

Statistical Failure Prediction with an Account for Prior Information

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

VON

Lennart Kann

WÜRZBURG 2020



Abstract

Prediction intervals are needed in many industrial applications. Frequently in mass production, small subgroups of unknown size with a lifetime behavior differing from the remainder of the population exist. A risk assessment for such a subgroup consists of two steps: i) the estimation of the subgroup size, and ii) the estimation of the lifetime behavior of this subgroup. This thesis covers both steps. An efficient practical method to estimate the size of a subgroup is presented and benchmarked against other methods. A prediction interval procedure which includes prior information in form of a Beta distribution is provided. This scheme is applied to the prediction of binomial and negative binomial counts. The effect of the population size on the prediction of the future number of failures is considered for a Weibull lifetime distribution, whose parameters are estimated from censored field data. Methods to obtain a prediction interval for the future number of failures with unknown sample size are presented. In many applications, failures are reported with a delay. The effects of such a reporting delay on the coverage properties of prediction intervals for the future number of failures are studied. The total failure probability of the two steps can be decomposed as a product probability. One-sided confidence intervals for such a product probability are presented.

Zusammenfassung

Vorhersageintervalle werden in vielen industriellen Anwendungen benötigt. In Massenproduktionen entstehen regelmäßig kleine Untergruppen von unbekannter Größe, welche ein anderes Lebensdauerverhalten als die übrige Population besitzen. Eine Risikoeinschätzung für eine solche Untergruppe besteht aus zwei Schritten: i) der Schätzung der Größe dieser Untergruppe und ii) der Schätzung des Lebensdauerverhaltens dieser Untergruppe. Diese Arbeit behandelt diese beiden Schritte. Eine effiziente Methode zur Schätzung der Größe der Untergruppe wird vorgestellt und mit anderen Methoden verglichen. Vorhersageintervalle unter Vorinformation in Form einer Beta-Verteilung werden dargelegt. Das Schema wird für die Vorhersage binomialer und negativ binomialer Zufallsvariablen angewandt. Der Effekt der Populationsgröße auf die Vorhersage der Anzahl von zukünftigen Ausfällen wird für eine Weibull-Verteilung betrachtet, deren Parameter auf Basis von zensierten Felddaten geschätzt werden. Methoden um Vorhersageintervalle bei unbekannter Populationsgröße zu bestimmen werden dargelegt. In vielen Anwendungen werden Ausfälle mit einem Verzug gemeldet. Die Effekte eines solchen Meldeverzugs auf die Überdeckungseigenschaften von Vorhersageintervallen für die Anzahl an zukünftigen Ausfällen werden untersucht. Die Gesamtausfallwahrscheinlichkeit aus den zwei Schritten kann in eine Produktwahrscheinlichkeit zerlegt werden. Einseitige Konfidenzintervalle für eine solche Produktwahrscheinlichkeit werden dargelegt.

Contents

Abstract	i
Zusammenfassung	ii
List of Figures	vii
List of Tables	x
1 Introduction	1
2 The Test Gate Method	7
2.1 Introduction	7
2.2 Industrial Applications	9
2.2.1 Solder Balls on Electric Control Units	9
2.2.2 Displaced Bond Wires	10
2.3 Technical Quality and Usage Quality	11
2.4 Formalizing the Test Gate Method	13
2.4.1 Proficiency Analysis of the Test Gate	14
2.4.2 Prediction	15
2.4.3 Prerequisites	15
2.5 Stochastic Aspects of the Test Gate Method	16
2.5.1 Basic Stochastic Modelling	16
2.5.2 Uncertainty of Slip Rate	17
2.5.3 Uncertainty due to Dispersion of the Bernoulli Process	19
2.5.4 Test Gate with no Detected Failures	20
2.6 Benchmarking with Other Methods	21
2.6.1 Test Gate Method versus Control Sampling	21
2.6.2 System Analytical Approach to Estimate the Volume Deviating . .	23
2.7 Conclusion	24

3	Discrete Prediction Intervals with Prior Information	25
3.1	Introduction	25
3.2	Minimum Weighted Volume Prediction Spaces	27
3.3	The Beta Prior Model for Inference on Binomially Distributed Variables	30
3.4	Properties of Distributions under the Beta Prior Model for Binomial Counts	32
3.5	Construction of Exact and Approximate MWV Prediction Intervals for Binomial Counts	35
3.6	The Sensitivity of Prediction Spaces in Prior Information for a Binomial Prediction Target	36
3.7	Coverage Properties for the Binomial Case	40
3.8	The Beta Prior Model for Inference on Negative Binomially Distributed Variables	44
3.9	Properties of some Discrete Distributions under the Beta Prior Model	45
3.9.1	Two Negative Binomial Counts	45
3.9.2	Binomial Predictor, Negative Binomial Prediction Target	46
3.10	Construction of Exact Prediction Intervals for a Negative Binomial Prediction Target	48
3.11	The Sensitivity of Prediction Spaces in Prior Information for a Negative Binomial Prediction Target	51
3.12	Coverage Properties for a Negative Binomial Prediction Target	52
3.13	Conclusion	55
	Appendix of Chapter 3	56
3.A1	Proof of Proposition 3.4.1	56
3.A2	Proof of Proposition 3.4.2	57
3.A3	Proof of Proposition 3.4.3	58
3.A4	Proof of Proposition 3.9.1	59
3.A5	Proof of Proposition 3.9.2	60
3.A6	Proof of Proposition 3.9.3	61
3.A7	Proof of Proposition 3.9.4	62
4	Prediction Intervals with Unknown Sample Size	63
4.1	Introduction	63
4.2	Background	64
4.2.1	Prerequisites	64
4.2.2	Model	65

4.2.3	Coverage Concepts	65
4.3	Prediction Intervals with Known Sample Size	67
4.3.1	Naive Prediction Intervals	67
4.3.2	Prediction of the Number of Future Failures	67
4.3.3	Calibration	68
4.4	Prediction Intervals with Unknown Population Size	69
4.4.1	Monotonicity of the ML Estimate for the Exponential Distribution	69
4.4.2	Effect of Population Size	70
4.4.3	Monotonicity Approach	72
4.4.4	Distribution-based Approach	75
4.5	Concluding Remarks and Future Work	77
5	Failure Prediction with Reporting Delay	78
5.1	Introduction	78
5.2	Model	79
5.3	Prediction Intervals	80
5.3.1	Types of Observations	80
5.3.2	Failure Prediction for a Population	81
5.4	Effect of Truncation on Prediction Interval	82
5.5	Approaches	83
5.5.1	Shifting the Censoring Time	84
5.5.2	Adjusting the Number of Failures	85
5.6	Simulation	86
5.7	Results	88
5.8	Conclusion	92
6	Confidence Intervals for a Product of Two Binomial Proportions	94
6.1	Introduction	94
6.2	Problem Description and Stochastic Model	96
6.3	Product of Clopper-Pearson Limits	97
6.4	Upper Confidence Limits under Prior Information based on Inverting a Test of Significance	98
6.5	Obtaining Prior Information	101
6.6	Computation of Confidence Limits under Prior Information	102
6.7	Comparison of Methods	102
6.8	Conclusion	107

Appendix of Chapter 6	109
6.A1 Proof of Proposition 6.4.1	109
6.A1.1 Proof of Assertion (i) of Proposition 6.4.1	109
6.A1.2 Proof of Assertion (ii) of Proposition 6.4.1	109
6.A2 Proof of Theorem 6.4.3	111
6.A3 Proof of Proposition 6.5.1	112
7 Conclusion	113
Bibliography	117

List of Figures

2.1	Exemplary illustration of a production process with a test gate	8
2.2	Solder ball between two pins	9
2.3	ECU with three ASICs	10
2.4	Displaced bond wires on ECU	11
2.5	Coverage for $p_1 = 0.20$ and different n_1, n_2 with nominal confidence level $\gamma = 0.95$	18
3.1	Comparison between the endpoints of exact intervals by algorithm 3.5.1 and approximate intervals by algorithm 3.5.2. $a = 1 = b, \gamma = 0.90$	36
3.2	Pointwise length of intervals A_{x_1} and weighted volumes $V_{X_2 X_1}(A)$ of exact x_1 -pointwise level γ prediction spaces with $\gamma = 0.90$, sample sizes $n_1 = 20, n_2 = 50$	37
3.3	Length of prediction intervals A_{x_1} of level $\gamma = 0.90$ for various observations $X_1 = x_1$ under the beta prior $Beta(1, b)$ with various values b . Sample sizes $n_1 = 20, n_2 = 50$	38
3.4	Weighted volume $V_{X_2 X_1}(A)$ as a function of the parameter b of the beta distribution, $a = 1$ fixed.	39
3.5	100% quantiles of $Beta(1, b)$ as a function of b	39
3.6	Required sample size $n_{1,\min}$ to undercut a prescribed weighted volume of level $\gamma = 0.90$ as a function of b	40
3.7	Pointwise x_1 coverage for the level-90 x_1 -pointwise MWV for the binomial case	41
3.8	Comparison of the coverage probabilities between the Score PI, Nelson's PI, Thatcher's PI and the MWV PI with $a = 1, b = 1$ for different sample sizes.	42
3.9	Coverage probability pointwise in x_1 for misspecified priors under sample sizes $n_1 = 20, n_2 = 100$ and confidence level $\gamma = 0.90$. Priors $Beta(a, b)$ with $a = 1$, parameter b determined by quantile prescription.	42

3.10	Pointwise coverage probability for the approximate MWV prediction interval, prior $Beta(1, 1)$, nominal confidence level $\gamma = 0.90$	43
3.11	Exact MWV prediction intervals, prior $Beta(1, 1)$ (left), $Beta(1, 99)$ (right), nominal confidence level $\gamma = 0.90$ for the binomial-negative binomial case (top) and the negative binomial case (bottom).	50
3.12	Conditional PMFs for prior Beta distribution with parameters $a = 1$, $b = 1$ (blue), and $a = 1$, $b = 99$ (red) for the binomial-negative binomial case (left) and the negative binomial case (right).	50
3.13	Pointwise length of the level-90 x_1 -pointwise MWV for a negative binomial count. Predictor binomial (top) with $n_1 = 20$, $r_2 = 5$ and predictor negative binomial (bottom) $r_1 = 5$, $r_2 = 5$, nominal confidence level $\gamma = 0.90$	51
3.14	Pointwise x_1 coverage of the level-90 x_1 -pointwise MWV for a negative binomial prediction target. Predictor binomial (top) and predictor negative binomial (bottom), nominal confidence level $\gamma = 0.90$	53
3.15	Coverage probability pointwise in x_1 for misspecified priors under sample size $n_1 = 20$, and number of required successes $r_2 = 5$ and confidence level $\gamma = 0.90$. Priors $Beta(a, b)$ with $a = 1$, parameter b determined by quantile prescription.	54
3.16	Coverage probability pointwise in x_1 for misspecified priors under required number of successes $r_1 = 5$, $r_2 = 5$ and confidence level $\gamma = 0.90$. Priors $Beta(a, b)$ with $a = 1$, parameter b determined by quantile prescription.	54
5.1	Coverage versus expected number of failures in $(t_c, t_p]$ with lognormal delay time distribution, $n = 1000$, $E_r = 10$, $E_{\text{trunc}} = 0.10$	89
5.2	Coverage versus expected number of failures in $(t_c, t_p]$ with lognormal delay time distribution, $n = 1000$, $E_r = 10$, $E_{\text{trunc}} = 0.20$	89
5.3	Coverage versus expected number of failures in $(t_c, t_p]$ with lognormal delay time distribution, $n = 1000$, $E_r = 10$, $E_{\text{trunc}} = 0.30$	90
5.4	Coverage versus expected number of failures in $(t_c, t_p]$ with lognormal delay time distribution, $n = 67$, $E_r = 20$, $E_{\text{trunc}} = 0.10$	90
5.5	Coverage versus expected number of failures in $(t_c, t_p]$ with lognormal delay time distribution, $n = 67$, $E_r = 20$, $E_{\text{trunc}} = 0.20$	91
5.6	Coverage versus expected number of failures in $(t_c, t_p]$ with lognormal delay time distribution, $n = 67$, $E_r = 20$, $E_{\text{trunc}} = 0.30$	91

5.7	Coverage versus expected number of failures in $(t_c, t_p]$ with exponential delay time distribution.	92
6.1	Coverage $C_{p_1}(p_1p_2)$ of product C-P (left) and test inversion limit (right), $p_1 = 0.20$ fixed, different n_1, n_2 , under nominal confidence level $\gamma = 0.95$	103
6.2	Coverage $C_{p_1}(p_1p_2)$ of test inversion limit, different prior bounds $p_{i\ max}$, $p_1 = 0.20$ fixed, $n_1 = n_2 = 30$ (top), $p_1 = 0.50$ fixed, $n_1 = n_2 = 50$ (bottom), nominal confidence level $\gamma = 0.95$	104
6.3	Upper confidence limit for q conditional under $\hat{p}_1 = s_1/n_1 = 0.002$ fixed, sample sizes $n_1 = 2n_2$ (left), and $n_1 = 5n_2$ (right).	104
6.4	Upper confidence limit at level $\gamma = 0.95$ under $s_1 = 0 = s_2$ (left) and $s_1 = 3, s_2 = 5$ (right). Prior information from Clopper-Pearson limits with $\gamma_1 = 0.975 = \gamma_0$ (one C-P) and $\gamma_i = \sqrt{0.975}$ (two C-P).	105
6.5	Upper confidence limit at level $\gamma = 0.95$ under $s_1 = 0.1n_1, s_2 = 1$ (upper-left), $s_1 = 1, s_2 = 0.10n_2$ (upper-right) and $s_1 = 0.10n_1, s_2 = 0.10n_2$ (lower-center), prior information from Clopper-Pearson limits with $\gamma_1 = 0.975$ (one C-P) and $\gamma_i = \sqrt{0.975}$ (two C-P).	106
6.6	Upper confidence limit at level $\gamma = 0.95$ under $s_i = 0$ (top) and $s_i = 2$ (bottom), different choices of γ_0 and γ'	107

List of Tables

2.1	Exemplary scheme of 4 characteristics and test gates for characteristics 1, 2, 3 with detection probabilities.	14
5.1	Censored and truncated information constructed from a sample Z_1, \dots, Z_n with ordered values $Z_{(1,n)} \leq \dots \leq Z_{(n,n)}$	81

1 Introduction

Predictions of random quantities are a core task in industrial quality assessments. On a macro-level, a manufacturer is interested in the number of parts which will fail at the customer's end or do not conform to specifications. Modern development and design processes enable designs such that the product failure rates over the designated lifetimes are extremely small, despite increasing product complexity. Similarly, modern production is capable of manufacturing the products according to its specified design with high reliability. Comprehensive process quality control is used to monitor that the production process is running cleanly. Immediate responses are initiated if critical boundary values are breached.

Nevertheless, even the most advanced production processes are unable to have a yield of 100 %. Hence, small subgroups with an unintended deviation which pass by the quality gates are a prevalent problem. The deviation may increase the failure probability within the designated lifetime.

Although these subgroups are very small compared to the overall production volume, depending on the effect of the technical deviation on the lifetime, the number of failures may reach unacceptable levels. Therefore, a risk assessment in this subgroup context involves two steps: i) an estimation of the size of the subgroup and ii) the effect of the technical deviation on the parameters of the lifetime distribution of the part.

Historically, predicting the number of failures has been considered in the context of warranty claims of the total population. Due to the high production numbers in modern mass production, even small failure rates can accumulate to a sizeable number of failures over the designated lifetime. The most prevalent prediction is a point prediction, the expected number of failures. Point predictions do not give any indications on the underlying uncertainty of the prediction. This shortcoming can be compensated with an interval forecast. For forecasting, this interval is termed prediction interval. The target of a prediction interval is to contain the quantity of interest in $\gamma \times 100$ % of the cases.

Another commonly encountered interval estimation is a confidence interval. A con-

confidence interval is used to provide an estimate for an unknown distribution parameter. Confidence intervals have been studied extensively, whereas prediction intervals have received relatively little attention, as has been noted by Patel (1989) and Wang (2010), among others. The preference for confidence intervals has even prevailed in industrial quality reporting, where oftentimes ppm (parts per million) numbers are reported instead of the expected number of failures. This is somewhat surprising, as prediction intervals have favorable characteristics to practitioners: i) they are easier to interpret than intervals estimates for some distribution parameters, ii) the prediction of a random quantity is often the primary interest, whereas the estimation of a distribution parameter is only a mean to perform a prediction and iii) it is empirically possible to measure the realization of the random quantity, whereas it not possible to observe a distribution parameter.

While confidence and prediction intervals have different objectives, they share the same theoretical basis. An interval, or in higher dimensions a region, should cover the yet unknown quantity of interest with a probability of at least γ . γ is called the nominal confidence level. The interval is the result from a procedure which uses the result of an observable sample and the nominal level. There is one major conflict for all interval procedures, namely precision vs reliability. The longer the resulting interval, the lower the precision and the higher the reliability. The precision is a decreasing function of the interval length, whereas the reliability is an increasing function of the interval length. The reliability is measured by the coverage probability.

Considerable research has been done to predict the number of failures in a warranty claim context. Oftentimes, repairable systems are considered and modelled by a Poisson process, see for example Kalbfleisch & Lawless (1988). Regression analysis has been considered to allow for estimates of influences on the lifetime of products.

Using field data to estimate the parameters of the lifetime distribution poses several challenges due to the complicated nature of this data: Field data is heavily censored, as only the exact lifetimes of the failures are known. The remaining population has censored lifetimes, meaning that they are only known to have lifetimes larger than the censoring time. Furthermore, due to continuous production and delivery times, the units are not taken into service at the same time. This is called staggered entry (Escobar & Meeker (1999)). Another important aspect to consider when dealing with field data is the reporting process. First, due to logistic and geographic reasons, there is a delay between the occurrence of the failure and the reporting of it. This delay can amount to

several months. Second, it is common that not all failures are reported, but, for example, only for certain markets. The reporting process leads to a truncation of the data. Some failures cannot be observed, if they do not fulfill some criteria.

Some research has been done in the realm of complicated field data. Escobar & Meeker (1999) and Hong et al. (2009) considered cases with staggered entry, and product retirement, where a product is potentially not used anymore before it fails. Kalbfleisch & Lawless (1992) made suggestions on how to deal with reporting delays.

The additional pitfall when considering failure prediction in the subgroup context instead of warranty analysis is the uncertainty of the population size.

This thesis covers several aspects of failure prediction with focus on the subgroup context. The first two chapters consider the estimation of the population size. From a practical point of view, the first valid question to be asked is what possibilities exist to estimate the size of the subgroup, when most of the parts are already shipped. The customary approach to this estimation is sampling. However, sampling methods have considerable flaws such as dealing with clusters and prohibitive sample size requirements for small subgroups. As an alternative, a powerful practical approach to estimate the number of parts affected by a technical deviation is presented in chapter 2. The test gate method is regularly applicable, as it uses existing screening methods designated for a different purpose within a production line. Large parts of this chapter have been accepted for publication by the journal *Quality Engineering* (Kann et al. (2018a)).

Chapter 3 deals with discrete prediction intervals under prior information. It is well-known that exact methods for discrete intervals are often too conservative, i.e. their coverage probability exceeds the desired nominal confidence level significantly. This results in unnecessarily long intervals, especially for small underlying probability parameters, which are common in many industrial settings. In many applications, there is prior information available. Chapter 3 presents an approach to use prior information encoded in a beta distribution to shorten the resulting prediction interval. These prediction intervals are of minimum weighted volume, while having pointwise x_1 coverage of at least the nominal level. This scheme is applied to three cases: the observed X_1 being binomial and the prediction target X_2 being either binomial or negative binomial; and the observed X_1 and the prediction target X_2 both being negative binomial.

The influence of the subgroup size is considered in chapter 4. To obtain a prediction interval for the number of failures based on the number of failures at some censoring time,

the procedure from Escobar & Meeker (1999) is adapted to the setting of unknown sample size. Furthermore, by using a population size interval with a probabilistic structure as in chapter 3, a method to shorten the prediction interval for the number of future failures is presented. Parts of this chapter have been presented at the *Reliability and Maintenance Symposium* 2018 in Reno, Nevada and will be published in the proceedings (Kann et al. (2018b)).

The influence of reporting delays on the coverage probability is considered in chapter 5. It is known that ignoring the reporting delay leads to an underestimation of the number of failures (Kalbfleisch & Robinson (1991)). The setting from Meng & Meeker (2011) is adapted with a lognormal delay time distribution as well as an exponential delay time distribution. Methods to adjust for the reporting delay are studied with respect to their coverage probabilities. The majority of this chapter has been accepted for publication by the journal *Quality and Reliability Engineering International* (Kann et al. (2018c)).

A different perspective of the failure prediction problem is taken on in chapter 6. A one-sided confidence interval procedure for a product probability $q = p_1 p_2$ is presented. This can be considered as the total proportion failing, where p_1 is the proportion of the subgroup within in the total population and p_2 is the failure probability over some designated lifetime, given a part carries the deviation. The setting is such that no direct inference on q is possible.

The thesis concludes with a summary and comments on future work in chapter 7.

Terminology

Throughout this thesis, $\mathbb{E}[\cdot]$ denotes the expectation, $V[\cdot]$ the variance and $\mathbb{1}$ denotes the indicator function. The probability density function is abbreviated by PDF and is denoted by a lowercase letter (e.g. f). The probability mass function is abbreviated by PMF and is also denoted by a lowercase letter. The cumulative distribution function is abbreviated CDF and denoted by an uppercase letter (e.g. F). \mathbb{P} denotes some probability measure.

Statistical Distributions

A continuous random variable X has the two-parametric *Beta distribution* $Beta(a, b)$ with shape parameters $a, b > 0$ if it has the PDF

$$f_X(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} & \text{for } x \in (0, 1), \\ 0 & \text{otherwise,} \end{cases}$$

where

$$B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$$

is the Beta function.

A discrete random variable X has the *binomial distribution* $Bin(n, p)$ with sample size $n \in \mathbb{N}$ and probability parameter $p \in [0, 1]$ if it has the PMF

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, \dots, n.$$

A continuous random variable has the *exponential distribution* $Exp(\lambda)$ with parameter $\lambda > 0$ if it has the pdf

$$f_X(x) = \lambda \exp(-\lambda x) \quad \text{for } x \geq 0.$$

A continuous random variable has the *lognormal distribution* $logn(\mu, \sigma)$ with parameters $\mu \in \mathbb{R}, \sigma > 0$ if it has the PDF

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right) \quad \text{for } x > 0.$$

A discrete random variable X has the *negative binomial distribution* $negBin(r, p)$ with $r \in \mathbb{N}$ and probability parameter $p \in (0, 1)$ if it has the PMF

$$f_X(x) = \binom{x+r-1}{x} p^r (1-p)^x \quad \text{for } x = 0, 1, 2, \dots$$

A continuous random variable has the *uniform distribution* $U(a, b)$ with support $[a, b] \subseteq \mathbb{R}$ if it has the PDF

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b], \\ 0 & \text{otherwise.} \end{cases}$$

A continuous random variable has the *Weibull distribution* $Wbl(\eta, b)$ with scale parameter $\eta > 0$ and shape parameter $b > 0$ if it has the PDF

$$f_X(x) = \frac{b}{\eta} \left(\frac{x}{\eta}\right)^{b-1} \exp\left(-\left(\frac{x}{\eta}\right)^b\right) \quad \text{for } x \geq 0.$$

The *exponential* distribution is a special case of the Weibull distribution for $b = 1$.

2 The Test Gate Method

2.1 Introduction

In industrial mass production, high quality in the production is ensured by product release and production process control. If nonetheless a deviation has been found in the product during production or within a complaint process, analyzing the properties of the already produced and released volume is important for several reasons.

The effects of the deviation on the lifetime behavior of the product have to be evaluated. The deviation might lead to a malfunction of some products in the field before a desired minimum lifetime has been reached. The crucial quantity of interest is the number of products which both carry the deviation and have been delivered to the customer. This number limits the maximum number of failures due to the technical deviation and thus, should be the first step of any risk assessment.¹ Typically, the number of products carrying the deviation is very small. The second risk assessment step is then to evaluate the effect of the deviation on the lifetime of the affected parts. The industrial statistics literature focuses on the second risk assessment step. The major topics of interest are the choice of an appropriate lifetime distribution for failures times (Chahkandi & Ganjali (2009), Silva et al. (2013) and the references therein) and the prediction of the future number of failures from a known homogeneous group based on a certain observation (see for example Krishnamoorthy & Peng (2011), Wang (2008), Meeker & Escobar (1998) and the references therein). The problem of assessing the unknown size of a subgroup with deviation by mathematical methods has not been studied in detail.

In modern high volume mass production (with production volumes exceeding 1000 products per day), typically only a small proportion of the total volume is affected by the deviation. Once the produced volume has been delivered to customers, few information on the size of the subgroup with deviation is available for assessment. These

¹We do not assume that the parts in the subgroup will necessarily fail over lifetime. We use the term volume with deviation or subgroup with deviation. The term defective subgroup is often used to describe the phenomenon that a small group of a large volume is quickly failing.

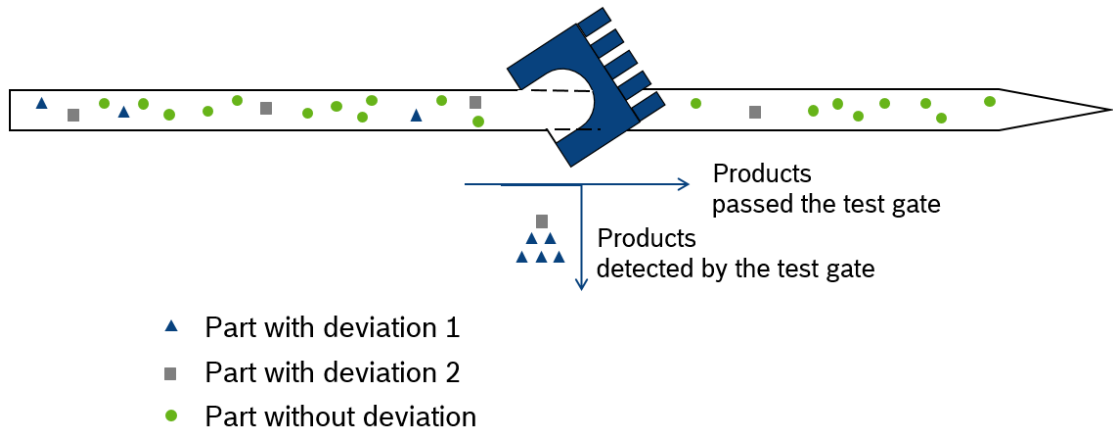


Figure 2.1: Exemplary illustration of a production process with a test gate

available pieces of information often include production volume still in stock, claims by customers and imperfect internal detection mechanisms which were not prioritized until the problem occurred on a bigger scale, e.g. a customer noticing parts with deviation. Control sampling is a well-known method to extract quality information from the volume still in stock. Based on the results obtained for a sample of parts in stock, the number of parts with deviation in the field is estimated. However, this approach has its limitations, especially in high quality production processes where most quality issues are in the parts per million (ppm) range. Reliable results in the ppm region would require enormous sample sizes which are unfeasible in industrial practice. Methods are required which can be used in contexts where control sampling is not suitable.

As an alternative to customary control sampling, we suggest the *test gate method*. This approach has been used by experienced practitioners in the past under very specific circumstances. We generalize the approach for a wider range of applications. The present chapter serves to describe the basic features of the test gate method, and to establish a statistical model, yet without stepping into details of statistical inference.

The subsequent study is structured as follows: In section 2.2 we describe two industrial applications where the efficiency of the test gate method is shown. Section 2.3 formally establishes the notion of volume with deviation and the difference between technical and usage quality. In section 2.4 we describe the requirements and introduce the mathematical description of the test gate method. Section 2.5 discusses stochastic aspects of the test gate method. We then benchmark the test gate method against well-established methods in section 2.6. The benchmarked methods are: control sampling of still existing

stock and system analytical approaches which use data from the manufacturing system. Section 2.7 summarizes the results and discusses open problems and extensions.

2.2 Industrial Applications

2.2.1 Solder Balls on Electric Control Units

Soldering is a standard process to assemble parts onto printed circuit boards (PCBs) in electronic mass production. To join the parts and the PCB, a solder is placed into the joint. During the process, the solder material has to be melted to create a permanent connection between the PCB and the part. Under some circumstances, a solder ball may form during the process on the PCB. One such solder ball is depicted in figure 2.2. With optimally adjusted soldering process parameters, almost no solder balls occur. Generally, solder balls cannot be avoided completely.

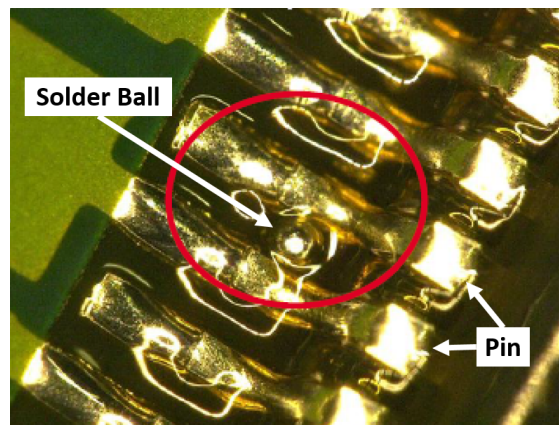


Figure 2.2: Solder ball between two pins

A solder ball between two pins might lead to a short, causing for example a highly increased consumption current of the ECU. As a consequence, the ECU cannot communicate with the diagnosis tool of the car, leading to a warning lamp.

Solder balls can be electronically active (short) or non-active. Active solder balls are typically detected at in-circuit testing (ICT) or end of line (EOL) testing. Non-active solder balls cannot be detected in that way, therefore an automated optical inspection (AOI) is additionally in place. During AOI, a camera scans the search domain and marks potential abnormalities. An operator then looks at the images and tags provided by the AOI for optical abnormalities on the inspected unit. The operator decides whether a unit

is proceeded or scrapped. The search domain of the AOI is defined by expert knowledge. Based on this prior knowledge, the scanned areas are known to cover almost 100 % of solder ball occurrences.

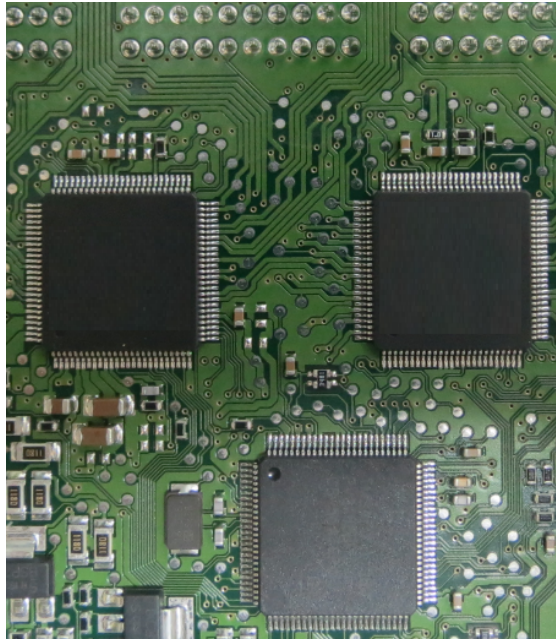


Figure 2.3: ECU with three ASICs

In figure 2.3, you can see an ECU with three ASICs (application-specific integrated circuit). Due to the relative positioning of the ASICs to the parts subjected to the soldering process, only the two closest of the three ASICs were controlled via AOI. After receiving a claim with a solder ball in area three, the AOI was immediately extended to all areas as a corrective action. The slip before the corrective action had to be estimated. After the containment, the ratio of occurrences in the area of ASIC one and two to the occurrences in the area of ASIC three can be measured. Using this ratio it was possible to assess the number of slipped parts which were already sent to field before the corrective action.

2.2.2 Displaced Bond Wires

Gelling of sensor elements is used as a protection against shorts and free moving particles. The gelling process consists of a needle movement to a specified location and an injection of gel. The trajectory of the needle has to be adjusted such that the needle does not

touch any part of the product.

In the here described case, the trajectory of the needle was insufficiently adjusted. Therefore, the needle sporadically touched and bent some bond wires during movement on its trajectory (see figure 2.4). A displacement of one bond wire reduces the distance



Figure 2.4: Displaced bond wires on ECU

between this wire and one of its neighbors.² A reduced distance can lead to a short between the affected bond wires over lifetime.

If this distance is smaller than a threshold an electronic detection at end of line (EOL) measurement is possible. The slip is estimated by the ratio of displaced bonds with a distance below the threshold to the number of displaced bonds with a distance above the threshold for detection. To estimate this ratio, a measurement of affected blocked parts was performed. The result was then used to quantify the number of parts already in the field.

2.3 Technical Quality and Usage Quality

A risk assessment is based on the relation between *technical quality* as considered in the manufacturing environment and *usage quality* perceived by the customer in the field. The test gate method is used to estimate the size of a subgroup with a technical deviation in a specific quality characteristic.

Definition 2.3.1 (technical quality, technical deviation). *The technical quality of a unit i of a certain type of product at time t is determined by the technical quality vector $x_i(t) = (x_{i1}(t), \dots, x_{im}(t))^T$ of the levels $x_{i1}(t), \dots, x_{im}(t)$ of univariate quality characteristics $1, \dots, m$ with corresponding technical target ζ_1, \dots, ζ_m assembled in the*

²Due to aging (thermomechanical impact or vibration), the distance between two neighboring bond wires may reduce over lifetime.

technical target vector $\zeta = (\zeta_1, \dots, \zeta_m)^\top$. Let $z_i(t) = (|x_{i1}(t) - \zeta_1|, \dots, |x_{im}(t) - \zeta_m|)^\top$ be the vector of deviations from target. The level of each quality characteristic j is evaluated by comparing the deviation $z_{i,j}(t)$ from target with the technical nominal condition interval R_j . The unit of product is technically conforming at time t iff the deviation vector $z_i(t)$ lies in the multivariate technical nominal condition range $R = R_1 \times \dots \times R_m$. Otherwise the unit is technically nonconforming at time t or technically deviating at time t .

For a set $\{1, \dots, n\}$ of product units let

$$S_{dev,\ell}(t) = |\{1 \leq i \leq n | z_{i\ell}(t) \notin R_\ell\}|, \quad (2.1)$$

be the volume of parts deviating in the particular characteristic ℓ , and let

$$S_{dev}(t) = |\{1 \leq i \leq n | z_i(t) \notin R\}| \quad (2.2)$$

be the volume of parts deviating in at least one of the characteristics $1, \dots, m$.

Example 2.3.2 (technical quality, solder balls). Consider the case of the solder ball in section 2.2.1. In this example, there is a binary characteristic with range $\{0, 1\}$. The pins are supposed to be electrically decoupled, i.e. have value 1. The solder ball leads to an electric coupling of two pins, and thus this characteristic has value 0. Therefore, if the value is 0, the unit is deviating.

Example 2.3.3 (technical quality, bond wires). In section 2.2.2, the technical characteristic of unit i is the distance $x_i(t)$ between the bond wires. The target is a sufficiently large distance value ζ . In this case, only downside deviations $x_i(t) \leq \zeta$ are relevant. The unit is technically deviating at time t if the downside deviation $\zeta - x_i(t) = |x_i(t) - \zeta|$ exceeds a specified threshold ι .

Technical product quality is an issue in the control of the manufacturing process, but not congruent with usage quality.

In particular, a unit being technically deviating is not equivalent with the unit being nonconforming from the customer's point of view. Subsequently, we define usage quality in a reliability context.

Definition 2.3.4 (usage quality, usage nonconforming). The usage quality characteristic of a unit i is its lifetime T_i . The unit i is usage nonconforming at time t iff $T_i < t$. The stochastic usage quality indicator for a population of units with identical lifetime CDF F_T is the usage proportion nonconforming at time t $p_t = F_T(t)$.

The levels of technical quality characteristics affect usage quality. In the present context, the usage proportion nonconforming is to be considered as a function $p_{t_{LT}}(z)$ of the technical deviation vector z . The functional relationship can be specified via a proportional hazards regression model

$$\lambda(s, z(s)) = \lambda_0(s) \exp(\psi(z(s))). \quad (2.3)$$

ψ is a regression function satisfying $\psi(0) = 0$ for the case when all technical quality characteristics are perfectly on target, $\lambda_0(s) = \lambda(s, 0)$ is the corresponding baseline non-negative hazard function. For an increasing link function ψ a deviation of a technical characteristic from target increases the hazard, and thus the usage proportion nonconforming by virtue of the relation

$$p_t(z(t)) = F_T(t, z(t)) = 1 - \exp\left(-\int_0^t \lambda(s, z(s)) ds\right). \quad (2.4)$$

In manufacturing design, the technical nominal condition interval R_j for each technical characteristic $j = 1, \dots, m$ is chosen in a way such that deviations $z \in R$ have negligible effect on the failure probability. This understanding can be expressed by the *R-insensitive* regression function

$$\psi(z) = \beta^\top z \mathbb{1}_R(z) = \beta_1 z_1 \mathbb{1}_{R_1}(z_1) + \dots + \beta_m z_m \mathbb{1}_{R_m}(z_m). \quad (2.5)$$

The classical model established by Cox (1972) considers the linear regression function $\psi(z(s)) = \beta^\top z(s)$.

2.4 Formalizing the Test Gate Method

This section formalizes the setting and defines the requirements for the test gate method.

Definition 2.4.1 (test gate). *A test gate for the quality characteristic j is a mechanism acting on units i , providing states $D_{ij} = 1$ (deviation signal for unit i) and $D_{ij} = 0$ (no deviation signal for unit i) such that $p_{ij} = P(D_{ij} = 1 | z_{ij} \notin R_j) > 0$ for the correct-positive probability p_{ij} . The false-negative probability $1 - p_{ij} = P(D_{ij} = 0 | z_{ij} \notin R_j)$ is called slip rate. A test gate is called perfect for characteristic j iff $p_{ij} = P(D_{ij} = 1 | z_{ij} \notin R_j) = 1$, or equivalently, iff the slip rate equals $1 - p_{ij} = 0$. For a unit i ,*

$$D_i = 1 - \prod_{\ell=1}^m (1 - D_{i\ell})$$

denotes the overall deviation signal at the test gate where $D_i = 1$ iff $D_{i\ell} = 1$ for at least one $\ell \in \{1, \dots, m\}$ and 0 otherwise.

In practice, it can be assumed that the total false-positive probability satisfies

$$P(D_i = 1 | z_i \in R) = 0, \quad (2.6)$$

i.e., that the test gate never signals a non-existing deviation.

The test gate procedure implies two statistical inference problems: i) An estimation problem in analyzing the proficiency of the test gate. ii) The prediction of the volume $S_{\text{dev},j}$ with deviation, based on the number of parts K_j with deviation in characteristic j found during the production at a test gate. The two inference tasks are discussed subsequently in subsections 2.4.1 and 2.4.2. The practical prerequisites are reviewed in subsection 2.4.3.

2.4.1 Proficiency Analysis of the Test Gate

Modern production environments integrate many test gates which are perfect for specific characteristics j by design. However, most frequently a test gate which is perfect by design for some particular characteristic j is also capable of checking further characteristics ℓ such that $p_{i\ell} = P(D_{i\ell} = 1 | z_{i\ell} \notin R_\ell) < 1$ for the correct-positive probability. This is illustrated in table 2.1. In such cases, the proficiency of the test gate as applied to the characteristic ℓ has to be evaluated by estimating the actual unknown correct-positive detection rate $p_{i\ell} = P(D_{i\ell} = 1 | z_{i\ell} \notin R_\ell)$. Depending on the deviation and the test gate, the value of $p_{i\ell}$ may differ significantly.

Characteristic	TG_1	TG_2
1	100 %	0 %
2	0 %	100 %
3	0 %	20 %
4	0 %	0 %

Table 2.1: Exemplary scheme of 4 characteristics and test gates for characteristics 1, 2, 3 with detection probabilities.

The evaluation of the test gate is separated from the production process. The objective of the proficiency analysis is to estimate the correct-positive detection rate $p_{i\ell} = P(D_{i\ell} = 1 | z_{i\ell} \notin R_\ell)$. Because the correct-positive probability is conditional on an existing deviation in characteristic ℓ , the estimation has to be based on a sample of units which are known to have the deviation in the considered characteristic ℓ . Two possibil-

ities of obtaining units satisfying $z_{i\ell} \notin R_\ell$ are: i) use an available stock from previous production, or ii) create items with deviation on purpose.

2.4.2 Prediction

We consider empirical aspects of the problem of predicting the volume $S_{dev,\ell}$ with deviation relative to a characteristic ℓ among n produced units. A stochastic analysis follows in section 2.5.

$S_{dev,\ell}$ can be predicted by combining i) the number of correctly identified parts K_ℓ with deviation relative to characteristic ℓ , i.e.,

$$K_\ell = \left| \{1 \leq i \leq n \mid D_i = 1, z_{i\ell} \notin R_\ell\} \right| = \sum_{i=1}^n \mathbb{1}_{\mathbb{R} \setminus R_\ell}(z_{i\ell}) D_i. \quad (2.7)$$

and ii) the result of the proficiency analysis of the test gate.

In practice, obtaining K_ℓ is not straightforward. At a test gate, all units i with overall signal $D_i = 1$ will be sorted out. Thus, the test gate provides the number of correctly identified parts with some deviation, i.e.

$$K = \left| \{1 \leq i \leq n \mid D_i = 1, z_{ih} \notin R_h, \text{ for some } h \in \{1, \dots, m\}\} \right|. \quad (2.8)$$

A follow-up analysis is required to identify the parts with deviation in characteristic l . K_l will often be much smaller than K . In particular, if the test gate is designed for characteristic $j \neq \ell$, the overall signal $D_i = 1$ will most of the time be caused by $D_{ij} = 1$.

2.4.3 Prerequisites

Four practical prerequisites have to be met for the application of the test gate method.

- 1) There has to be a test gate in the production process which can detect the considered deviation in characteristic k .
- 2) The data regarding the parts detected at the particular test gate has to be available. This means that for all units with $D_i = 1$, K_k has to be known. Parts where the detection from the test gate stems from some characteristic $j \neq k$ are excluded from the estimation because only a deviation in the focused characteristic k is considered at one time.
- 3) All produced parts have to be tested by the test gate during the production process.

- 4) Parts with the considered deviation must be available to test the proficiency of the test gate. These parts can be either blocked parts which have not been checked by the test gate or specifically produced for testing. It is essential that the parts used in the test are not the same parts found by the test gate in the production process to ensure the statistical independence of the detection.

In practice, condition 2) is attainable because all parts sorted out are undergoing a root cause analysis. Condition 3) can be loosened, if detailed knowledge about the distribution of the volume with deviation within the total population exists. In this case, it may be warranted that only a proportion of the total volume is checked by the test gate.

Revisiting the examples from section 2.2, prerequisites 3) and 4) are fulfilled. Furthermore, prerequisite 2) is met, as solder balls and displayed bond wires are clearly distinguishable phenomena. The first prerequisite is the most interesting one.

In the bond wire example from section 2.2.2, the test gate is the end of line test with the electronic measurement. In the solder ball example from section 2.2.1, the test gate for the detecting the solder ball is the optical inspection. The example somewhat differs, as solder ball in outside the search domain could initially not be detected. It appears therefore, that prerequisite 1) is not met. However, due to fact that after the extension of the search domain you can detect solder balls in the whole ECU, you can use this information to estimate the proportion of solder balls in the previously unchecked area. Therefore, you can estimate the proficiency of your original search domain.

2.5 Stochastic Aspects of the Test Gate Method

2.5.1 Basic Stochastic Modelling

Consider a production process where units i are technically deviating with respect to a technical characteristic j with a constant probability $q_j = P(z_{ij} \notin R_j)$. All produced parts are examined by a test gate for a deviation relative to the characteristic j . After the end of some time period $[0, t]$, n parts have been produced, K_j parts with deviation in j have correctly been identified by the test gate, $S_{\text{dev},j} \geq K_j$ parts with deviation in j have been produced, and an unknown number

$$S_{\text{dev},j} - K_j = \sum_{i=1}^n \mathbb{1}_{\mathbb{R} \setminus R_j}(z_{ij})(1 - D_i) \quad (2.9)$$

of parts with deviation in j have passed the test gate. The volume deviating $S_{\text{dev},j}$ or, equivalently, the *volume slipped* $S_{\text{dev},j} - K_j$ is the target of prediction in the test gate analysis, see section 2.4.2.

We assume that the correct-positive probability p_{ij} for the detection is identical for every produced part i with the same deviation in characteristic j , i.e. $p_{ij} = p_j$. This assumption is justified because in most cases the variation in the test gate process is very small. For a reasonably designed test gate, successive tests can be assumed to be independent among each other. By these assumptions, the test gate process can be described as a Bernoulli process where the detection probability p_j is the characteristic parameter. Then the volume slipped $S_{\text{dev},j} - K_j$ is distributed by the binomial distribution $\text{Bin}(n, q_j(1 - p_j))$. By Feller's (1968) inequalities we obtain the bounds

$$\text{P}(S_{\text{dev},j} - K_j \geq d) \leq \frac{d[1 - q_j(1 - p_j)]}{[d - nq_j(1 - p_j)]^2} \quad (2.10)$$

for $d > nq_j(1 - p_j)$, and

$$\text{P}(S_{\text{dev},j} - K_j \geq d) \leq \frac{(n - d)q_j(1 - p_j)}{[nq_j(1 - p_j) - d]^2} \quad (2.11)$$

for $d < nq_j(1 - p_j)$.

In practice, the exact value of the probability deviating q_j is unknown. The conditional distribution of K_j under $S_{\text{dev},j}$ is the binomial distribution $\text{Bin}(S_{\text{dev},j}, p_j)$. Since K_j is known, $S_{\text{dev},j}$ can be estimated using the conditional distribution, if p_j can be estimated sufficiently precise.

2.5.2 Uncertainty of Slip Rate

In practice, the correct-positive probability p_j unknown, and has to be estimated by a suitable experiment. The experiment result leads straightforward to a point estimator for p_j :

$$\hat{p}_j = \frac{x}{m}, \quad (2.12)$$

where x is the number of detected parts in the experiment and m the sample size. This point estimator does not convey information about the underlying uncertainty about the true value of the parameter p_j . The uncertainty is especially large, if the sample size used in the test is small, and p_j is small. In statistical inference, the large uncertainty is visible by an excessive length of the confidence interval such that no reasonable conclusions can be drawn. The result of the excessive length is illustrated in the subsequent example.

The example is general and independent of the particular product considered and the deviation. The examples mentioned in section 2.2 could be considered, for instance.

Example 2.5.1 (Uncertainty of Slip Ratio). *Consider a test with $m = 100$ parts with deviation of which $x = 20$ are detected. The resulting 90 % Clopper-Pearson confidence interval is $[0.1367, 0.2772]$. The expected ratio of slipped parts to detected parts for the lower bound is $\frac{1-0.1367}{0.1367} = 6.3174$, whereas the ratio for the upper bound is $\frac{1-0.2772}{0.2772} = 2.6075$. Thus, there is a significant difference in the expected ratio, depending on whether the detection probability is near the lower or near the upper bound of the confidence interval.*

Directly related to the problem of estimating the total proportion usage nonconforming is the problem of a confidence interval for a product probability. Suppose that each slipped part fails with a probability $p_{fail} > 0$ during the agreed lifetime in the field and that parts without deviation do not fail over the agreed lifetime. Furthermore, assume that the failure probability is independent of the detection at the test gate. Then the proportion nonconforming from the subgroup with deviation is $(1-p) \times p_{fail}$. A confidence interval for an upper bound with confidence level γ of this product probability can be obtained by taking the product of the individual Clopper-Pearson upper bounds $\gamma_{slip}, \gamma_{p_{fail}}$ such that $\gamma = \gamma_{slip} \times \gamma_{p_{fail}}$. However, research on the optimal choice of γ_{slip} and $\gamma_{p_{fail}}$ is missing. Additionally, the conservativeness of the individual Clopper-

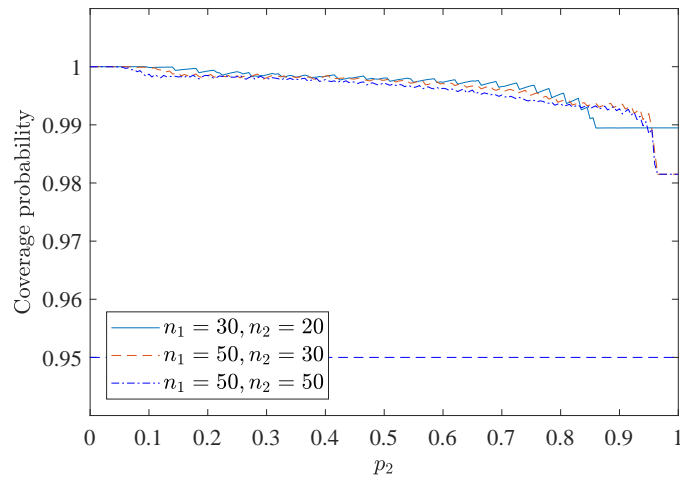


Figure 2.5: Coverage for $p_1 = 0.20$ and different n_1, n_2 with nominal confidence level $\gamma = 0.95$.

Pearson upper bounds lead to an even more conservative product upper bound as the

possibility of offsets between the upper bounds of the first and the second probability is not considered. This conservativeness is shown in figure 2.5. For more details on one-sided confidence intervals for a product probability, see chapter 6.

For a single binomial proportion, conservative frequentist two-sided confidence intervals with prior information exist, see G6b & Lurz (2014). If prior information is available, these intervals are considerably shorter than intervals without prior information. This suggests that something similar should be possible for the product probability problem. Using a Bayesian framework for the product probability is also a possibility worth exploring.

2.5.3 Uncertainty due to Dispersion of the Bernoulli Process

In risk assessments, the underlying parameter p is actually not the primary variable of interest. The primary quantity of interest is the random number of parts with deviation which slipped through the test gate and were sent to customers. Therefore, the uncertainty due to the dispersion of this random variable has to be taken into account. To predict the realization of a random variable and thus, also capturing the uncertainty due to the dispersion is known in statistics as a prediction interval problem.

Prediction intervals have received considerable less attention in the literature than confidence intervals, particularly for discrete distributions. For the binomial distribution, among the existing intervals the ‘exact’ prediction interval from Thatcher (1964), the approximate interval by Nelson (1982), and more recently, approximate interval procedures from Wang (2010) and Krishnamoorthy & Peng (2011). For the negative binomial and the Poisson distribution, there is even less literature, see for example Bain & Patel (1993) and Krishnamoorthy & Peng (2011).

The problem of having different distributions in the learning sample and in the prediction target which share the same parameter has also not been studied. This problem is relevant for the test gate method without existing part detections, see section 2.5.4. There, the learning sample is binomial, and the prediction target is a negative binomial random variable. The learning sample and prediction are different in distribution but share the same parameter in p . Prediction interval procedures encompass similar problems as confidence intervals, most notably the conflicting interests in obtaining short intervals and in guaranteeing a prescribed nominal confidence level. The Thatcher interval exceeds the nominal confidence requirement significantly, whereas approximate intervals violate the confidence requirement for a subset of the parameter space.

No attempt has been made to include prior information for prediction intervals in the frequentist framework. In many applications, prior information about the underlying distribution parameter is often existing due to experience or expert knowledge, or a subset of the parameter space is of no interest as any result in this range is deemed unacceptable. Wang (2008) proposes a general framework, which can also be used for the Bayesian framework. However, this framework is based on an already chosen prior and does not evaluate its choice.

As mentioned in section 2.1, a statistical risk assessment contains the two steps of estimating the volume with deviation and of estimating the effect of the deviation on the lifetimes of the affected volume. This was also illustrated in section 2.3, where technical and usage quality were distinguished and the effect of the technical deviation on the lifetime was described by a proportional hazards model. Relative to the second step, many applications of prediction intervals consider lifetime problems, where due to censoring usually only approximate solutions exist. Literature focuses on different procedures to obtain prediction intervals and their respective coverage properties. These approaches usually improve upon the naive (or plug-in) estimate by calibration, see Escobar & Meeker (1999) and Lawless & Fredette (2005). The focus of the application is on reliability either in lifetime testing or warranty analysis.

There exists also literature about Bayesian prediction intervals for the Weibull distribution. The Bayesian literature often focuses on different ways to estimate the parameters and the adequate choice of the prior distribution for each parameter, see Kundau & Raqab (2012).

All these applications consider a fixed known sample size, usually in the context of laboratory conditions such as lifetime tests or lifetime behavior of the total population. In a large class of problems in applications, a subpopulation of interest exists, whose size is unknown. How the methods for known sample sizes work for this problem class and how they need to be adapted has not been studied. Risk assessments in the context presented in this chapter belongs to this class. An unknown population size imposes additional uncertainty which needs to be accounted for. This will be discussed in chapter 4.

2.5.4 Test Gate with no Detected Failures

In some cases, the test gate method is applicable even though no part with deviation has yet been detected during production. More precisely, this means that there exists a

test gate which can detect $z_{ij} \notin R_j$, even though at time t the number of detections K_j by this test gate is 0.

Since $K_j = 0$, the knowledge about the existence of $z_{ij} \notin R_j$ for some products has to come from another source. This could be, for example, an internal failure at a parallel production line which uses the same supply chain. Now, the question is how many ‘failures’, i.e. slipped parts, have occurred given that 0 ‘successes’, i.e. detections, have occurred. This can be interpreted as a discrete waiting time problem, which can be described with a geometric distribution. The PMF of a geometric distribution is given by

$$P(Y = k) = 1 - p_j(1 - p_j)^k. \quad (2.13)$$

Here, p_j denotes the detection probability and therefore $1 - p_j$ the slip rate. Y is the number of slipped parts before the first detection. For a known detection probability, a prediction interval can be obtained by using the quantiles of (2.13). If p_j can be estimated with sufficient precision, the same procedure using (2.13) provides a good approximation. Commonly, p_j is estimated by independent sampling and binomial distribution inference. For feasible sample sizes, the uncertainty with respect to p_j can be large, as has been shown in example 2.5.1. In this case, no exact prediction interval procedure for Y is known. The predictor is binomially distributed and the prediction target has a geometric distribution. Exact procedures are only available if both predictor and prediction target are either binomial or negative binomial.

2.6 Benchmarking with Other Methods

2.6.1 Test Gate Method versus Control Sampling

A standard method to determine the size of the subgroup with deviation is control sampling, as defined subsequently.

Definition 2.6.1 (control sampling plan). *A control sampling plan is defined by the subsequent components:*

1. *A sample of size $n_{sample} \in \mathbb{N}$ of product units of the same nature as the parts $1, \dots, n$ in the field.*
2. *The number X_{sample} of deviating parts with $x_i \notin R_i$ in the sample of units $1, \dots, n_{sample}$.*

3. An estimator $\widehat{S}_{dev,i}(t)$ of the volume $S_{dev,i}(t)$ of deviating parts with $x_i \notin R_i$ in the field.

Control sampling and the test gate method are fundamentally different. Control sampling considers a part of the population, whereas the test gate method is a screening procedure. A potential advantage of control sampling can be the more subtle investigation of the sampled items, since the number of considered items is small. The test gate method as an automated screening procedure involves a misclassification with rate $1 - p$. However, control sampling has various shortcomings in comparison to the test gate method.

The test gate method is implemented within the production whereas control sampling is an additional intervention, external to the production process. The control sampling method needs to be representative in the sense that the sample population i.e. sample in stock exhibits the same proportion of products with deviation as the volume sent to customers. In general, it is not possible to verify this requirement in industrial practice. Even more, in several cases it is violated, for instance if the deviation occurs in clusters over time. Then, control sampling leads to inappropriate estimates.

The potential accuracy of the control sampling method can go along with high testing expenses, up to destructive testing. Another shortcoming of control sampling in modern practice is its incompatibility with rigorous quality limits in the ppm range. For 0 observed ‘successes’ in a sample of large size, the heuristic ‘rule of three’ states (see van Belle (2008)) that $\frac{3}{n_{sample}}$ provides an approximate 95 % confidence interval for p . If p is in the ppm range and the upper bound is supposed to be of the same magnitude, it is obvious that the required sample size is prohibitive. For instance, if an upper bound of 10 ppm is desired, the required sample size is

$$n_{sample} = \frac{3}{10ppm} = 300000. \quad (2.14)$$

The implementation requirements for the test gate method listed in section 2.4 are often met in practice. Being process internal, the test gate can be used to provide a fast response in small quality problems in the ppm (parts per million) range. In high quality mass production, even a first failure triggers a risk assessment. Fast responses are required for communication and an efficient allocation of resources. Being able to show that the subgroup with deviation is small is therefore crucial. The test gate method is a powerful and cost-efficient way to accomplish this task, whereas control sampling is unfeasible for these ppm cases.

Being a screening method, the test gate method is able to cope with the occurrence of temporally coherent clusters of items with deviation. The latter is a serious problem of control sampling. Control sampling can be adapted to the case only under very restrictive conditions.

Prior information can be used to improve control sampling, e.g. by the following approach. A sample is taken at time t and provides an estimate $\widehat{S}_{\text{dev}}(t)$ of the volume $S_{\text{dev}}(t)$ with deviation. Based on a model of the evolution of the volume with deviation along the time axis, an inference from $S_{\text{dev}}(t)$ to $S_{\text{dev}}(s)$, $s \neq t$ may be made. The uncertainty of the prediction can be quantified if the distribution of $S_{\text{dev}}(s)$ is sufficiently known. A simple version of the required time series model assumes that the proportion $p_{\text{dev}}(t)$ of the volume with deviation in the production volume at time t is constant over a certain time window where a production problem was prevalent. However, the validity of predictions from control sampling based on such assumptions strongly depend on the reliability and accuracy of the prior information.

2.6.2 System Analytical Approach to Estimate the Volume Deviating

Sometimes, you may try to predict the volume $S_{\text{dev}}(t)$ of parts deviating at time t based on insights into the relation between product characteristics and manufacturing system. Let ε_t be the vector of manufacturing system characteristics which is related to the vector z of deviations of the product quality characteristic by a regression equation

$$z(t) = f(v(t), \varepsilon_t) \quad (2.15)$$

where ε_t is a noise variable. Specific instances of the model (2.15) are studied by Kuruüzüm & Akyüz (2009) and Yao et al. (2016). If the distribution of v_t is sufficiently known, the volume $S_{\text{dev}}(t)$ of parts deviating at time t can be predicted.

The system analytical approach is similar to the test gate method in the sense that both predict the volume deviating based on a comprehensive insight into the manufacturing system instead of considering samples as in control sampling. However, the system analytical approach imposes much stronger requirements than the test gate approach. Whereas knowledge on the distribution of manufacturing system characteristics is available in many cases, a sufficiently accurate model of the type (2.15) which relates manufacturing system and product characteristics is difficult to obtain. The prerequisites are a combination of engineering knowledge and extensive empirical studies. In industrial practice, these strong requirements on analysis and prior information are satisfied in special cases only.

2.7 Conclusion

We have described a method to estimate the size of a subgroup with deviation based on an imperfect detection mechanism at a test gate during the production. The test gate method is a powerful and cost-efficient way to show that a potential new quality problem is actually minor. This is crucial, since in modern high quality mass production even a single failure triggers a risk assessment. Fast responses are required for communication and efficient allocation of resources. Even in case of small detection rates, the method has proven to be very powerful in applications.

The presented test gate method fills an existing gap in industrial practice to determine small volumes with deviation within large production volumes. Standard sampling methods are often not applicable, or do not lead to useful results. A detailed treatment of the statistical problems mentioned in section 2.5 is desirable for future work. Another area of future work is the time series of detections which provides information on the underlying deviation. This information could be used to limit the time period where the deviation was present depending on the time structure of the detections.

3 Discrete Prediction Intervals with Prior Information

3.1 Introduction

Chapter 2 presented the test gate method to estimate the size of a subgroup in industrial production and provided a comparison to control sampling. A key prerequisite to use sampling methods in industrial applications is the ability to analyze the technical root cause to enable clustering into subgroups. Apart from this technical requirement, the estimation of the subgroup size is crucial.

For practitioners, prediction regions are much easier to use and understand than confidence regions for distribution parameters. Predicting a phenomenon is the practitioner's primary interest, whereas the study of distributions is only a secondary theoretical instrument for making predictions. A distribution parameter is an unobservable model abstraction, whereas predictions target directly the empirical observable phenomenon the practitioner is dealing with. However, prediction regions have received considerably less attention in statistical literature than confidence regions for distribution parameters.

Despite the different empirical objectives, the theoretical basis of confidence and prediction regions is the same. A yet unknown quantity of interest should be covered by a region calculated from a sample with a probability larger or equal to a prescribed nominal level γ . This construct is subject to two conflicting requirements relative to the information provided by the region: i) reliability, which is measured by the coverage probability; ii) precision, which is a decreasing function of the volume of the region. High reliability goes along with poor precision, and vice versa. So-called "exact" regions guarantee compliance with the nominal level, however at the expense of considerable exceedance of the actual coverage probability over the nominal level, and thus poor precision. So-called "approximate" regions often tend to undercut the nominal level, but they may be attractive due to better precision.

The clue to better precision of exact regions is in exploiting prior information. Classical confidence and prediction regions exploit only the information provided by the sample data. This is an unnecessary limitation for many industrial and business environments where usually some partial prior knowledge on the unobserved quantity of interest exists, due to experience or historical data or an analytic understanding of the considered phenomenon. In the stochastic model, such prior knowledge can be expressed via prior distributions. For the confidence region problem, a general scheme for the use of prior information has been established by Göb & Lurz (2014). It has been demonstrated that prior information can substantially increase the precision of exact confidence regions. Subsequently, we will adopt the prior information scheme for the construction of prediction intervals, namely for prediction intervals for a binomially distributed counting variable and for a negative binomially distributed counting variable.

In particular, we consider the problem of predicting a counting variable X_2 based on an observed counting variable X_1 for the following cases: i) both X_1 and X_2 are binomially distributed with the same probability parameter p , ii) X_1 is binomially distributed and X_2 is negative binomially distributed with the same probability parameter p and iii) both X_1 and X_2 are negative binomially distributed with the same probability parameter p . The objective is to provide a prediction interval $B = B(X_1)$ based on the observed sample X_1 such that the unobserved X_2 is covered by B with a sufficiently high probability.

Several approaches to binomial prediction intervals are suggested in the literature, all without taking prior information into account. Thatcher's (1964) exact interval guarantees a prescribed nominal level of the coverage probability for all values of p . However, the coverage is exceeding the nominal level significantly even for large sample sizes. Nelson's (1982) approximate interval is easy to compute, but exhibits severe undercoverage, especially near the boundaries. Recently, Krishnamoorthy & Peng (2011) and Wang (2010) suggested closed-form approximate intervals, with improved coverage properties compared to Nelson's interval. Krishnamoorthy & Peng (2011) showed that their proposed "Joint Sampling" interval is either equal to or included in Wang's "Score" interval. However, both intervals are very conservative near the boundaries and have undercoverage for a significant part of the parameter space.

There are even less approaches for negative binomial prediction. Bain & Patel (1993) applied Faulkenberry's (1973) approach to the negative binomial distribution. They also provided a prediction interval based on a normal distribution. An exact interval was introduced by Patel & Samaranayake (1991). However, this interval is not easily accessible.

Wang (2008) provided a coverage study for discrete prediction intervals and proposed a method to improve the approximate procedures. Again, none of these approaches take prior information into account.

We use prior information on p to establish a model for the conditional distribution of the prediction target X_2 under the observation X_1 . This association model enables the construction of minimum volume prediction intervals which pointwise in $X_1 = x_1$ obey a prescribed nominal prediction confidence level. The remainder of this chapter is structured as follows. Section 3.2 introduces the general concept of minimum weighted volume prediction spaces. Section 3.3 considers minimum weighted volume prediction intervals for the binomial prediction problem. Section 3.4 studies the properties of joint and conditional distributions of the observation binomial variable X_1 and the binomial prediction target X_2 . Algorithms for the numerical calculation of minimum weighted volume prediction spaces are presented in section 3.5. Section 3.6 considers the sensitivity of prediction spaces relative to prior information for the binomial case. Section 3.7 evaluates coverage properties of prediction intervals under prior information for the binomial case.

Subsequently, the cases of an observation binomial variable X_1 and a negative binomial prediction target X_2 and an observation negative binomial variable X_1 and a negative binomial prediction target X_2 are treated similarly. Section 3.8 considers minimum weighted volume prediction intervals for a negative binomial prediction target. The properties of the required distributions are presented in section 3.9. The algorithm presented in section 3.5 needs minor adaptations, which are stated in section 3.10. Section 3.11 and section 3.12 discuss sensitivity and coverage aspects the case of a negative binomial prediction target. The chapter ends with a conclusion and an outline of future work.

3.2 Minimum Weighted Volume Prediction Spaces

This paragraph introduces the general concept of minimum volume X_1 - X_2 prediction spaces of level γ pointwise in $X_1 = x_1$. The setting adopts the theory of minimum volume level γ confidence regions established by GÖb & Lurz (2014) for the prediction region problem.

Let (Ω, \mathcal{F}, P) be a probability space, $R_1 \subseteq \mathbb{R}^{d_1}$, $R_2 \subseteq \mathbb{R}^{d_2}$. Let \mathcal{A}_i be a σ -field in $R_1 \subseteq \mathbb{R}^{d_1}$, $R_2 \subseteq \mathbb{R}^{d_2}$, $R \subseteq \mathbb{R}^d$, with corresponding measures ν_1, ν_2, ν . Let $X_1: \Omega \rightarrow R_1$,

$X_2: \Omega \rightarrow R_2$ be random variables. Let f_{X_1, X_2} be the joint density of X_1 and X_2 with respect to a product measure $\mu_1 \otimes \mu_2$ on the product σ -field $\mathcal{A}_1 \otimes \mathcal{A}_2$; let f_{X_i} be the marginal density and $f_{X_i|X_j=x_j}$ be the conditional density.

We interpret a set $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ as a *prediction space* in the sense that the projection $A_{x_1} := \{x_2 | (x_1, x_2) \in A\}$ is a prediction region for X_2 under the observed value $X_1 = x_1$, and the projection $A_{x_2} := \{x_1 | (x_1, x_2) \in A\}$ is a prediction region for X_1 under the observed value $X_2 = x_2$. The projections satisfy

$$x_1 \in A_{x_2} \Leftrightarrow x_2 \in A_{x_1} \Leftrightarrow (x_1, x_2) \in A \text{ for } x_1 \in R_1, x_2 \in R_2, \quad (3.1)$$

i.e., x_2 is an element of the prediction region for X_2 under the observation $X_1 = x_1$ iff x_1 is an element of the prediction region for X_1 under the observation $X_2 = x_2$.

Subsequently, we consider X_1 as the predictor, X_2 the prediction target. Then for each observation $X_1 = x_1$, the prediction target X_2 should be covered by the prediction region A_{x_1} with a prescribed minimum probability, as stipulated by the subsequent definition 3.2.1.

Definition 3.2.1 (x_1 -pointwise level γ PS). Let $0 < \gamma < 1$. A set $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ is called x_1 -pointwise level γ prediction space (PS) for (X_1, X_2) iff

$$\begin{aligned} \gamma &\leq P_{x_1}(X_2 \in A_{x_1}) = P_{x_1}(X_1 \in A_{x_2}) \\ &= \int_{A_{x_1}} f_{X_2|X_1=x_1}(x_2) d\mu_2(x_2) \quad \text{for all } x_1 \in R_1. \end{aligned} \quad (3.2)$$

Many customary approaches to prediction regions assume independent observation X_1 and prediction target X_2 , and consider the coverage probability $P_y(X_2 \in A_{X_1})$ indexed in a distribution parameter common to X_1 and X_2 . The coverage probability $P_{x_1}(X_2 \in A_{x_1})$ considered in (3.2) is a useful quantity only in the case of dependent X_1 and X_2 . We obtain an association model by considering the distribution parameter y in the classical model as a realization $Y = y$ of a random variable $Y: \Omega \rightarrow R, R \subseteq \mathbb{R}^d$. The density f_Y of Y with respect to some dominating measure μ is assumed to be known from prior information, and X_1 and X_2 are assumed to be conditionally independent under $Y = y$. Then the joint PDF f_{X_1, X_2} is obtained from

$$f_{X_1, X_2}(x_1, x_2) = \int_R f_{X_1|Y=y}(x_1) f_{X_2|Y=y}(x_2) f_Y(y) d\mu(y). \quad (3.3)$$

The above model suits to many practical cases. For instance, consider a batch manufacturing process with an output of units 1 through n , with a corresponding sequence

Z_1, \dots, Z_n of quality characteristics of the items. A sample Z_1, \dots, Z_{n_1} from the batch is analyzed by a statistic $X_1 = T_1(Z_1, \dots, Z_{n_1})$ to make a prediction on a statistic $X_2 = T_2(Z_{n_1+1}, \dots, Z_n)$ of the remainder of the batch. Let Y be a parameter of the probability distribution of Z_i . From the statistical analysis of previously produced batches, the density f_Y of the parameter Y is sufficiently well known so that the joint density f_{X_1, X_2} can be calculated by (3.3).

For given $X_1 = x_1$, the precision of the prediction region A_{x_1} is expressed by its volume $\nu_2(A_{x_1})$ where ν_2 is a suitable volume measure on the σ -field \mathcal{A}_2 in R_2 , e.g., the Lebesgue-Borel measure or a counting measure. To express the precision of all possible projection volumes A_{x_1} in a unique indicator, we consider the *weighted volume*, which averages all projection volumes $\nu_2(A_{x_1})$ weighted by the PDF $f_{X_1}(x_1)$ of X_1 , and we seek *minimum weighted volume* prediction spaces, as characterized by the subsequent definition.

Definition 3.2.2 (weighted volume, MWV PS). *The weighted volume of a prediction space A is defined by*

$$V_{X_2|X_1}(A) = \int_{R_1} \nu_2(A_{x_1}) f_{X_1}(x_1) d\mu_1(x_1). \quad (3.4)$$

A prediction space A^* is called minimum weighted volume (MWV) x_1 -pointwise level γ prediction space iff A^* satisfies the level requirement (3.2) and $V_{X_2|X_1}(A^*)$ is smallest among the weighted volumes $V_{X_2|X_1}(A)$ of all prediction spaces A which satisfy the level requirement (3.2).

Since the level requirement (3.2) is stated for each condition $X_1 = x_1$ separately, it is obvious from the definition (3.4) of the weighted volume $V_{X_2|X_1}(A)$ that an MWV x_1 -pointwise level γ prediction space A^* can be constructed by the following algorithm 3.2.3.

Algorithm 3.2.3 (MWV x_1 -pointwise level γ prediction space).

- S1) For each x_1 , construct a set $A(x_1) \in \mathcal{A}_2$ which minimizes $\nu_2(A(x_1))$ under $P_{x_1}(X_2 \in A(x_1)) \geq \gamma$.
- S2) Let $A^* = \cup_{x_1 \in R_1} \{x_1\} \times A(x_1)$.

In the case of a univariate X_2 with values in a discrete or continuous segment $R_2 \subseteq \mathbb{R}$, particular interest is in prediction spaces A where the projections A_{x_1} , $x_1 \in R_1$, are

intervals

$$A_{x_1} = [x_L(x_1), x_U(x_1)] \cap R_2 = \{x_2 \in R_2 | x_L(x_1) \leq x_2 \leq x_U(x_1)\}. \quad (3.5)$$

From the representation (3.2) of the level requirement it is obvious that a sufficient condition for obtaining interval projections under algorithm 3.2.3 is that for each $x_1 \in R_1$ the conditional PDF $f_{X_2|X_1=x_1}$ is unimodal. The prediction space A^* which minimizes the weighted volume $V_{X_2|X_1}(A)$ among the prediction spaces A with interval projections A_{x_1} is provided by the following modified algorithm 3.2.4.

Algorithm 3.2.4 (MWV x_1 -pointwise level γ PS with interval projections).

S1) For each $x_1 \in R_1$, minimize $\nu_2([x_L(x_1), x_U(x_1)] \cap R_2)$ under the constraint $F_{X_2|X_1=x_1}(x_U(x_1)) - F_{X_2|X_1=x_1}(x_L(x_1)^-) \geq \gamma$.

S2) Let $A^ = \cup_{x_1 \in R_1} \{x_1\} \times A(x_1)$.*

3.3 The Beta Prior Model for Inference on Binomially Distributed Variables

This section expands on a special case of the model established by the previous paragraph 3.2. We consider counting variables X_i where the distribution parameter Y is the probability parameter of a binomial distribution. Conditional on $Y = y$, the observation variable X_1 and the prediction target X_2 are assumed to be independent, where X_1 has the binomial distribution $Bin(n_1, y)$, and the prediction target X_2 has the binomial distribution $Bin(n_2, y)$. We choose $R_i = \{0, \dots, n_i\}$, $i = 1, 2$, and consider counting measures μ_1 , μ_2 and ν_2 with $\mu_i(B) = |B|$ for $B \subseteq R_i$, $\nu_2(B) = |B|$ for $B \subseteq R_2$. Then the confidence requirement (3.2) amounts to

$$\begin{aligned} \gamma &\leq P_{x_1}(x_1 \in A_{x_2}) = P_{x_1}(X_2 \in A_2) \\ &= \sum_{x_2 \in A_{x_1}} f_{X_2|X_1=x_1}(x_2) \text{ for all } x_1 \in R_1, \end{aligned} \quad (3.6)$$

and the weighted volume defined by (3.4) is

$$V_{X_2|X_1}(A) = \sum_{x_1 \in R_1} |A_{x_1}| f_{X_1}(x_1) = \sum_{x_2 \in R_2} \sum_{x_1 \in A_{x_2}} f_{X_1}(x_1). \quad (3.7)$$

The probability parameter Y is assumed to have a beta distribution $Beta(a, b)$ with shape parameters $a, b > 0$, i.e.,

$$f_Y(Y = y) = \begin{cases} \frac{1}{B(a,b)} y^{a-1} (1-y)^{b-1} & \text{if } 0 < y < 1, \\ 0 & \text{otherwise,} \end{cases} \quad (3.8)$$

where

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad (3.9)$$

is the beta function. With the generalized binomial coefficients defined by

$$\binom{x}{y} = \frac{\Gamma(x+1)}{\Gamma(y+1)\Gamma(x-y+1)} = \binom{x}{x-y} \quad (3.10)$$

for $x, y \in (-1, +\infty)$, $x+1 > y$, the beta function can be expressed as

$$B(s, t) = \frac{1}{s \binom{s+t-1}{s}} = \frac{1}{t \binom{s+t-1}{t}} \quad \text{for } s, t > 0. \quad (3.11)$$

The beta distribution model has several appealing characteristics which made it the preferred distribution for expressing prior information on a probability y , particularly in Bayesian statistics, see Hald (1981) in acceptance sampling in quality control where y is the probability of manufacturing a nonconforming unit, Godfrey & Andrews (1982) or Berg (2006) in audit sampling where y is the probability of a misspecification in an account: flexibility; sparse parametrization; the property of being the conjugate prior for the binomial distribution; the potential to express various density shapes like bathtub, inverted bathtub, strictly decreasing, strictly increasing, constant (equidistribution).

The equidistribution case with $a = 1 = b$ corresponds to complete uncertainty on the binomial parameter y . In many industrial applications, the binomial parameter y is a probability nonconforming, e.g., the probability of producing a nonconforming item in manufacturing or the probability of a booking error in accounting. In such contexts, y is usually very small. The relevant beta priors have parameters $a \leq 1 < b$ or $a < 1 \leq b$ with PDFs strictly decreasing on $[0, 1]$. The more a or b deviate from 1, the higher the concentration of the probability mass on values close to 0.

The two parameters a and b of the beta distribution are uniquely determined by the mean and one quantile, or by two quantiles. In the case of repetitive sampling, the latter parameters may be estimated from historical data, for instance in audit sampling or quality control where data from past inspections may be exploited. However, often appropriate reference data is not available. In this case, the features of the distribution

have to be elicited from expert opinions in interviews or panels, see the studies by Corless (1972), Hogarth (1975), Kadane et al. (1980), Chaloner & Duncan (1983), O'Hagan O'Hagan (1998), Walls & Quigley Walls & Quigley (2001), for instance. Software assisted eliciting schemes are considered by Blocher & Robertson (1976) or Garthwaite & O'Hagan (2000).

3.4 Properties of Distributions under the Beta Prior Model for Binomial Counts

Under the beta prior information model (3.8) on the probability parameter, this section derives and studies the joint, marginal, and conditional distributions of the counts X_1 and X_2 . The following proposition 3.4.1 establishes the joint, marginal, and conditional PMFs.

Proposition 3.4.1 (PMFs of X_1, X_2). *Let Y have the beta distribution $Beta(a, b)$. Conditional on $Y = y$, let $X_i, i = 1, 2$ have the binomial distribution $Bin(n_i, y)$. Then we have:*

(i) *The unconditional joint PMF f_{X_1, X_2} of (X_1, X_2) is given by*

$$\begin{aligned} f_{X_1, X_2}(x_1, x_2) &= \frac{\binom{n_1}{x_1} \binom{n_2}{x_2} B(a + x_1 + x_2, b + n_1 + n_2 - x_1 - x_2)}{B(a, b)} \\ &= \frac{\binom{n_1}{x_1} \binom{n_2}{x_2} \binom{a+x_1+x_2-1}{x_1+x_2} \binom{b+n_1-x_1+n_2-x_2-1}{n_1+n_2-x_1-x_2}}{\binom{n_1+n_2}{x_1+x_2} \binom{a+b+n_1+n_2-1}{n_1+n_2}} \quad (3.12) \\ &= \frac{\binom{n_1}{x_1} \binom{n_2}{x_2} \binom{-a}{x_1+x_2} \binom{-b}{n_1+n_2-x_1-x_2}}{\binom{n_1+n_2}{x_1+x_2} \binom{-a-b}{n_1+n_2}} \end{aligned}$$

for $x_i = 0, \dots, n_i$.

(ii) *The unconditional PMF f_{X_i} of X_i satisfies*

$$\begin{aligned} f_{X_i}(x_i) &= \frac{\binom{n_i}{x_i} B(x_i + a, n_i - x_i + b)}{B(a, b)} \\ &= \frac{\binom{a+x_i-1}{x_i} \binom{b+n_i-x_i-1}{n_i-x_i}}{\binom{a+b+n_i-1}{n_i}} = \frac{\binom{-a}{x_i} \binom{-b}{n_i-x_i}}{\binom{-a-b}{n_i}} \quad (3.13) \end{aligned}$$

for $x_i = 0, \dots, n_i$.

(iii) The conditional PMF $f_{X_2|X_1=x_1}$ of X_2 under $X_1 = x_1$ satisfies

$$\begin{aligned} f_{X_2|X_1=x_1}(x_2) &= \frac{\binom{n_2}{x_2} B(a+x_1+x_2, b+n_1+n_2-x_1-x_2)}{B(a+x_1, b+n_1-x_1)} \\ &= \frac{\binom{a+x_1+x_2-1}{x_1} \binom{b+n_1-x_1+n_2-x_2-1}{n_2-x_2}}{\binom{a+b+n_1+n_2-1}{n_2}} = \frac{\binom{-a-x_1}{x_1+x_2} \binom{-b-n_1+x_1}{n_2-x_2}}{\binom{-a-b-n_1}{n_2}} \end{aligned} \quad (3.14)$$

for $x_2 = 0, \dots, n_2$.

The proof of proposition 3.4.1 is provided in appendix 3.A1. The last representation of the PMF in (3.12), (3.13), (3.14), respectively, shows the relation to the generalized hypergeometric distribution introduced by Kemp & Kemp (1956), see also Johnson et al. (1993), with parameters α, β, n and the probabilities

$$\frac{\binom{\alpha}{r} \binom{\beta}{n-r}}{\binom{\alpha+\beta}{n}} \quad \text{for } r = 0, 1, 2, \dots \quad (3.15)$$

In the classification by Kemp & Kemp (1956), the marginal distribution of X_i provided by (3.13) is the generalized hypergeometric distribution of type II A with parameters $\alpha = -a, \beta = -b, n = n_i$. The conditional distribution of X_2 under $X_1 = x_1$ is the generalized hypergeometric distribution of type II A with parameters $\alpha = -a - x_1 < 0, \beta = -b - n_1 + x_1 < 0, n = n_2$. Using the relationships of the distributions considered by proposition 3.4.1 with the generalized hypergeometric distribution, we can infer expressions for the moments from the respective formulae provided by Kemp & Kemp (1956). Hence,

$$E[X_i] = n_i \frac{a}{a+b+n_i}, \quad E[X_2|X_1=x_1] = n_2 \frac{a+x_1}{a+b+n_1}, \quad (3.16)$$

$$V[X_i] = \frac{n_i ab(a+b+n_i)}{(a+b)^2(a+b+1)}, \quad (3.17)$$

$$V[X_2|X_1=x_1] = \frac{n_2(a+x_1)(b+n_1-x_1)(a+b+n_1+n_2)}{(a+b+n_1)^2(a+b+n_1+1)}. \quad (3.18)$$

A sufficient condition for obtaining prediction spaces with interval projections is the unimodality of the conditional PDF $f_{X_2|X_1=x_1}$ for each $x_1 \in R_1$. For the conditional density (3.14) the unimodality is stated by the subsequent proposition 3.4.2 whose proof is provided by appendix 3.A2.

Proposition 3.4.2 (unimodality of $f_{X_2|X_1=x_1}$). *Under the beta prior with parameters $a, b > 0$, consider the conditional PMF $f_{X_2|X_1=x_1}$ provided by (3.14).*

(i) *For $x_2 \in \{0, \dots, n_2 - 1\}$ we have*

$$f_{X_2|X_1=x_1}(x_2 + 1) = \frac{n_2 - x_2}{x_2 + 1} \frac{a + x_1 + x_2}{b + n_1 + n_2 - x_1 - x_2 - 1} f_{X_2|X_1=x_1}(x_2). \quad (3.19)$$

(ii) *For $x_2 \in \{0, \dots, n_2 - 1\}$ we have*

$$f_{X_2|X_1=x_1}(x_2 + 1) \begin{cases} > f_{X_2|X_1=x_1}(x_2) & \text{if } x_2 < \frac{(n_2+1)(a+x_1-1)}{a+b+n_1-2} - 1, \\ = f_{X_2|X_1=x_1}(x_2) & \text{if } x_2 = \frac{(n_2+1)(a+x_1-1)}{a+b+n_1-2} - 1, \\ < f_{X_2|X_1=x_1}(x_2) & \text{if } x_2 > \frac{(n_2+1)(a+x_1-1)}{a+b+n_1-2} - 1. \end{cases} \quad (3.20)$$

Due to the occurrences of binomial coefficients and beta functions in the conditional PMF $f_{X_2|X_1=x_1}$ and the repeated calculation of the CDF $F_{X_2|X_1=x_1}$, a search algorithm like algorithm 3.2.4 is computationally intensive for larger sample sizes. A simple approximation of the conditional CDF $F_{X_2|X_1=x_1}$ can be inferred from the subsequent proposition 3.4.3 on the asymptotics of the beta function. The proof of proposition 3.4.3 is given in appendix 3.A3.

Proposition 3.4.3 (asymptotics of beta function). *Let $\alpha, \beta > 0$, and let $p: (0, \infty) \rightarrow (0, 1)$ be a function with $\lim_{n \rightarrow \infty} p(n) = s \in (0, 1)$. Then we have*

$$\lim_{n \rightarrow \infty} \frac{\binom{n}{np(n)} B(\alpha + np(n), \beta + n[1 - p(n)])}{\frac{1}{n} p(n)^{\alpha-1} [1 - p(n)]^{\beta-1}} = 1. \quad (3.21)$$

Consider proposition 3.4.3 with $\alpha = a + x_1$, $\beta = b + n_1 - x_1$. From the representation (3.14) of the conditional density $f_{X_2|X_1=x_1}$ and from the limiting relation (3.21) we infer for large n_2 the approximation

$$\begin{aligned} F_{X_2|X_1=x_1}(x_2) &\approx \frac{\int_0^{x_2} \left(\frac{z}{n_2}\right)^{a+x_1-1} \left(1 - \frac{z}{n_2}\right)^{b+n_1-x_1-1} dz}{B(a+x_1, b+n_1-x_1)} \\ &= \frac{\int_0^{x_2/n_2} p^{a+x_1-1} (1-p)^{b+n_1-x_1-1} dp}{B(a+x_1, b+n_1-x_1)} = G_{a+x_1, b+n_1-x_1} \left(\frac{x_2}{n_2}\right) \end{aligned} \quad (3.22)$$

where $G_{a+x_1, b+n_1-x_1}$ is the CDF of the beta distribution $Beta(a+x_1, b+n_1-x_1)$.

3.5 Construction of Exact and Approximate MWV Prediction Intervals for Binomial Counts

As a special case of the algorithm 3.2.4 considered in section 3.2, the present section establishes an exact and an approximative algorithm to calculate MWV x_1 -pointwise level γ prediction spaces for the prediction target count X_2 under the predictor count $X_1 = x_1$.

By proposition 3.4.2, the conditional PMF $f_{X_2|X_1=x_1}$ expressed by formula (3.14) is unimodal. In this case, the MWV x_1 -pointwise level γ prediction spaces constructed by the basic algorithm 3.2.3 necessarily have interval projections A_{x_1} . Hence, we obtain MWV x_1 -pointwise level γ prediction spaces from the algorithm 3.2.4 where $R_i = \{0, \dots, n_i\}$ and ν_2 is the counting measure with support R_2 , see the subsequent algorithm 3.5.1.

Algorithm 3.5.1 (exact MWV x_1 -pointwise level γ PS under beta prior).

S1) For each $x_1 \in \{0, \dots, n_1\}$:

S1.1) Determine $u(x_1) = \max\{x_2 \in R_2 | 1 - F_{X_2|X_1=x_1}(x_2 - 1) \geq \gamma\}$.

S1.2) For each $x_2 = 0, \dots, u(x_1)$ determine

$$k(x_2) = \min \{k \in \mathbb{N}_0 | F_{X_2|X_1=x_1}(x_2 + k) - F_{X_2|X_1=x_1}(x_2 - 1) \geq \gamma\}.$$

S1.3) Determine $\hat{x}_2(x_1) \in \{0, \dots, u(x_1)\}$ which minimizes $k(x_2)$.

S1.4) Let $A(x_1) = \{\hat{x}_2(x_1), \dots, \hat{x}_2(x_1) + k(\hat{x}_2(x_1))\}$.

S2) Let $A^* = \{(x_1, x_2) | x_2 \in A(x_1)\}$.

For large sample sizes n_2 , the search algorithm can be accelerated by using the approximation (3.22) of the conditional CDF $F_{X_2|X_1=x_1}$, see the subsequent algorithm 3.5.2.

Algorithm 3.5.2 (approximate MWV x_1 -pointwise level γ PS under beta prior). For $x_1 \in \{0, \dots, n_1\}$, let $G_{a+x_1, b+n_1-x_1}$ be the CDF of the beta distribution $Beta(a+x_1, b+n_1-x_1)$.

S1) For each $x_1 \in \{0, \dots, n_1\}$:

S1.1) Determine $u(x_1) = n_2 G_{a+x_1, b+n_1-x_1}^{-1}(1 - \gamma)$.

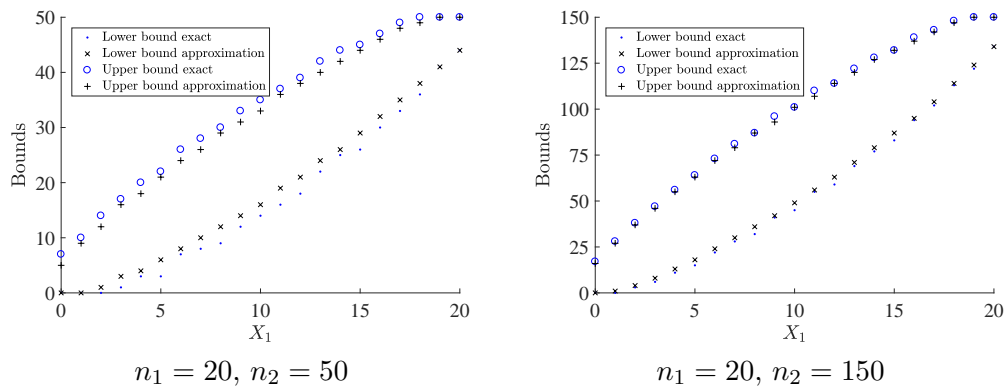


Figure 3.1: Comparison between the endpoints of exact intervals by algorithm 3.5.1 and approximate intervals by algorithm 3.5.2. $a = 1 = b$, $\gamma = 0.90$.

S1.2) For each $x_2 = 0, \dots, u(x_1)$ determine

$$k(x_2) = -x_2 + n_2 G_{a+x_1, b+n_1-x_1}^{-1} \left(\gamma + G_{a+x_1, b+n_1-x_1}^{-1} \left(\frac{x_2}{n_2} \right) \right).$$

S1.3) Determine $\hat{x}_2(x_1) \in [0, u(x_1)]$ which minimizes $k(x_2)$.

S1.4) Let $A(x_1) = \{\hat{x}_2(x_1), \dots, \hat{x}_2(x_1) + k(\hat{x}_2(x_1))\}$.

S2) Let $A = \{(x_1, x_2) | x_2 \in A(x_1)\}$.

Figure 3.1 illustrates the accuracy of the approximate prediction intervals by algorithm 3.5.2 in comparison with the exact intervals by algorithm 3.5.1. For a small prediction sample size of $n_2 = 50$, minor differences are visible. The approximate intervals tend to be slightly shorter than the exact intervals. For a moderate sample size of $n_2 = 150$, the difference between the bounds are hardly noticeable. However, due to the discreteness of the prediction target, a small difference in the length can lead to a considerable drop in coverage. The latter effect will be analyzed in section 3.7.

3.6 The Sensitivity of Prediction Spaces in Prior Information for a Binomial Prediction Target

To be useful, the prior information expressed by a beta distribution $Beta(a, b)$ of the underlying probability y should take effect on properties of the prediction space. However,

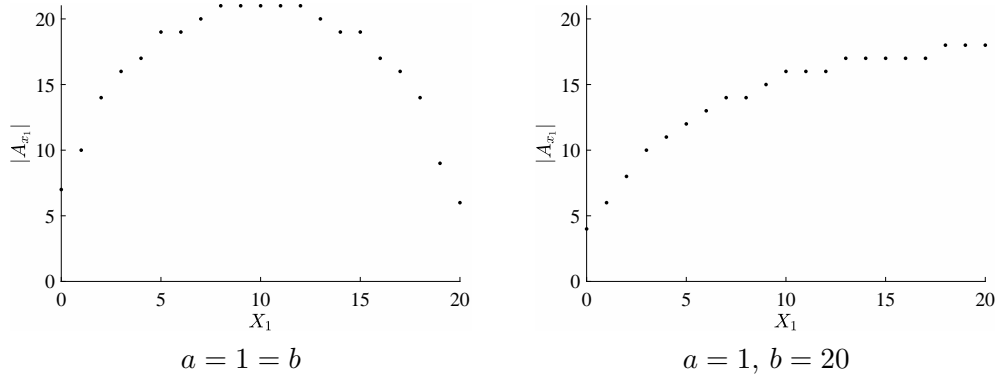


Figure 3.2: Pointwise length of intervals A_{x_1} and weighted volumes $V_{X_2|X_1}(A)$ of exact x_1 -pointwise level γ prediction spaces with $\gamma = 0.90$, sample sizes $n_1 = 20$, $n_2 = 50$.

the characteristics of the prediction space should not be oversensitive in the parameters a and b so as to avoid serious inference errors resulting from minor misspecifications.

In this section, we study the effect of prior information on three core characteristics of x_1 -pointwise level γ prediction spaces: i) the length of single prediction intervals, ii) the weighted volume of the entire prediction space, iii) the necessary sample size to achieve a prescribed weighted volume. Throughout, we assume the confidence level $\gamma = 0.90$, and compare priors $Beta(1.0, b)$ with varying b under fixed $a = 1.0$. In all subsequent studies, we assume the size n_1 of the predictor sample to be considerably smaller than the size n_2 of the prediction target group. The latter relation holds in most applications, particularly in prediction problems in industrial manufacturing where a small predictor sample, e.g., obtained from a controlled experiment, is exploited to provide information on a large prediction target group, e.g., a lot of manufactured units released to the field. The opposite relation, as considered by Krishnamoorthy & Peng (2011) and Wang (2010), is much less common.

Figure 3.2 illustrates the effect of prior information on interval length by comparing the length of single prediction intervals under the noninformative prior $Beta(1.0, 1.0)$ (equidistribution) with the prior $Beta(1.0, 20.0)$. The latter prior has a right-skewed strictly decreasing PDF which puts large weight on small $y = p$. For instance, the 95 % quantile of $Beta(1.0, 20.0)$ is 0.14 only. Under $Beta(1, 1)$, the intervals for medium size realizations of X_1 are longest, and the intervals for small and large realizations are

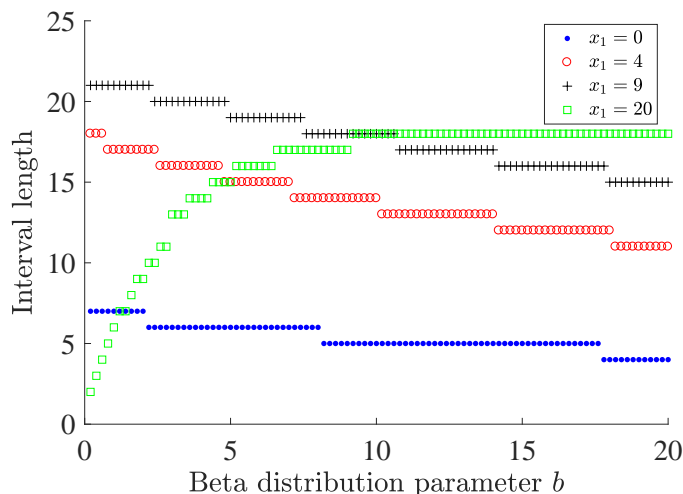


Figure 3.3: Length of prediction intervals A_{x_1} of level $\gamma = 0.90$ for various observations $X_1 = x_1$ under the beta prior $Beta(1, b)$ with various values b . Sample sizes $n_1 = 20$, $n_2 = 50$.

shorter. Under $Beta(1, 20)$, corresponding to the shape of the PDF the intervals are significantly shorter for small x_1 than in the $Beta(1, 1)$ case. For large x_1 , the intervals are longer in the $Beta(1, 20)$ than in the $Beta(1, 1)$ case.

Figure 3.3 compares two extreme cases of prior information. Figure 3.3 illustrates the effect of the parameter b ranging over the interval $[1, 20]$ in detail. $a = 1$ is fixed. The length of intervals A_{x_1} is considered for observations $X_1 = 0, 4, 9, 20 = n_1$. A strong local effect of small changes in b is visible only under the observation $x_1 = 20 = n_1$ where values of b moderately above 1.0 rapidly inflate the interval length. However, such observations X_1 close to n_1 occur with minor probability for moderate to large b . For the relevant observations $X_1 = 0, 4, 9$ a small increase in b takes only minor effect on the interval length.

Figure 3.4 considers the sensitivity of the weighted volume $V_{X_2|X_1}(A)$ in b under fixed $a = 1.0$. For other values of a , the shape of the sensitivity curve is similar but shifted along the horizontal axis. The weighted volume is highly sensitive in b in the region around 1.0 where small deviations of b away from 1.0, i.e., small deviations away from the noninformative case $a = 1 = b$, entail strong reductions in volume whereas changes in larger values b affect the descent in the weighted volume much less.

For most applications, the sensitivity pattern exhibited by figure 3.4 is not problematic

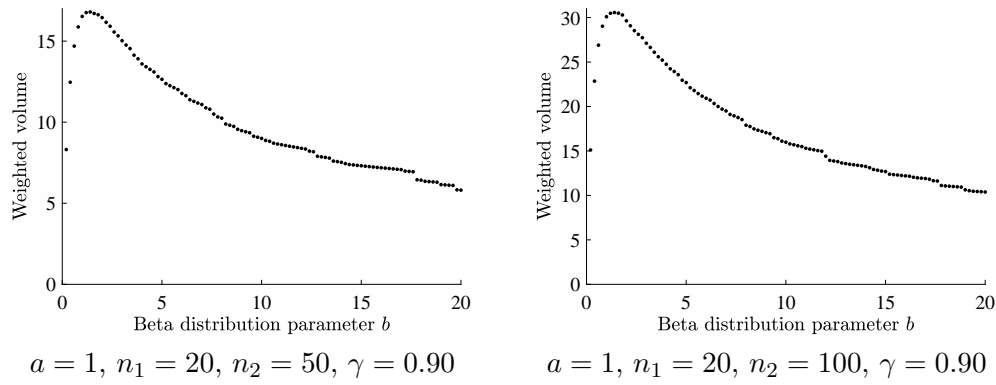


Figure 3.4: Weighted volume $V_{X_2|X_1}(A)$ as a function of the parameter b of the beta distribution, $a = 1$ fixed.

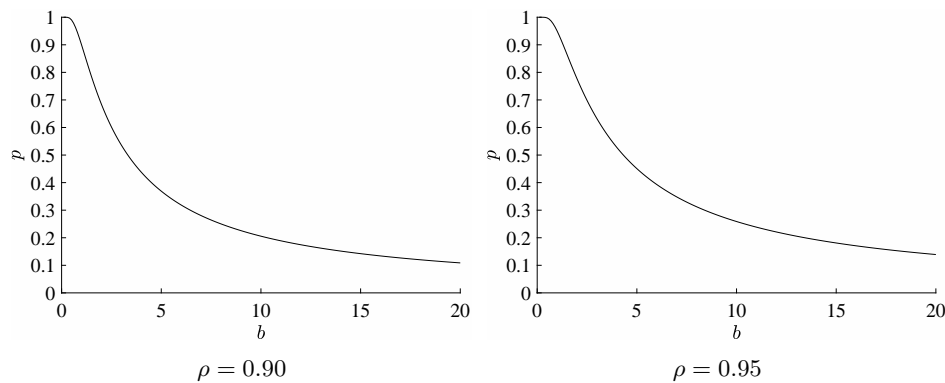


Figure 3.5: $100\rho\%$ quantiles of $Beta(1, b)$ as a function of b .

but rather useful. Figure 3.6 displays the 90 % and 95 % quantiles of $Beta(1.0, b)$ as functions of $1 \leq b \leq 20$. The region of step descent of the quantile functions coincides with the region of step descent of the weighted volume. Increasing b from 1.0 to 5.0 decreases the 90 % quantile from 0.90 down to 0.37, and the 95 % quantile from 0.95 down to 0.45. However, for most applications where Y is a probability nonconforming, in particular in quality control or auditing, the latter values of quantiles of Y are still extremely conservative. This means: Rather conservative increases in b away from 1.0 lead to substantial reductions in weighted volume; under less conservative assumptions of higher values of b which are subject to a higher risk of misspecification the gain in weighted volume is much smaller.

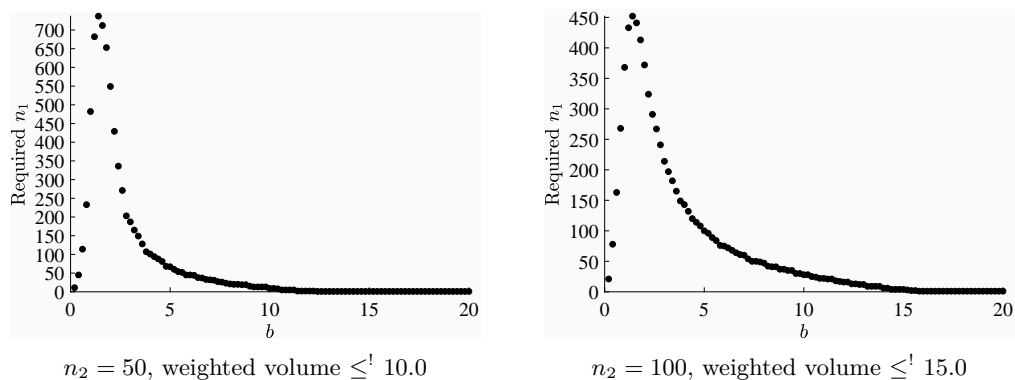


Figure 3.6: Required sample size $n_{1,\min}$ to undercut a prescribed weighted volume of level $\gamma = 0.90$ as a function of b .

The effect of prior information on the sample size required to undercut a prescribed weighted volume is complementary to the effect on the weighted volume. Figure 3.6 displays the minimum sample size $n_{1,\min}$ which undercuts a prescribed weighted volume of the prediction space as a function of b . $n_{1,\min}$ is rapidly decreasing when b moves away from 1.0. Changes in large b have a weak effect on $n_{1,\min}$ only.

3.7 Coverage Properties for the Binomial Case

The present section defines concepts of the *coverage* of a prediction interval for the prediction target X_2 ; it i) compares the coverage of exact MWV x_1 -pointwise level γ prediction intervals with the coverage of prediction intervals without prior information, ii) studies the effect of misspecified prior information on the coverage of exact MWV x_1 -pointwise level γ prediction intervals, and iii) evaluates the coverage of approximate MWV x_1 -pointwise level γ prediction intervals constructed by algorithm 3.5.2.

Under the beta prior model of section 3.3, the pointwise coverage $P_{x_1}(X_2 \in A_{x_1})$ considered by definition 3.2.1 amounts to

$$\begin{aligned}
 P_{x_1}(X_2 \in A_{x_1}) &= \sum_{x_2=0}^{n_2} f_{X_2|X_1=x_1}(x_2) \mathbb{1}_{A_{x_1}}(x_2) \\
 &= \sum_{x_2=0}^{n_2} \frac{\binom{n_2}{x_2} B(a+x_1+x_2, b+n_1+n_2-x_1-x_2)}{B(a+x_1, b+n_1-x_1)} \mathbb{1}_{A_{x_1}}(x_2),
 \end{aligned} \tag{3.23}$$

where $\mathbb{1}_B$ denotes the indicator function of a set B , see equation (3.14) for the condi-

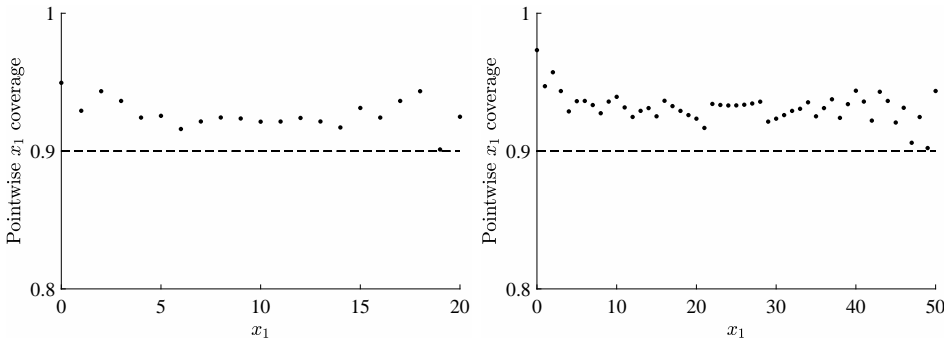


Figure 3.7: Pointwise x_1 coverage for the level-90 x_1 -pointwise MWV for the binomial case

tional PMF $f_{X_2|X_1=x_1}(x_2)$. (3.23) is the appropriate coverage concept under a model of randomly varying $y = p$ as established by (3.3). For a correct prior, the coverage (3.23) is at least the nominal level. This is illustrated in figure 3.7. In the absence of a prior on $y = p$, the coverage (3.23) is not meaningful. The customary coverage concept in prediction region theory without priors considers the probability $P_p(X_2 \in A_{X_1})$ indexed in the parameter $y = p$, here

$$P_p(X_2 \in A_{X_1}) = \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} p^{x_1+x_2} (1-p)^{n_1+n_2-x_1-x_2} \mathbb{1}_{A_{x_1}}(x_2). \quad (3.24)$$

To compare the coverage properties of the MWV prediction interval with the coverage of customary prediction intervals we have to use the coverage concept (3.24). The indicated MWV prediction intervals for a fair comparison with intervals without priors are the MWV intervals under the uninformative uniform prior $Beta(1.0, 1.0)$ which expresses total uncertainty on the probability parameter $y = p$. Figure 3.8 shows that the MWV prediction interval under $Beta(1.0, 1.0)$ has good properties in the sense of the coverage (3.24) pointwise in p . The prescribed level $\gamma = 0.90$ is undercut in some areas, but the violations are minor.

We proceed to the topic ii) of this section, the effect of misspecified prior information on the coverage. For this purpose we consider the coverage concept (3.23) pointwise in x_1 which is inherent to prediction spaces under prior information.

Figure 3.9 illustrates the problem of violations of the nominal level γ by the pointwise coverage $P_{x_1}(X_2 \in A_{x_1})$ under priors misspecified relative to a true quantile. Throughout, violations occur for medium to large x_1 only. In practice, particular caution is required relative to misspecifications underestimating the true quantile, as in the cases

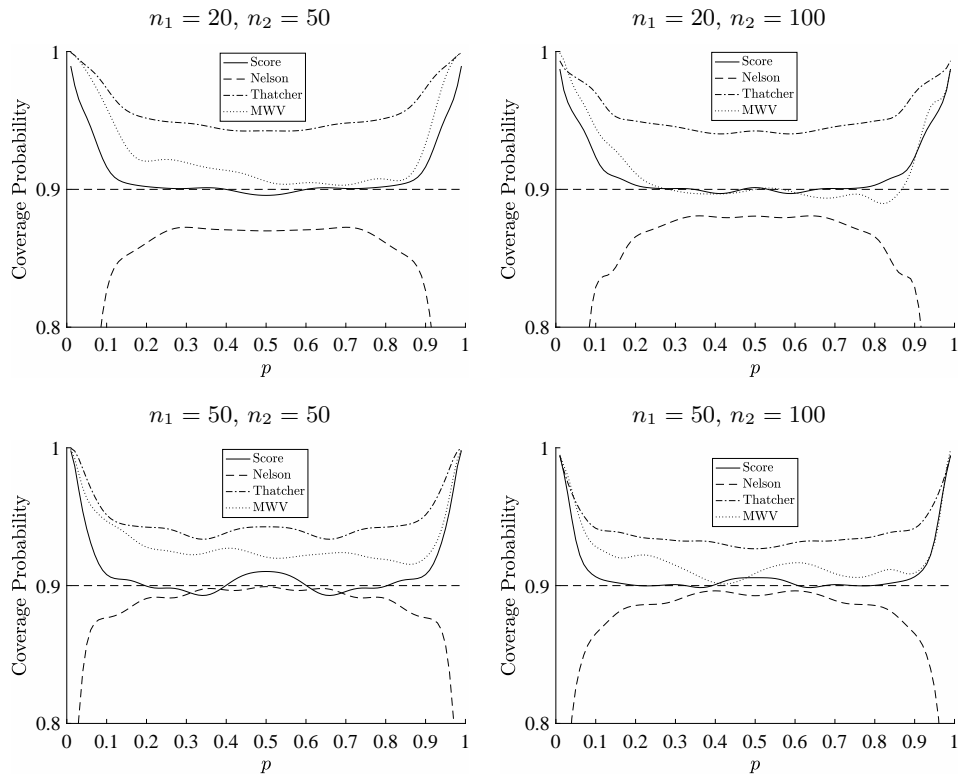


Figure 3.8: Comparison of the coverage probabilities between the Score PI, Nelson's PI, Thatcher's PI and the MWV PI with $a = 1, b = 1$ for different sample sizes.

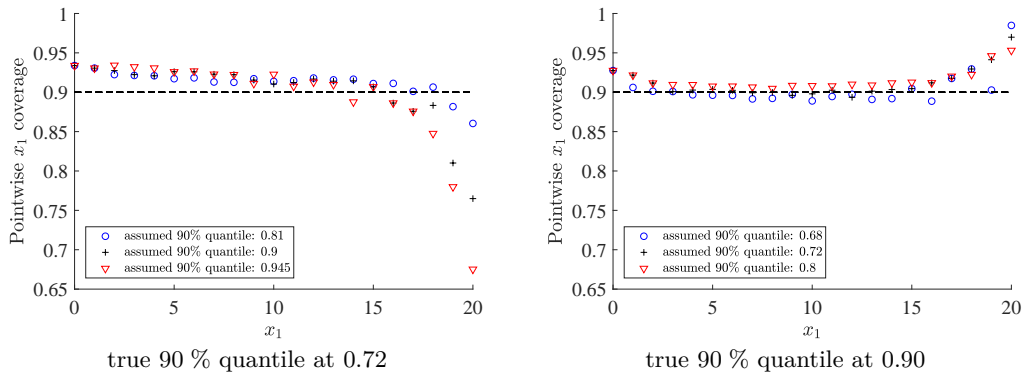


Figure 3.9: Coverage probability pointwise in x_1 for misspecified priors under sample sizes $n_1 = 20, n_2 = 100$ and confidence level $\gamma = 0.90$. Priors $Beta(a, b)$ with $a = 1$, parameter b determined by quantile prescription.

on the right-hand side of figure 3.9: assumed 90 % quantile 0.68, assumed 90 % quantile 0.72 and assumed 90 % quantile 0.80 versus the true 90 % quantile 0.90. In these cases, large values of p and, consequently, medium to large values of x_1 with potential coverage shortfalls are actually much more likely than assumed by the misspecified priors. However, the shortfalls observed for the underestimating misspecifications are actually minor. Conservative misspecifications overestimating the true quantile are less seri-

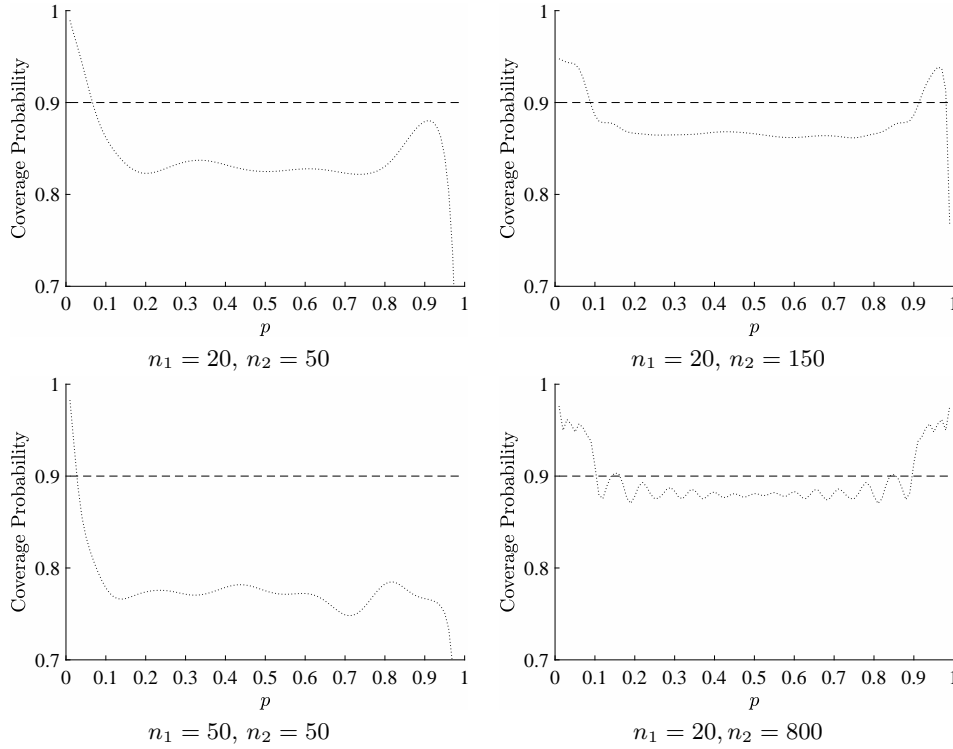


Figure 3.10: Pointwise coverage probability for the approximate MWV prediction interval, prior $Beta(1, 1)$, nominal confidence level $\gamma = 0.90$

ous since medium to large values of x_1 with potential coverage shortfalls are actually less likely than assumed by the misspecified priors. The coverage shortfalls observed at medium to large x_1 under the three conservative misspecifications (assumed 90 % quantile 0.81, assumed 90 % quantile 0.90 and assumed 90 % quantile 0.945 versus the true 90 % quantile 0.72) are partially serious. However, the critical values of x_1 are actually much less likely to occur under the true prior than under the stipulated misspecification. For example, the probability of $x_1 = 19$ with the true prior is 0.0120, whereas in the assumed cases the resulting probabilities are 0.0248, 0.0476, and 0.0661 respectively. The

conclusive rule for practice is to bet on conservative prior assumptions. Then, coverage shortfalls are rather negligible.

Finally we consider the topic iii) of the coverage of type (3.23) of approximate MWV x_1 -pointwise level γ prediction intervals constructed by algorithm 3.5.2. Figure 3.10 indicates that the coverage properties of approximate intervals depends strongly on the ratio of the sample sizes n_1 and n_2 . For $n_1 = 20$, $n_2 = 50$ the coverage is considerably below the nominal level 0.90 with a minimum around 0.82. Increasing n_2 to 150 with $n_1 = 20$ fixed increases the coverage, the minimum coverage is around 0.87. Leaving $n_2 = 50$ fixed and increasing n_1 to 50 leads to a further drop in coverage to values below 0.80. Figure 3.1 shows that approximate intervals tend to be shorter than exact intervals where, however, the bounds on both sides have only minor differences between approximate and exact. Nevertheless, the missing points can have strong effect on the coverage particularly under small sample sizes where the points have a relatively large probability mass.

3.8 The Beta Prior Model for Inference on Negative Binomially Distributed Variables

This section modifies the case for binomially distributed variables presented in section 3.3 for the case where conditional on $Y = y$, the prediction target X_2 has the negative binomial distribution $negBin(r_2, y)$, with PMF

$$f_X(x) = \binom{x + r_2 - 1}{x} y^{r_2} (1 - y)^x \quad \text{for } x = 0, 1, 2, \dots, \quad (3.25)$$

for $y \in (0, 1)$, $r_2 \in \mathbb{N}$. We consider the two cases where, conditional on $Y = y$, X_1 has i) the binomial distribution $Bin(n_1, y)$ and ii) the negative binomial distribution $negBin(r_1, y)$. As in section 3.3, conditional on $Y = y$, X_1 and X_2 are assumed to be independent. Therefore, the distribution parameter Y in case i) is the shared probability parameter of a binomial and negative binomial distribution, in case ii) is the probability parameter of a negative binomial distribution.

In the framework of section 3.2, we choose $R_1 = \{0, \dots, n_1\}$ and $R_2 = \mathbb{N}_0$ in case i) and we choose $R_i = \mathbb{N}_0$, $i = 1, 2$ in case ii). The confidence requirement (3.2) and the weighted volume defined by (3.4) result in the same expressions as in section 3.3.

3.9 Properties of some Discrete Distributions under the Beta Prior Model

Under the beta prior information model (3.8) on the probability parameter, this section derives and studies the joint, marginal, and conditional distributions of the random variables X_1 and X_2 . We consider the two cases: i) both X_i are negative binomial random variables and ii) X_1 is a binomial, and X_2 is a negative binomial random variable.

3.9.1 Two Negative Binomial Counts

First, we consider minimum volume prediction spaces with prescribed level pointwise in x_1 for two negative binomial counts. The following proposition 3.9.1 establishes the joint, marginal, and conditional PMFs. The proof of proposition 3.9.1 is given in appendix 3.A4.

Proposition 3.9.1 (PMFs of X_1, X_2). *Let Y have the beta distribution $Beta(a, b)$, $a, b > 0$, with PDF*

$$f_Y(y) = \begin{cases} \frac{1}{B(a,b)} y^{a-1} (1-y)^{b-1} & \text{if } 0 < y < 1, \\ 0 & \text{otherwise,} \end{cases}$$

where the Beta function is defined as in (3.9). Conditional under $Y = y$, let X_i have the negative binomial distribution $negBin(r_i, y)$ with $r_1, r_2 \in \mathbb{N}$. Then we have:

(i) *The unconditional joint PMF f_{X_1, X_2} of X_1, X_2 satisfies*

$$f_{X_1, X_2}(x_1, x_2) = \frac{\binom{x_1+r_1-1}{x_1} \binom{x_2+r_2-1}{x_2} B(r_1+r_2+a, x_1+x_2+b)}{B(a, b)}, \quad (3.26)$$

for $x_i = 0, 1, 2, \dots$

(ii) *The unconditional PMF f_{X_i} of X_i satisfies*

$$f_{X_i}(x_i) = \frac{\binom{x_i+r_i-1}{x_i} B(r_i+a, x_i+b)}{B(a, b)}, \quad (3.27)$$

for $x_i = 0, 1, 2, \dots$

(iii) *The conditional PMF $f_{X_2|X_1=x_1}$ of X_2 under $X_1 = x_1$ satisfies*

$$f_{X_2|X_1=x_1}(x_2) = \binom{x_2+r_2-1}{x_2} \frac{B(r_1+r_2+a, x_1+x_2+b)}{B(r_1+a, x_1+b)}, \quad (3.28)$$

for $x_2 = 0, 1, 2, \dots$

(iv) The prediction likelihood ratio is given by

$$Q_{x_1}(x_2) = \frac{B(a, b)B(r_1 + r_2 + a, x_1 + x_2 + b)}{B(r_1 + a, x_1 + b)B(r_2 + a, x_2 + b)},$$

for $x_i = 0, 1, 2, \dots$

3.9.2 Binomial Predictor, Negative Binomial Prediction Target

We now consider minimum volume prediction spaces with prescribed level pointwise in x_1 , where the observed sample is the realization of a binomial random variable, but the prediction target is a negative binomial random variable. Again, we establish the joint, marginal and conditional PMFs. The proof of proposition 3.9.2 is provided in appendix 3.A5.

Proposition 3.9.2 (PDFs of X_1, X_2). . Let Y have the beta distribution $Beta(a, b)$, $a, b > 0$, with PDF

$$f_Y(y) = \begin{cases} \frac{1}{B(a, b)} y^{a-1} (1-y)^{b-1} & \text{if } 0 < y < 1, \\ 0 & \text{otherwise,} \end{cases}$$

where the Beta function is defined as in (3.9). Conditional under $Y = y$, let X_1 have the binomial distribution $Bin(n_1, y)$ with $n_1 \in \mathbb{N}$, $0 < y < 1$, and conditional under $Y = y$, let X_2 have the negative binomial distribution $negBin(r_2, y)$ with $r_2 \in \mathbb{N}$. Then we have:

(i) The unconditional joint PMF f_{X_1, X_2} of X_1, X_2 satisfies

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{B(a, b)} \binom{n_1}{x_1} \binom{x_2 + r_2 - 1}{x_2} B(x_1 + r_2 + a, n_1 - x_1 + x_2 + b), \quad (3.29)$$

for $x_1 = 0, \dots, n_1, x_2 = 0, 1, 2, \dots$

(ii) The unconditional PMF f_{X_1} of X_1 satisfies

$$f_{X_1}(x_1) = \frac{\binom{n_1}{x_1} B(x_1 + a, n_1 - x_1 + b)}{B(a, b)}, \quad (3.30)$$

for $x_1 = 0, \dots, n_1$.

(iii) The unconditional PMF f_{X_2} of X_2 satisfies

$$f_{X_2}(x_2) = \frac{\binom{x_2 + r_2 - 1}{x_2}}{B(a, b)} B(r_2 + a, x_2 + b), \quad (3.31)$$

for $x_2 = 0, 1, 2, \dots$

(iv) The conditional PMF $f_{X_2|X_1=x_1}$ of X_2 under $X_1 = x_1$ satisfies

$$f_{X_2|X_1=x_1}(x_2) = \binom{x_2 + r_2 - 1}{x_2} \frac{B(x_1 + r_2 + a, n_1 - x_1 + x_2 + b)}{B(x_1 + a, n_1 - x_1 + b)}, \quad (3.32)$$

for $x_1 = 0, \dots, n_1$, $x_2 = 0, 1, 2, \dots$.

(v) The prediction likelihood ratio is given by

$$Q_{x_1}(x_2) = \frac{B(a, b)B(x_1 + r_2 + a, n_1 - x_1 + x_2 + b)}{B(r_2 + a, x_2 + b)B(x_1 + a, n_1 - x_1 + b)}, \quad (3.33)$$

for $x_1 = 0, \dots, n_1$, $x_2 = 0, 1, 2, \dots$.

As in section 3.4, we prove the unimodality of the conditional PMFs $f_{X_2|X_1=x_1}$ which is a sufficient condition to obtain prediction spaces with interval projections. Proposition 3.9.3 states the unimodality for the conditional PMF (3.28), the proof is provided in appendix 3.A6. The unimodality of the conditional PMF (3.32) is subsequently stated in proposition 3.9.4 and the proof is presented in appendix 3.A7.

Proposition 3.9.3 (unimodality of $f_{X_2|X_1=x_1}$). Consider the conditional PMF $f_{X_2|X_1=x_1}$ given by (3.28) under the beta prior.

(i) For $x_2 \in \mathbb{N}_0$ we have

$$f_{X_2|X_1=x_1}(x_2 + 1) = \frac{x_2 + r_2}{x_2 + 1} \frac{x_1 + x_2 + b}{x_1 + x_2 + b + r_1 + r_2 + a} f_{X_2|X_1=x_1}(x_2). \quad (3.34)$$

(ii) For $x_2 \in \mathbb{N}_0$ we have

$$f_{X_2|X_1=x_1}(x_2 + 1) \begin{cases} > f_{X_2|X_1=x_1}(x_2) & \text{if } x_2 < \frac{r_2(x_1+b-1)-x_1-b-a-r_1}{r_1+a+1}, \\ = f_{X_2|X_1=x_1}(x_2) & \text{if } x_2 = \frac{r_2(x_1+b-1)-x_1-b-a-r_1}{r_1+a+1}, \\ < f_{X_2|X_1=x_1}(x_2) & \text{if } x_2 > \frac{r_2(x_1+b-1)-x_1-b-a-r_1}{r_1+a+1}. \end{cases}$$

Proposition 3.9.4 (unimodality of $f_{X_2|X_1=x_1}$). Consider the conditional PMF $f_{X_2|X_1=x_1}$ given by (3.32) under the beta prior.

(i) For $x_2 \in \mathbb{N}_0$ we have

$$f_{X_2|X_1=x_1}(x_2 + 1) = \frac{x_2 + r_2}{x_2 + 1} \frac{n_1 - x_1 + x_2 + b}{n_1 + x_2 + r_2 + a + b} f_{X_2|X_1=x_1}(x_2). \quad (3.35)$$

(ii) For $x_2 \in \mathbb{N}_0$ we have

$$f_{X_2|X_1=x_1}(x_2 + 1) \begin{cases} > f_{X_2|X_1=x_1}(x_2) & \text{if } x_2 < \frac{r_2(n_1-x_1+b-1)-n_1-a-b}{x_1+a+1}, \\ = f_{X_2|X_1=x_1}(x_2) & \text{if } x_2 = \frac{r_2(n_1-x_1+b-1)-n_1-a-b}{x_1+a+1}, \\ < f_{X_2|X_1=x_1}(x_2) & \text{if } x_2 > \frac{r_2(n_1-x_1+b-1)-n_1-a-b}{x_1+a+1}. \end{cases}$$

3.10 Construction of Exact Prediction Intervals for a Negative Binomial Prediction Target

As in section 3.8 and section 3.9, we consider two cases for a negative binomial prediction target: i) the predictor X_1 is binomial and ii) the predictor is also negative binomial.

We have established the unimodality of the conditional distribution $f_{X_2|X_1=x_1}$ for both cases in section 3.9. Therefore, algorithm 3.5.1 can be used with minimal adaptations. The adaptation is necessary as $R_i = \mathbb{N}$, $i = 1, 2$ is unbounded for the prediction target X_2 . In case i) step S1.1) in algorithm 3.5.1 has to be limited to a maximum number of possible values of $x_2 \in R_2$. In case ii), it is furthermore required to limit the number of possible values for $x_1 \in R_1$ to a finite number in step S1). The adapted algorithms for the two cases are subsequently stated in algorithm 3.10.1 and algorithm 3.10.2.

Algorithm 3.10.1 (exact MWV x_1 -pointwise level γ PS under beta prior for the binomial-negative binomial case).

S1) For each $x_1 \in \{0, \dots, n_1\}$:

S1.1) Choose a maximum value x_{2max} .

Determine $u(x_1) = \max\{x_2 \in R_2, x_2 \leq x_{2max} | 1 - F_{X_2|X_1=x_1}(x_2 - 1) \geq \gamma\}$.

S1.2) For each $x_2 = 0, \dots, u(x_1)$ determine

$$k(x_2) = \min\{k \in \mathbb{N}_0 | F_{X_2|X_1=x_1}(x_2 + k) - F_{X_2|X_1=x_1}(x_2 - 1) \geq \gamma\}.$$

S1.3) Determine $\hat{x}_2(x_1) \in \{0, \dots, u(x_1)\}$ which minimizes $k(x_2)$.

S1.4) Let $A(x_1) = \{\hat{x}_2(x_1), \dots, \hat{x}_2(x_1) + k(\hat{x}_2(x_1))\}$.

S2) Let $A^* = \{(x_1, x_2) | x_2 \in A(x_1)\}$.

Algorithm 3.10.2 (exact MWV x_1 -pointwise level γ PS under beta prior for the negative binomial case).

S1) Choose a maximum value $x_{1max} \in R_1$. For each $x_1 \in \{0, \dots, x_{1max}\}$:

S1.1) Choose a maximum value $x_{2max} \in R_2$.

Determine $u(x_1) = \max\{x_2 \in R_2, x_2 \leq x_{2max} | 1 - F_{X_2|X_1=x_1}(x_2 - 1) \geq \gamma\}$.

S1.2) For each $x_2 = 0, \dots, u(x_1)$ determine

$$k(x_2) = \min \{k \in \mathbb{N}_0 | F_{X_2|X_1=x_1}(x_2 + k) - F_{X_2|X_1=x_1}(x_2 - 1) \geq \gamma\}.$$

S1.3) Determine $\hat{x}_2(x_1) \in \{0, \dots, u(x_1)\}$ which minimizes $k(x_2)$.

S1.4) Let $A(x_1) = \{\hat{x}_2(x_1), \dots, \hat{x}_2(x_1) + k(\hat{x}_2(x_1))\}$.

S2) Let $A^* = \{(x_1, x_2) | x_2 \in A(x_1)\}$.

The maximum values x_{1max} and x_{2max} have to be chosen large enough that the values above them have a negligible effect on the algorithm.

Figure 3.11 depicts exact MWV intervals for several cases. Small realizations of X_1 indicate a small y for the binomial predictor and a large y for the negative binomial predictor. In both cases, contrary to the binomial case, a small y makes a large realization of X_2 more likely. The prediction intervals are affected by the prior information in the desired manner. The $Beta(1, 99)$ prior on the right of figure 3.11 puts much weight on a small $y = p$. The resulting intervals have larger lower bounds and larger upper bounds compared to the intervals with the $Beta(1, 1)$ prior on the left. For realizations of X_1 indicating a small to medium y , the upper bound is increased significantly for the $Beta(1, 99)$ prior compared to the $Beta(1, 1)$ prior.

There are two main observations for the lower bound. The first observation is that for realizations of X_1 indicating a small $y = p$, the lower bounds are fairly small. The conditional PMFs (3.28) and (3.32) which are plotted in figure 3.12 are right-skewed, whereas the conditional PMF (3.14) in the binomial case is rather symmetric. Thus, a considerable amount of probability mass lies at the lower end, which can explain the observed phenomenon. The second observation is that despite strong changes in the prior information, the lower bound does not move significantly, even if x_1 indicates a small $y = p$. In figure 3.12 you can see that the dispersion of the conditional PMF is larger for the $Beta(1, 99)$ prior than for the $Beta(1, 1)$ prior. This is opposite to the binomial case. Thus, the length reduction observed in the binomial case cannot be observed for a negative binomial prediction target.

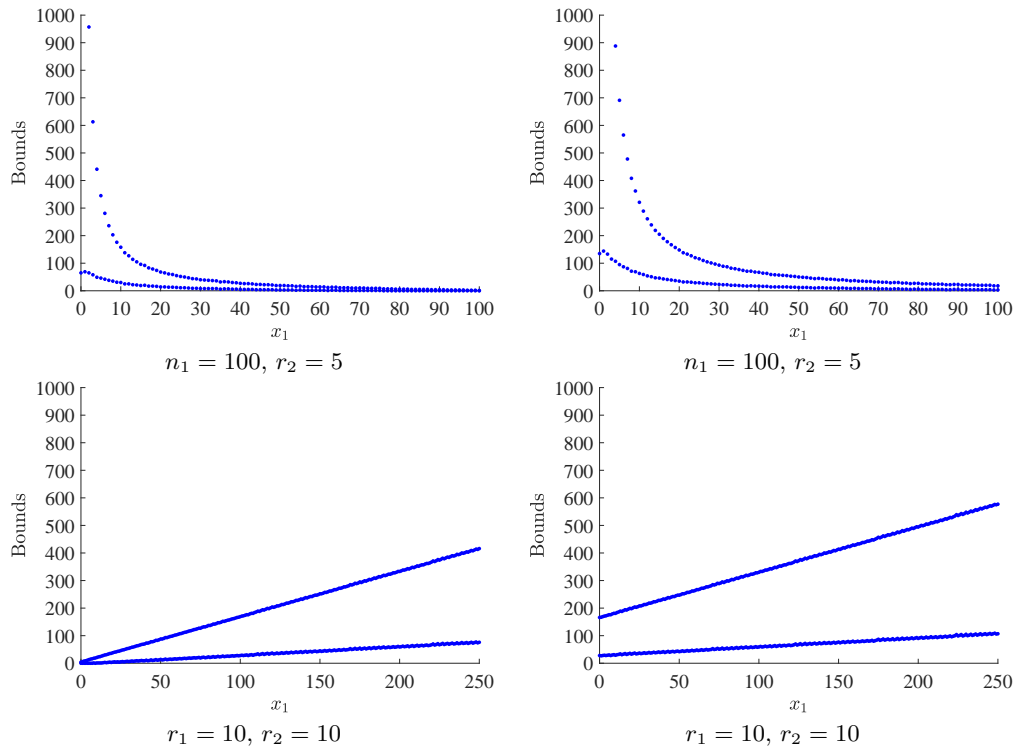


Figure 3.11: Exact MWV prediction intervals, prior $Beta(1, 1)$ (left), $Beta(1, 99)$ (right), nominal confidence level $\gamma = 0.90$ for the binomial-negative binomial case (top) and the negative binomial case (bottom).

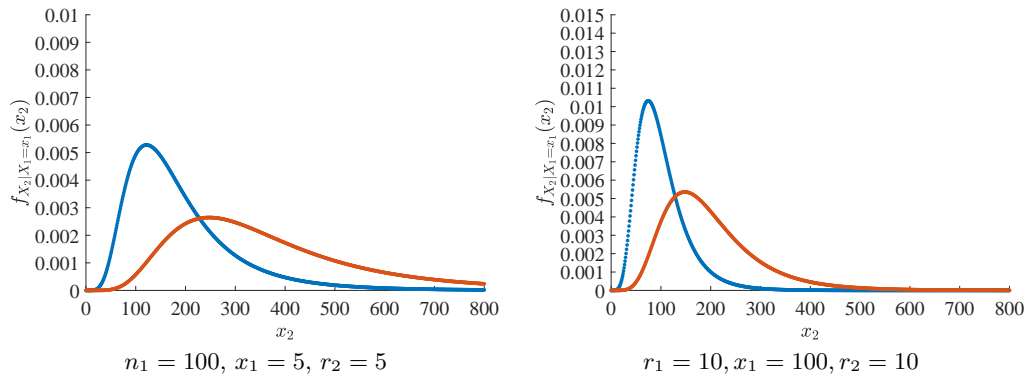


Figure 3.12: Conditional PMFs for prior Beta distribution with parameters $a = 1$, $b = 1$ (blue), and $a = 1$, $b = 99$ (red) for the binomial-negative binomial case (left) and the negative binomial case (right).

3.11 The Sensitivity of Prediction Spaces in Prior Information for a Negative Binomial Prediction Target

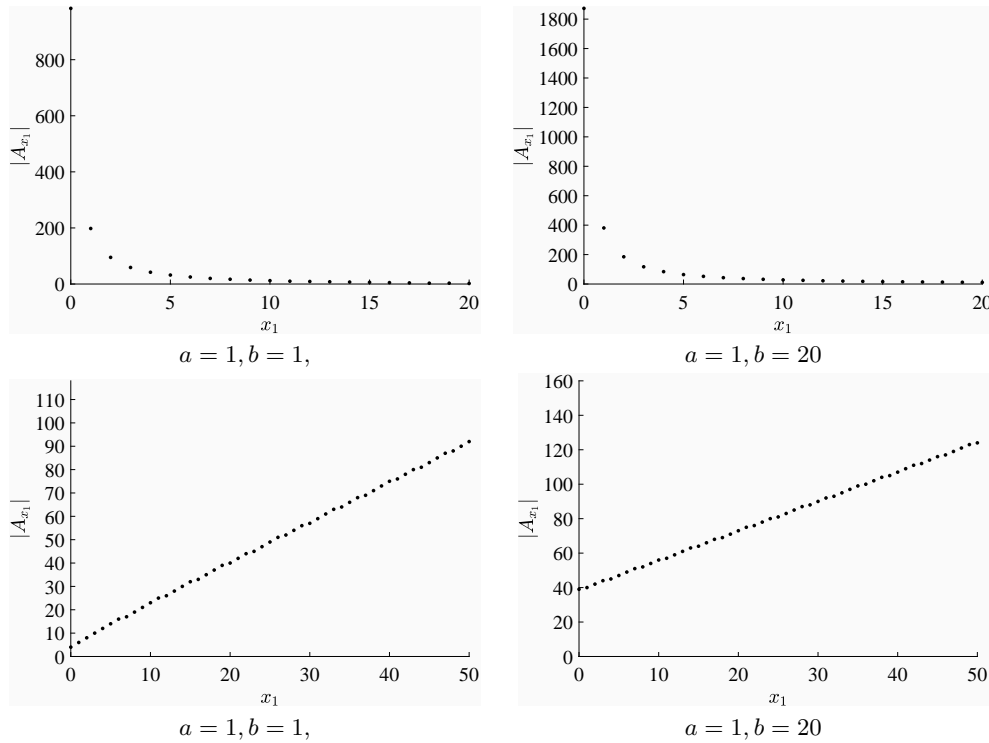


Figure 3.13: Pointwise length of the level-90 x_1 -pointwise MWV for a negative binomial count. Predictor binomial (top) with $n_1 = 20, r_2 = 5$ and predictor negative binomial (bottom) $r_1 = 5, r_2 = 5$, nominal confidence level $\gamma = 0.90$.

As in section 3.6 the underlying probability y should take effect on properties of the prediction space, without being oversensitive in the parameters a and b . These requirements on the prior information expressed by a beta distribution $\text{Beta}(a,b)$ are independent of the prediction target. We proceed similarly to section 3.6 to study the core characteristics of x_1 -pointwise level prediction spaces for a negative binomial prediction target. Throughout, we assume the confidence level $\gamma = 0.90$, and compare priors $\text{Beta}(1.0, b)$ with varying b under fixed $a = 1.0$.

Figure 3.13 illustrates the effect of prior information on interval length by comparing the length of single prediction intervals under the noninformative prior $\text{Beta}(1.0, 1.0)$ with the prior $\text{Beta}(1.0, 20.0)$ for a binomial and a negative binomial predictor X_1 . The

Beta(1.0, 20.0) prior has a right-skewed strictly decreasing PDF which puts large weight on small $y = p$. The interval lengths differ significantly for the different predictors. For the binomial predictor in the upper part of figure 3.13, the interval length decreases for increasing realizations x_1 . The decrease in the interval length is weaker for the Beta(1.0, 20.0) prior as this prior indicates that small values of p are more likely. Therefore, it is more likely that a high number of trials is needed to reach the required number of successes. For the negative predictor in the lower part of figure 3.13, the interval length increases for increasing realizations of X_1 . The interval for any realization is longer for the Beta(1.0, 20.0) prior, again for the reason that this prior indicates smaller values of $y = p$ to be more likely.

3.12 Coverage Properties for a Negative Binomial Prediction Target

As in section 3.7, we study the x_1 -pointwise coverage of the prediction interval for a negative binomial prediction target X_2 . Under the beta prior model of section 3.8, the pointwise coverage $P_{x_1}(X_2 \in A_{x_1})$ considered by definition 3.2.1 amounts to

$$\begin{aligned} P_{x_1}(X_2 \in A_{x_1}) &= \sum_{x_2=0}^{\infty} f_{X_2|X_1=x_1}(x_2) \mathbb{1}_{A_{x_1}}(x_2) \\ &= \sum_{x_2=0}^{\infty} \binom{x_2 + r_2 - 1}{x_2} \frac{B(x_1 + r_2 + a, n_1 - x_1 + x_2 + b)}{B(x_1 + a, n_1 - x_1 + b)} \mathbb{1}_{A_{x_1}}(x_2) \end{aligned} \quad (3.36)$$

for the case of a binomial predictor, where $\mathbb{1}_B$ denotes the indicator function of a set B . For the case of a negative binomial predictor, we have

$$\begin{aligned} P_{x_1}(X_2 \in A_{x_1}) &= \sum_{x_2=0}^{\infty} f_{X_2|X_1=x_1}(x_2) \mathbb{1}_{A_{x_1}}(x_2) \\ &= \sum_{x_2=0}^{\infty} \binom{x_2 + r_2 - 1}{x_2} \frac{B(r_1 + r_2 + a, x_1 + x_2 + b)}{B(r_1 + a, x_1 + b)} \mathbb{1}_{A_{x_1}}(x_2). \end{aligned} \quad (3.37)$$

For a correct prior, the coverage of (3.36) and (3.37) is at least the nominal level. This is shown in figure 3.14. In the top of figure 3.14, large realizations x_1 indicate a high success probability and therefore very small realizations x_2 , whereas in the bottom large realizations x_1 indicate a small success probability and thus large realizations x_2 . In the bottom part, the coverage is very close to the nominal level for large realizations x_1 .

This can be explained by the unbounded support of the negative binomial distribution.

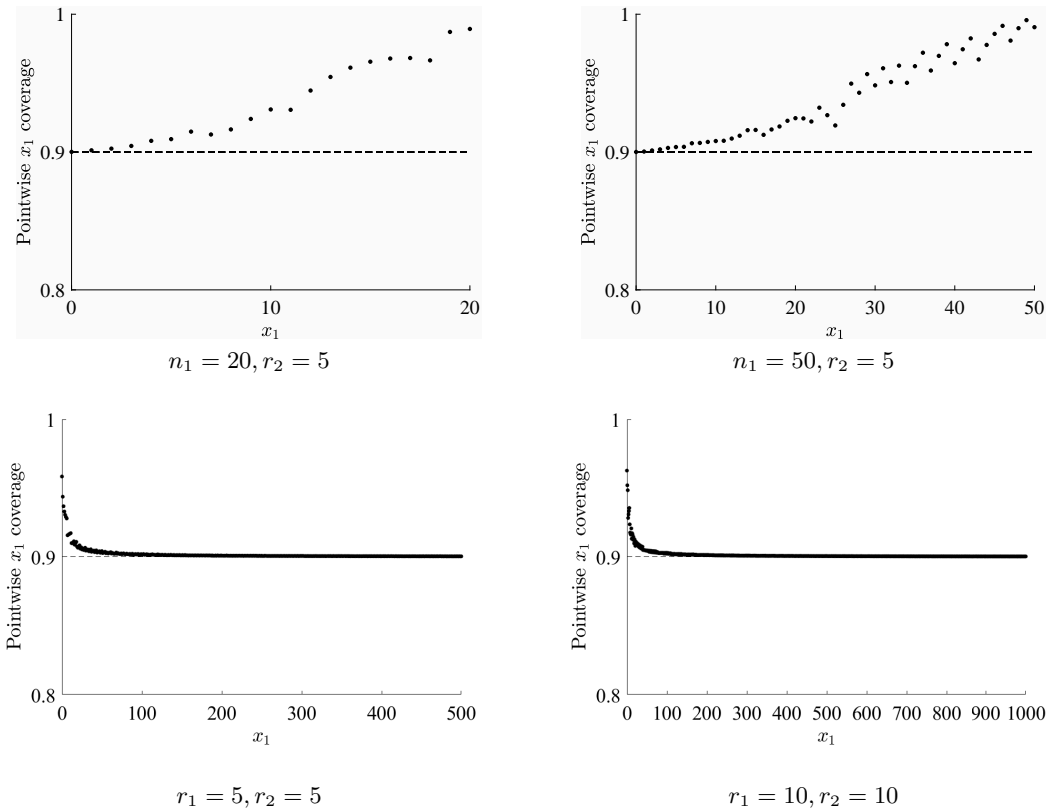


Figure 3.14: Pointwise x_1 coverage of the level-90 x_1 -pointwise MWV for a negative binomial prediction target. Predictor binomial (top) and predictor negative binomial (bottom), nominal confidence level $\gamma = 0.90$.

We now study the effect of misspecified prior information on the coverage, using the coverage concept (3.36) for the binomial-negative binomial case and (3.37) for the negative binomial case pointwise in x_1 .

In figure 3.15 and figure 3.16 the coverage is plotted for the same misspecified priors that were used in section 3.7, see figure 3.9, in particular. As in section 3.7, using conservative priors, i.e. overestimating the true quantile is the safe option. Underestimating the true quantile can result in coverage violations in relevant areas of the parameter spaces, as can be seen on the left of figure 3.15. Overall, the sensitivity with respect to misspecification is not significant, resulting in only minor coverage level violations.

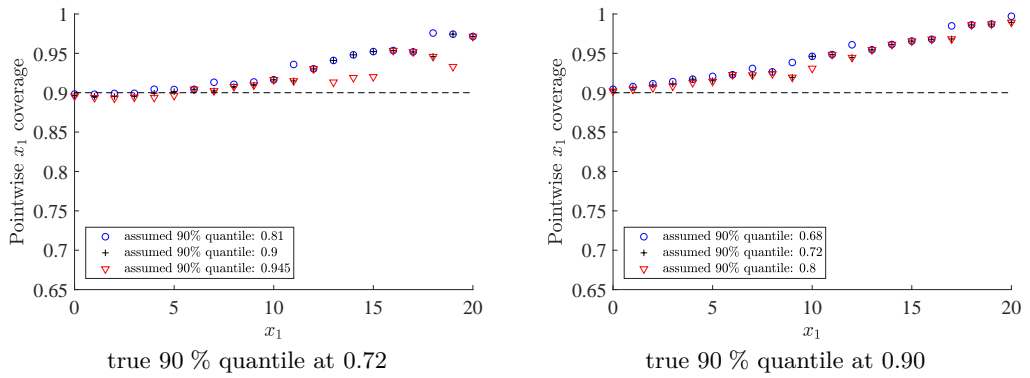


Figure 3.15: Coverage probability pointwise in x_1 for misspecified priors under sample size $n_1 = 20$, and number of required successes $r_2 = 5$ and confidence level $\gamma = 0.90$. Priors $Beta(a, b)$ with $a = 1$, parameter b determined by quantile prescription.

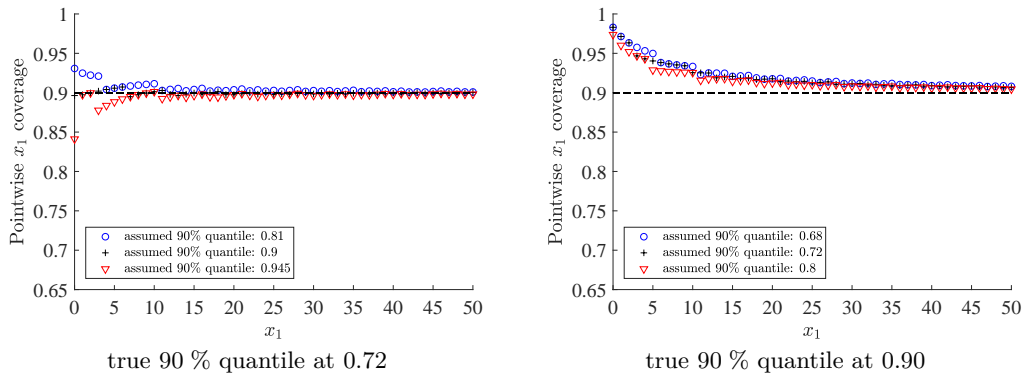


Figure 3.16: Coverage probability pointwise in x_1 for misspecified priors under required number of successes $r_1 = 5$, $r_2 = 5$ and confidence level $\gamma = 0.90$. Priors $Beta(a, b)$ with $a = 1$, parameter b determined by quantile prescription.

3.13 Conclusion

We have introduced a general framework of x_1 -pointwise level γ minimum weighted volume prediction spaces under prior information, and we have applied the framework to binomial counts and negative binomial counts. In the former instance, the properties of the prediction spaces are very satisfactory. The prediction intervals react on changing prior information, but they are not oversensitive. The usage of prior information enables strong reductions in the sample size.

For the case of predicting a negative binomial random variable, the prior information also affects the prediction interval in the desired manner. However, due to the properties of the conditional PMFs $f_{X_2|X_1=x_1}$ for the binomial-negative binomial and the negative binomial case, the lower bound does not change considerably for large expected X_2 , even when using strong prior information. Therefore, stronger prior information does not necessarily lead to a shorter interval.

Many topics remain for future research. The established framework of x_1 -pointwise level γ MWV prediction spaces can be applied to many other prediction problems on discrete or continuous random quantities. It is also important to develop and apply MWV prediction spaces under prior information where the confidence level is prescribed pointwise in a relevant distribution parameter instead of pointwise in the predictor x_1 .

Appendix of Chapter 3

3.A1 Proof of Proposition 3.4.1

From the well-known recursion $\Gamma(x+1) = x\Gamma(x)$ for the gamma function, from the definition (3.9) of the beta function, and from the equation

$$\binom{-x}{k} = (-1)^k \binom{x+k-1}{k} \quad \text{for } x \in \mathbb{R}, k \in \mathbb{Z} \quad (3.38)$$

for binomial coefficients we obtain for the beta function the recursion

$$\begin{aligned} B(s+k, t+m) &= B(s, t) \prod_{i=0}^{k-1} \frac{s+i}{s+t+m+i} \prod_{j=0}^{m-1} \frac{t+j}{s+t+j} \\ &= B(s, t) \frac{\binom{s+k-1}{k} \binom{t+m-1}{m}}{\binom{s+t+k+m-1}{k+m}} \frac{1}{\binom{m+k}{m}} = B(s, t) \frac{\binom{-s}{k} \binom{-t}{m}}{\binom{-s-t}{k+m}} \frac{1}{\binom{m+k}{m}} \end{aligned} \quad (3.39)$$

for $s, t > 0, k, m \in \mathbb{N}_0$.

Consider assertion (i) of proposition 3.4.1. Using the integral definition of the beta function, recursion (3.39) with $s = a, k = x_1 + x_2, t = b, m = n_1 + n_2 - x_1 - x_2$, and equation (3.38) we obtain

$$\begin{aligned} f_{X_1, X_2}(x_1, x_2) &= \int_{[0,1]} f_{X_1, X_2|Y=y}(x_1, x_2) f_Y(y) dy \\ &= \frac{\binom{n_1}{x_1} \binom{n_2}{x_2}}{B(a, b)} \int_{[0,1]} y^{x_1+x_2+a-1} (1-y)^{n_1+n_2-x_1-x_2+b-1} dy \\ &= \frac{\binom{n_1}{x_1} \binom{n_2}{x_2}}{B(a, b)} \frac{B(a+x_1+x_2, b+n_1+n_2-x_1-x_2)}{B(a, b)} \\ &= \frac{\binom{n_1}{x_1} \binom{n_2}{x_2}}{\binom{n_1+n_2}{x_1+x_2}} \frac{\binom{a+x_1+x_2-1}{x_1+x_2} \binom{b+n_1-x_1+n_2-x_2-1}{n_1+n_2-x_1-x_2}}{\binom{a+b+n_1+n_2-1}{n_1+n_2}} \\ &= \frac{\binom{n_1}{x_1} \binom{n_2}{x_2}}{\binom{n_1+n_2}{x_1+x_2}} \frac{\binom{-a}{x_1+x_2} \binom{-b}{n_1+n_2-x_1-x_2}}{\binom{-a-b}{n_1+n_2}} \end{aligned}$$

as asserted by equation (3.12).

Equation (3.13) in assertion (ii) is obtained analogously to the above proof of equation

(3.12). Equation (3.14) in assertion (iii) is obtained from (3.12) and (3.13) by

$$\begin{aligned}
f_{X_2|X_1}(x_2) &= \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_1}(x_1)} \\
&= \frac{\binom{n_1}{x_1} \binom{n_2}{x_2} B(a + x_1 + x_2, b + n_1 + n_2 - x_1 - x_2)}{B(a, b)} \\
&\quad \times \frac{B(a, b)}{\binom{n_1}{x_1} B(x_1 + a, n_1 + b - x_1)} \\
&= \binom{n_2}{x_2} \frac{B(a + x_1 + x_2, b + n_1 + n_2 - x_1 - x_2)}{B(a + x_1, b + n_1 - x_1)}.
\end{aligned}$$

□

3.A2 Proof of Proposition 3.4.2

Consider assertion (i) of proposition 3.4.2. Let $x_2 \in \{0, \dots, n_2 - 1\}$. We have

$$\binom{n_2}{x_2 + 1} = \frac{n_2!}{(x_2 + 1)!(n_2 - x_2 - 1)!} = \frac{n_2 - x_2}{x_2 + 1} \binom{n_2}{x_2}.$$

Using recursion (3.39) with $s = a + x_1 + x_2$, $t = b + n_1 + n_2 - x_1 - x_2$, $k = 1$, $m = 0$ and the relation $B(x, y - 1) = \frac{x+y-1}{y-1} B(x, y)$ we obtain

$$\begin{aligned}
&B(a + x_1 + x_2 + 1, b + n_1 + n_2 - x_2 - 1 - x_1) \\
&= \frac{a + x_1 + x_2}{b + n_1 + n_2 - x_1 - x_2 - 1} B(a + x_1 + x_2, b + n_1 + n_2 - x_1 - x_2).
\end{aligned}$$

Inserting the latter two results into (3.14) we obtain

$$\begin{aligned}
f_{X_2|X_1=x_1}(x_2 + 1) &= \frac{\binom{n_2}{x_2+1} B(a + x_1 + x_2 + 1, b + n_1 + n_2 - x_1 - x_2 - 1)}{B(a + x_1, b + n_1 - x_1)} \\
&= \frac{n_2 - x_2}{x_2 + 1} \frac{a + x_1 + x_2}{b + n_1 + n_2 - x_1 - x_2 - 1} \\
&\quad \times \frac{\binom{n_2}{x_2} B(a + x_1 + x_2, b + n_1 + n_2 - x_1 - x_2)}{B(a + x_1, b + n_1 - x_1)} \\
&= \frac{n_2 - x_2}{x_2 + 1} \frac{a + x_1 + x_2}{b + n_1 + n_2 - x_1 - x_2 - 1} f_{X_2|X_1=x_1}(x_2).
\end{aligned}$$

Assertion (ii) follows from assertion (i) by elementary calculation. □

3.A3 Proof of Proposition 3.4.3

By Stirling's formula, there is a function $0 < r(x) < \frac{1}{12x}$ such that

$$\Gamma(x) = \sqrt{2\pi} x^{x-0.5} \exp(-x) \exp(r(x)) \quad \text{for } x > 0. \quad (3.40)$$

With definition (3.10) of binomial coefficients we obtain from (3.40) for $n > 0$

$$\begin{aligned} \binom{n}{np(n)} &= \frac{(n+1)^{n+0.5} \exp(1) \exp(r(n))}{\sqrt{2\pi} (np(n)+1)^{np(n)+0.5} (n(1-p(n))+1)^{n(1-p(n))+0.5}} \\ &= \left\{ \sqrt{2\pi np(n)^{np(n)+0.5}} [1-p(n)]^{n[1-p(n)]+0.5} \right\}^{-1} \\ &\quad \times \frac{\left(1 + \frac{1}{n}\right)^{n+0.5} \exp(1) \exp(r(n))}{\left(1 + \frac{1}{np(n)}\right)^{np(n)+0.5} \left(1 + \frac{1}{n[1-p(n)]}\right)^{n[1-p(n)]+0.5}}, \end{aligned} \quad (3.41)$$

where

$$\frac{-1}{12(n(1-p(n))+1)} + \frac{-1}{12(np(n)+1)} < r(n) < \frac{1}{12(n+1)}. \quad (3.42)$$

Let

$$A_1(n) = \frac{1}{\sqrt{2\pi n}} \frac{1}{p(n)^{np(n)+0.5}} \frac{1}{(1-p(n))^{n(1-p(n))+0.5}}. \quad (3.43)$$

With $\lim_{u \rightarrow \infty} (1 + 1/u)^u = \exp(1)$ we obtain from (3.41), (3.42), (3.43)

$$\lim_{n \rightarrow \infty} \frac{\binom{n}{np(n)}}{A_1(n)} = 1. \quad (3.44)$$

With the definition (3.9) of the beta function and by applying Stirling's formula (3.40) for $n > 0$ we obtain

$$\begin{aligned} &B(\alpha + np(n), \beta + n[1-p(n)]) \\ &= \frac{\sqrt{2\pi} [\alpha + np(n)]^{\alpha+np(n)-0.5} [\beta + n[1-p(n)]]^{\beta+n[1-p(n)]-0.5}}{(\alpha + \beta + n)^{\alpha+\beta+n-0.5}} \exp(r(n)) \\ &= \sqrt{\frac{2\pi}{n}} \exp(r(n)) \frac{\left[\frac{\alpha}{np(n)} + 1\right]^{\alpha+np(n)-0.5} \left(\frac{\beta}{n[1-p(n)]} + 1\right)^{\beta+n[1-p(n)]-0.5}}{\left(\frac{\alpha+\beta}{n} + 1\right)^{\alpha+\beta+n-0.5}} \\ &\quad \times p(n)^{\alpha+np(n)-0.5} [1-p(n)]^{\beta+n[1-p(n)]-0.5}, \end{aligned} \quad (3.45)$$

where

$$\frac{-1}{12(\alpha + \beta + n)} < r(n) < \frac{1}{12[\alpha + np(n)]} + \frac{1}{12[\beta + n[1 - p(n)]]}. \quad (3.46)$$

Let

$$A_2(n) = \sqrt{\frac{2\pi}{n}} p(n)^{\alpha-0.5+np(n)} [1 - p(n)]^{\beta-0.5+n[1-p(n)]}. \quad (3.47)$$

With $\lim_{u \rightarrow \infty} (1 + 1/u)^u = \exp(1)$ we obtain from (3.45), (3.46), and (3.47)

$$\lim_{n \rightarrow \infty} \frac{B(\alpha + np(n), \beta + n[1 - p(n)])}{A_2(n)} = \frac{\exp(\alpha) \exp(\beta)}{\exp(\alpha + \beta)} = 1. \quad (3.48)$$

From (3.44) and (3.48) we obtain the limiting relation (3.21) stipulated by proposition 3.4.3. \square

3.A4 Proof of Proposition 3.9.1

Let $x_1, x_2 \in \mathbb{N}_0$, $r_1, r_2 \in \mathbb{N}$ and let X_1, X_2 , and Y be as stated in proposition 3.9.1. Then we obtain

(i) $f_{X_1, X_2}(x_1, x_2)$

$$\begin{aligned} &= \int_{[0,1]} f_{X_1, X_2|Y=y}(x_1, x_2) f_Y(y) dy \\ &= \int_{[0,1]} \binom{x_1 + r_1 - 1}{x_1} \binom{x_2 + r_2 - 1}{x_2} \\ &\quad \times (1 - y)^{x_1} y^{r_1} (1 - y)^{x_2} y^{r_2} \frac{1}{B(a, b)} y^{a-1} (1 - y)^{b-1} dy \\ &= \frac{1}{B(a, b)} \binom{x_1 + r_1 - 1}{x_1} \binom{x_2 + r_2 - 1}{x_2} \int_{[0,1]} y^{r_1+r_2+a-1} (1 - y)^{x_1+x_2+b-1} dy \\ &= \frac{\binom{x_1+r_1-1}{x_1} \binom{x_2+r_2-1}{x_2} B(r_1 + r_2 + a, x_1 + x_2 + b)}{B(a, b)} \end{aligned}$$

(ii)

$$\begin{aligned} f_{X_i}(x_i) &= \int_{[0,1]} f_{X_i|Y=y}(x_i) f_Y(y) dy \\ &= \int_{[0,1]} \binom{x_i + r_i - 1}{x_i} (1 - y)^{x_i} y^{r_i} \frac{1}{B(a, b)} y^{a-1} (1 - y)^{b-1} dy \\ &= \binom{x_i + r_i - 1}{x_i} \frac{1}{B(a, b)} B(r_i + a, x_i + b) \end{aligned}$$

(iii)

$$\begin{aligned} f_{X_2|X_1=x_1}(x_2) &= \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_1}(x_1)} \\ &= \binom{x_2 + r_2 - 1}{x_2} \frac{B(r_1 + r_2 + a, x_1 + x_2 + b)}{B(r_1 + a, x_1 + b)} \end{aligned}$$

(iv)

$$\begin{aligned} Q_{x_1}(x_2) &= \frac{f_{X_2|X_1=x_1}(x_2)}{f_{X_2}(x_2)} \\ &= \frac{B(a, b)B(r_1 + r_2 + a, x_1 + x_2 + b)}{B(r_1 + a, x_1 + b)B(r_2 + a, x_2 + b)}. \end{aligned}$$

□

3.A5 Proof of Proposition 3.9.2

Let $x_1 \in \{0, \dots, n_1\}$, $x_2 \in \mathbb{N}_0$, $n_1, r_2 \in \mathbb{N}$ and let X_1, X_2 , and Y be as stated in proposition 3.9.2. Then we have

(i)

$$\begin{aligned} &f_{X_1, X_2}(x_1, x_2) \\ &= \int_{[0,1]} f_{X_1, X_2|Y=y}(x_1, x_2) f_Y(y) dy \\ &= \int_{[0,1]} \binom{n_1}{x_1} y^{x_1} (1-y)^{n_1-x_1} \\ &\quad \times \binom{x_2 + r_2 - 1}{x_2} y^{r_2} (1-y)^{x_2} \frac{1}{B(a, b)} y^{a-1} (1-y)^{b-1} dy \\ &= \frac{1}{B(a, b)} \binom{n_1}{x_1} \binom{x_2 + r_2 - 1}{x_2} \int_{[0,1]} y^{x_1+r_1+a-1} (1-y)^{n_1-x_1+x_2+b-1} dy \\ &= \frac{1}{B(a, b)} \binom{n_1}{x_1} \binom{x_2 + r_2 - 1}{x_2} B(x_1 + r_2 + a, n_1 - x_1 + x_2 + b). \end{aligned}$$

(ii) See proof of proposition 3.4.1 for the unconditional distribution of X_i for two binomial samples.

(iii) See proof of proposition 3.9.1 for the unconditional distribution of X_i for two negative binomial samples.

(iv)

$$\begin{aligned}
f_{X_2|X_1=x_1}(x_2) &= \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_1}(x_1)} \\
&= \frac{\binom{n_1}{x_1} \binom{x_2+r_2-1}{x_2} B(x_1+r_2+a, n_1-x_1+x_2+b)}{B(a, b)} \\
&\quad \times \frac{B(a, b)}{\binom{n_1}{x_1} B(x_1+a_1, n_1-x_1+b)} \\
&= \binom{x_2+r_2-1}{x_2} \frac{B(x_1+r_2+a, n_1-x_1+x_2+b)}{B(x_1+a_1, n_1-x_1+b)}
\end{aligned}$$

(v)

$$\begin{aligned}
Q_{x_1}(x_2) &= \frac{f_{X_2|X_1=x_1}(x_2)}{f_{X_2}(x_2)} \\
&= \frac{B(a, b) B(x_1+r_2+a, n_1-x_1+x_2+b)}{B(r_2+a, x_2+b) B(x_1+a, n_1-x_1+b)}.
\end{aligned}$$

□

3.A6 Proof of Proposition 3.9.3

Let $x_2 \in \mathbb{N}_0$. Then we have for the conditional PMF $f_{X_2|X_1=x_1}$ given by (3.28)

(i)

$$\begin{aligned}
&f_{X_2|X_1=x_1}(x_2+1) \\
&= \binom{x_2+1+r_2-1}{x_2+1} \frac{B(r_1+r_2+a, x_1+x_2+1+b)}{B(r_1+a, x_1+b)} \\
&= \binom{x_2+1+r_2-1}{x_2+1} \\
&\quad \times \frac{B(r_1+r_2+a, x_1+x_2+b)}{B(r_1+a, x_1+b)} \frac{x_1+x_2+b}{x_1+x_2+b+r_1+r_2+a} \\
&= \binom{x_2+r_2-1}{x_2} \frac{x_2+r_2}{x_2+1} \\
&\quad \times \frac{x_1+x_2+b}{x_1+x_2+b+r_1+r_2+a} \frac{B(r_1+r_2+a, x_1+x_2+b)}{B(r_1+a, x_1+b)} \\
&= \frac{x_2+r}{x_2+1} \frac{x_1+x_2+b}{x_1+x_2+b+r_1+r_2+a} f_{X_2|X_1}(x_2),
\end{aligned}$$

using (3.39) with $s = r_1 + r_2 + a$, $t = x_1 + x_2 + b$, $k = 0$, $m = 1$.

(ii) Follows from (i) by straightforward calculations. \square

3.A7 Proof of Proposition 3.9.4

Let $x_2 \in \mathbb{N}_0$. Then we obtain for the conditional PMF $f_{X_2|X_1=x_1}$ given by (3.32)

(i)

$$\begin{aligned}
 & f_{X_2|X_1=x_1}(x_2 + 1) \\
 &= \binom{x_2 + 1 + r_2 - 1}{x_2 + 1} \frac{B(x_1 + r_2 + a, n_1 - x_1 + x_2 + 1 + b)}{B(x_1 + a, n_1 - x_1 + b)} \\
 &= \binom{x_2 + r_2 - 1}{x_2} \frac{x_2 + r_2}{x_2 + 1} \\
 &\quad \times \frac{n_1 - x_1 + x_2 + b}{r_2 + a + n_1 + x_2 + b} \frac{B(x_1 + r_2 + a, n_1 - x_1 + x_2 + b)}{B(x_1 + a, n_1 - x_1 + b)} \\
 &= \frac{x_2 + r_2}{x_2 + 1} \frac{n_1 - x_1 + x_2 + b}{r_2 + a + n_1 + x_2 + b} f_{X_2|X_1=x_1}(x_2),
 \end{aligned}$$

using (3.39) with $s = x_1 + r_2 + a$, $t = n_1 - x_1 + x_2 + b$, $k = 0$, $m = 1$.

(ii) Follows from (i) by straightforward calculations. \square

4 Prediction Intervals with Unknown Sample Size

4.1 Introduction

Chapter 3 studied discrete prediction intervals, where the prediction interval is obtained from a learning sample, prior information and the confidence level. The future number of failures is also a discrete prediction target, however the setting and therefore the approach is quite different. In this case, each member of the population has a continuous lifetime distribution. The parameters of the lifetime distribution are unknown. The most commonly used lifetime distributions are the Weibull and the lognormal distribution. Oftentimes, a prediction can only be made based on observed field data. The prediction target Y is connected to the data by an unknown vector of parameters θ . The prediction target can be either within the sample or in a new sample, see Escobar & Meeker (1999).

For a continuous distribution, a prediction interval $[Y_l, Y_u]$ which satisfies

$$P_{\theta}(Y \in [Y_l, Y_u]) = \gamma \quad \text{for all } \theta \quad (4.1)$$

exists, if a pivotal quantity is available. A random variable $Z(X, Y)$ is called pivotal, if it does not depend on the unknown θ . For the location-scale family, which includes the Weibull and the lognormal distribution, pivotal quantities are, for example, known for uncensored data. However, these pivots are no longer pivots under certain censoring schemes, see Lawless (2003). As such censoring schemes are customary in field data, the pivotal approach cannot be used. A customary approach is the so-called the naive or plug-in approach. The resulting prediction interval does not take into account the uncertainty of the parameter estimates and therefore often has coverage below the desired nominal confidence level.

Much of the existing literature focuses on improving the naive prediction intervals, either by analytical means, see Beran (1990), Barndorff-Nielsen & Cox (1996), or by

using simulation, see Escobar & Meeker (1999), Meng & Meeker (2011), and Mee & Kushary (1994). This problem is also related to finding a predictive distribution, see Lawless & Fredette (2005). Escobar & Meeker (1999), and Meng & Meeker (2011) study the problem of predicting the future number of failures based on censored data. In their studies, initially n units were put into service at time 0. We extend their setting to the subgroup problem described below.

Frequently, in modern mass production a specific deviation occurs for a small subgroup within the total population. These parts have a different lifetime distribution than the rest of the population. Therefore, the number of future failures from this subpopulation is the quantity of interest. However, the exact size of this subpopulation is unknown, as the parts are in the field and cannot be distinguished unless they fail. Hence, prediction intervals for this problem have to account for additional uncertainty, since the population size is not known. We consider a parametric framework, where individual lifetimes are given by a Weibull distribution, although the methodology may be applied to other continuous lifetime distributions.

The rest of this chapter is structured as follows. In section 4.2 the prerequisites and the model of the study are discussed. Then, section 4.3, provides the concepts for prediction intervals in the framework of a known population size. Section 4.4 discusses prediction intervals with unknown sample size. The final section 4.5 provides a summary and points out future work.

4.2 Background

4.2.1 Prerequisites

We want to predict the future number of failures of a population, which is a realization of a random quantity. To predict such a random quantity, two requirements must be fulfilled.

- (i) A statistical model has to be defined which describes the process of interest for the population. There is an important distinction between parametric models, where the model consists of a probability distribution depending on a vector θ of variables and nonparametric models.
- (ii) Information about the parameters θ and the population must exist. The information about θ can come from experiments, field data or expert knowledge.

We only consider parametric models, where the lifetimes are described by a continuous probability distribution with CDF $F(t, \theta)$, and a corresponding PDF $f(t, \theta)$. For prediction intervals for nonparametric models, see for example Hahn & Meeker (1991). We will focus on the case where θ is unknown and has to be estimated. In many cases, θ has to be estimated from censored data. For censored data, there are two standard assumptions which we will also use, see Lawless (2003).

Assumption 4.2.1. (i) *The times of the events of interest are statistically independent.*

(ii) *The censoring times are independent of any event times.*

4.2.2 Model

We consider the situation where n units enter the field at time 0. Of these n units, S have a deviation from their nominal state. A unit with deviation cannot be distinguished from a unit without deviation unless it fails. S is unknown, but a $\gamma_1 \times 100\%$ prediction interval $[S_l, S_u]$ is available. We consider a parametric framework where the lifetime of a member of S is determined by a Weibull distribution with CDF

$$F(t, \eta, b) = 1 - \exp\left(-\left(\frac{t}{\eta}\right)^b\right). \quad (4.2)$$

The lifetimes for the $n - S$ units without the deviation are of no interest and can be considered as ∞ for simplicity. At some censoring time $t_c > 0$, $r > 1$ failures occurred, whose exact failure times are recorded. The remaining $n - r$ units are still in operation and consequently have censored lifetimes. Based on this censored data, the goal is to obtain a $\gamma \times 100\%$ prediction interval for the additional number of failures in a future time period $(t_c, t_p]$ from the $S - r$ remaining units with deviation.

We consider S to be unknown, but a level γ_1 prediction interval $[S_l, S_u]$ is available for S , i.e. $P(S \in [S_l, S_u]) \geq \gamma_1$.

4.2.3 Coverage Concepts

For a given CDF $F(t, \theta)$ of continuous type with known θ , an equal-tailed two-sided $100\gamma\%$ prediction interval $[T_l, T_u]$ for a future observation is obtained by

$$[T_l, T_u] = [q_{\frac{1-\gamma}{2}}, q_{\frac{1+\gamma}{2}}], \quad (4.3)$$

where q_α denotes the 100α % quantile of $F(t, \theta)$. Consequently,

$$P_\theta(T \in [T_l, T_u]) = \gamma.$$

In practice, the prediction interval (4.3) is unavailable since θ is unknown and needs to be estimated. Hence, it is required to obtain prediction intervals procedures, when θ is estimated.

A given data set, which includes the sample result and the setting such as censoring times, leads to a parameter estimate $\hat{\theta}$. We focus on $\hat{\theta}$ being the maximum likelihood estimator (MLE) of θ , although other estimation procedures would also be possible and can be adapted in a straightforward manner. The target is to obtain an interval $[T_l, T_u]$ for a future observation with a nominal level of γ . The nominal level $\gamma \in (0, 1)$ is a lower bound for the actual coverage probability of the prediction interval, i.e.

$$P_\theta(T \in [T_l, T_u]) \geq \gamma \quad \text{for all } \theta. \quad (4.4)$$

We determine intervals $[T_l, T_u] = [T_l(\hat{\theta}), T_u(\hat{\theta})]$ as functions of $\hat{\theta}$. Conditional on the parameter estimate $\hat{\theta}$, a prediction interval $[T_l, T_u] = [T_l(\hat{\theta}), T_u(\hat{\theta})]$ has the conditional coverage probability

$$P_{\theta, \hat{\theta}}(T \in [T_l, T_u]) = P(T_l \leq T \leq T_u | \theta, \hat{\theta}) \quad (4.5)$$

$$= F(T_u(\hat{\theta}), \theta) - F(T_l(\hat{\theta}), \theta). \quad (4.6)$$

The latter coverage probability cannot be analyzed because θ is unknown. The estimated data will differ from sample to sample for the same setting. Thus, the estimate $\hat{\theta}$ will also vary, leading to a different prediction interval $[T_l, T_u]$. The prediction interval procedure can be evaluated using the unconditional coverage probability

$$P_\theta(T \in [T_l, T_u]) = P(T_l \leq T \leq T_u | \theta) \quad (4.7)$$

$$= \mathbb{E}_{\hat{\theta}}[P(T_l \leq T \leq T_u | \theta, \hat{\theta})] \quad (4.8)$$

$$= \int_{\hat{\theta}} P(T_l \leq T \leq T_u | \theta, \hat{\theta}) h(\hat{\theta}) d\hat{\theta}, \quad (4.9)$$

where h is the PDF of $\hat{\theta}$. This unconditional coverage probability can be evaluated or approximated using numerical procedures, see for example Escobar & Meeker (1999).

4.3 Prediction Intervals with Known Sample Size

4.3.1 Naive Prediction Intervals

A straightforward way of obtaining a prediction interval is to use the approach from equation (4.3) and substitute the unknown θ by the ML estimate $\hat{\theta}$. The resulting prediction interval is thus

$$[T_l, T_u] = \left[\hat{q}_{\frac{1-\gamma}{2}}, \hat{q}_{\frac{1+\gamma}{2}} \right], \quad (4.10)$$

where \hat{q}_α denotes the α level quantile of $F(t, \hat{\theta})$. This is called the “plug-in” or “naive” prediction interval. It does not account for the uncertainty of the ML estimate $\hat{\theta}$ with respect to θ . Therefore, the unconditional coverage can be significantly below the nominal level. If $\hat{\theta}$ has good asymptotic properties, the naive prediction interval will be a good approximate prediction interval for large sample sizes.

4.3.2 Prediction of the Number of Future Failures

So far, we have only considered individual lifetimes, but not yet the aggregated population of size n which we are interested in. A unit which is still operating at the censoring time t_c can either fail or survive the time period $(t_c, t_p]$ and all units belonging to the population of size n have the same lifetime distribution (4.2). Hence, conditional on the number of failures r at time t_c , the number of failures Y in $(t_c, t_p]$ has a binomial distribution with parameters $n - r$ and ρ , where

$$\rho = \frac{F(t_p, \eta, b) - F(t_c, \eta, b)}{1 - F(t_c, \eta, b)}. \quad (4.11)$$

Therefore,

$$P(Y \leq k) = \sum_{j=0}^k \binom{n-r}{j} \rho^j (1-\rho)^{n-r-j}. \quad (4.12)$$

The ML estimate $\hat{\rho}$ of ρ is obtained by plugging the ML estimates $\hat{\eta}$ and \hat{b} of the Weibull parameters η and b into (4.11). It is important to note, that the number of failures r has an effect on (4.12) in two ways: i) on the remaining population $n - r$ which is directly visible in (4.12), whereas ii) is more implicit, as the ML estimates $\hat{\eta}$ and \hat{b} are based on the r exact failure times and the $n - r$ survivors and thus change with r . Therefore, $\hat{\rho}$ changes with r .

Let $\delta_l, \delta_u \in (0, 1)$. Plugging $\hat{\rho}$ into (4.12), we solve for Y_l , the largest k which satisfies

$$P_{\hat{\rho}}(Y < k) = \sum_{j=0}^{k-1} \binom{n-r}{j} \hat{\rho}^j (1 - \hat{\rho})^{n-r-j} \leq \delta_l \quad (4.13)$$

and for Y_u , the smallest k which satisfies

$$P_{\hat{\rho}}(Y \leq k) = \sum_{j=0}^k \binom{n-r}{j} \hat{\rho}^j (1-\hat{\rho})^{n-r-j} \geq \delta_u. \quad (4.14)$$

For $\delta_l = \frac{1-\gamma}{2}$, $\delta_u = \frac{1+\gamma}{2}$, $[Y_l, Y_u]$ is called the naive two-sided level γ prediction interval. Consequently, for $\delta_l = 1-\gamma$, Y_l is the naive one-sided level γ lower prediction bound and for $\delta_u = \gamma$, Y_u is the naive one-sided level γ upper prediction bound. The inequalities in (4.13) and (4.14) are caused by the discreteness of the binomial distribution which generally prevents an exact solution of (4.13) and (4.14) as an equation.

4.3.3 Calibration

As mentioned before, the naive level γ prediction interval does not have actual coverage γ . Let $H(\gamma, \eta, b)$ be the actual coverage of the naive prediction interval. The idea of calibration is to find an approximation \bar{H} of H , depending only on γ and not on the unknown parameters η and b . In the framework of Escobar & Meeker (1999), this is done by using the ML estimates and finding a γ' such that the level γ' naive prediction interval has coverage of approximately γ , i.e. $\bar{H}(\gamma') = \gamma$. The following algorithm by Escobar & Meeker (1999) can be used to calibrate a naive level γ prediction interval for a population of size S . This is done by combining the lower and upper one-sided level $(1+\gamma)/2$ prediction bounds. Let $r > 1$ be the number of failures at the censoring time t_c , let t_p be the prediction time, let $\theta = (\eta, b)$ be the Weibull parameters, and let $\hat{\theta} = (\hat{\eta}, \hat{b})$ be the ML estimate.

Algorithm 4.3.1 (Calibration). (i) Choose a number of simulation runs B .

(ii) Choose a confidence level $\gamma' > 0$.

(iii) For each simulation j , $j = 1, \dots, B$, sample S failure times from $F(t, \hat{\theta})$ and apply the censoring pattern at t_c . Obtain the number of failures r'_j at the censoring time t_c , and the ML estimates $\hat{\theta}'_j$ of the simulated data.

(iv) Calculate $\hat{\rho}'_j$ and the level γ' naive prediction interval $[S'_{l,j}, S'_{u,j}]$ from the $\text{Bin}(S - r'_j, \hat{\rho}'_j)$ distribution.

(v) For each simulation j , calculate

$$P'_{l,j} = 1 - \sum_{k=0}^{S'_{l,j}-1} \binom{S-r'_j}{k} \hat{\rho}^k (1-\hat{\rho})^{S-r'_j-k}, \quad (4.15)$$

$$P'_{u,j} = \sum_{k=0}^{S'_{u,j}} \binom{S-r'_j}{k} \hat{\rho}^k (1-\hat{\rho})^{S-r'_j-k}. \quad (4.16)$$

(vi) The unconditional coverage can be estimated by

$$\bar{H}_l(\gamma') = \frac{1}{B} \sum_{j=1}^B P'_{l,j}, \quad (4.17)$$

$$\bar{H}_u(\gamma') = \frac{1}{B} \sum_{j=1}^B P'_{u,j}. \quad (4.18)$$

Steps 1 – 6 are repeated until values γ'_l, γ'_u are found such that (4.17) and (4.18) are equal to $\frac{1+\gamma}{2}$.

4.4 Prediction Intervals with Unknown Population Size

4.4.1 Monotonicity of the ML Estimate for the Exponential Distribution

We will show, that for a fixed number of exact observations, the ML estimate $\hat{\theta}$ is increasing in the sample size n . Let $t_c > 0$ be the censoring time. Let

$$t_i^* = \begin{cases} t_i, & \text{if } t_i \leq t_c, \\ t_c, & \text{if } t_i > t_c. \end{cases}$$

For type I censored data with r exact observations and sample size n , the ML estimate is $\hat{\theta} = \frac{\sum_{i=1}^n t_i^*}{r}$, see for example Lawless (2003). Let $n_1, n_2 \in \mathbb{N}, n_1 < n_2$ be two sample sizes. Let $r \in \mathbb{N}$ be the number of exact observations at some censoring time $t_c > 0$. Then we have

$$\begin{aligned} \hat{\theta}_2 &= \sum_{i=1}^{n_2} \frac{t_i^*}{r} = \sum_{i=1}^{n_1} \frac{t_i^*}{r} + \frac{(n_2 - n_1)t_c}{r} \\ &\geq \sum_{i=1}^{n_1} \frac{t_i^*}{r} = \hat{\theta}_1. \end{aligned}$$

Let $X \sim \text{Wbl}(\eta, b)$, then $X^b =: t \sim \text{Exp}(\theta)$, with $\theta = \eta^b$. Therefore, for a Weibull distribution with known shape parameter b , with failure times $x_i, i = 1, \dots, n$ we have the ML estimators

$$\hat{\theta} = \frac{\sum_{i=1}^n t_i}{r} = \frac{\sum_{i=1}^n x_i^b}{r}, \quad (4.19)$$

$$\hat{\eta} = \left(\frac{\sum_{i=1}^n t_i}{r} \right)^{\frac{1}{b}} = \left(\frac{\sum_{i=1}^n x_i^b}{r} \right)^{\frac{1}{b}}. \quad (4.20)$$

Thus, the monotonicity of the estimate $\hat{\eta}$ of the Weibull scale parameter η for a known shape parameter b follows from the monotonicity of the estimate of the exponential parameter θ .

For the ML estimate $\hat{\theta}$ of the parameter θ of an exponential distribution, we have for $\hat{\rho}$ as introduced in (4.11) and $t_c < t_p$

$$\begin{aligned} \hat{\rho} &= \frac{F(t_p, \hat{\theta}) - F(t_c, \hat{\theta})}{1 - F(t_c, \hat{\theta})} = \frac{\left(1 - \exp\left(-\frac{t_p}{\hat{\theta}}\right)\right) - \left(1 - \exp\left(-\frac{t_c}{\hat{\theta}}\right)\right)}{1 - \left(1 - \exp\left(-\frac{t_c}{\hat{\theta}}\right)\right)} \\ &= \frac{\exp\left(-\frac{t_c}{\hat{\theta}}\right) - \exp\left(-\frac{t_p}{\hat{\theta}}\right)}{\exp\left(-\frac{t_c}{\hat{\theta}}\right)} = 1 - \frac{\exp\left(-\frac{t_p}{\hat{\theta}}\right)}{\exp\left(-\frac{t_c}{\hat{\theta}}\right)}. \end{aligned}$$

Differentiating with respect to $\hat{\theta}$ yields

$$\begin{aligned} \frac{d\hat{\rho}}{d\hat{\theta}} &= \frac{-\frac{t_p}{\hat{\theta}^2} \exp\left(-\frac{t_p}{\hat{\theta}}\right) \exp\left(-\frac{t_c}{\hat{\theta}}\right) - \left(-\exp\left(-\frac{t_p}{\hat{\theta}}\right)\right) \frac{t_c}{\hat{\theta}^2} \exp\left(-\frac{t_c}{\hat{\theta}}\right)}{\left(\exp\left(-\frac{t_c}{\hat{\theta}}\right)\right)^2} \\ &= \frac{-\exp\left(-\frac{t_p}{\hat{\theta}}\right) \left(\frac{t_p - t_c}{\hat{\theta}^2}\right)}{\exp\left(-\frac{t_c}{\hat{\theta}}\right)} < 0. \end{aligned}$$

It follows that $\hat{\rho} = \hat{\rho}(\hat{\theta})$ is monotone decreasing in $\hat{\theta}$. Consequently, $\hat{\rho} = \hat{\rho}(n, r)$ is monotone decreasing in n for fixed $r \in \mathbb{N}$.

4.4.2 Effect of Population Size

As can be seen in equation (4.12), the resulting prediction interval depends on the population size S . Indirectly, the size of the population becomes also present in equation (4.11), when substituting the ML estimates $\hat{\eta}$ and \hat{b} into the Weibull distribution. For the ML estimation of the Weibull parameters, the number of exact and censored observations is required. While the number of exact observation is fixed, the number of

censored observations requires the population size. Thus, for different population sizes, the resulting ML estimates are different and consequently the resulting $\hat{\rho}$ is different.

To adopt the calibration approach for an unknown S , the effect of the population size on the resulting prediction interval has to be considered. The first step is to adapt the censored data appropriately to the new setting. If the assumed population S_l size is smaller than the actual population size S , then $S - S_l$ censored observations have to be removed. Thus, there are r exact observations and $S_l - r$ censored observations at t_c in the modified data. Accordingly, if the assumed population S_u is larger than S , then $S_u - S$ censored observations have to be added to the data. The modified data consists of r exact and $S_u - r$ censored observations at t_c in this case.

The first observation from our simulations is that the ML estimate of the Weibull parameters for different population sizes shows little variation in the shape parameter b , while the variation in the scale parameter η is high. However, the variation in the estimate of η also depends on the number of failures and the value of the censoring time t_c . For the same number of failures and the same censoring time, the ML estimate $\hat{\eta}$ is increasing with increasing population size. The ML estimate \hat{b} is decreasing with increasing population size.

The behavior of the ML estimates $\hat{\eta}$ and \hat{b} with respect to the population size determines the effect of the population size on $\hat{\rho}$. For a fixed number of failures at a censoring time t_c , a larger population size S will lead to a decrease in the value of $\hat{\rho}$. The latter observation imposes the question of whether or not, the increase in the population size will outweigh the decrease in $\hat{\rho}$.

An obvious starting point to answer this question would be to look at the expected number of failures in the interval $(t_c, t_p]$ for different population sizes and the resulting ML estimates $\hat{\rho} = \hat{\rho}(S)$ for the binomial proportion ρ . The expected value of a binomial random variable Y with parameters S and ρ is $E[Y] = S\rho$. If the ML estimate $\hat{\rho}(S)$ of the true population S is small, the (absolute) variation in $\hat{\rho}$ for different population sizes is relatively small. This causes $S_l \times \hat{\rho}(S_l)$, and $S_u \times \hat{\rho}(S_u)$ to be similar even if the interval $[S_l, S_u]$ for the population size is quite wide. For larger values of $\hat{\rho}(S)$, the variation in $\hat{\rho}$ is also larger. In this case,

$$E[Y_l(S_l)] = S_l \times \hat{\rho}(S_l) < S_u \times \hat{\rho}(S_u) = E[Y_u(S_u)]. \quad (4.21)$$

The difference of the expected values in (4.21) can be significant, if the interval $[S_l, S_u]$ is wide. It should be noted that (4.21) might not hold in general. For cases with small

variation in $\hat{\rho}$ it can occur that $E[Y_l(S_l)] \approx E[Y_u(S_u)]$, although we have not observed a case where the order is reversed.

4.4.3 Monotonicity Approach

The idea of the monotonicity approach is simple. Based on (4.21), the resulting naive prediction intervals should be such that the lower bound from the smaller population S_l is smaller than the lower bound from the larger population S_u . Similarly, the naive upper bound from the larger population S_u should be larger than the upper bound from the smaller population S_l . Since the naive interval is considered as the starting interval for the calibration, the monotonicity should be kept within the calibration. Our studies have confirmed this with one minor exception, which we will discuss below.

For a given interval $[S_l, S_u]$, the following approach can be used to obtain a prediction interval for the number of failures.

Algorithm 4.4.1 (Monotonicity Approach). (i) Choose a nominal confidence level γ for the calibration procedure.

(ii) Obtain the level γ calibrated prediction interval for $S = S_l$ and $S = S_u$.

(iii) Choose the smaller calibrated lower bound $Y_l = \min(Y_l(S_l), Y_l(S_u))$, and the larger calibrated upper bound $Y_u = \max(Y_u(S_l), Y_u(S_u))$ to obtain the prediction interval $[Y_l, Y_u]$.

As mentioned above, the value $S \hat{\rho}(S)$ is not monotone in S . In the case of a small variation in $\hat{\rho}$, which can occur for small values of $\hat{\rho}$, it can occur that the naive prediction intervals are the same for the different population sizes. In this case, it can occur that calibrated lower prediction bound of the larger population is smaller than the calibrated lower bound of the smaller population. Similarly, it can happen that the calibrated upper bound of the smaller population is larger than the calibrated upper bound of the larger population.

However, we have not observed a case where the calibrated upper bound of the smaller population is larger than the calibrated upper bound of the larger population by more than 1. We have also not observed a case, where the calibrated lower bound of the larger population is smaller than the calibrated lower bound of the smaller population by more than 1. Furthermore, when this effect occurs, the influence of the population size on the final prediction interval is small, or sometimes even not existing. This implies that the

exact size of the subgroup with deviation does not have a significant influence on the number of failures in the time period of interest.

In applications, this is valuable information as investing more resources in reducing the uncertainty of the affected population will offer little benefit as long as the prediction period will not have to be extended or the ML estimation changes considerably, e.g. due to additional failures which were not reported or were delayed. We have not observed a case where the monotonicity is violated for the naive prediction intervals.

If the difference in (4.21) is increasing, so does the effect of the population size. In this case, reducing the uncertainty of the population size, which will result in a shorter prediction interval for the population size, will lead to a much shorter prediction interval for the number of failures in $(t_c, t_p]$. Thus, obtaining a shorter prediction interval for the population size should be targeted in this case. We now provide two examples to illustrate the effect of the population size on the calibrated prediction interval.

Example 4.4.2 (Monotonicity approach). *Escobar & Meeker (1999) considered the case where $S = 10000$ units were put into service. After $t_c = 48$ months, $r = 80$ units had failed. They reported the ML estimates of the Weibull distribution as $\hat{\eta} = 1152$, and $\hat{b} = 1.51$. We adopt this case and take $[8202, 12038]$ as the prediction interval for the population size. For illustration purposes, this interval was chosen such that $S = 10000$ is approximately in the center of the interval. The interval is the result from Thatcher's (1964) procedure with a learning sample of size 1000 with 100 'successes' and a confidence level of $\gamma_1 = 0.95$ and a total population of 100000. We created a data set from the true Weibull parameters $\eta = 1152, b = 1.51$ and obtained $r = 71$ failures at $t_c = 48$. The prediction time period is the next 12 months, thus $t_p = 60$. The resulting ML estimates for the binomial parameter ρ are*

$$\hat{\rho}(S_l) = 0.0039, \hat{\rho}(S) = 0.0032, \hat{\rho}(S_u) = 0.0027.$$

The resulting $\gamma = 0.9$ naive prediction intervals are $[23, 42]$ for all cases and the calibrated prediction intervals are $[20, 45]$, which was also the result in the original example. Therefore, $[20, 45]$ would be the calibrated prediction interval.

Example 4.4.2 illustrates the case, where the population size has no influence on the final prediction interval, as long as it is contained in the prediction interval for the population size.

Example 4.4.3 (Monotonicity approach). Consider a true population size of $S = 100$. Again, we chose a suitable Thatcher prediction interval of level $\gamma_1 = 0.95$ such that the true population size is approximately in the center of the interval. The prediction interval for the population is $[47, 180]$. This is the Thatcher prediction interval for a sample size of 100, 10 ‘successes’ and a total population of 1000. At the censoring time $t_c = 36, r = 20$ failures were reported. We used the true parameters, $\eta = 76.8$, and $b = 1.518$. The prediction period was taken as $(36, 72]$. The resulting ML estimates for ρ are

$$\hat{\rho}(S_l) = 0.8423, \hat{\rho}(S) = 0.4756, \hat{\rho}(S_u) = 0.2774.$$

The level $\gamma = 0.9$ calibrated prediction intervals are $[16, 27], [22, 53], [25, 64]$. Therefore, the final prediction interval is $[16, 64]$.

In example 4.4.3, the population size has a significant influence on the final prediction interval. For the lower estimate of the population size 47, at the censoring time only 27 units are still operating. Since the proportion failing in $(t_c, t_p]$ is large, the total remaining population limits the total number of failures. Therefore, it is not surprising that the resulting upper bound is significantly smaller than upper bound for $S = 100$ or $S_u = 180$. However, for $S = 100$ and for $S_u = 180$, the resulting calibrated upper bounds are not close to the number of operating units at the censoring time. Nevertheless, the upper bound from S_u is significantly larger than the upper bound from S .

The advantages of the monotonicity approaches are obvious. If a prediction interval for the population size is available, the calibration procedure to improve the naive prediction intervals can be adopted in a simple manner.

Before executing the computationally intensive calibration, calculating the expected value, i.e. the point estimate as in (4.21), can offer a first insight whether the population size has a sizeable effect on the resulting prediction interval. If the effect of the population size is small, the coverage properties of the resulting interval are similar to those in the case with a known population size. See Meng & Meeker (2011) for a study on the coverage properties for a known population size.

If the effect on the population size is large, then the coverage properties are also changed significantly. As calibration aims to bring the coverage for a given population size close to the desired nominal level, combining the smaller lower bound and the larger upper bound to obtain a prediction interval results in a wide and often highly conservative prediction interval. This is caused by the fact that the true population will

only in a very few cases take the boundary values. In many cases, the population lies in the interior of the interval where, if the population size S were known, a much shorter prediction interval would be obtained.

4.4.4 Distribution-based Approach

In this section, we suggest a method to reduce the length of the resulting prediction interval of the monotonicity approach. This approach is of course only reasonable, if the length of the prediction interval has a significant influence on the final prediction interval. In example 4.4.2 and example 4.4.3, we computed Thatcher prediction intervals for the population size. While the Thatcher interval is a conservative procedure to predict a binomial random variable based on a random sample from the same binomial distribution, it provides no information on the location of the predicted random variable within the interval. In chapter 3, we introduced a framework for frequentist prediction intervals with prior information for some discrete distributions. Unlike the Thatcher interval, these intervals contain information on the location of the predicted random variable within the prediction interval. This information is given by the conditional distribution given $X_1 = x_1$.

The approach we introduce in the remainder of this sections makes use of such a conditional distribution. It does not use a prior and posterior distribution as a Bayesian approach would do and is therefore a frequentist approach.

Consider we have a discrete probability distribution with PMF g , which describes the distribution of the population size within the prediction interval. In practice, this distribution could be specified by expert knowledge or by obtaining a prediction interval from a sample with the same distribution, where prior information about the underlying distribution parameter exists.

For example, if a binomial sample is used to predict a binomial random variable, and there exists prior information on the underlying parameter p in form of a beta distribution, then g would be the PMF of the beta-binomial distribution, see section 3.4. However, basically any discrete probability distribution, such as the binomial, the negative binomial, or the Poisson distribution, are possible candidates.

Now, given $[S_l, S_u]$ should be a level γ_1 prediction interval, the sum of g over all elements should ideally be equal to γ_1 . However, this is in general not possible due to the discreteness of the distribution for the population size. We now propose a procedure

to use the probabilistic structure of the interval $[S_l, S_u]$ to obtain a prediction interval for the number of failures in $(t_c, t_p]$.

Algorithm 4.4.4 (Distribution-based approach). (i) For each $S \in [S_l, S_u]$, calculate the calibrated level γ prediction interval $[Y_l(S), Y_u(S)]$ for the available data.

(ii) For each $S \in [S_l, S_u]$, calculate

$$W_S = \frac{g(S, \cdot)}{\sum_{W \in [S_l, S_u]} g(W, \cdot)}. \quad (4.22)$$

(iii) For each $S \in [S_l, S_u]$, obtain the products $W_S \times Y_l(S)$ and $W_S \times Y_u(S)$.

(iv) The lower bound Y_l is given by the largest integer smaller than

$$Z_l = \sum_{S \in [S_l, S_u]} W_S Y_l(S). \quad (4.23)$$

(v) The upper bound Y_u is given by the smallest integer larger than

$$Z_u = \sum_{S \in [S_l, S_u]} W_S Y_u(S). \quad (4.24)$$

In step 1, the calibrated prediction interval is obtained for every population size within the interval $[S_l, S_u]$. In step 2, the weight for each $S \in [S_l, S_u]$ is calculated. The weights are normalized to the overall probability of $[S_l, S_u]$ according to g . Thus, the sum of the weights over $[S_l, S_u]$ equals 1. In step 3, the calibrated prediction bounds are weighted with their respective weights. The resulting prediction bounds are the sum of the weighted prediction bounds, rounded to the next smaller / larger integer.

Example 4.4.5 (Distribution-based approach). Consider a population of 100, and a prediction interval of $[37, 136]$ for the population size S . This prediction is obtained using the procedure from section 3.5 with $X_1 = 10$, $n_1 = 100$, $n_2 = 800$, $\gamma = 0.95$ and the prior information is given by the beta distribution with parameters $a = 1, b = 2$. The true Weibull parameters are $\eta = 76.8$ and $b = 2.3$. At the censoring time $t_c = 36, r = 17$ had failed. The prediction period is $(36, 72]$. The calibrated level $\gamma = 0.90$ prediction interval using the monotonicity approach yields the interval $[11, 49]$. The beta-binomial weighted prediction interval using the distribution based approach results in the interval $[14, 41]$.

Example 3 shows the benefit of a distribution-based approach. The interval length is considerably shorter due to the usage of the probabilistic structure of the interval $[S_l, S_u]$. The size of the reduction in interval length depends on two factors.

The first factor is the same as for the monotonicity approach: the variation in $\hat{\rho}$ and its influence on the intervals length of the monotonicity approach. The second factor is the shape of the distribution of the population size prediction interval. This distribution determines which of the calibrated intervals are weighted most heavily, and thus have the highest influence on the resulting prediction interval.

The weighting of the calibrated intervals points out the importance of the correct choice and characterization of g when using this approach. Furthermore, there are also natural limits to this approach, as for example, a very extreme case of a point prediction for the population size reduces the approach to the interval for an assumed known population size.

4.5 Concluding Remarks and Future Work

This chapter has established a framework to predict the number of failures in a future time period with an unknown population size.

There are several topics for future work. For the monotonicity approach, analytical results regarding the influence of the population size on the ML estimates would be desirable. Furthermore, the effect of the variation in $\hat{\rho}$ and the population size should be studied in more detail. The transition, when the length of the population size interval starts to influence the final prediction interval is of particularly interest.

The overall coverage of the final prediction interval should be studied in detail for both presented approaches. Other possibilities to use the distribution of the prediction interval for the population size to obtain a prediction interval can be considered. For example, the sampling of the population size could be included in the simulations.

5 Failure Prediction with Reporting Delay

5.1 Introduction

In the previous chapter 4, failure prediction based on censored field data with unknown sample size was discussed. The main problem in this setting was the unknown size of the subgroup, which affects the parameter estimation of the lifetime distribution and the remaining population at the censoring time. In the setting of the previous chapter, no additional noise in the field data was present. However, most of the time additional noise is present in the setting of field data, which needs to be accounted for.

Nelson (2000) and Nordman & Meeker (2002) studied Weibull prediction intervals for a future number of failures with a given shape parameter. Escobar & Meeker (1999) studied a failure prediction problem for time-censored data in two situations: single time of field entry where all units enter the field at the same time, and staggered field entry where units enter the field over a longer period of time. Staggered entry leads to different ages of the units in the field at the time of censoring. In particular, Escobar & Meeker (1999) focused on improving the naive prediction interval using simulation-based calibration. The coverage properties of this problem were subsequently studied by Meng & Meeker (2011) for the Weibull distribution. Extensive literature exists on forecasting in the context of warranty claims, see for example Kalbfleisch & Robinson (1991), Wu & Akbarov (2012), Akbarov & Wu (2012) and Majeske (2007).

Another important problem in industry is forecasting in the context of risk assessments, where the number of failures resulting from one specific technical root cause is of interest. In this case, the observed number of failures at some point in time is smaller than in the warranty case. Furthermore, these risk assessments are time-critical and require fast responses. This has not received much attention in the literature. Wu & Akbarov (2012) consider warranty forecasts for a recently launched product with little field experience, where similar past products with longtime field experience exist. However, this is a significantly different situation than in risk assessments, as in many modern production environments one specific technical root cause changes the lifetime

behavior over a limited production period. Therefore, no comparable products with longtime field data exist.

Reporting delays occur frequently and can be significant in length. This becomes particularly crucial, if the time in field is short. Reporting delays have been studied in the warranty claim framework, e.g., by Kalbfleisch & Lawless (1992) and Kalbfleisch & Robinson (1991). Estimating the delay time distribution has received particular attention, see Kalbfleisch & Lawless (1992). Recently, Hong et al. (2009) suggested a model for failure prediction with reporting delays and product retirement. Hong et al. (2009) also noted the lack of studies in their context.

The topics of the present chapter are the evaluation of the effect of reporting delays in terms of the coverage properties of the resulting prediction interval, and the study of the efficiency of methods to account for the truncation in terms of coverage. These topics have not yet been considered by the relevant literature. This chapter is structured as follows: Section 5.2 states the scope which is subsequently studied. In section 5.3, some important concepts in the realm of prediction intervals are reviewed. Then, the influence of truncation on the prediction intervals is studied in section 5.4. Methods to account for the data truncation are presented in section 5.5. The parameters for the simulation study are introduced in section 5.6. The methods are evaluated in terms of their coverage properties in section 5.7. Section 5.8 concludes.

5.2 Model

In this section, we define the model under consideration. For both, failures times and delay times, we use a parametric framework. We consider a total population of size n where each member $i = 1, \dots, n$ with lifetimes x_1, \dots, x_n fails according to a continuous lifetime distribution with CDF $F(x, \theta)$. We focus on a Weibull model, i.e., $\theta = (\eta, b)$ and CDF

$$F(x, \eta, b) = 1 - \exp\left(-\left(\frac{x}{\eta}\right)^b\right), \quad (5.1)$$

where η is the scale parameter and b the shape parameter. We consider a fixed censoring time t_c at which the number of failures r is recorded. From this data, we want to estimate the total number of failures in an interval $(t_c, t_p]$, where $t_p > t_c$. Often, t_p is the warranty time or contractually stipulated lifetime, but other choices such as a one year time period are also common.

Failures are not reported instantaneously, but with a reporting delay $\tau_i, i = 1, \dots, n$. We model the delay times as iid lognormal random variables, $\tau_i \sim \text{logn}(\mu, \sigma)$ with CDF

$$G(\tau, \mu, \sigma) = \Phi\left(\frac{\log(\tau) - \mu}{\sigma}\right), \quad (5.2)$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution. Modelling the reporting delay by a lognormal distribution provides a good fit to real world data. When a failure is reported, the exact failure time x_j and the reporting time $x_j + \tau_j$ are observed. Reporting delays induce a right truncation on the data. At the censoring time, the failure of unit $j \in \{1, \dots, n\}$ is only observable if $x_j + \tau_j \leq t_c$. If $x_j \leq t_c$, but $x_j + \tau_j > t_c$, the unit has already failed before the censoring time, but the failure time cannot be observed. In the latter situation, the unit j seemingly has the censored lifetime $x_j = t_c$, although in reality the unit failed before t_c with actual lifetime $x_j < t_c$. In practice, mutual influences between the reporting and the failure process are out of scope. We can safely assume the independence of x_i and τ_i .

5.3 Prediction Intervals

5.3.1 Types of Observations

There are three classes of observations. Exact observations, censored observations and truncated observations. Censoring and truncation can be combined. The right censored and right truncated observation types relevant for our problem are listed in table 5.1. Censoring has been studied extensively in the literature, e.g., see Lawless (2003) for a detailed treatment of the matter. Type I censoring, also called time censoring, uses a predetermined time threshold. Thus the number of exact observations in the sample is random. Type II censoring stops when a predetermined number of exact observations has been recorded. In this case, the censoring time is random.

Right truncation occurs when an observation is only possible below a threshold value. Above the threshold, the existence of the observation is not known. For a right truncated random variable T_r with truncation at τ_r , we have

$$P(T_r \leq t | T_r < \tau_r) = \frac{F(t, \theta)}{F(\tau_r, \theta)} \quad \text{for } 0 \leq t < \tau_r.$$

When considering the reporting times $x_j + \tau_j$, the data is type I right censored with threshold t_c as in table 5.1. However, only the lifetimes x_j are of interest for the inference. x_j is only observable if $x_j + \tau_j \leq t_c$. The lifetimes x_j can also be considered as truncated

Table 5.1: Censored and truncated information constructed from a sample Z_1, \dots, Z_n with ordered values $Z_{(1,n)} \leq \dots \leq Z_{(n,n)}$

type I right censored with threshold t_c	$Y_1 = Z_{(1,n)}, \dots, Y_{k_{t_c}} = Z_{(k_{t_c},n)}, Z_\ell = t_c$ for $\ell = k_{t_c} + 1, \dots, n$, where $Z_{(k_{t_c},n)} \leq t_c < Z_{(k_{t_c}+1,n)}$
type II right censored with threshold $1 \leq k < n$	$Y_1 = Z_{(1,n)}, \dots, Y_k = Z_{(k,n)}, Y_\ell = Z_{(k,n)}$ for $\ell = k + 1, \dots, n$
right truncation with threshold t	$Y_1 = Z_{(1,n)}, \dots, Y_{k_t} = Z_{(k_t,n)}$, where $Z_{(k_t,n)} \leq t < Z_{(k_t+1,n)}$

at random thresholds $t_c - \tau_j$. This is a truncation condition analogous to the one stated in row three of table 5.1 with random threshold $t_c - \tau_j$, for each x_j . Besides, we assume the total number of units n to be known. Thus, we differ substantially from both of the classical concepts in table 5.1. In the present context, we refer to the above described sample setting as truncated.

5.3.2 Failure Prediction for a Population

So far, we have only considered individual lifetimes, but not yet the aggregated population of size n which we are interested in. A unit which is still operating at the censoring time t_c can either fail or survive the time period $(t_c, t_p]$ and all units belonging to the population of size n have the same lifetime distribution (5.1). Hence, conditional on the number of failures r at time t_c , the number of failures Y in $(t_c, t_p]$ has a binomial distribution with parameters $n - r$ and ρ , where

$$\rho = \frac{F(t_p, \eta, b) - F(t_c, \eta, b)}{1 - F(t_c, \eta, b)}. \quad (5.3)$$

Therefore,

$$P(Y \leq k) = \sum_{j=0}^k \binom{n-r}{j} \rho^j (1-\rho)^{n-r-j}. \quad (5.4)$$

The ML estimate $\hat{\rho}$ of ρ is obtained by plugging the ML estimates $\hat{\eta}$ and \hat{b} of the Weibull parameters η and b into (5.3). When plugging the ML estimate $\hat{\rho}$ into (5.4), it is important to note, that the number of failures r has an effect on (5.4) in two ways: i) on the remaining population $n - r$ which is directly visible in (5.4), whereas ii) is more implicit, as the ML estimates $\hat{\eta}$ and \hat{b} are based on the r exact failure times and the $n - r$ survivors and thus change with r . Therefore, $\hat{\rho}$ changes with r .

Let $\delta_l, \delta_u \in (0, 1)$. Plugging $\hat{\rho}$ into (5.4), we solve for y_l , the largest k which satisfies

$$P(Y < k) = \sum_{j=0}^{k-1} \binom{n-r}{j} \hat{\rho}^j (1 - \hat{\rho})^{n-r-j} \leq \delta_l \quad (5.5)$$

and for y_u , the smallest k which satisfies

$$P(Y \leq k) = \sum_{j=0}^k \binom{n-r}{j} \hat{\rho}^j (1 - \hat{\rho})^{n-r-j} \geq \delta_u. \quad (5.6)$$

For $\delta_l = \frac{1-\gamma}{2}, \delta_u = \frac{1+\gamma}{2}$, $[y_l, y_u]$ is called the naive two-sided level γ prediction interval. For $\delta_l = 1 - \gamma, y_l$ is the naive one-sided level γ lower prediction bound and for $\delta_u = \gamma, y_u$ is the naive one-sided level γ upper prediction bound. The inequalities in (5.5) and (5.6) are caused by the discreteness of the binomial distribution which generally prevents an exact solution of (5.5) and (5.6) as an equation. Since the truncation leads to an underestimation of the number of failures at the censoring time and subsequently to a smaller upper prediction bound, we will focus on the one-sided upper prediction interval.

5.4 Effect of Truncation on Prediction Interval

The truncation has effects on the (naive) prediction interval for the number of failures as described in section 5.3.2. The truncation has effects on the (naive) prediction interval for the number of failures as described in section 5.3.2. We will discuss the effect of the ML estimation in this section

The failure probability $\hat{\rho} = \hat{\rho}(\hat{\eta}, \hat{b})$ as defined in (5.3) is a function of the Weibull ML estimates. The question is which likelihood function should be used. For a fixed truncation time t_f , an ML estimation can be done on the so-called truncated conditional likelihood

$$L_{trunc}(\theta) = \prod_{i:x_i \leq t_f} \frac{f(x_i, \theta)}{F(t_f, \theta)}. \quad (5.7)$$

The estimate resulting from (5.7) is however quite uninformative, as has been reported by Kalbfleisch & Lawless (1988) for the classical truncation case, particularly in cases where only a small proportion of the total population has exact lifetimes. This is the common case in many applications in industry. Furthermore, the truncation time in our case is not fixed, but varies on the unit index i .

The censored likelihood

$$L_{cens}(\theta) = \prod_{i:x_i \leq t_c} f(x_i, \theta) \prod_{i:x_i > t_c} \bar{F}(t_c, \theta) \quad (5.8)$$

uses information on the remaining population at the censoring time t_c . Therefore, this information should not be disregarded, if possible. Without the presence of delay times, this estimator contains all existing information. However, if the number of truncated observations is > 0 , (5.8) treats some units which have a lifetime less than the censoring time as censored observations, which induces bias. If (5.8) is used regardless of the truncation, the proportion of units failing until t_p is underestimated.

Furthermore, the binomial prediction interval derived from (5.4) uses the remaining population at the censoring time. Therefore, using only the reported failures will lead to a different prediction interval than the one which would be available without reporting delays.

5.5 Approaches

Fundamentally, there are at least two possible approaches to adjust for possible missed observations due to reporting delay: A time shift of the censoring time or a probabilistic adjustment of the reported number of failures. These two approaches will be presented in this section.

In the simplest case, the reporting delay is a constant, known time period d . In this case, the adjusted censoring time t'_c is the original censoring time t_c minus the constant reporting delay, i.e. $t'_c = t_c - d$. Thus, the prediction period $(t_c, t_p]$ would be extended to the left by the reporting delay, i.e. the prediction period becomes $(t'_c, t_p]$. The parameters can thus be estimated on the censored data set in the sense of table 5.1, where all failures up to the adjusted censoring time t'_c are known.

However, if there is either uncertainty with regard to the length of reporting delay, or the reporting delay is completely stochastic in its nature, described through a probability distribution with CDF $G(t, \mu, \sigma)$, the shift in the censoring time cannot be done exact anymore. However, the idea of shifting the censoring time such that no failures have occurred between the shifted and the actual censored time can be adapted. This will be explored in section 5.5.1.

The alternative idea is, instead of shifting the censoring time, to estimate the number of unreported failures which already occurred, thus estimating the actual number of failures at the censoring time. This can be done by estimating the probability of a failure with a delay time such that a truncation occurs. This approach will be explored in section 5.5.2.

5.5.1 Shifting the Censoring Time

We explore the idea to shift back the censoring time such that the data at the new censoring time does not contain any truncated observations. It is worth mentioning that a forecast at t_c could be done at time $t_c + d$, where all failures prior to t_c will be reported. However, in many applications the analysis is time critical and d may amount to several months, thus we do not consider this option further.

Let t_c be the original censoring time and let r_{obs} be the number of reported failures at t_c . For $i = 1, \dots, r_{obs}$, we have ordered failure times $x_1, \dots, x_{r_{obs}}$, reporting times $y_1, \dots, y_{r_{obs}}$, and delay times $\tau_1 = y_1 - x_1, \dots, \tau_{r_{obs}} = y_{r_{obs}} - x_{r_{obs}}$. Let $x_{r_{obs}+1}$ be the failure time for the first future reported failure. Ideally, if $x_{r_{obs}+1}$ were known, the maximum available amount of information is used for the parameter estimation. The shifted censoring time in this case is

$$\tilde{t}_c = \max_{x_{r_{obs}+1} \leq t \leq t_c} \{x_{r_{obs}+1} > t\}. \quad (5.9)$$

If the reporting delay is a constant $d > 0$ and known, all failures before $t_c - d$ are reported at t_c and the data can be considered as censored in the sense of row 1 of table 5.1 with censoring time $t_c - d$. The idea of the approach presented in this section is to approximate this setting, if d is stochastic. In the sequel, we consider the stochastic delay time with a delay time distribution $G(\mu, \sigma)$. Estimating a nonparametric delay time distribution has been considered by Kalbfleisch & Lawless (1992).

There are several ways to handle the uncertainty. If a meaningful maximum upper delay time can be estimated, a conservative approach is to use this upper bound and consider this as a constant delay time. Two of the most easily available and used quantities of a distribution are the mean and the median. Subsequently, we will use these parameters for shifting the censoring time.

Additional uncertainty is introduced, if the parameters of the delay time distribution are also unknown and thus have to be estimated. This uncertainty is usually smaller because reporting schemes tend to be similar, regardless of the product in consideration, whereas the failure distribution is not. Thus, the database of the manufacturer may contain comprehensive and representative information on the reporting delay, such that parameters can be reliably estimated. Here, we consider the parameters of the delay time distribution to be known. Let $G(t) = G(t, \mu, \sigma)$ be the CDF of a lognormal distribution with $\mu \in (-\infty, \infty), \sigma > 0$. Then for $\tau \sim G(t, \mu, \sigma)$ we have

$$E[\tau] = e^{\mu + \frac{1}{2}\sigma^2}, V[\tau] = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}, \text{median}(\tau) = e^\mu. \quad (5.10)$$

We define the shifted censoring times

$$t_{c_e} = t_c - E[\tau], \quad (5.11)$$

$$t_{c_m} = t_c - \text{median}(\tau). \quad (5.12)$$

From (5.10) it is obvious that $E[\tau] \geq \text{median}(\tau)$. Thus, $t_{c_e} \leq t_{c_m}$. Of course, t_{c_m}, t_{c_e} should be greater than 0. The prolonged prediction period then becomes $(t_{c_e}, t_p]$, and $(t_{c_m}, t_p]$ respectively. Then, the parameter estimates $\hat{\eta}$ and \hat{b} for η and b are based on (5.8). The estimates for the failure probability ρ are

$$\hat{\rho}_{c_e} = \frac{F(t_p, \hat{\eta}, \hat{b}) - F(t_{c_e}, \hat{\eta}, \hat{b})}{1 - F(t_{c_e}, \hat{\eta}, \hat{b})}, \quad (5.13)$$

$$\hat{\rho}_{c_m} = \frac{F(t_p, \hat{\eta}, \hat{b}) - F(t_{c_m}, \hat{\eta}, \hat{b})}{1 - F(t_{c_m}, \hat{\eta}, \hat{b})}. \quad (5.14)$$

The naive prediction intervals can now be obtained as described in section 5.3.2 using $\hat{\rho}_{c_e}$ and $\hat{\rho}_{c_m}$. The approach can be easily adopted for any α -quantile q_α for the delay time distribution, which for example could be specified by an expert. In this case,

$$t_{c_q} = t_c - q_\alpha, \quad (5.15)$$

and

$$\hat{\rho}_{c_q} = \frac{F(t_p, \hat{\eta}, \hat{b}) - F(t_{c_q}, \hat{\eta}, \hat{b})}{1 - F(t_{c_q}, \hat{\eta}, \hat{b})}. \quad (5.16)$$

5.5.2 Adjusting the Number of Failures

The probabilistic approach is based on the idea of estimating the actual number of failures at t_c , i.e. $|\{j \in \{1, \dots, n\} : x_j \leq t_c\}|$, by using the reported number of failures and the delay time distribution.

Again, we consider a censoring time t_c . Let r_{obs} be the number of reported failures at t_c and let r_{t,t_c} be the number of failures which occurred at $t, t \leq t_c$ and were reported before t_c . Let $G(t) = G(t, \mu, \sigma)$ be the CDF of the reporting delay distribution, and let $F(t, \eta, b)$ be the CDF of the Weibull failure distribution. A point estimate for the actual number of failures at a time $t, t \leq t_c$ is given by

$$\tilde{r}_t = \frac{r_{t,t_c}}{G(t_c - t, \mu, \sigma)}. \quad (5.17)$$

The estimator is motivated by the fact that only $G(t_c - t, \mu, \sigma) \times 100\%$ of the failures at time t have a reporting delay $\tau \leq t_c - t$ and are thus reported before t_c . This estimator

has also been suggested by Kalbfleisch & Lawless (1992). It follows from (5.17) that any time t , where a failure occurs will receive a sort of weight. For reported failures times $x_1, \dots, x_{r_{\text{obs}}}$, the actual number of failures before t_c can thus be estimated by

$$\hat{r}_{t_c} = \left\lceil \sum_{k=1}^{r_{\text{obs}}} \frac{r_{x_k, t_c}}{G(t_c - x_k, \mu, \sigma)} \right\rceil. \quad (5.18)$$

The standard procedure in the prediction problem without reporting delay is to estimate the parameters of the lifetime distribution based on the censored lifetime data (5.8). This is usually done using maximum likelihood. Therefore, it is required to have the exact lifetimes of the failures before t_c and the censoring time t_c . Adjusting the total number of failures reported at t_c upwards creates a sort of artificial failures, where no failure data is available at the time of the analysis. To use these failures in the maximum likelihood estimation, failure data has to be created. This can be done by using the delay time CDF G and create a failure time between 0 and t_c ,

$$\hat{x}_j = t_c - \tau_j, \quad (5.19)$$

where τ_j is a random variable with CDF $G(\mu, \sigma)$. (5.19) implies, that all artificial failures are reported at t_c . Of course, a reporting time $> t_c$ could be chosen but this has two drawbacks. First, additional assumptions about the distribution of the reporting dates need to be made. Second, for the same delay time τ_j , the resulting failure time x_j is larger, leading to an ML estimate which results in a less conservative prediction bound.

Now the data set contains \hat{r}_{t_c} failures with failure times $x_1, \dots, x_{r_{\text{obs}}}, \hat{x}_{r_{\text{obs}}+1}, \dots, \hat{x}_{\hat{r}_{t_c}}$, where $\hat{r}_{t_c} - r_{\text{obs}}$ failures have not been reported and are artificially created. This adjusted data set is then treated as a censored data set, but without any truncation. Thus the ML estimation is done as in the classical censored problem (5.8). The prediction interval is then obtained as described in section 5.3.2. The estimator \hat{r}_{t_c} is used in (4.12) as the number of failures r .

5.6 Simulation

In this section, we present the setup for the simulation which was done to evaluate coverage properties of the naive prediction interval in our model. As a basis, we used the factors from the simulation study by Meng & Meeker (2011) and added additional factors for truncation and the delay time distribution.

- (i) p_f : the expected proportion failing before the censoring time t_c

- (ii) E_r : the expected number of failures before the censoring time t_c
- (iii) E_m : the expected number of failures between the censoring time and the future prediction time t_p
- (iv) η : the Weibull scale parameter
- (v) b : the Weibull shape parameter

We set the expected value and the variance in (5.10) for a lognormal random variable τ modelling the delay time to a proportion of the censoring time t_c to scale the sizes of the delay times properly to the sizes of the lifetimes.

- (vi) $E_{\text{trunc}} \in (0, 1)$: proportion of the censoring time for the expected value. The expected number of truncated observations (= unreported failures) increases with E_{trunc}
- (vii) $V_{\text{trunc}} \in (0, 1)$: proportion of the censoring time for the variance

From these factors, some additional factors have to be derived for a suitable statistical model.

- (i) the population size n : $n = \frac{E_r}{p_f}$
- (ii) the censoring time t_c : $t_c = F_{\text{Weibull}}^{-1}(p_f, \eta, b)$, where $F_{\text{Weibull}}(\cdot)^{-1}$ denotes the inverse of the CDF defined in (5.1)
- (iii) the prediction time t_p : $t_p = F_{\text{Weibull}}^{-1}\left(\frac{E_r + E_m}{n}, \eta, b\right)$
- (iv) the lognormal parameter μ : $\mu = \log\left(\frac{E_{\text{trunc}} t_c}{\sqrt{\frac{V_{\text{trunc}}}{E_{\text{trunc}}^2 t_c} + 1}}\right)$
- (v) the lognormal parameter σ : $\sigma = \sqrt{\log\left(\frac{V_{\text{trunc}}}{E_{\text{trunc}}^2 t_c} + 1\right)}$

The lognormal parameters μ and σ are obtained by setting $E[\tau] = E_{\text{trunc}} t_c$ and $V[\tau] = V_{\text{trunc}} t_c$ in (5.10) and solve the two equations for the two parameters.

Furthermore, we used an exponential delay time distribution with parameter λ as an alternative to the lognormal distribution. The exponential distribution is used to model a situation where short reporting delays are more likely than large reporting delays.

We evaluate the coverage in section 5.7 by using the following scheme.

- (i) Use the true parameters of the lifetime distribution and the true parameters of the delay time distribution to simulate $m = 10000$ samples of lifetimes and delay times.
- (ii) Apply the censoring and truncation scheme described in section 5.2. In particular, obtain the number of failures r_{obs} with $x_j + \tau_j \leq t_c$ and the actual number of failures r with $x_j \leq t_c$.
- (iii) For each sample $j = 1, \dots, m$, obtain the naive upper bound $u_{a,j}$ for the approaches a described in section 5.5. Furthermore, obtain the naive prediction intervals using the censored case without reporting delays and the naive prediction intervals treating the case with reporting delays without any adjustments.
- (iv) For each sample, $j = 1, \dots, m$ and each approach a , obtain $p_{a,j} = P(Y \leq u_{a,j})$ for the binomial CDF (5.4) with the remaining population size and true binomial parameter ρ obtained from (5.3) by using the true Weibull parameters.
- (v) The unconditional coverage probability, as defined by (4.7) in section 4.2.3, for each approach is approximated by $\frac{1}{m} \sum_{j=1}^m p_{a,j}$.

The justification for the final step is given by Escobar & Meeker (1999), as mentioned in section 4.2.3.

5.7 Results

We present the most relevant results of the simulation study in this section. For the failure time distribution, we considered only $\eta = 1$, as it is a scale parameter and other values η can be transformed to this case. Furthermore, we report the results only for $b = 1$. Approximate invariance of the coverage results have been reported by Meng & Meeker (2011) and Genschel & Meeker (2010).

As a benchmark, we use the coverage of the censored sample, which is available in a simulation study. This is justified as the censored case is the result without any delay times and thus the maximum attainable information at the censoring time in the considered setting. Even for the censored case, the coverage of the naive prediction intervals can be significantly below the nominal coverage level, especially for long prediction periods.

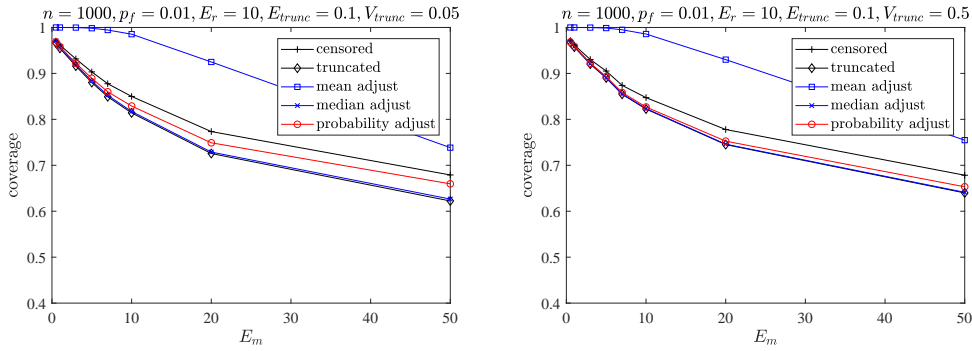


Figure 5.1: Coverage versus expected number of failures in $(t_c, t_p]$ with lognormal delay time distribution, $n = 1000$, $E_r = 10$, $E_{\text{trunc}} = 0.10$.

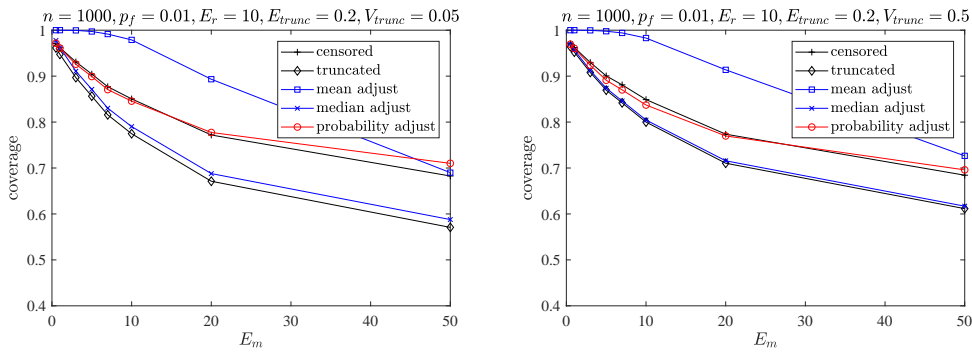


Figure 5.2: Coverage versus expected number of failures in $(t_c, t_p]$ with lognormal delay time distribution, $n = 1000$, $E_r = 10$, $E_{\text{trunc}} = 0.20$.

However, the coverage properties of the naive prediction interval in the censored case can be improved towards the nominal coverage by using calibration, see Meng & Meeker (2011). Therefore, it seems likely that this can also be achieved for the truncated case, if the coverage in truncated case is close to the coverage in the censored case. For comparison we also provide the coverage of the upper bound which is obtained by ignoring the data truncation and proceeding as described in section 5.3.2 without any adjustment. If the coverage of a method is smaller than in the censored case, the loss in information due to the reporting delay is not compensated sufficiently.

In figure 5.1, the coverage of all methods is only slightly worse than the benchmark coverage of the censored sample. This can be explained by the small expected truncation

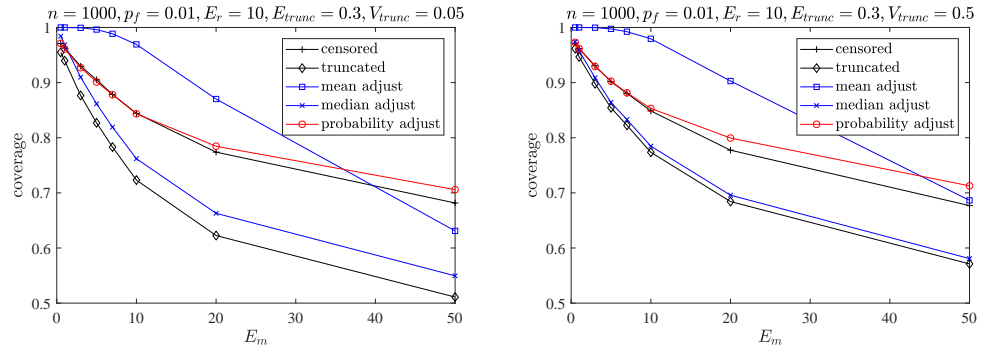


Figure 5.3: Coverage versus expected number of failures in $(t_c, t_p]$ with lognormal delay time distribution, $n = 1000$, $E_r = 10$, $E_{\text{trunc}} = 0.30$.

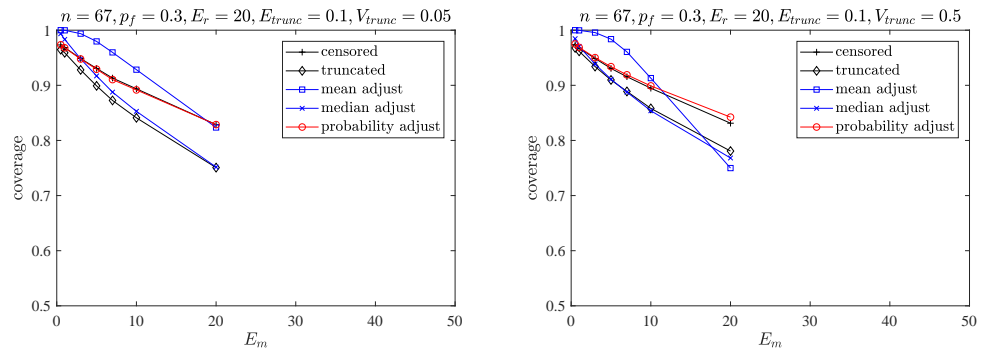


Figure 5.4: Coverage versus expected number of failures in $(t_c, t_p]$ with lognormal delay time distribution, $n = 67$, $E_r = 20$, $E_{\text{trunc}} = 0.10$.

and the overall small proportion of failing units. The mean adjustment in this case prolongs the prediction period to such an extent, that the upper bound is larger than in the censored case.

With increasing truncation, the coverage decreases for all methods. The gap between the censored coverage and the nonadjusted coverage widens considerably. This can be seen in figure 5.2 and figure 5.3. The coverage decrease is larger for the time adjusted methods than for the probability adjustment. For a large number of failures in the prediction period, the mean adjustment is no longer more conservative than the benchmark. The time shift by the median adjustment is small, the coverage is only slightly better than in the nonadjusted case.

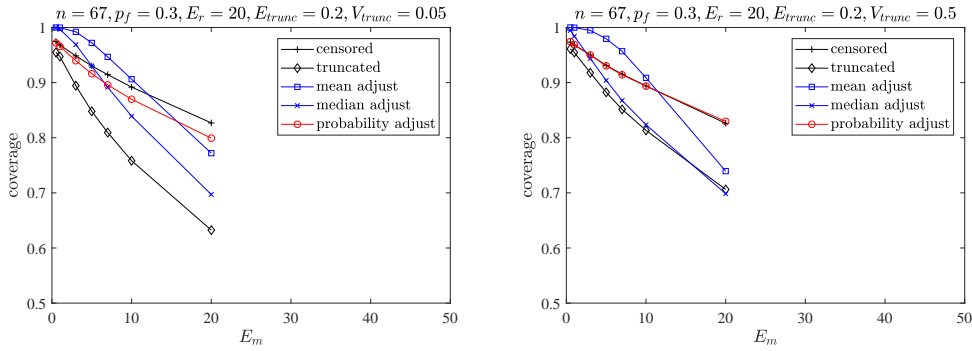


Figure 5.5: Coverage versus expected number of failures in $(t_c, t_p]$ with lognormal delay time distribution, $n = 67$, $E_r = 20$, $E_{\text{trunc}} = 0.20$.

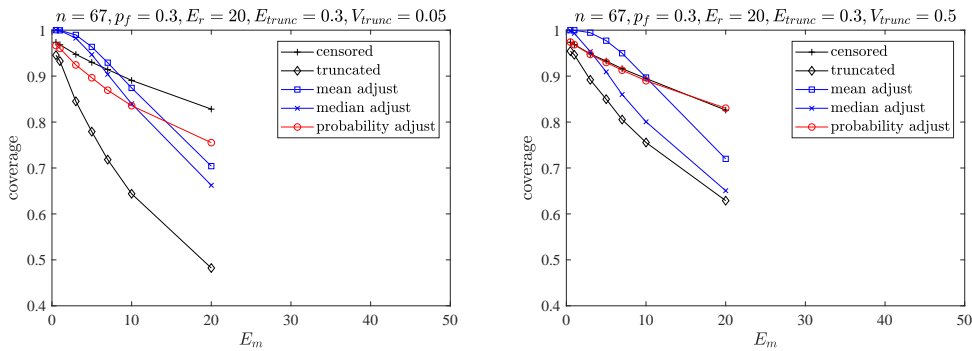


Figure 5.6: Coverage versus expected number of failures in $(t_c, t_p]$ with lognormal delay time distribution, $n = 67$, $E_r = 20$, $E_{\text{trunc}} = 0.30$.

If the proportion of units failing is larger, the coverage drops more strongly. This can be seen in figures 5.4, 5.5 and 5.6. In particular, the coverage of the mean adjustment drops for smaller prediction periods. For a large proportion failing, the variance influences the probability adjustment. For a larger variance, the number of unreported failures is scaled upwards more strongly. Therefore, the resulting upper bounds are higher leading to a better coverage. This can be seen in figure 5.6.

Finally, the effect of an exponential delay time distribution is shown in figure 5.7. The coverage of mean and median shift are much more similar than in the lognormal case. For both cases, the coverage is better than in the nonadjusted case. For a short prediction period, the time shift appears superior to the probability adjustment. The probability adjustment is not as close to the benchmark case as in the lognormal case,

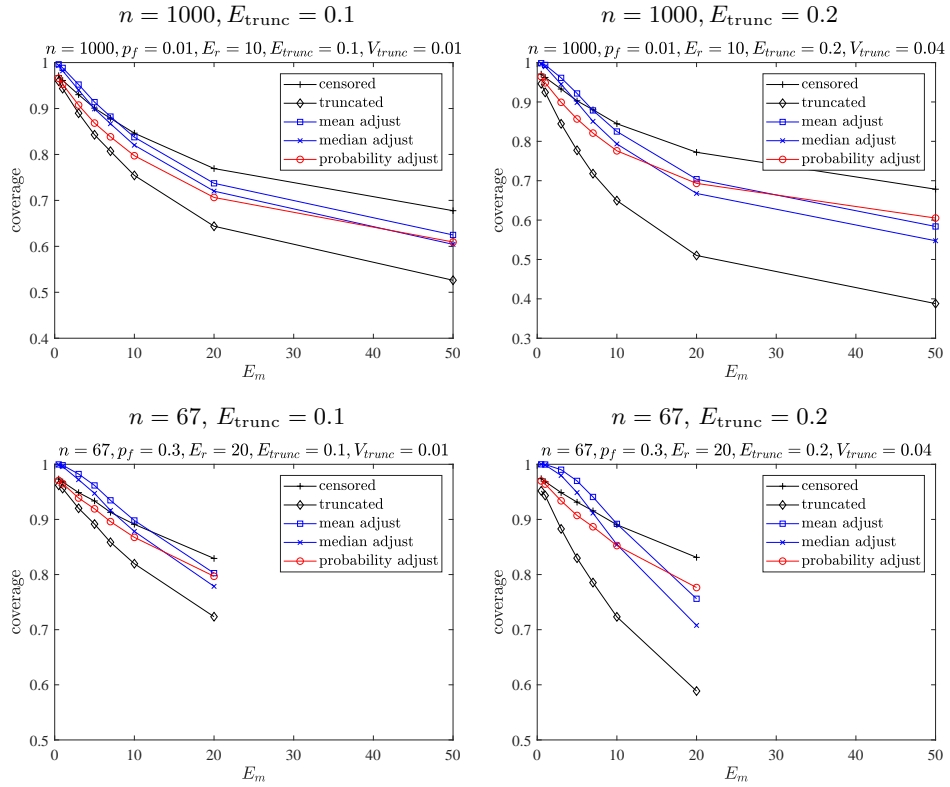


Figure 5.7: Coverage versus expected number of failures in $(t_c, t_p]$ with exponential delay time distribution.

but is superior to the chosen time adjustments for longer prediction periods.

5.8 Conclusion

We have studied the effect of reporting delays on Weibull prediction intervals for the future number of failures. We proposed two methods to account for the truncation which is induced by the delay between failure and reporting. Both methods improve the coverage compared to the unmodified prediction interval.

For the time adjustment, the length of the adjustment is crucial. A longer shift backwards results in a more conservative upper bound. An adequate choice depends on the specific case under consideration.

For the practitioner, the time adjustment is easier to apply and to communicate. How-

ever, in many cases, when there is little field data available, a meaningful shift will be near the time 0. The probability adjustment approach does not suffer from the shortcomings of the time shift approach and is less dependent on the specific case of interest. It also shows dependency on the variance and the delay time distribution. Besides, probability adjustment is difficult to communicate, as artificial failures have to be created and used.

There are several areas for future research. The improvement of the naive prediction interval through calibration should be studied in the case of truncation. Furthermore, the sensitivity of the coverage to the parameters of the delay time distribution should also be considered, as in applications there will be at least some uncertainty about the parameters of delay time distribution as well. As an alternative to the parametric approach presented in this chapter, the probability adjustment could be done in a nonparametric way by using a suitable inequality, e.g. Camp-Meidell, for the values of the delay distribution function.

Finally, another crucial truncation often present in prediction problems is caused by incomplete reporting. Incomplete reporting can occur for various reasons; the customer might only report a certain percentage of the total number of failures to a supplier, or warranty claims are only recorded for certain markets. A suitable stochastic model for this situation is also desirable for future work.

6 Confidence Intervals for a Product of Two Binomial Proportions

6.1 Introduction

In the previous chapters, prediction intervals were considered in various settings related to the occurrence of small subgroups with different lifetime behavior. In this chapter, we will look at the problem from a different perspective in form of a total probability failing for the whole population, which can be decomposed into a probability of belonging to the subgroup and a failure probability, conditional on belonging to the subgroup. In this case, the lifetime distribution is transferred into one relevant failure probability. This approach can be suitable if information on the failure probability can be determined in an laboratory environment, for example.

Consider the following more concrete example from risk assessment in the production of wafers for electronic components in automobiles. The wafers are produced in large batches under fixed machine settings. After a part of a specific batch has been processed in automobile manufacturing, end-of-line (EOL) testing of finished cars reveals occasional wafer failures. The EOL test results are reported to the wafer manufacturer. Diagnostics undertaken by the wafer manufacturer reveal a unique failure root cause C_1 occurring on a proportion p_1 of the wafers. The root cause C_1 leads to a failure C_2 with a small conditional probability p_2 . Then the total failure probability of a wafer is the product $q = p_1p_2$. However, the presented scheme is general and in many industrial applications, a failure probability q can be decomposed into a product $q = p_1p_2$ of two factor probabilities.

In contemporary industrial environments, the total failure probability q is very small. Feasible sample sizes are insufficient to estimate small q with due precision directly. The decomposition $q = p_1p_2$ enables sufficiently accurate estimation. The probability p_1 of having the root cause C_1 is usually large enough to be estimated reliably from a sample of practicable size n_1 from the process. The conditional failure probability p_2 can be

estimated from a targeted experiment where a sufficiently large group of units bearing the characteristic C_1 is intentionally generated to observe the failure rate in this group.

Similar problems and decompositions of failure probabilities occur in various contexts beyond manufacturing, e.g., in life sciences. Yet, corresponding methods of sample inference, in particular confidence intervals for q based on a decomposition $q = p_1 p_2$, have not been considered in the literature. The present chapter studies such confidence intervals based on binomial samples.

Confidence intervals for a single binomial proportion have received wide attention in the literature. The Clopper & Pearson (1934) interval is exact in the sense that pointwise in $p \in [0, 1]$ the coverage does not undercut a prescribed nominal level. However, on large segments of the unit interval the coverage exceeds the nominal level considerably. The research on less conservative and thus shorter exact intervals has been concentrating exclusively on the two-sided case, see the work by Sterne (1954), Crow (1956), Blyth & Still (1983). von Collani & Dräger (2001) and Göb & Lurz (2014) obtained considerable reductions in the length of exact frequentist two-sided intervals by exploiting prior information on p .

In many applications one-sided intervals are more relevant than two-sided intervals. In particular, for failure probabilities or probabilities nonconforming in manufacturing upper confidence limits are of prevailing interest. We will concentrate on this case in our study. The case of a one-sided interval for $q = p_1 p_2$ with a lower limit can be obtained in an obvious manner by considering the complementary probability $1 - q$. Our aim is to construct exact intervals where the actual coverage does not undercut the nominal level. By the lack of relevant research in the literature, the only one-sided exact interval we can make use of as a starting point for constructing an interval for a factorized probability $q = p_1 p_2$ is the Clopper-Pearson interval. One of the rare contributions on one-sided intervals is due to Cai (2005), however considering approximate intervals only.

The subsequent study presented in this chapter is structured as follows. Section 6.2 introduces the underlying stochastic model for the empirical analysis of a factorized probability $q = p_1 p_2$. Section 6.3 combines two individual Clopper-Pearson upper confidence limits to obtain a limit for q . Section 6.4 obtains intervals for q by inverting a one-sided test of significance, and exploits prior information on the factors p_1 and p_2 so as to obtain less conservative limits. Section 6.5 combines the methods considered by sections 6.3 and 6.4 into one scheme. We compare all presented methods in section 6.7 in terms of length and coverage. Section 6.8 concludes and points out areas of future

work.

6.2 Problem Description and Stochastic Model

In this section, we describe the sampling inference on decomposed probabilities $q = p_1 p_2$ in a stochastic model.

Consider a population of units $i = 1, 2, \dots$, e.g., the output of a manufacturing process. Each item i exhibits a characteristic C_1 with probability p_1 . Conditional on having the characteristic C_1 , item i can have the characteristic C_2 with conditional probability p_2 . Let X_{1i}, X_{2i} be the indicators of the characteristics C_1, C_2 relative to item i , i.e.,

$$X_{\ell i} = \begin{cases} 1 & \text{if item } i \text{ has characteristic } C_\ell, \\ 0 & \text{if item } i \text{ has not characteristic } C_\ell. \end{cases} \quad (6.1)$$

We assume that the series $(X_{1i}, X_{2i}), i = 1, 2, \dots$ is i.i.d. The characteristic C_2 can only occur on items with characteristic C_1 . Thus for all i

$$p_1 = P(X_{1i} = 1), p_2 = P(X_{2i} = 1 | X_{1i} = 1), 0 = P(X_{2i} = 1 | X_{1i} = 0). \quad (6.2)$$

For the unconditional probability $q = P(X_{2i} = 1)$ of an item i having the characteristic C_2 we obtain

$$\begin{aligned} q &= P(X_{2i} = 1 | X_{1i} = 1)P(X_{1i} = 1) + P(X_{2i} = 1 | X_{1i} = 0)P(X_{1i} = 0) \\ &= P(X_{2i} = 1 | X_{1i} = 1)P(X_{1i} = 1) = p_1 p_2. \end{aligned} \quad (6.3)$$

The objective is to construct a one-sided confidence interval with upper limit for q . By the decomposition (6.3), inference on q can be obtained by combining separate inferences on p_1 and p_2 . Consider an independent sample of size n_1 from the total process, and a second sample, independent of the first and independent within itself, of size n_2 from a population of items with characteristic C_2 . Let S_1 be the number of items with characteristic C_1 in the first sample, and let S_2 be the number of items with characteristic C_2 in the second sample. By assumption, S_ℓ is binomially distributed by $\text{Bin}(n_\ell, p_\ell)$. Due to the mutual independence of the two samples, the joint PMF of S_1 and S_2 is

$$f_{n_1, n_2, p_1, p_2}(s_1, s_2) = \binom{n_1}{s_1} p_1^{s_1} (1 - p_1)^{n_1 - s_1} \binom{n_2}{s_2} p_2^{s_2} (1 - p_2)^{n_2 - s_2}. \quad (6.4)$$

Using (6.3), we can substitute $p_2 = \frac{q}{p_1}$, to obtain

$$f_{n_1, n_2, p_1, q}(s_1, s_2) = \binom{n_1}{s_1} p_1^{s_1} (1 - p_1)^{n_1 - s_1} \binom{n_2}{s_2} \left(\frac{q}{p_1}\right)^{s_2} \left(1 - \frac{q}{p_1}\right)^{n_2 - s_2}. \quad (6.5)$$

6.3 Product of Clopper-Pearson Limits

In this section we construct an upper confidence limit for the product probability by taking the product of upper confidence limits for the involved factor probabilities.

The sample statistic S_ℓ of the sample for the characteristic C_ℓ is binomially distributed by $\text{Bin}(n_\ell, p_\ell)$. Exact upper limits for p_ℓ can be obtained by the well-known Clopper & Pearson (1934) procedure. Let $0 < \gamma_\ell < 1$ be the prescribed nominal level for sample ℓ . For sample results $s_\ell < n_\ell$, the Clopper-Pearson upper limit $p_{U,\ell} = p_{U,\ell}(s_\ell)$ is the unique solution $0 < p < 1$ of the equation

$$\sum_{k=0}^{s_\ell} \binom{n_\ell}{k} p^k (1-p)^{n_\ell-k} \stackrel{!}{=} 1 - \gamma_\ell, \quad (6.6)$$

for $s_\ell = n_\ell$ the trivial limit $p_{U,\ell} = 1$ is used. The solution of equation (6.6) equals the quantile of level γ_ℓ of the beta distribution $\text{Beta}(s_\ell + 1, n_\ell - s_\ell)$.

For $\ell = 1, 2$, the Clopper-Pearson limit satisfies $P(p \leq p_{U,\ell}) \geq \gamma_\ell$. By the independence of the two samples we obtain for the product limit $q_U = p_{U,1}p_{U,2}$

$$\begin{aligned} P(q \leq q_U) &= P(p_1 p_2 \leq p_{U,1} p_{U,2}) \\ &= P(p_1 \leq p_{U,1}, p_2 \leq p_{U,2}) + P(p_2 p_{U,1} < p_1 p_2 \leq p_{U,1} p_{U,2}) \\ &\quad + P(p_1 p_{U,2} < p_1 p_2 \leq p_{U,1} p_{U,2}) \quad (6.7) \\ &\geq P(p_1 \leq p_{U,1}) P(p_2 \leq p_{U,2}) \geq \gamma_1 \gamma_2. \end{aligned}$$

A total confidence level of at least γ can thus be reached with the limit $q_U = p_{U,1}p_{U,2}$ by any choice of γ_1, γ_2 such that $\gamma = \gamma_1 \gamma_2$. A natural candidate is $\gamma_1 = \gamma_2 = \sqrt{\gamma}$.

However, the product limit $q_U = p_{U,1}p_{U,2}$ is very conservative, for two reasons. First, the individual limits $p_{U,\ell}$ are very conservative, i.e., for a large range of true values of p_ℓ , the actual coverage $P(p_\ell \leq p_{U,\ell})$ is considerably larger than a prescribed level γ_ℓ , see for example Agresti & Coull (1998) and Newcombe (1998). Second, the inequality (6.7) is based on the trivial lower bound 0 for two of the three probabilities involved in the second line of (6.7).

6.4 Upper Confidence Limits under Prior Information based on Inverting a Test of Significance

The product limit $q_U = p_{U,1}p_{U,2}$ considered in section 6.3 is overconservative, and does not allow for exploiting prior information. These two related issues will be approached in the present section by constructing one-sided confidence limits from inverting a one-sided test of significance on p .

In modern industrial environments, the root cause probability p_1 and the conditional failure probability p_2 are small. It can safely be assumed that reliable upper bounds $p_{\ell \max}$ for the probabilities p_ℓ are known. The parameter space is thus reduced to $p_\ell \in [0, p_{\ell \max}]$, $\ell = 1, 2$. The resulting upper bound for the product probability is $q \leq q_{\max} = p_{1 \max}p_{2 \max}$. Subsequently, we consider the parametrization in p_1 and in the product probability q directly with prescribed upper bounds $0 < q_{\max} \leq p_{1 \max} \leq 1$. Then the relevant parameter space is

$$\Theta_{\max} = \{(p_1, q) | 0 < q \leq q_{\max}, q \leq p_1 \leq p_{1 \max}\}. \quad (6.8)$$

We derive an upper confidence limit for q by inverting tests for one-sided null hypotheses of the type

$$H_q = \{(p_1, \tilde{q}) \in \Theta_{\max} | \tilde{q} \geq q\}, \quad 0 < q \leq q_{\max}. \quad (6.9)$$

The natural test statistic is the product $S = S_1S_2$ of the number S_1 of items with characteristic C_1 in the sample of size n_1 from the total process and the number S_2 of items with characteristic C_2 in the sample of size n_2 from a population with C_1 . Intuitively, a hypothesis H_q should be rejected iff $S = S_1S_2$ is too small, i.e., we consider rejection regions

$$\mathcal{R}_c = \{s \in \{0, \dots, n_1n_2\} | s \leq c\}. \quad (6.10)$$

In the design of a critical value $c_\alpha(q)$ of a test of significance level $0 < \alpha < 1$ for H_q we have to control the nuisance parameter p_1 . For $0 \leq q \leq 1$, $c \in \mathbb{R}$, consider the function $R_{q,c}: [q, 1] \rightarrow [0, 1]$ defined by

$$R_{q,c}(p_1) = \sum_{\substack{0 \leq s_\ell \leq n_\ell \\ s_1s_2 \leq c}} \binom{n_1}{s_1} p_1^{s_1} (1-p_1)^{n_1-s_1} \binom{n_2}{s_2} \left(\frac{q}{p_1}\right)^{s_2} \left(1 - \frac{q}{p_1}\right)^{n_2-s_2}. \quad (6.11)$$

The function $\mathbb{R} \ni c \mapsto R_{q,c}(p_1)$ is the CDF of $S = S_1S_2$, see the reparametrized joint PMF of S_1, S_2 provided by (6.5). $R_{q,c}$ is a continuous function on $[q, 1]$ with

$$R_{0,c}(p_1) = 1 \text{ for } 0 \leq p_1 \leq 1, c \geq 0, \quad (6.12)$$

$$R_{q,c}(q) = \sum_{\substack{0 \leq s_1 \leq n_1 \\ s_1 n_2 \leq c}} \binom{n_1}{s_1} q^{s_1} (1-q)^{n_1-s_1} \text{ for } c \in \mathbb{R}, 0 \leq q \leq 1. \quad (6.13)$$

For $0 \leq q \leq 1$, $c \in \mathbb{R}$, $R_{q,c}$ adopts its absolute maximum

$$\text{MAX}_{c,p_{1\max}}(q) = \max_{q \leq p_1 \leq p_{1\max}} R_{q,c}(p_1) \quad (6.14)$$

on any compact prior information interval $[q, p_{1\max}]$ for p_1 . The subsequent proposition 6.4.1 clarifies properties of the maximum.

Proposition 6.4.1 (maxima of $R_{q,c}$). *Let $0 < p_{1\max} \leq 1$, $0 \leq c < n_1 n_2$. Then we have:*

(i) *For $0 < q \leq p_{1\max}$, $R_{q,c}$ adopts its global maximum $\text{MAX}_{c,p_{1\max}}(q)$ on $[q, p_{1\max}]$ in a finite number of points $p_{1i}(q) \in [q, p_{1\max}]$.*

(ii) *The function $[0, p_{1\max}] \ni q \mapsto \text{MAX}_{c,p_{1\max}}(q)$ is continuous and strictly decreasing from $\text{MAX}_{c,p_{1\max}}(0) = 1$ down to*

$$\text{MAX}_{c,p_{1\max}}(p_{1\max}) = \sum_{0 \leq s_1 \leq c/n_2} \binom{n_1}{s_1} p_{1\max}^{s_1} (1-p_{1\max})^{n_1-s_1}, \quad (6.15)$$

and in particular invertible on its image interval

$$\left[\sum_{0 \leq s_1 \leq c/n_2} \binom{n_1}{s_1} p_{1\max}^{s_1} (1-p_{1\max})^{n_1-s_1}, 1 \right]. \quad (6.16)$$

(iii) *For $q \in (0, p_{1\max})$ we have $\text{MAX}_{c,p_{1\max}}(q) < \text{MAX}_{c+1,p_{1\max}}(q)$ where $\text{MAX}_{c',p_{1\max}}(q) = 1$ for $c' \geq n_1 n_2$.*

The proof of assertions (i) and (ii) of proposition 6.4.1 is provided in appendix 6.A1, assertion (iii) is trivial. By the results of proposition 6.4.1, the next proposition 6.4.2 establishes a significance test of level α for the hypothesis H_q as defined by (6.9).

Proposition 6.4.2 (level α test for hypothesis H_q). *Let $0 < q \leq q_{\max} \leq p_{1\max} \leq 1$, let $0 < \alpha < 1$, and let the hypothesis H_q be defined by (6.9). By assertion (iii) of proposition 6.4.1, a critical bound $c_\alpha(q)$ is defined by*

$$c_\alpha(q) = \begin{cases} \max\{c \in \{0, \dots, n_1 n_2 - 1\} | \text{MAX}_{c,p_{1\max}}(q) \leq \alpha\} & \text{if } \text{MAX}_{0,p_{1\max}}(q) \leq \alpha, \\ -1 & \text{otherwise.} \end{cases} \quad (6.17)$$

Consider the test for H_q with the test statistic $S = S_1 S_2$ and the rejection region $\mathcal{R}_{c_\alpha(q)} = \{s | s \leq c_\alpha(q)\}$, see (6.10). Then we have:

(i) The power function of the test satisfies

$$\mathcal{G}(p_1, \tilde{q}) = P_{p_1, \tilde{q}}(S_1 S_2 \leq c) = R_{\tilde{q}, c_\alpha(q)}(p_1) \quad (6.18)$$

for $0 < \tilde{q} \leq q_{\max}$, $\tilde{q} \leq p_1 \leq p_{1 \max}$.

(ii) The test is of level α , i.e.,

$$\mathcal{G}(p_1, \tilde{q}) \leq \alpha \quad \text{for all } q \leq \tilde{q} \leq q_{\max}, \tilde{q} \leq p_1 \leq p_{1 \max}. \quad (6.19)$$

Assertion (i) of proposition 6.4.2 follows immediately from the definition (6.11) of the function $R_{q,c}$. Assertion (ii) follows from the definition (6.17) of the critical bound $c_\alpha(q)$ and the monotonicity result of assertion (ii) of proposition 6.4.1.

Exploiting the duality between tests of significance and confidence regions, the subsequent main theorem 6.4.3 provides one-sided confidence intervals with prescribed nominal level γ for the product probability q .

Theorem 6.4.3 (level γ upper confidence limit). *Let $0 < \gamma < 1$. For $0 < q_{\max} \leq p_{1 \max} \leq 1$ let the upper level γ limit $q_{U,\gamma}(s)$ for $s \in \{0, \dots, n_1 n_2\}$ be defined by*

$$q_{U,\gamma}(s) = \begin{cases} MAX_{s, p_{1 \max}}^{-1}(1 - \gamma) & \text{if } 1 - \gamma \geq MAX_{s, p_{1 \max}}(q_{\max}), \\ q_{\max} & \text{otherwise.} \end{cases} \quad (6.20)$$

(i) The random interval

$$B_\gamma = B_{\gamma, p_{1 \max}, q_{\max}}(S_1, S_2) = \begin{cases} (0, q_{U,\gamma}(S_1 S_2)) & \text{if } 1 - \gamma \geq MAX_{s, p_{1 \max}}(q_{\max}), \\ (0, q_{U,\gamma}(S_1 S_2)] & \text{otherwise} \end{cases} \quad (6.21)$$

is a level γ confidence interval for the product probability q relative to the prior information parameter space Θ_{\max} defined by (6.8), i.e.,

$$P_{p_1, q}(q \in B_\gamma) \geq \gamma \quad \text{for } 0 < q \leq q_{\max}, q \leq p_1 \leq p_{1 \max}. \quad (6.22)$$

In particular, we have $B_\gamma(S_1, S_2) = (0, q_{\max}]$ if $S_1 S_2 = n_1 n_2$.

(ii) Let $0 < q_{\max} \leq p_{1\max} \leq 1$, $0 < q'_{\max} \leq p'_{1\max} \leq 1$ with $q'_{\max} \leq q_{\max}$, $p'_{1\max} \leq p_{1\max}$.
Then we have $B_{\gamma, p'_{1\max}, q'_{\max}} \subset B_{\gamma, p_{1\max}, q_{\max}}$.

The proof of theorem 6.4.3 is provided by appendix 6.A2. The critical prior limit in the construction of theorem 6.4.3 is $p_{1\max}$. As visible from definition (6.20), the prior limit q_{\max} only serves as an upper cut-off for the confidence limit, whereas the confidence limit is a continuous function of $p_{1\max}$. To avoid ambiguities in the prior limit q_{\max} one may choose the trivial limit $q_{\max} = p_{1\max}$ without substantially affecting the confidence interval.

6.5 Obtaining Prior Information

In a modern stable high quality manufacturing environment, conservative prior bounds for the root cause probability p_1 and the conditional failure probability p_2 are easily obtained. Bounds $p_{i\max}$ in the parts per thousand range can be considered as reliable and conservative. In exploratory contexts with little previous experience on the study matter, it may be difficult to obtain prior bounds. In such cases, the analysis of the two samples of sizes n_1 and n_2 can proceed in two steps. First, obtain confidence limits $p_{U,1}(S_1)$, $p_{U,2}(S_2)$, e.g., the Clopper-Pearson upper confidence limits $p_{U,1}(S_1)$, $p_{U,2}(S_2)$ as described in section 6.3. Second, use $q_{\max} = p_{U,1}p_{U,2}$ and $p_{1\max} = p_{U,1}$ as prior bounds to obtain a confidence interval of the type (6.21) established by theorem 6.4.3. The subsequent proposition 6.5.1 provides a lower bound for the coverage of this method.

Proposition 6.5.1 (lower coverage bound). Let $0 < \gamma_0, \gamma' < 1$. For $\ell = 1, 2$, let $p_{U,\ell}(S_\ell)$ be an upper confidence limit such that

$$P_{p_1, p_2}(p_1 \leq p_{U,1}(S_1), p_2 \leq p_{U,2}(S_2)) \geq \gamma_0 \quad \text{for all } 0 \leq p_1, p_2 \leq 1. \quad (6.23)$$

Let $q_{\max}(S_1 S_2) = p_{U,1}(S_1)p_{U,2}(S_2)$, $p_{1\max}(S_1) = p_{U,1}(S_1)$, and let $B_{\gamma'} = B_{\gamma', q_{\max}(S_1 S_2), p_{1\max}(S_1)}(S_1 S_2)$ be the confidence interval defined by (6.21). Then we have

$$P_{p_1, q}(q \in B_{\gamma'}) \geq \gamma_0 + \gamma' - 1 \quad \text{for all } 0 < q \leq p_1 \leq 1. \quad (6.24)$$

If prior bounds are prestated without reference to the actual sample, the crucial bound for determining the confidence limit (6.20) is $p_{1\max}$. Without a substantial limitation of the result, you can choose the trivial bound $q_{\max} = p_{1\max}$, see the remarks following theorem 6.4.3. An analogous clear pattern does not hold for sample based Clopper-Pearson bounds $p_{1\max}(S_1)$, $q_{\max}(S_1 S_2)$, see the results of section 6.7, below.

6.6 Computation of Confidence Limits under Prior Information

We now provide an algorithm to obtain the upper confidence limit $q_{U,\gamma}(s)$ defined by (6.20).

Algorithm 6.6.1. *Consider the prior bounds $0 < q_{\max} \leq p_{1\max} \leq 1$. Let $0 < \gamma < 1$ be the nominal confidence level, and let $\varepsilon > 0$ be the accuracy bound. Let $s_i \in \{0, \dots, n_i\}$ be the sample observations, and let $s = s_1 s_2$.*

S1) If $s = n_1 n_2$ let $q_{U,\gamma}(s) = q_{\max}$, return. If $s < n_1 n_2$ go to step S2).

S2) Initialize $q_{\text{lower}} = 0$, $q_{\text{upper}} = q_{\max}$. Go to step S3).

S3) Let $q_a = \frac{1}{2}(q_{\text{lower}} + q_{\text{upper}})$.

S4) Determine the maximum $MAX_{s,p_{1\max}}(q_a)$.

S4.1) If $|MAX_{s,p_{1\max}}(q_a) - (1 - \gamma)| \leq \varepsilon$ let $q_{U,\gamma}(s) = q_a$, return.

S4.2) If $MAX_{s,p_{1\max}}(q_a) > 1 - \gamma + \varepsilon$ let $q_{\text{lower}} = q_a$. Go back to step S3).

S4.3) If $MAX_{s,p_{1\max}}(q_a) < 1 - \gamma - \varepsilon$ let $q_{\text{upper}} = q_a$. Go back to step S3).

The algorithm 6.6.1 uses a bisection procedure on the interval $[0, q_{\max}]$ to close in on the confidence limit $q_{U,\gamma}(s)$. The hard task is determining the maximum $MAX_{s,p_{1\max}}(q_a)$ in step S4). The function $R_{q_a,s}$ is in general not unimodal, so that simple bisection procedures to obtain an absolute maximizer p_a fail. However, by proposition 6.4.1 $R_{q_a,s}$ adopts its global maximum in a finite number of points. Therefore, we proceed by a brute force scheme and maximize over many different subintervals, which is a standard heuristic procedure used for global optimization, see Press (2007).

6.7 Comparison of Methods

This section compares the properties of the three previously developed upper confidence limits: i) the product $q_{U,C-P}(s_1, s_2) = p_{U,C-P}(s_1)p_{U,C-P}(s_2)$ of two independent Clopper-Pearson limits, see section 6.3; ii) the limit $q_{U,\text{prior}}(s_1 s_2)$ under prestated prior upper bounds for p_1 and p_2 , see theorem 6.4.3 in section 6.4; iii) the limit $q_{U,\text{combine}}(s_1 s_2)$ suggested in section 6.5 where the sample based prior bounds are Clopper-Pearson limits $p_{U,C-P}(s_1)$, $p_{U,C-P}(s_2)$. We analyze two characteristics of the respective confidence

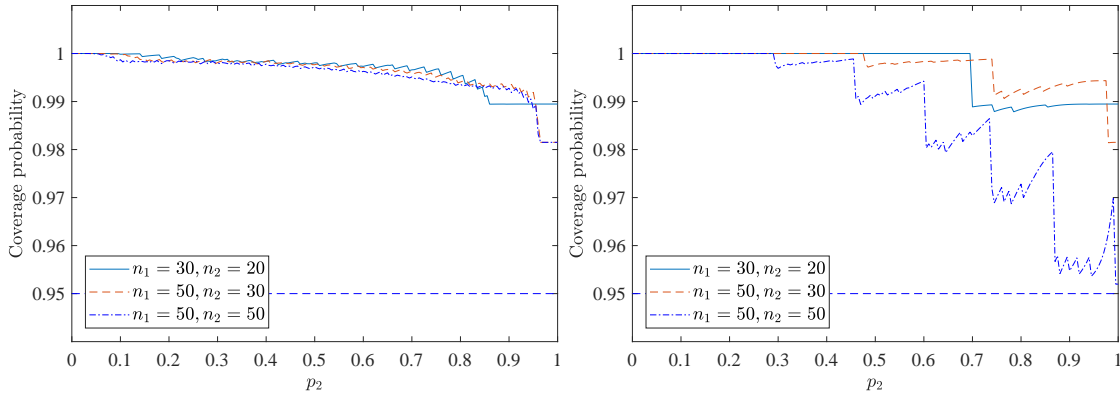


Figure 6.1: Coverage $C_{p_1}(p_1 p_2)$ of product C-P (left) and test inversion limit (right), $p_1 = 0.20$ fixed, different n_1, n_2 , under nominal confidence level $\gamma = 0.95$.

intervals B : 1) the actual coverage provided by

$$C_{p_1}(q) = \sum_{s_1=0}^{n_1} \sum_{s_2=0}^{n_2} \binom{n_1}{s_1} p_1^{s_1} (1-p_1)^{n_1-s_1} \binom{n_2}{s_2} \left(\frac{q}{p_1}\right)^{s_2} \left(1 - \frac{q}{p_1}\right)^{n_2-s_2} \mathbb{1}_B(q), \quad (6.25)$$

where $\mathbb{1}_B(\cdot)$ denotes the indicator function of the confidence interval B . 2) The width of the intervals for selected observations s_1, s_2 .

The type i) Clopper-Pearson product interval and the type ii) prior information interval with trivial prior bounds $q_{\max} = p_{1 \max} = 1$ are exact relative to a prescribed nominal level γ on the unrestricted parameter space, i.e., $C_{p_1}(q) \geq \gamma$ for all $0 < q \leq p_1 \leq 1$. However, both intervals are very conservative, as illustrated by figure 6.1. The actual coverage largely exceeds γ throughout, and is close to 1 for small values of $q = p_1 p_2$. While somewhat less conservative for small p_2 , the coverage of the product Clopper-Pearson upper limit never comes close to the desired nominal level. The type ii) interval by theorem 6.4.3 is less conservative for moderate to large values of p_2 .

The prior information type ii) interval with at least one nontrivial prior bound $p_{1 \max} < 1$ or $p_{2 \max} < 1$ is exact relative to the restricted parameter space Θ only, i.e., $C_{p_1}(p_1 p_2) \geq \gamma$ for all $0 < p_i \leq p_{i \max}$, $0 < q < q_{\max}$. Beyond the restricted parameter space the coverage is uncontrolled. Figure 6.2 shows that nontrivial prior bounds push the actual coverage towards the nominal confidence level inside Θ whereas outside Θ the nominal level can be undercut.

Prior information exploits the coverage budget provided by the nominal level more efficiently. Correspondingly, the upper confidence limits under prior information are

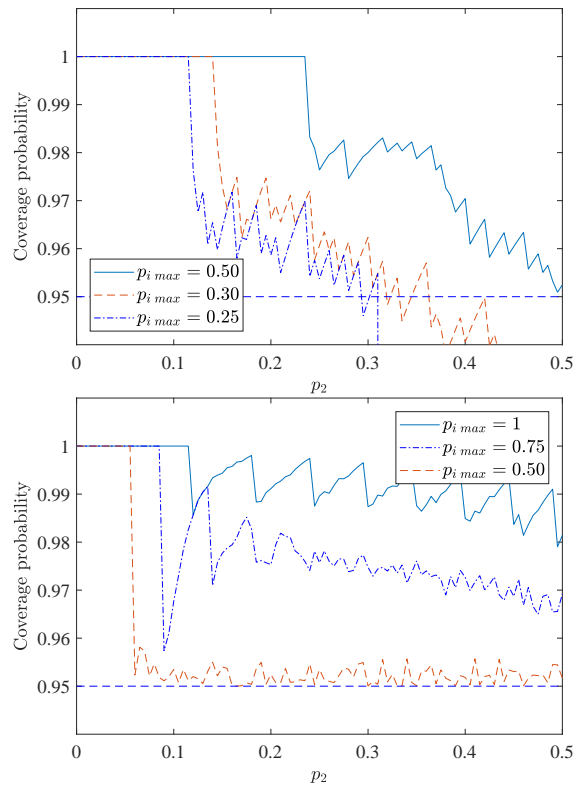


Figure 6.2: Coverage $C_{p_1}(p_1 p_2)$ of test inversion limit, different prior bounds $p_{i max}$, $p_1 = 0.20$ fixed, $n_1 = n_2 = 30$ (top), $p_1 = 0.50$ fixed, $n_1 = n_2 = 50$ (bottom), nominal confidence level $\gamma = 0.95$.

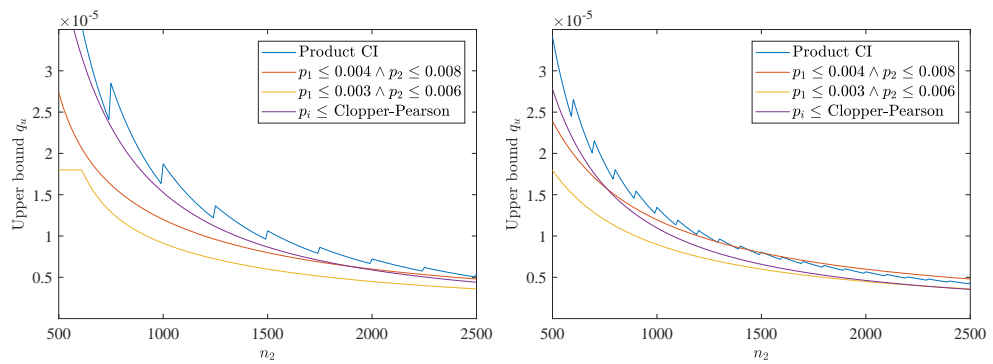


Figure 6.3: Upper confidence limit for q conditional under $\hat{p}_1 = s_1/n_1 = 0.002$ fixed, sample sizes $n_1 = 2n_2$ (left), and $n_1 = 5n_2$ (right).

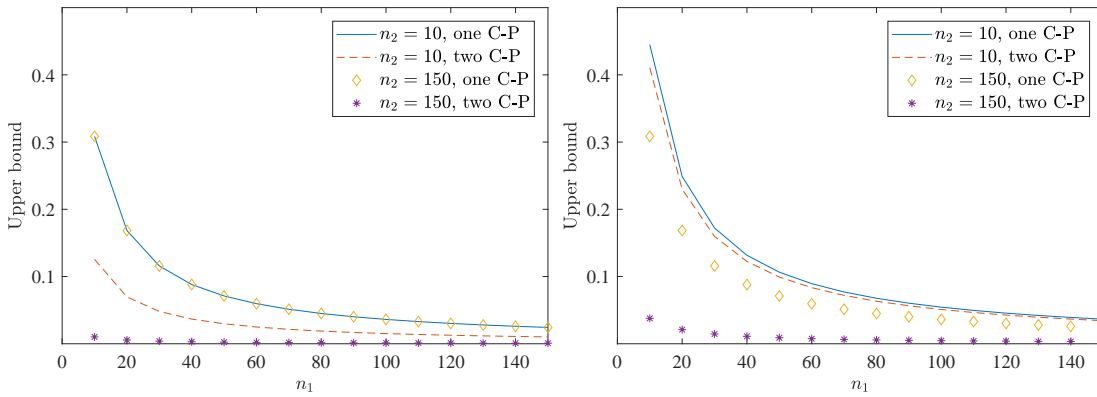


Figure 6.4: Upper confidence limit at level $\gamma = 0.95$ under $s_1 = 0 = s_2$ (left) and $s_1 = 3, s_2 = 5$ (right). Prior information from Clopper-Pearson limits with $\gamma_1 = 0.975 = \gamma_0$ (one C-P) and $\gamma_i = \sqrt{0.975}$ (two C-P).

shorter, and thus more informative, than without prior information, as illustrated for selected cases by figure 6.3. Subsequently, we consider the limit $q_{U,combine}(s_1 s_2)$ of type iii) suggested in section 6.5 where the sample based prior bounds are Clopper-Pearson limits. By proposition 6.5.1, any choice of levels γ_0 for the Clopper-Pearson part and γ' for the follow-up part resting on theorem 6.4.3 provides a confidence interval B of nominal level γ as long as $\gamma_0 + \gamma' - 1 = \gamma$. For tuning the method, two questions are relevant: a) How should the level γ_0 be factorized to $\gamma_0 = \gamma_1 \gamma_2$ over the two Clopper-Pearson steps? b) How should the total weight $1 + \gamma$ be allocated over γ_0 and γ' ?

Figures 6.4 and 6.5 consider the question a) for the total level $\gamma = 0.95$ under equally weighted $\gamma_0 = 0.975 = \gamma'$. We consider the two alternatives of 1) loading the Clopper-Pearson level fully on the first sample, i.e., $\gamma_1 = \gamma_0 = 0.975$, $\gamma_2 = 1.0$, and 2) allocating the Clopper-Pearson level equally to both samples by $\gamma_1 = \sqrt{\gamma_0} = \gamma_2$. The target characteristic is the width of the confidence interval. Figure 6.4 considers fixed sample results s_i under varying sample sizes. Figure 6.5 considers at least one sample result proportional to the sample size.

For $s_1 = 0 = s_2$ in the left part of figure 6.4 the allocation 2) is more favorable. For larger $s = s_1 s_2$, particularly larger s_2 , the right part of figure 6.4 shows that the allocation 2) is not necessarily the best. In figure 6.5 the values of the s_i are not both fixed. Either one s_i is fixed and the other one is a fixed proportion of the sample size n_i , or both are a fixed proportion. For expected small values of s_2 , the evidence suggests that it is favorable to use two Clopper-Pearson.

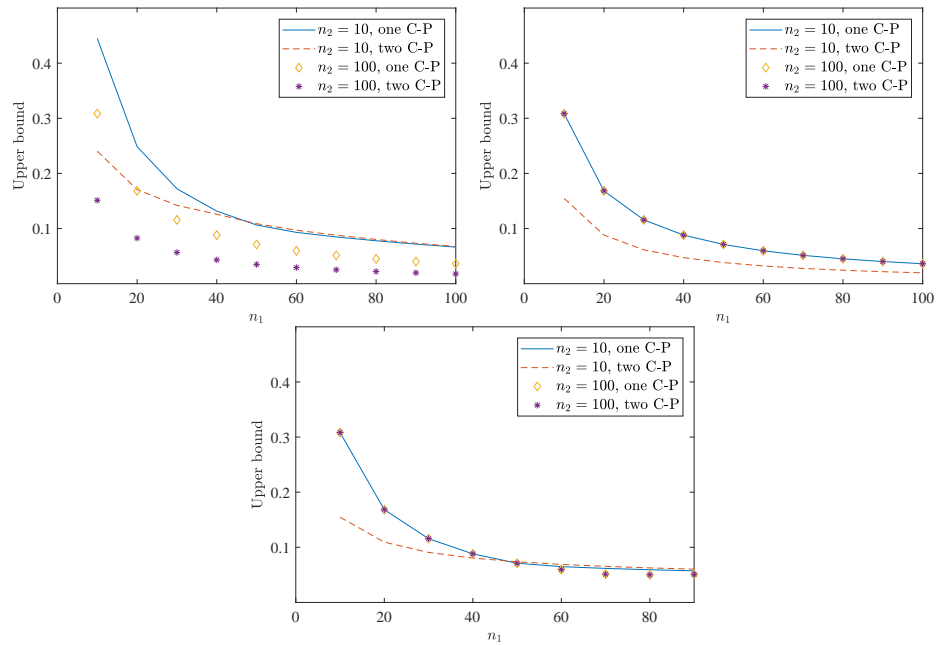


Figure 6.5: Upper confidence limit at level $\gamma = 0.95$ under $s_1 = 0.1n_1, s_2 = 1$ (upper-left), $s_1 = 1, s_2 = 0.10n_2$ (upper-right) and $s_1 = 0.10n_1, s_2 = 0.10n_2$ (lower-center), prior information from Clopper-Pearson limits with $\gamma_1 = 0.975$ (one C-P) and $\gamma_i = \sqrt{0.975}$ (two C-P).

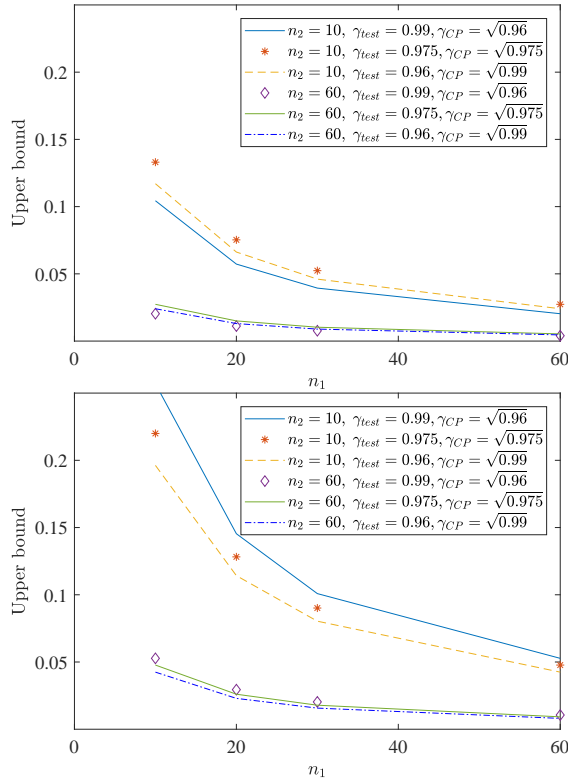


Figure 6.6: Upper confidence limit at level $\gamma = 0.95$ under $s_i = 0$ (top) and $s_i = 2$ (bottom), different choices of γ_0 and γ' .

For fixed total confidence level $\gamma = 0.95$, different choices of the test confidence level γ' and the prior total confidence level γ_0 with $\gamma_0 + \gamma' - 1 = \gamma = 0.95$ are shown in figure 6.6. The results vary with the value of $s = s_1 s_2$. For $s_1 = 0 = s_2$, choosing a higher level results in a lower upper limit, whereas for $s_1 = 2 = s_2$ choosing a higher level for the Clopper-Pearson bounds seems favorable. For increasing sample sizes, the differences in the upper confidence limits diminish. Overall, a clear general effect of the particular choice of γ' and γ_0 cannot be observed.

6.8 Conclusion

We presented a novel approach to obtain an upper bound for a product probability which cannot be directly estimated. The test inversion approach without prior information leads to very conservative upper bounds, especially near the left boundary.

The strength of the test inversion is its flexibility, as it can be used in combination

with prior information. Using prior information considerably reduces the resulting upper bound while maintaining the nominal confidence level. If prior information on the factor probabilities can be obtained by sampling, the individual Clopper-Pearson bounds can be used as prior information in the test inversion. This combined approach also shows a considerable length reduction for the confidence interval.

Several areas remain for future research. Since a product S can be often be decomposed in more than one set of factors $S_1 S_2$, the effect of different sample results with respect to their sample sizes and the same S should be studied. Furthermore, extending the procedure to more than two factor probabilities can be considered.

Appendix of Chapter 6

6.A1 Proof of Proposition 6.4.1

6.A1.1 Proof of Assertion (i) of Proposition 6.4.1

By definition (6.11) $R_{q,c}$ is a continuous rational function on $[q, 1]$. Hence $R_{q,c}$ adopts its maximum $\text{MAX}_{c,p_1 \text{ max}}(q) = \max_{q \leq p_1 \leq p_1 \text{ max}} R_{q,c}(p_1)$ in a finite number of points $p_i(q) \in [q, p_1 \text{ max}]$. \square

6.A1.2 Proof of Assertion (ii) of Proposition 6.4.1

For $p_1, p_2 \in [0, 1]$, $c \in \mathbb{R}$, let the distribution function of $S_1 S_2$ be denoted by

$$L_{n_1, n_2, c}(p_1, p_2) = \sum_{\substack{0 \leq y_1 \leq n_1 \\ 0 \leq y_2 \leq n_2 \\ y_1 y_2 \leq c}} \binom{n_1}{y_1} \binom{n_2}{y_2} p_1^{y_1} (1-p_1)^{n_1-y_1} p_2^{y_2} (1-p_2)^{n_2-y_2}. \quad (6.26)$$

For fixed $p_1 \in (0, 1]$, the partial derivative of $L_{n_1, n_2, c}$ in p_2 is

$$\begin{aligned} \frac{\partial}{\partial p_2} L_{n_1, n_2, c}(p_1, p_2) = \\ -n_2 \sum_{\substack{0 \leq y_1 \leq n_1 \\ 0 \leq y_2 \leq n_2-1 \\ c-y_1+1 \leq y_1 y_2 \leq c}} \binom{n_1}{y_1} \binom{n_2-1}{y_2} p_1^{y_1} (1-p_1)^{n_1-y_1} p_2^{y_2} (1-p_2)^{n_2-1-y_2}. \end{aligned} \quad (6.27)$$

Let $a_{n_1, s_1}(p_1) = \binom{n_1}{s_1} p_1^{s_1} (1-p_1)^{n_1-s_1}$. (6.27) is obtained by calculating

$$\begin{aligned}
& \frac{\partial}{\partial p_2} L_{n_1, n_2, c}(p_1, p_2) \\
&= \sum_{\substack{0 \leq s_i \leq n_i \\ s_1 s_2 \leq c}} a_{n_1, s_1}(p_1) \binom{n_2}{s_2} p_2^{s_2-2} (1-p_2)^{n_2-s_2-2} [s_2(1-p_2) - (n-s_2)p_2] \\
&= \sum_{\substack{0 \leq s_i \leq n_i \\ s_1 s_2 \leq c}} a_{n_1, s_1}(p_1) n_2 \binom{n_2-1}{s_2-1} p_2^{s_2-2} (1-p_2)^{n_2-s_2} \\
&\quad - \sum_{\substack{0 \leq s_i \leq n_i \\ s_1 s_2 \leq c}} a_{n_1, s_1}(p_1) n_2 \binom{n_2-1}{n_2-s_2-1} p_2^{s_2} (1-p_2)^{n_2-s_2-1} \\
&= n_2 \left\{ \sum_{\substack{-1 \leq s_2 \leq n_2-1, 0 \leq s_1 \leq n_1 \\ (s_2+1)s_1 \leq c}} a_{n_1, s_1}(p_1) \binom{n_2-1}{s_2} p_2^{s_2} (1-p_2)^{n_2-1-s_2} \right. \\
&\quad \left. - \sum_{\substack{0 \leq s_2 \leq n_2, 0 \leq s_1 \leq n_1 \\ s_1 s_2 \leq c}} a_{n_1, s_1}(p_1) n_2 \binom{n_2-1}{n_2-s_2-1} p_2^{s_2} (1-p_2)^{n_2-s_2-1} \right\} \\
&= -n_2 \left\{ - \sum_{\substack{0 \leq s_2 \leq n_2, 0 \leq s_1 \leq n_1 \\ s_2 s_1 \leq c-s_1}} a_{n_1, s_1}(p_1) \binom{n_2-1}{s_2} p_2^{s_2} (1-p_2)^{n_2-1-s_2} \right. \\
&\quad \left. + \sum_{\substack{0 \leq s_2 \leq n_2, 0 \leq s_1 \leq n_1 \\ s_1 s_2 \leq c}} a_{n_1, s_1}(p_1) n_2 \binom{n_2-1}{s_2} p_2^{s_2} (1-p_2)^{n_2-s_2-1} \right\} \\
&= -n_2 \sum_{\substack{0 \leq s_i \leq n_i \\ c-s_1+1 \leq s_2 s_1 \leq c}} a_{n_1, s_1}(p_1) \binom{n_2-1}{s_2} p_2^{s_2} (1-p_2)^{n_2-1-s_2} \\
&= -n_2 \sum_{\substack{0 \leq s_2 \leq n_2-1, 0 \leq s_1 \leq n_1 \\ c-s_1+1 \leq s_2 s_1 \leq c}} a_{n_1, s_1}(p_1) \binom{n_2-1}{s_2} p_2^{s_2} (1-p_2)^{n_2-1-s_2}.
\end{aligned}$$

For fixed $0 < p_1 \leq 1$ we obtain for $0 < q < p_1$, $0 \leq c < n_1 n_2$ from (6.11), (6.26), (6.27)

$$\frac{d}{dq} R_{q,c}(p_1) = \frac{1}{p_1} \frac{\partial}{\partial p_2} L_{n_1, n_2, c}(p_1, p_2) \Big|_{p_2=q/p_1} < 0. \quad (6.28)$$

Hence for fixed $0 < p_1 \leq 1$ the function $(0, p_1] \ni q \mapsto R_{q,c}(p_1)$ is strictly decreasing.

Now we can complete the proof of assertion (ii) of the proposition. Let $0 < q_1 < q_2 \leq p_{1 \max} \leq 1$, and let $R_{q_i, c}$ adopt its global maximum $\max_{q \leq p_1 \leq p_{1 \max}} R_{q_i, c}$ in the point $p_1(q_i) \in [q_i, p_{1 \max}]$ for $i = 1, 2$. Then because of $q_1 < q_2 \leq p_1(q_2)$

$$\begin{aligned} \text{MAX}_{c, p_{1 \max}}(q_1) &= \max_{q_1 \leq p_1 \leq p_{1 \max}} R_{q_1, c}(p_1) = R_{q_1, c}(p_1(q_1)) \geq \\ R_{q_1, c}(p_1(q_2)) &> R_{q_2, c}(p_1(q_2)) = \max_{q_2 \leq p_1 \leq p_{1 \max}} R_{q_2, c}(p_1) = \text{MAX}_{c, p_{1 \max}}(q_2). \end{aligned}$$

The latter shows that $[0, p_{1 \max}] \ni q \mapsto \text{MAX}_{c, p_{1 \max}}(q)$ is strictly decreasing. $[0, p_{1 \max}] \ni q \mapsto \text{MAX}_{c, p_{1 \max}}(q)$ is continuous since $(q, p_1) \mapsto R_{q, c}(p_1)$ is continuous. The image interval (6.16) is now obtained from the results (6.12) and (6.13). \square

6.A2 Proof of Theorem 6.4.3

We consider assertion (i) of theorem 6.4.3.

First, we demonstrate

$$q \in B \iff s > c_{1-\gamma}(q) \quad \text{for all } 0 < q \leq q_{\max} \quad (6.29)$$

where $c_{1-\gamma}(q)$ is defined by (6.17).

Consider the case $1 - \gamma < \text{MAX}_{s, p_{1 \max}}(q_{\max})$. By definition we have $B = (0, q_{\max}]$. Hence $q \in B$ for all $0 < q \leq q_{\max}$. For all $0 < q \leq q_{\max}$ we have by assumption

$$1 - \gamma < \text{MAX}_{s, p_{1 \max}}(q_{\max}) \leq \text{MAX}_{s, p_{1 \max}}(q),$$

hence by the definition (6.17) $s > c_{1-\gamma}(q)$. Hence (6.29) holds.

Consider the case $1 - \gamma \geq \text{MAX}_{s, p_{1 \max}}(q_{\max})$. By definition we have $B = (0, \text{MAX}_{s, p_{1 \max}}^{-1}(1 - \gamma))$. For all $0 < q \leq q_{\max}$ we have by assertion (ii) of proposition 6.4.1 and by definition (6.17)

$$\begin{aligned} q \in B &\iff q < \text{MAX}_{s, p_{1 \max}}^{-1}(1 - \gamma) \iff \\ \text{MAX}_{s, p_{1 \max}}(q) &> 1 - \gamma \iff s > c_{1-\gamma}(q), \end{aligned}$$

i.e., (6.29).

Applying (6.29) we obtain for all $0 < q \leq q_{\max}$, $q \leq p_1 \leq p_{1 \max}$ by proposition 6.4.2

$$\text{P}_{p_1, q}(q \in B) = \text{P}_{p_1, q}(s > c_{1-\gamma}(q)) \geq 1 - (1 - \gamma) = \gamma.$$

Assertion (ii) of theorem 6.4.3 is obvious from proposition 6.4.1 and the definitions (6.20), (6.21). \square

6.A3 Proof of Proposition 6.5.1

Let $0 < q \leq p_1 \leq 1$. By the monotonicity result of assertion (ii) of theorem 6.4.3 and by the Bonferroni inequality $P(E_1 \cap E_2) \geq P(E_1) + P(E_2) - 1$ we obtain

$$\begin{aligned}
& P_{p_1, q}(q \in B_{\gamma', p_{U,1}(S_1)p_{U,2}(S_2), p_{U,1}(S_1)}(S_1 S_2)) \\
& \geq P_{p_1, q}(q \in B_{\gamma', p_{U,1}(S_1)p_{U,2}(S_2), p_{U,1}(S_1)}(S_1 S_2), p_1 \leq p_{U,1}(S_1), q/p_1 \leq p_{U,2}(S_2)) \\
& \geq P_{p_1, q}(q \in B_{\gamma', q, p_1}(S_1 S_2), p_1 \leq p_{U,1}(S_1), q/p_1 \leq p_{U,2}(S_2)) \\
& \geq P_{p_1, q}(q \in B_{\gamma', q, p_1}(S_1 S_2)) + P_{p_1, q}(p_1 \leq p_{U,1}(S_1), q/p_1 \leq p_{U,2}(S_2)) - 1 \\
& \geq \gamma' + \gamma_0 - 1.
\end{aligned}$$

□

7 Conclusion

Various topics on failure prediction were covered in this thesis. The focus was on prediction intervals in various contexts occurring in industrial risk assessment. The first two chapters dealt with the estimation of the size of the subgroup. Although still widely used, control sampling for the subgroup size estimation has serious disadvantages. The test gate method presented in chapter 2 provides an efficient alternative to control sampling. This method is based on existing screening within the production for specific characteristics, which is often capable of detecting deviations in other characteristics as well, although less reliably than the one it was intended for. In particular, the test gate method is applicable for clusters and problems in the ppm range.

An exact prediction procedure for the test gate method is still an open problem. Furthermore, the method may be extended to a time series context, where the timing of the detections is used to identify clusters where the deviation was present. This would enable the practitioner to not only make a more accurate prediction about the size of the subgroup, it would also enable him to identify unaffected populations with respect to the desired confidence level.

Estimating the size of the subgroup with deviation requires using discrete prediction intervals. A general procedure to obtain prediction intervals exploiting prior information was presented in chapter 3. The scheme was applied to the binomial and negative binomial distribution. The prior information on the underlying parameter is expressed by a beta distribution. These prediction intervals are the shortest with respect to the weighted volume while fulfilling the coverage demand pointwise in x_1 . This coverage criterion is not meaningful for prediction intervals without prior information, where the learning sample and prediction target are independent.

Despite not being exact in the classical pointwise in $Y = p$ sense, the MWV prediction intervals show good coverage properties in the classical sense. Due to the construction, the intervals rely on the correct specification of the prior distribution. Although they are not overly sensitive to a wrong specification, the coverage requirement can be violated

in this case. Therefore, pointwise $Y = p$ prediction intervals under prior information are desirable, as they would overcome the reliance on the correct specification of the prior information. Furthermore, the framework can be extended to other discrete distributions such as the Poisson distribution. This will require the use of a different prior distribution.

In chapter 4 the effect of the sample size on the resulting prediction interval for the future number of failures was studied. Using the scheme from Escobar & Meeker (1999), the underlying Weibull lifetime distribution for the individual parts was used to obtain a prediction interval for the future number of failures. Two methods to obtain a prediction interval for the future number of failures based on the censored field data were presented. These methods adapt the procedure for a known sample size.

The monotonicity approach uses a distribution free prediction interval for the population size, whereas the distribution-based approach uses a prediction interval with a probabilistic structure as presented in chapter 3 for the population size. The latter approach is able to shorten the prediction interval significantly, if the population size has an effect on the resulting prediction interval for the number of failures. Since the population size is only specified in the form of a prediction interval, the study of the overall coverage is an interesting topic for future work. Furthermore, more work on the influence of the population size is desirable, as this question determines whether a more accurate estimation of the subgroup size yields an effect on the resulting prediction interval for the future number of failures.

The effect of a reporting delay on the failure prediction was studied in chapter 5. It has been shown, that neglecting the reporting delay severely impacts the coverage probability for the future number of failures. Two methods to account for the reporting delay achieve much better results. The first method shifts the censoring time back to an estimated time. All failures prior to this estimated time should be reported before the original censoring time. The prediction period is adjusted accordingly. Different time shifts have been studied, in particular the mean and the median. The second method estimates the number of unseen failures, which have already occurred, but have not been reported yet. For this estimated number, artificial failure times are assigned based on the delay time distribution and the censoring time. The prediction is then based on the adjusted number of failures and the, partially artificial, failure times.

The delay times have been modelled by a lognormal and an exponential distribution. The results vary depending on the distribution choice. Therefore, more work on

the distribution choice is desirable. In applications, the parameters of the delay time distribution are not known and therefore need to be estimated. This adds additional challenges to the presented approaches. Therefore, the coverage should be studied, when the parameters of the delay time distribution are estimated.

In chapter 6 one-sided confidence intervals for a product probability $q = p_1 p_2$ were studied. This problem was motivated by the subgroup context, where p_1 is the proportion with deviation and p_2 is the proportion of units failing, conditional on having the deviation. In this context, direct inference on q is not feasible. We showed that combining two Clopper-Pearson confidence intervals to obtain a confidence interval for the product probability is even more conservative than the Clopper-Pearson procedure for a single proportion.

We constructed a test of significance to obtain an exact level γ one-sided confidence interval for q . However, this interval is also very conservative. It is possible to use prior information on p_1 and p_2 for the test of significance. This shortens the resulting confidence interval considerably. As obtaining prior information is often difficult, the Clopper-Pearson bounds resulting from the individual samples can be used as prior information. This combined approach also showed significant improvements in the resulting upper bound. Using sampling results as prior information, multiple confidence levels on the individual steps are used to obtain a final confidence level of γ .

An optimal choice of the individual confidence levels was not found. More research on the influence of the individual levels should be carried out. This should also include different decompositions of $S = S_1 S_2$. Furthermore, the method can be extended to more than two characteristics. Due to the conservativeness already observed with two factor probabilities, the prior information required for more than two factors to obtain intervals close to the nominal level should be studied.

Some aspects which are adjoined to the topics of the individual chapters of this thesis are worth mentioning for future work. In industrial applications, the final goal is to obtain a $100 \times \gamma$ % prediction interval for the future number of failures which incorporates all existing uncertainties. Therefore, the aspects covered in the individual chapters should be considered in total.

First, an overall procedure should be developed. This procedure takes all available information to obtain a prediction interval for the subgroup size. Then, it uses this prediction interval and all available field data to estimate all required parameters. Finally,

the procedure would provide a prediction interval for the future number of failures. Afterwards, the coverage of this procedure should be studied. Particular attention should be given to the sensitivity analysis with respect to the individual uncertainties.

Furthermore, additional effects which were not covered in this thesis, but are also common industrial settings, should be studied. In particular, this includes staggered entry and partial reporting in form of reporting behavior. These two components add new difficulties, as staggered entry leads to different ages at the censoring time and partial reporting leads to a permanent truncation of the number of failures. However, as both components are present in industrial settings, including them in a model will be highly valuable.

Bibliography

- Agresti A, Coull B (1998) Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician* 52 (2): 119–126
- Akbarov A, Wu S (2012) Warranty claim forecasting based on weighted maximum likelihood estimation. *Quality and Reliability Engineering International* 28 (6): 663–669
- Bain LJ, Patel JK (1993) Prediction intervals based on partial observations for some discrete distributions. *IEEE Transactions on Reliability* 42 (3): 459–463
- Barndorff-Nielsen O, Cox DR (1996) Prediction and asymptotics. *Bernoulli* 2 (4): 319 – 340
- van Belle G (2008) *Statistical Rules of Thumbs*. Wiley
- Beran R (1990) Calibrating prediction regions. *Journal of the American Statistical Association* 85 (411): 715–723
- Blocher E, Robertson JC (1976) Bayesian sampling procedures for auditors: Computer-assisted instruction. *The Accounting Review* 51 (2): 359–363
- Blyth C, Still H (1983) Binomial confidence intervals. *Journal of the American Statistical Association* 78 (381): 108–116
- Cai T (2005) One-sided confidence intervals in discrete distributions. *Journal of Statistical Planning and Inference* 131: 63 – 88
- Chahkandi M, Ganjali M (2009) One some lifetime distributions with decreasing failure rate. *Computational Statistics and Data Analysis* 53 (1): 4433–4440
- Chaloner KM, Duncan GT (1983) Assessment of a beta prior distribution: Pm elicitation. *Journal of the Royal Statistical Society Series D (The Statistician)* 32 (1/2): 174–180

- Clopper C, Pearson ES (1934) The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 26 (4): 404–413
- von Collani E, Dräger K (2001) *Binomial Distribution Handbook for Scientists and Engineers*. Boston: Birkhäuser
- Corless J (1972) Assessing prior distributions for applying bayesian statistics in auditing. *The Accounting Review* 47 (3): 556–566
- Cox D (1972) Regression models and life-tables. *Journal of the Royal Statistical Society Series B (Methodological)* 34 (2): 187–220
- Crow E (1956) Confidence interval for a proportion. *Biometrika* 43 (3-4): 423–435
- Escobar LA, Meeker WQ (1999) Statistical prediction based on censored life data. *Technometrics* 41 (2): 113–124
- Faulkenberry GD (1973) A method of obtaining prediction intervals. *Journal of the American Statistical Association* 68 (342): 433–435
- Feller W (1968) *An Introduction to Probability Theory and Its Applications*. Wiley, New York
- Garthwaite PH, O’Hagan A (2000) Quantifying expert opinion in the uk water industry: An experimental study. *Journal of the Royal Statistical Society: Series D (The Statistician)* 49 (4): 455–477
- Göb R, Lurz K (2014) Design and analysis of shortest two-sided confidence intervals for a probability under prior information. *Metrika* 77 (3): 389–413
- Genschel U, Meeker WQ (2010) A comparison of maximum likelihood and median-rank regression for weibull estimation. *Quality Engineering* 22 (4): 236
- Hahn GJ, Meeker WQ (1991) *Statistical Intervals: A Guide for Practitioners*. Wiley-Blackwell
- Hogarth R (1975) Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association* 70 (350): 271–289
- Hong Y, Meeker WQ, McCalley JD (2009) Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. *Annals of Applied Statistics* 3 (2): 857–879

- Johnson NL, Kotz S, Kemp AW (1993) *Univariate Discrete Distributions*. Wiley, New York
- Kadane JB, Dickey JM, Winkler RL, Smith WS, Peters SC (1980) Interactive elicitation of opinion for a normal linear model. *Journal of the American Statistical Association* 75 (372): 845–854
- Kalbfleisch J, Lawless J (1988) Estimation of reliability in field–performance studies. *Technometrics* 30 (4): 365–378
- Kalbfleisch JD, Lawless JF (1992) Some useful statistical methods for truncated data. *Journal of Quality Technology* 24 (3): 145–152
- Kalbfleisch JF, Dand Lawless, Robinson JA (1991) Methods for the analysis and prediction of warranty claims. *Technometrics* 33 (3): 273–285
- Kann L, Herrmann D, Göb R (2018a) On the estimation of the size of a subgroup in industrial production: The test gate method. *Quality Engineering* 1–19, forthcoming
- Kann L, Sobotta B, Göb R (2018b) Failure prediction from censored data with uncertain population size. 2018 Annual Reliability and Maintainability Symposium (RAMS) 1–5, forthcoming
- Kann L, Sobotta B, Göb R (2018c) Failure prediction with reporting delay. *Quality and Reliability Engineering International* 1–10, forthcoming
- Kemp CD, Kemp AW (1956) Generalized hypergeometric distributions. *Journal of the Royal Statistical Society Series B (Methodological)* 18 (2): 202
- Krishnamoorthy K, Peng J (2011) Improved closed-form prediction intervals for binomial and poisson distributions. *Journal of Statistical Planning and Inference* 141 (5): 1709–1718
- Kundau D, Raqab MZ (2012) Bayesian inference and prediction of order statistics for a type-ii censored weibull distribution. *Journal of Statistical Planning and Inference* 142 (1): 41–47
- Kuruüzüm O, Akyüz G (2009) Modelling the relationship between product quality and process characteristics in process industry: An application. *International Journal of Productivity and Quality Management* 4 (4): 461–475

- Lawless JF (2003) *Statistical Models and Methods for Lifetime Data*. Wiley, New York
- Lawless JF, Fredette M (2005) Frequentist prediction intervals and predictive distributions. *Biometrika* 92 (3): 529. URL <http://dx.doi.org/10.1093/biomet/92.3.529>
- Majeske KD (2007) A non-homogeneous poisson process predictive model for automobile warranty claims. *Reliability Engineering and System Safety* 92: 243 – 251
- Mee R, Kushary D (1994) Prediction limits for the weibull distribution utilizing simulation. *Computational Statistics and Data Analysis* 17 (3)
- Meeker W, Escobar L (1998) *Statistical Methods for Reliability Data*. Wiley, New York
- Meng F, Meeker WQ (2011) Coverage properties of weibull prediction interval procedures to contain a future number of failures. Technical report, Iowa State University
- Nelson W (1982) *Applied Life Data Analysis*. Wiley, New York
- Nelson W (2000) Weibull prediction of a future number of failures. *Quality and Reliability Engineering International* 16 (1): 23–26
- Newcombe R (1998) Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine* 17 (8): 857–872
- Nordman DJ, Meeker WQ (2002) Weibull prediction intervals for a future number of failures. *Technometrics* 44 (1): 15–23
- O’Hagan A (1998) Eliciting expert beliefs in substantial practical applications. *The Statistician* 47 (1): 21–35
- Patel J (1989) Prediction intervals - a review. *Communications in Statistics - Theory and Methods* 18 (7): 2393–2465
- Patel JK, Samaranayake VA (1991) Prediction intervals for some discrete distributions. *Journal of Quality Technology* 23 (4): 270–278
- Press WH (2007) *Numerical recipes : the art of scientific computing*. Cambridge, UK ; New York : Cambridge University Press, 2007.
- Silva R, Bourguignon M, Cicero C, Gauss M (2013) The compound class of extended weibull power series distributions. *Computational Statistics and Data Analysis Third Special Issue on Statistical Signal Extraction and Filtering*: 352–367

-
- Sterne T (1954) Some remarks on confidence or fiducial limits. *Biometrika* 41 (1-2): 275–278
- Thatcher AR (1964) Relationships between bayesian and confidence limits for predictions. *Journal of the Royal Statistical Society Series B (Methodological)* 26 (2): 176–210
- Walls L, Quigley J (2001) Building prior distributions to support bayesian reliability growth modelling using expert judgement. *Reliability Engineering and System Safety* 74 (2): 117–128
- Wang H (2008) Coverage probability of prediction intervals for discrete random variables. *Computational Statistics and Data Analysis* 53 (1): 17–26
- Wang H (2010) Closed form prediction intervals applied for disease counts. *The American Statistician* 64 (3): 250–256
- Wu S, Akbarov A (2012) Forecasting warranty claims for recently launched products. *Reliability Engineering and System Safety* 106: 160 – 164
- Yao J, Xiong B, Zhou Y, He Y (2016) Research on relationship between manufacturing process quality variation and product reliability. *Proceedings of 2016 Prognostics and System Health Management Conference, PHM-Chengdu 2016*