

[21] Hausdorff, F. (1921): Summationsmethoden und Momentenfolgen. *Math. Z.* 9, 74-108, 280-299.

[22] Kotz S. and Johnson, N.L. (eds) (1988): *Encyclopedia of Statistical Sciences.* Wiley, New York.

[23] Kolmogorov, A.N. (1956): *Foundadtions of the Theory of Probability.* 2nd ed., translated by N. Morrison, Chelsea, New York.

[24] Levy, H. (19): *The Universe of Science.* Watts & Co., London.

[25] Meehl, P. (1997): The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. In L.L. Harlow, S.A. Mulaik, J.H. Steiger (Eds.), *What if there were no significance tests?*, Erlbaum, Mahwah, NJ, 293 – 426.

[26] Pearson, K. (1900): *The Grammar of Science.* Second edition, A. and C. Black, London.

[27] Polanyi, M. (1964): *Science, Faith and Society.* The University of Chikago Press, Chikago.

[28] Schilling, P. (1991): Chance and Order in Theology. *Theology Today* 47. 1-11.

[29] Schmidt, F. (1996): Statistical significance testing and cumulative knowledge in psychology: Implications for the training of resaerchers. *Psychological Methods* 1, 115 – 129.

[30] Searle, J.R. (1999): The Future of Philosophy. *Phil. Trans. R. Soc.* 354, 2069 – 2080.

[31] Shulman, L.S. (1970): Reconstruction of educational research. *Review of Educational Research* 40, 371 – 393.

[32] Wulff, K. (2006): *Naturwissenschaften im Kulturvergleich. Europa – Islam – China.* Verlag Harry Deutsch, Frankfurt am Main.

Index