2020

STATISTICAL INTERVALS FOR VARIOUS DISTRIBUTIONS BASED ON DIFFERENT INFERENCE METHODS

Yixuan Zou
University of Kentucky, yixuanzou@gmail.com
Author ORCID Identifier:
https://orcid.org/0000-0003-2477-6906
Digital Object Identifier: https://doi.org/10.13023/etd.2020.231

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Yixuan Zou, Student

Dr. Derek Young, Major Professor

Dr. William Rayens, Director of Graduate Studies

STATISTICAL INTERVALS FOR VARIOUS DISTRIBUTIONS BASED ON DIFFERENT INFERENCE METHODS

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of
Philosophy in the College of Arts and Sciences at
the University of Kentucky

By
Yixuan Zou
Lexington, Kentucky
Director: Dr. Derek Young, Associate Professor of Statistics
Lexington, Kentucky
2020

https://orcid.org/0000-0003-2477-6906

ABSTRACT OF DISSERTATION

STATISTICAL INTERVALS FOR VARIOUS DISTRIBUTIONS BASED ON DIFFERENT INFERENCE METHODS

Statistical intervals (e.g., confidence, prediction, or tolerance) are widely used to quantify uncertainty, but complex settings can create challenges to obtain such intervals that possess the desired properties. My thesis will address diverse data settings and approaches that are shown empirically to have good performance. We first introduce a focused treatment on using a single-layer bootstrap calibration to improve the coverage probabilities of two-sided parametric tolerance intervals for non-normal distributions. We then turn to zero-inflated data, which are commonly found in, among other areas, pharmaceutical and quality control applications. However, the inference problem often becomes difficult in the presence of excess zeros. When data are semicontinuous, the log-normal and gamma distributions are often considered for modeling the positive part of the model. The problems of constructing a confidence interval for the mean and calculating an upper tolerance limit of a zero-inflated gamma population are considered using generalized fiducial inference. Furthermore, we use generalized fiducial inference on the problem of constructing confidence intervals for the population mean of zero-inflated Poisson distribution. Birnbaum–Saunders distribution is widely used as a failure time distribution in reliability applications to model failure times. Statistical intervals for Birnbaum–Saunders distribution are not well developed. Moreover, we utilize generalized fiducial inference to obtain the upper prediction limit and upper tolerance limit for Birnbaum–Saunders distribution. Simulation studies and real data examples are used to illustrate the effectiveness of the proposed methods.

KEYWORDS: Tolerance Interval, Prediction Interval, Confidence Interval, Bootstrap Calibration, Generalized Fiducial Inference, Zero-inflated Model

Yixuan Zou

May 7, 2020

STATISTICAL INTERVALS FOR VARIOUS DISTRIBUTIONS BASED ON DIFFERENT INFERENCE METHODS

By
Yixuan Zou

Dr. Derek Young

Director of Dissertation

Dr. Katherine Thompson

Director of Graduate Studies

May 7, 2020

To my beloved parents and wife.

ACKNOWLEDGMENTS

First, I would like to acknowledge with my deep gratitude to Dr. Derek Young for his guidance on my research. I have been very fortunate to be his student. Without his dedicated guidance, this dissertation could never be finished. Furthermore, I want to thank my other committee members Dr. Solomon Harrar, Dr. Jianrong Wu, Dr. Arnold Stromberg and Dr. William Griffith for their service and suggestions. Moreover, I am thankful to Department of Statistics for both training me and supporting me during my years at University of Kentucky, and to all the faculty members, staffs, peers and friends, who are always ready to help.

Next, I could never accomplish this work without the support and love of my dearest parents and wife. I have been blessed with have such wonderful parents. No matter what happens, you always stand behind me, give me the advice and trust me in every choice I make. Thank you to my beautiful and lovely wife Dr. Tang, your encouragement helps me overcome the obstacle in pursing this Ph.D, and your suggestion on my writing skill and the discussion with you regarding to my thesis couldn't be more helpful.

Lastly, I want to thank Dr. Chee Ng and the lab members Dr. Chin Ng, Dr. Deni Hardiansyah, Yao Chen and James Li from College of Pharmacy, University of Kentucky. It was a unforgettable and enjoyable experience working with you. Dr. Chee Ng has been a such great mentor for both research and personal life. I could not make up my mind to go back to pursue my Ph.D without listening to your advice.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

Statistical intervals such as confidence interval, prediction interval, and tolerance interval are widely used to quantify uncertainty. Which statistical interval is more appropriate for a particular application relies on the purpose. A confidence interval is an interval that will, with a specified degree of confidence, contain true parameter of a distribution. If we want to estimate a parameter of interest, e.g., mean or variance, then we can approximate the true parameter by estimating it from the sample data. Furthermore, a confidence interval can be used to provide a quantification of the precision of the approximation. The most common confidence interval is the confidence interval for the population mean. On the other hand, a prediction interval for a single future observation is an interval that will, with certain confidence, contain a future observation from a distribution. Such interval can be used when our goal is to predict individual data rather than the population data. Moreover, a tolerance interval is an interval that we have some confidence that the interval will contain at least a specified proportion of a distribution. The end points of a statistical interval are upper and lower statistical limits. Ideally, no matter what statistical interval we use, we want the developed interval has nominal coverage probability as well as good precision. There are already extensive studies of statistical inference for normal distribution. However, for non-normal distributions, complex distributional settings can create challenges to obtain such intervals that possess the desired properties such as normal coverage probability and good precision.

Our research mainly focuses on constructing statistical intervals for a variety of distributions based on different inference methods: bootstrap calibration and generalized fiducial inference. Bootstrap calibration has been successfully applied to numerous problems to improve the coverage accuracy of the confidence interval procedure. [1, 2, 3, 4, 5] Generalized fiducial inference, proposed by Hannig [6] and modified from Fisher's original idea of fiducial inference in 1930, [7] has been proven to often generate attractive solution with asymptotically correct frequentist coverage probabilities and has very good small sample properties.[6, 8]

In this chapter, we will introduce the follow concepts and methods:

- Statistical Intervals
- Parametric Bootstrap Calibration
- Generalized Fiducial Inference
- Zero-inflated Data
- Birnbaum–Saunders distribution

1.1 Statistical Intervals

Let X be a random variable and $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^{\mathrm{T}}$ be an i.i.d. random sample.

Confidence Interval

Let θ be the parameter of interest. A $(1-\alpha)$ two-sided confidence interval $(L(\boldsymbol{X}), U(\boldsymbol{X}))$ satisfies the condition

$$P_X(L(\boldsymbol{X}) \leq \theta \leq U(\boldsymbol{X})) = 1 - \alpha,$$

where $(1-\alpha)$ is the *confidence level*. When $L(\boldsymbol{X}) = -\infty$, $(1-\alpha)$ upper confidence limit is $(-\infty, U(\boldsymbol{X})]$. Similarly, when $U(\boldsymbol{X}) = \infty$, $(1-\alpha)$ lower confidence limit is $[L(\boldsymbol{X}), +\infty)$.

Prediction Interval for a Single Measurement

A future measurement X_{n+1} is from the same BS distribution. A $(1-\alpha)$ two-sided prediction interval $(PL(\boldsymbol{X}), PU(\boldsymbol{X}))$ for a single measurement satisfies the condition

$$P_X(PL(\boldsymbol{X}) \leq X_{n+1} \leq PU(\boldsymbol{X})) = 1-\alpha,$$

where $(1-\alpha)$ is the *confidence level*. When $PL(\boldsymbol{X}) = -\infty$, $(1-\alpha)$ upper prediction limit is $(-\infty, PU(\boldsymbol{X})]$. Similarly, when $PU(\boldsymbol{X}) = \infty$, $(1-\alpha)$ lower prediction limit is $[PL(\boldsymbol{X}), +\infty)$.

Tolerance Interval

A $(p, 1-\alpha)$ two-sided tolerance interval $(TL(\boldsymbol{X}), TU(\boldsymbol{X}))$ satisfies the condition

$$P_{\boldsymbol{X}}\{P_X\big(TL(\boldsymbol{X}) \leq X \leq TU(\boldsymbol{X})|\boldsymbol{X}\big) \geq p\} = 1-\alpha,$$

where p is the the tolerance interval's *content* and $(1-\alpha)$ is the *confidence level*.

When $TL(\boldsymbol{X}) = -\infty$, $(p, 1-\alpha)$ one-sided upper tolerance limit is $(-\infty, TU(\boldsymbol{X})]$. Similarly, when $TU(\boldsymbol{X}) = \infty$, $(p, 1-\alpha)$ one-sided lower tolerance limit is $[TL(\boldsymbol{X}), +\infty)$.

The upper tolerance limit can be considered as the upper confidence limit of the $100p$th percentile W_p which satisfy the follow condition:

$$P_{\boldsymbol{X}}\{W_p \leq TU(\boldsymbol{X})\} = 1-\alpha. \tag{1.1}$$

Similarly, the lower tolerance limit can be considered as the lower confidence limit of the $(1-p)100$th percentile.

An Example to Illustrate the Three Intervals

Let's illustrate the different usage of three statistical intervals by a real world example.[9] In the Environmental Protection Agency mileage test, several nominally identical autos of a particular model are tested to produce mileage data $y_1, y_2, ..., y_n$. We could calculate a confidence interval for the mean mileage of the model to inform the mean or total gasoline consumption for the manufactured fleet of such autos. If a person wants to know whether a full tank of gas will be sufficient to cover 350 miles, it is better to use a prediction interval for a single future observation for this purpose. On the other hand, a tolerance interval is more appropriate than the other two intervals, when for example, a design engineer needs to determine the size of a gas tank that can guarantee that a specified proportion of the autos produced will have a certain cruising range. A tolerance interval with some confidence that it contains at least a certain fraction of mileages of such autos will fit the engineer needs.

1.2 Parametric Bootstrap Calibration

The bootstrap calibration as discussed in Chapter 18, Section 3 of Efron and Tibshirani,[10] was proposed for improving the coverage of a $(1-\alpha)$ confidence interval when the original procedure does not achieve the nominal coverage level. Their method can be summarized as follows. Firstly they use the classic nonparametric bootstrap and compute the confidence points (since their algorithm was presented in the context of calibrating one-sided confidence limits) for each bootstrap sample at each confidence level, say $(1-\gamma)$, in a set of candidate confidence levels. Then, for each γ, the coverage probability is calculated using the B bootstrap resamples. Finally, they select the γ^* such that the $(1-\gamma^*)$ confidence interval achieves, or at least comes close to, the nominal coverage level. This $(1-\gamma^*)$ confidence interval is then the calibrated $(1-\alpha)$ confidence interval.

Classic bootstrap calibration is also properly termed an *iterated bootstrap* procedure because there is an outer layer of constructing a bootstrap-based interval (e.g., the bootstrap percentile interval or the bootstrap t interval) followed by an inner layer of bootstrap samples to calibrate the confidence coefficient of those intervals such that they yield coverages close to the nominal level. Loh [1, 2] found that it was often enough to do a one-step calibration followed by linear interpolation in order to obtain a calibrated confidence interval with better coverage properties. Diciccio et al [3] constructed double bootstrap confidence intervals that involves replacing the inner level of resampling by an analytical approximation. We shall apply Loh's single layer bootstrap calibration on improving the coverage probability of two-sided tolerance interval in Chapter 2.

The bootstrap calibration has been proposed to be performed on achieving better coverage probability mostly for confidence interval. For tolerance intervals, however, one could potentially do calibration on the content level p. Some authors have explored this approach. Here, we will also use bootstrap calibration to improve the performance of coverage probability of tolerance interval. However, different from what has been explored in the past, the bootstrap calibration will be applied on the confidence level $(1 - \alpha)$ instead of the content level p. More dissuasion on the methodology used and the reason why we choose confidence level $(1 - \alpha)$ over the content level p will be given in Chapter 2.

1.3 Generalized Fiducial Inference

Fiducial inference was first introduced by Fisher in 1930.[7] He criticized the use of prior distribution of Bayesian inference when there was no prior information available, and he proposed a fiducial argument to obtain a distribution of the parameter similar to Bayesian posterior distribution, but not relying on a prior distribution. However, Fisher's idea was only limited to single parameter of continuous distribution, but did not apply to multiple parameters and parameters of discrete distribution. Furthermore, Fisher did not give a rigorous definition of the fiducial distribution. Fiducial inference failed to gain wide acceptance among statisticians due to inadequacies in the general approach. However, later works demonstrate the use of fiducial inference could be appropriate under some circumstances. For example, Weerahandi [11] introduced generalized confidence intervals constructed by generalized pivotal quantities, and Hannig et al [12] further established the connection between generalized confidence intervals and the fiducial argument of Fisher. Hannig further extended Fisher's idea and proposed a generalized fiducial inference, which also works for multiple parameters and discrete distributions.[6]

The aim of generalized fiducial inference is to identity a distribution for parameters of interest utilizing all the information from data. Based on the fiducial distribution, inference for the parameters can be made. The basic idea of the generalized fiducial inference is to switch the role of the parameters and the data by transforming the randomness from data to parameters. The philosophy of generalized fiducial inference is briefly explained as follows. Suppose data Y is generated through the structural equation $Y = G(\xi, U)$ where ξ is a vector of parameters and U are some random variable with a known distribution independent of the parameter ξ. The structural equation can be regarded as data generation process that noise process U and the signal ξ will produce observed data Y. Hence, the distribution of Y can be decided via structural equation given a fixed parameter ξ and the distribution U. After the data Y is observed, we can switch the position of data and parameters by solving the structural equation conditioning on that the solution to that equation exists, then we can get $\xi = Q(Y, U)$.

The generalized fiducial recipe is relatively easy to compute for complex distributional settings and possesses good small sample properties. Hence, we shall use it to construct approximate statistical intervals for zero-inflated distributions and Birnbaum–Saunders distribution.

1.4 Zero-Inflated Data

Data with excessive zeros are not uncommon in practice. For count data, the Poisson distribution is commonly used. When there are excess zero counts in the data, the classic Poisson distribution often underestimates the observed dispersion. Mullahy [13] first proposed a two-part model that can generate data with more flexibility to deal with the excess zeros: zeros follow a binomial distribution while positive values follow a truncated distribution. Such a model can accommodate under- and over-dispersed data. The model using a zero-truncated Poisson is often called the Poisson hurdle model. Lambert [14] extended the discussion of the count data with excess zeros to the regression setting. He also discussed how the zeros were generated: *random zeros*, in which case some zeros follow the assumed count distribution; *structural zeros*, in which case excess zeros follow a separate, degenerate process. Although excess zeros in a Poisson setting can be handled by both Poisson hurdle and zero-inflated Poisson models, they treat the generation of zeros very differently.

Moreover, when data are continuous with a high proportion of zeros, the excessive zeros need to be taken into consideration when being analyzed. When data are semicontinuous, since the data are skewed, the log-normal and gamma distributions are often considered for modeling the positive part of the model. Another question arises regarding how those zeros are generated. Similarly to count data, there are also two possible scenarios: *true zeros*, when the zeros are true observations; *censoring*, when the zeros are censored values because of the detection limit. In this dissertation, we will simply discuss semicontinuous data when all the zeros observed are true zeros.

Statistical intervals for zero-inflated count and continuous data will be developed in Chapter 3 and Chapter 4, respectively.

1.5 Birnbaum–Saunders distribution

Birnbaum–Saunders distribution was first proposed by Birnbaum and Saunders [15] as a failure time distribution for fatigue failure caused under cyclic loading. Desmond [16] derived the distribution in a more general way based on a biological model. Extensive work has been done on the statistical inference of the Birnbaum–Saunders distribution because of its wide range of applications. There has been a lot of research focusing on developing the point and interval estimation of the parameters for the Birnbaum–Saunders distribution, while for other statistical intervals such as tolerance interval and prediction interval, which are useful in reliability applications, are not well developed.

Tolerance interval and prediction interval for Birnbaum–Saunders distribution will be developed in Chapter 5.

1.6 Overview of the Dissertation

The rest of dissertation is organized as follows. Firstly, we introduce a focused treatment on using a single-layer bootstrap calibration to improve the coverage probabilities of two-sided parametric tolerance intervals for non-normal distributions. Secondly, we turn to zero-inflated data, which are commonly found in, among other areas, pharmaceutical and quality control applications. However, the inference problem often becomes difficult in the presence of excess zeros. When data are semicontinuous, the log-normal and gamma distributions are often considered for modeling the positive part of the model. The problems of constructing a confidence interval for the mean and calculating an upper tolerance limit of a zero-inflated gamma population are considered using generalized fiducial inference. Thirdly, we use generalized fiducial inference on the problem of constructing confidence intervals for the population mean of zero-inflated Poisson distribution. Lastly, Birnbaum–Saunders distribution is widely used as a failure time distribution in reliability applications to model failure times. Statistical intervals for Birnbaum–Saunders distribution are not well developed. Moreover, we utilize generalized fiducial inference to obtain the upper prediction limit and upper tolerance limit for Birnbaum–Saunders distribution. In the end, a brief summary and a discussion for future work will be given for this dissertation.

Chapter 2 Improving Coverage Probabilities for Parametric Tolerance Intervals via Bootstrap Calibration

2.1 Introduction

Meaningful comparisons and establishing statistically-based acceptance criteria are ubiquitous in biomedical and pharmaceutical research. For example, Zhai et al [17] studied ways to compare dissolution profiles between the reference and test formulations of a drug to ascertain similarities between the two formulations, which is required by regulatory authorities. Dong et al [18] addressed the assessment of pharmaceutical quality through the lens of traditional statistical quality control. Specifically, they evaluated batch quality by determining if a large proportion of the pharmaceutical product is within specification limits. Fedorov et al [19] addressed the issue of validation and characterization of deformable biomedical image registration for use in image-guided interventions. Specifically, their work was aimed at assessing results from an MRI-guided prostate biopsy research trial. Young et al [20] addressed the conformance to specifications during the design verification stage of a medical device that aids in the treatment of refractory epilepsy and treatment-resistant depression. Lizotte and Tahmasebi [21] assessed the variability in individual outcomes under a dynamic treatment regime to provide patient-centered data-driven sequential decisions.

The common statistical methodology employed in all of the research highlighted above is the tolerance interval. A *statistical tolerance interval* is used to capture a certain proportion p of the sampled population with a given confidence level $(1-\alpha)$, which we refer to as a $(p, 1-\alpha)$ tolerance interval. A one-sided $(p, 1-\alpha)$ upper (lower) tolerance limit is determined such that at least a proportion p is less (greater) than or equal to the upper (lower) tolerance limit with confidence $(1-\alpha)$. Two different criteria can be employed to construct two-sided tolerance intervals. One criterion is the interval will contain at least a proportion p of the population with confidence $(1-\alpha)$, which is the traditional definition of a tolerance interval. The other criterion is that the interval will contain at least a proportion p of the *center* of the population with confidence $(1-\alpha)$, which is usually called an *equal-tailed tolerance interval*. The importance of statistical tolerance intervals are not only highlighted by the numerous applications we cite in the present work, but also by their usage in documents published by regulatory agencies and international organizations, such as the Environmental Protection Agency,[22] the International Atomic Energy Agency,[23] and the International Organization for Standardization.[24] We refer to the text by Krishnamoorthy and Mathew [25] for a thorough treatment on statistical tolerance regions.

One-sided and two-sided tolerance intervals for normally-distributed data have been studied extensively in the literature,[26, 27, 28, 29, 30] while tolerance intervals constructed using data from other distributions have received far less treatment. However, non-normal distributions like the gamma, exponential, and Weibull distributions — which are all widely used in lifetime modelling — and the logistic distribution — which is primarily used for growth modelling — have received considerably less attention in the literature. Moreover, the difficulty of handling non-normal distributional forms in deriving their respective tolerance intervals nearly always necessitates the use of approximation methods, which vary in terms of performance metrics like coverage probabilities and expected lengths. Although analytical two-sided tolerance intervals are difficult to obtain, we can restrict the two-sided tolerance interval to be equal-tailed, and then apply the Bonferroni correction to the one-sided tolerance limits calculation using the aforementioned methods. The coverage probability of such a tolerance interval is the probability that it contains at least a proportion p of the population. However, an equal-tailed tolerance interval not only needs to include at least a proportion of p of the sampled population, but it is also required to cover equal proportions in the tails of the population's distribution. Since equal-tailed tolerance intervals have a more stringent condition than two-sided tolerance intervals, such an interval is very conservative in terms of actual coverage probability when the equal-tailed requirement does not have to be satisfied. To overcome the conservative behavior of a two-sided tolerance interval constructed using an equal-tailed restriction, we can then employ a bootstrap calibration [1, 2] to the confidence level $(1 - \alpha)$, resulting in coverage probabilities closer to the nominal level.

Bootstrap calibration has been successfully applied to numerous problems for improving the coverage accuracy of the confidence interval procedure.[1, 2, 3, 4, 5] Following the general approach in Chapter 18, Section 3 of Efron and Tibshirani,[10] the idea is that if the coverage of a $(1-\alpha)$ confidence interval procedure does not achieve the nominal coverage level, then bootstrapping can be employed to find a mapping $\alpha \to \gamma^*$ such that the $(1-\gamma^*)$ confidence interval does achieve, or at least come closer to, the nominal coverage level. The use of bootstrap calibration methodology, however, has only recently been employed to improve the coverage probabilities for some non-standard tolerance interval procedures, such as for random effects models,[31] the ratio of normal random variables,[32, 33] and the comparison of drug dissolution profiles.[17] Univariate distributions are clearly simpler settings than those just cited, but they are also arguably the most frequent settings for which tolerance intervals are constructed. This can be noted by the emphasis of the content in Krishnamoorthy and Mathew [25] and Young.[34] Moreover, most of those tolerance intervals are approximate, and hence are not guaranteed to achieve the nominal coverage level. This is especially true for small samples. The bootstrap calibration is clearly a viable approach for improving coverage probabilities of these approximate tolerance interval procedures for non-normal distributions, however, a general treatment of such an approach is, to our knowledge, not available in the literature. The present work fills that gap by studying the coverage probabilities and average widths of commonly used approximate tolerance interval calculations for some non-normal distributions, and then demonstrating the improvements of these performance metrics by employing the bootstrap calibration.

This chapter is organized as follows. In Section 2.2, we present the general definition of statistical tolerance intervals as well as formulas for tolerance intervals for some common non-normal continuous distributions. In Section 2.3, we present a single-layer bootstrap calibration algorithm and discuss how it addresses the conservative behavior of two-sided tolerance intervals created from equal-tailed tolerance intervals. In Section 2.4, we present results from numerical work about the proposed methodology, which includes summarizing coverage probabilities and average interval widths to evaluate the performance of the bootstrap calibration. In Section 2.5, we illustrate the methodology on three medical datasets. We conclude in Section 2.6 with a summary of our work and a discussion.

2.2 Calculation of Tolerance Intervals

General Definitions

Let X be a continuous random variable with distribution function $F_X(x)$ and $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^{\mathrm{T}}$ be an i.i.d. random sample from $F_X(x)$. A $(p, 1-\alpha)$ one-sided upper tolerance interval $(-\infty, U_1(\boldsymbol{X})]$ and lower tolerance interval $[L_1(\boldsymbol{X}), +\infty)$ are required to satisfy the respective conditions

$$P_{\boldsymbol{X}}\{P_X\big(X \leq U_1(\boldsymbol{X})|\boldsymbol{X}\big) \geq p\} = 1-\alpha$$

and

$$P_{\boldsymbol{X}}\{P_X\big(L_1(\boldsymbol{X}) \leq X|\boldsymbol{X}\big) \geq p\} = 1-\alpha,$$

where p is the tolerance interval's *content* and $(1-\alpha)$ is the *confidence level*. Similarly, a $(p, 1-\alpha)$ two-sided tolerance interval $(L(\boldsymbol{X}), U(\boldsymbol{X}))$ satisfies the condition

$$P_{\boldsymbol{X}}\{P_X\big(L(\boldsymbol{X}) \leq X \leq U(\boldsymbol{X})|\boldsymbol{X}\big) \geq p\} = 1-\alpha,$$

which is equivalent to

$$P_{\boldsymbol{X}}\{F_X(U(\boldsymbol{X})) - F_X(L(\boldsymbol{X})) \geq p\} = 1 - \alpha.$$

In our discussion, we emphasize an equal-tailed tolerance interval, which contains *at least* a proportion p of the *center* of the population with confidence level $(1-\alpha)$. A two-sided equal-tailed tolerance interval $(L_E(\boldsymbol{X}), U_E(\boldsymbol{X}))$ satisfies the condition

$$P_{\boldsymbol{X}}\left\{L_E(\boldsymbol{X}) \leq q_{\frac{1-p}{2}} \text{ and } q_{\frac{1+p}{2}} \leq U_E(\boldsymbol{X})\right\} = 1 - \alpha,$$

where $q(\cdot)$ is the quantile function of the distribution. For parametric tolerance intervals that have an exact one-sided tolerance limit, the two-sided equal-tailed tolerance limits can easily be obtained using the Bonferroni correction. The lower limit $L_E(\boldsymbol{X})$ and upper limit $U_E(\boldsymbol{X})$ satisfy the following conditions separately:

$$\begin{aligned}
P_{\boldsymbol{X}}\left\{L_E(\boldsymbol{X}) \leq q_{\frac{1-p}{2}}\right\} &= 1 - \alpha/2, \\
P_{\boldsymbol{X}}\left\{q_{\frac{1+p}{2}} \leq U_E(\boldsymbol{X})\right\} &= 1 - \alpha/2.
\end{aligned}$$

Next we briefly present existing methods to calculate tolerance limits for non-normal distributions that are frequently used in medical research, lifetime analysis, and quality control. The distributions we discuss are implemented in the `R` package `tolerance`.[34] More details on each of these tolerance intervals can be found in Krishnamoorthy and Mathew [25] and Young.[34]

Tolerance Intervals for the Weibull Distribution

A random variable X follows a Weibull distribution if it has the cumulative distribution function (CDF)

$$F_X(x;\theta,\beta) = 1 - e^{-\left(\frac{x}{\theta}\right)^{\beta}}, \quad x > 0,$$

where $\beta > 0$ is the shape parameter and $\theta > 0$ is the scale parameter. Let Y be the natural logarithm transformation on X. Then Y has an extreme-value distribution with CDF

$$F_Y(y; \xi, \delta) = 1 - \exp\left\{-e^{\left(\frac{y-\xi}{\delta}\right)}\right\}, \quad y \in \mathbb{R},$$

where $\xi = \ln(\theta)$, and $\delta = \beta^{-1}$. Tolerance limits for the extreme-value distribution are constructed based on the MLE $(\hat{\xi}, \hat{\delta})$, which is obtained numerically using the Newton-Raphson algorithm. The formulas for estimating the lower tolerance limit L_e and the upper tolerance limit U_e for the extreme-value distribution are

$$L_e = \hat{\xi} - \frac{\hat{\delta} t^*_{n-1;\alpha}\left(-\sqrt{n}\lambda_p\right)}{\sqrt{n-1}}$$
$$U_e = \hat{\xi} - \frac{\hat{\delta} t^*_{n-1;1-\alpha}\left(-\sqrt{n}\lambda_{1-p}\right)}{\sqrt{n-1}},$$

where $\lambda_a = \ln(-\ln(a))$ and $t^*_{n-1;\alpha}(\gamma)$ is the α quantile of the t-distribution with $n-1$ degrees of freedom and non-centrality parameter γ.[35]

Hence, by replacing α with $\alpha/2$ and p with $(p+1)/2$ in the tolerance limits L_e and U_e, we obtain the equal-tailed two-sided tolerance interval (L_e', U_e') as follows:

$$L_e' = \hat{\xi} - \frac{\hat{\delta} t^*_{n-1;\alpha/2}\left(-\sqrt{n}\lambda_{(p+1)/2}\right)}{\sqrt{n-1}}$$
$$U_e' = \hat{\xi} - \frac{\hat{\delta} t^*_{n-1;1-\alpha/2}\left(-\sqrt{n}\lambda_{(1-p)/2}\right)}{\sqrt{n-1}}.$$

Therefore, the equal-tailed two-sided tolerance interval (L_1, U_1) for a Weibull distribution is to exponentiate the above, which yields

$$L_1 = e^{L_e'}$$
$$U_1 = e^{U_e'}.$$

Tolerance Intervals for the Exponential Distribution

A random variable X follows an exponential distribution if it has the CDF

$$F_X(x;\lambda) = 1 - e^{-\lambda x}, \quad x > 0,$$

where $\lambda > 0$ is the scale parameter. For a sample of size n from an exponential distribution, the tolerance limits are constructed based on the maximum likelihood estimate (MLE) $\hat{\lambda} = \bar{x}_n^{-1}$. The formulas for estimating the lower tolerance limit L_1 and the upper tolerance limit U_1 are expressed as follows:

$$L_1' = -\frac{2n\hat{\lambda}\ln(1-p)}{\chi^2_{2n;1-\alpha}}$$
$$U_1' = -\frac{2n\hat{\lambda}\ln(p)}{\chi^2_{2n;\alpha}},$$

where $\chi^2_{2n;\alpha}$ is the α quantile of a χ^2 distribution with $2n$ degrees of freedom. Additional details on the derivation of these tolerance limits are discussed in Guenther [36] and Blischke and Murthy.[37] In fact, the general form of both L_1' and U_1' is $k\hat{\lambda}$, for which Guenther [36] provides sound mathematical justification in using this form for the tolerance limits. This k-factor can be explicitly derived and calculated as given in the formulas above, and it is in this sense that these tolerance limits are considered *exact*. Moreover, by replacing α with $\alpha/2$ and p with $(p+1)/2$ in the tolerance limits L_1' and U_1' we can obtain the equal-tailed two-sided tolerance interval (L_1, U_1) for an exponential distribution:

$$L_1 = -\frac{2n\hat{\lambda}\ln((1-p)/2)}{\chi^2_{2n;1-\alpha/2}}$$
$$U_1 = -\frac{2n\hat{\lambda}\ln((p+1)/2)}{\chi^2_{2n;\alpha/2}}.$$

Tolerance Intervals for the Logistic Distribution

A random variable X follows a logistic distribution if it has the CDF

$$F_X(x;\theta,\sigma) = \int_{-\infty}^{x} \left[1 + e^{-\frac{\pi(t-\theta)}{\sqrt{3\sigma^2}}}\right]^{-1} dt, \quad x \in \mathbb{R},$$

where the location parameter is $\theta \in \mathbb{R}$ and the scale parameter is $\sigma > 0$. Tolerance limits for the logistic distribution are constructed based on the MLE $(\hat{\theta}, \hat{\sigma})$, which can be obtained using Newton-Raphson to solve the necessary nonlinear equation. The formulas for estimating the lower tolerance limit L_1 and the upper tolerance limit U_1 are given below (Bain and Englehardt, Chapter 7)[38]:

$$L_1' = \hat{\theta} - k(\alpha, p)\hat{\sigma},$$
$$U_1' = \hat{\theta} + k(\alpha, p)\hat{\sigma}.$$

The above formulas depend on the tolerance factor $k(\alpha, p) = \sqrt{c/n} z_{1-\alpha} - z_{1-p}$, where $c = C_{11} + (z_{1-p})^2 C_{22} + 2z_{1-p}C_{12}$. Here, z_ξ is the ξ quantile of the standard normal distribution, $C_{11} = n\text{Var}(\hat{\theta})/\hat{\sigma}^2, C_{22} = n\text{Var}(\hat{\sigma})/\hat{\sigma}^2$, and $C_{12} = n\text{Cov}(\hat{\theta}, \hat{\sigma})/\hat{\sigma}^2$, where $\text{Var}(\hat{\theta}), \text{Var}(\hat{\sigma})$ and $\text{Cov}(\hat{\theta}, \hat{\sigma})$ are from the estimated variance-covariance matrix, which can be obtained by calculating the inverse of the negative Hessian evaluated at the MLE $(\hat{\theta}, \hat{\sigma})$.

Again, by replacing α with $\alpha/2$ and p with $(p+1)/2$ in the tolerance limits L_1' and U_1' we can obtain the equal-tailed two-sided tolerance interval (L_1, U_1) for a logistic distribution:

$$L_1 = \hat{\theta} - k(\alpha/2, (p+1)/2)\hat{\sigma},$$
$$U_1 = \hat{\theta} + k(\alpha/2, (p+1)/2)\hat{\sigma}.$$

Recall that if X is a log-logistic random variable, then $\log(X) = Y$ is a logistic random variable. Thus, tolerance intervals can be found for log-logistic data by proceeding with the above calculations, followed by transforming back to the log-logistic scale; i.e., by calculating e^{L_1} and e^{U_1} for the lower and upper limits, respectively.

2.3 Bootstrap Calibration

Most one-sided and two-sided tolerance interval formulas are approximate, and hence are not guaranteed to achieve exactly the nominal coverage level. This can often be demonstrated through coverage studies characterizing the procedure's behavior for small sample sizes. Moreover, one can construct two-sided tolerance intervals based on a Bonferroni correction applied to the formulas for the one-sided tolerance limits. However, because the Bonferroni correction is conservative, the actual coverage probabilities of the resulting two-sided tolerance interval will also typically be conservative relative to the nominal level. Hence, applying the bootstrap calibration on the confidence level for the two-sided tolerance interval can help improve the accuracy of the resulting interval closer to the nominal level.

The bootstrap calibration that was discussed in Chapter 18, Section 3 of Efron and Tibshirani,[10] was proposed for when the coverage of a $(1-\alpha)$ confidence interval procedure does not achieve the nominal coverage level. They employ the classic nonparametric bootstrap and compute the confidence points (since their algorithm was presented in the context of calibrating one-sided confidence limits) for each bootstrap sample at each confidence level, say $(1-\gamma)$, in a set of candidate confidence levels. Then, for each γ, the coverage probability is calculated using the B bootstrap resamples. Finally, they select the γ^* such that the $(1-\gamma^*)$ confidence interval achieves, or at least comes close to, the nominal coverage level. This $(1-\gamma^*)$ confidence interval is then the calibrated $(1-\alpha)$ confidence interval.

Classic bootstrap calibration is also properly termed an *iterated bootstrap* procedure because there is an outer layer of constructing a bootstrap-based interval (e.g., the bootstrap percentile interval or the bootstrap t interval) followed by an inner layer of bootstrap samples to calibrate the confidence coefficient of those intervals such that they yield coverages close to the nominal level. In our setting, we are calibrating approximate tolerance intervals directly. Thus, we do not require a bootstrap layer to construct a bootstrap-based interval, but only need to draw a parametric bootstrap sample to calibrate the confidence coefficient of the approximate tolerance interval. This notion of a single-layer calibration has been mentioned in the literature. Specifically, Loh [1] found that it was often enough to do a one-step calibration followed by linear interpolation in order to obtain a calibrated confidence interval with better coverage properties. This similar approach that we employ is validated by the numerical results presented in the next section.

The bootstrap calibration algorithm is implemented as follows:

Algorithm 1 Bootstrap Calibration Algorithm

1. Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ be a sample of size n from a distribution $F_{\boldsymbol{\theta}}$ parameterized by $\boldsymbol{\theta}$. Let $\hat{\boldsymbol{\theta}}$ denote the MLE of $\boldsymbol{\theta}$. We are interested in constructing a $(p, 1-\alpha)$ two-sided tolerance interval based on $\boldsymbol{x}$ such that it has nominal coverage $(1-\alpha)$.

2. For each $b = 1, 2, \ldots, B$, generate $\boldsymbol{x}_b^* = (x_{1b}^*, x_{2b}^*, \cdots, x_{nb}^*)$ using the parametric bootstrap method based on $F_{\hat{\boldsymbol{\theta}}}$. For an arbitrary confidence level, say $(1-\gamma)$, denote the $(p, 1-\gamma)$ two-sided tolerance interval for bootstrap sample as (L_b, U_b) and the content of the two-sided tolerance interval for bootstrap sample as $p_b^*(\gamma)$, where $p_b^*(\gamma) = F_{\hat{\boldsymbol{\theta}}}(U_b) - F_{\hat{\boldsymbol{\theta}}}(L_b)$. Note that this content is calculated relative to $F_{\hat{\boldsymbol{\theta}}}$.

3. Find the value γ^* such that

$$\gamma^* = \arg\min_{\gamma} \left| \frac{1}{B} \sum_{b=1}^{B} \mathtt{I}\{p \leq p_b^*(\gamma)\} - (1-\alpha) \right|,$$

   where $\mathtt{I}\{\cdot\}$ is the indicator function. Practically speaking, the above is implemented by choosing a set of candidate values $\gamma_1 < \gamma_2 < \cdots < \gamma_m$, where typically $\gamma_1 < \alpha < \gamma_m$.

4. The $(p, 1-\gamma^*)$ two-sided approximate tolerance interval using the original data, say $(L^*(\boldsymbol{x}), U^*(\boldsymbol{x}))$, is then the calibrated $(p, 1-\alpha)$ having coverage probability approximately equal to $(1-\alpha)$.

The algorithm is summarized as the follow diagram:

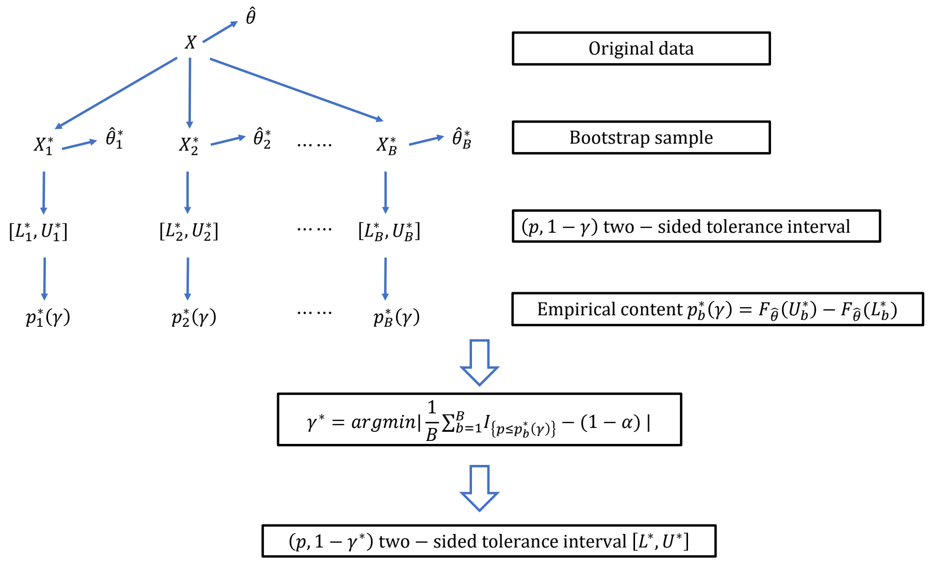


Figure 2.1: Diagram of the algorithm to calculate two-sided approximate tolerance interval on non-normal distributions with bootstrap calibration.

Calibrate $(1-\alpha)$ or p?

Bootstrap calibration is almost exclusively performed on confidence interval procedures.[1, 2, 3] In such settings, the calibration necessarily results in an adjustment on the confidence level in order to achieve confidence intervals with coverage probabilities that are close to or attain the nominal coverage level. For tolerance intervals, however, one could potentially adjust the content level p. Some authors have explored this approach. For example, Fernholz and Gillespie [39] developed a bootstrap approach where the content correction is achieved through adjusting the k-factor in symmetric tolerance intervals of the form $\bar{X} \pm k_{n,\alpha,p}S$, where n is the sample size. The adjusted content is $p_n^* = F_n(\bar{X} + k_{n,\alpha,p}S) - F_n(\bar{X} - k_{n,\alpha,p}S) - d_\alpha^*/\sqrt{n}$, where F_n is the empirical distribution function and d_α^* is a difference quantity estimated through the bootstrap. This procedure is limited to adjusting symmetric tolerance intervals in terms of the normal-based k-factor. Flouri et al [33] and Zhai et al [17] each applied a bootstrap calibration to adjust the content level p in order to improve the performance of the tolerance intervals for their respective problems. Flouri et al [33] further employed bootstrap calibration on the confidence level for constructing confidence intervals of the median of a ratio of normal random variables.

Our view is to focus on and advocate for calibration on the confidence level, which stems from a couple of pragmatic considerations. First, calibrating the confidence level is consistent with the setup that is used to traditionally construct the bootstrap-adjusted confidence intervals, such as those works discussed at the beginning of this section. Second, we view the content level p as being akin to a hard-to-change factor. When any statistical interval procedure is developed, it is desired to achieve the nominal coverage level, which is tied to the chosen confidence level. As long as the nominal coverage level is achieved, even through some sort of adjustment, then the interval procedure can, practically speaking, be considered efficacious. This includes the tolerance interval setting. However, in most tolerance interval applications, the value of p will be tied to some meaningful proportion of the sampled population that must be kept fixed in order to show compliance with, for example, regulatory guidelines by the FDA or the EPA. Thus, even though the roles of p and $(1-\alpha)$ can effectively be swapped in our bootstrap algorithm, we advocate for performing the calibration only on the confidence level.

2.4 Simulation Study

The performance of the bootstrap calibration is assessed on two-sided tolerance intervals for data generated from Weibull, exponential, and logistic distributions. The values for $(1-\alpha)$ and p were chosen from $\{0.90, 0.95\}$, which are levels typically used in practice.[25] We assessed the performance for sample sizes in the set $\{20, 50, 100, 200, 300\}$. The number of bootstrap samples to perform the calibration was set to $B = 500$. We then generated $M = 1000$ Monte Carlo samples, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$, from each generating distribution $F_{\boldsymbol{\theta}}$. We then ran the steps of Algorithm 1 for each sample and calculated the coverage probability as follows:

$$\text{CP} = \frac{1}{M}\sum_{j=1}^{M} \mathbb{I}\{F_{\boldsymbol{\theta}}(U^*(\boldsymbol{x}_m)) - F_{\boldsymbol{\theta}}(L^*(\boldsymbol{x}_m)) \geq p\}. \tag{2.1}$$

We also calculated the mean and standard deviation of the width of the calibrated tolerance interval. These metrics are also calculated for the uncalibrated tolerance interval for comparison.

The first set of simulation results is for the Weibull distribution with shape parameter $\beta = 5$ and scale parameter $\theta = 1$. The results are given in Table 2.1. As we can see, for the different simulation settings, the coverage probabilities using equal-tailed two-sided tolerance intervals without calibration are very conservative, while the calibrated tolerance intervals yield coverage probabilities close to the nominal level and also result in shorter average interval widths.

The second set of simulation results is for the exponential distribution with scale parameter $\lambda = 0.3$. The results are given in Table 2.2. For the different simulation scenarios, the coverage probabilities using equal-tailed two-sided tolerance interval without calibration are extremely conservative, in fact, quite close to 1. But again, the calibrated tolerance intervals have yielded coverage probabilities much closer to the nominal level and also shorter average interval widths.

The last set of simulation results we considered are for the logistic distribution with location parameter $\theta = 5$ and scale parameter $\sigma = 1$. The results are given in Table 2.3. The results here have slightly different performance compared to what was observed with the Weibull and exponential results. We see that when the sample size is relatively small, $n = 20$ or 50, the performance of the uncalibrated and calibrated tolerance intervals are both very close. For larger sample sizes, however, the uncalibrated tolerance intervals are clearly more conservative, while the calibrated tolerance intervals have yielded results coverage probabilities to nominal level and, again, shorter average interval widths.

Table 2.1: Estimated coverage probabilities, width mean, and standard deviation (sd) of two-sided tolerance intervals with/without calibration. n is the sample size of the simulation data set, $1-\alpha$ is the confidence level and p is the content.

Weibull			With Calibration		Without Calibration	
n	$1-\alpha$	p	Coverage	Mean (sd)	Coverage	Mean (sd)
20	0.95	0.90	0.952	0.933 (0.143)	0.979	1.007 (0.153)
20	0.95	0.95	0.950	1.093 (0.164)	0.974	1.163 (0.174)
20	0.90	0.90	0.899	0.873 (0.133)	0.960	0.944 (0.143)
20	0.90	0.95	0.900	1.024 (0.153)	0.954	1.093 (0.163)
50	0.95	0.90	0.943	0.821 (0.077)	0.990	0.871 (0.082)
50	0.95	0.95	0.939	0.965 (0.088)	0.985	1.013 (0.093)
50	0.90	0.90	0.893	0.791 (0.074)	0.962	0.838 (0.078)
50	0.90	0.95	0.892	0.930 (0.085)	0.958	0.977 (0.089)
100	0.95	0.90	0.946	0.777 (0.050)	0.992	0.814 (0.053)
100	0.95	0.95	0.943	0.915 (0.058)	0.985	0.950 (0.060)
100	0.90	0.90	0.895	0.758 (0.049)	0.968	0.792 (0.051)
100	0.90	0.95	0.895	0.892 (0.056)	0.964	0.927 (0.058)
200	0.95	0.90	0.943	0.751 (0.035)	0.992	0.778 (0.036)
200	0.95	0.95	0.942	0.885 (0.040)	0.989	0.911 (0.041)
200	0.90	0.90	0.900	0.738 (0.035)	0.978	0.763 (0.036)
200	0.90	0.95	0.899	0.870 (0.040)	0.970	0.895 (0.041)
300	0.95	0.90	0.942	0.739 (0.028)	0.989	0.761 (0.029)
300	0.95	0.95	0.942	0.871 (0.032)	0.985	0.892 (0.033)
300	0.90	0.90	0.899	0.728 (0.027)	0.974	0.749 (0.028)
300	0.90	0.95	0.897	0.859 (0.031)	0.969	0.880 (0.032)

2.5 Applications

In this section, we analyze three medical datasets using the tolerance intervals of Section 2.2 as well as applying a bootstrap calibration. Each analysis uses a distribution that has been shown to be appropriate for the respective dataset. Of course, further distributional assessments can be done, such as applying model selection criteria, but our focus is on constructing the tolerance intervals for the working model of the data, applying the bootstrap calibration, and then interpreting the resulting limits in the context of the application.

Table 2.2: Estimated coverage probabilities, width mean, and standard deviation (sd) of two-sided tolerance intervals with/without calibration. n is the sample size of the simulation data set, $1-\alpha$ is the confidence level and p is the content.

Exponential			With Calibration		Without Calibration	
n	$1-\alpha$	p	Coverage	Mean (sd)	Coverage	Mean (sd)
20	0.95	0.90	0.960	13.360 (2.966)	0.995	16.289 (3.604)
20	0.95	0.95	0.960	16.895 (3.751)	0.995	20.143 (4.457)
20	0.90	0.90	0.916	12.371 (2.744)	0.988	14.997 (3.318)
20	0.90	0.95	0.919	15.647 (3.470)	0.982	18.557 (4.106)
50	0.95	0.90	0.960	11.641 (1.641)	1.000	13.367 (1.828)
50	0.95	0.95	0.959	14.650 (2.066)	0.998	16.558 (2.264)
50	0.90	0.90	0.912	11.125 (1.569)	0.995	12.719 (1.739)
50	0.90	0.95	0.915	14.006 (1.975)	0.991	15.764 (2.156)
100	0.95	0.90	0.967	11.013 (1.076)	0.999	12.100 (1.186)
100	0.95	0.95	0.966	13.823 (1.350)	0.997	15.004 (1.471)
100	0.90	0.90	0.915	10.680 (1.043)	0.993	11.692 (1.146)
100	0.90	0.95	0.917	13.408 (1.309)	0.989	14.505 (1.422)
200	0.95	0.90	0.964	10.583 (0.716)	1.000	11.354 (0.788)
200	0.95	0.95	0.963	13.255 (0.896)	0.998	14.091 (0.978)
200	0.90	0.90	0.910	10.363 (0.701)	0.997	11.086 (0.770)
200	0.90	0.95	0.912	12.980 (0.878)	0.995	13.763 (0.956)
300	0.95	0.90	0.963	10.429 (0.589)	1.000	11.035 (0.611)
300	0.95	0.95	0.963	13.049 (0.737)	1.000	13.700 (0.758)
300	0.90	0.90	0.904	10.254 (0.580)	0.998	10.823 (0.599)
300	0.90	0.95	0.909	12.830 (0.725)	0.993	13.442 (0.744)

Example 1: Bladder Cancer Remission Times

Data about remission times (in months) of $n = 137$ bladder cancer patients [40] were fitted by the Weibull distribution. A histogram of these data with the estimated Weibull density curve overlaid is given in Figure 2.2. The MLEs of the shape and scale parameters for this distribution are 1.051 and 9.415, respectively. Construction of a two-sided tolerance interval for remission times can provide potential insight into various medical concerns, such as

1. how to develop personalized treatment plans based on remission times for the majority of such cancer patients;

2. informing capacity planning within a cancer unit; and

Table 2.3: Estimated coverage probabilities, width mean, and standard deviation (sd) of two-sided tolerance intervals with/without calibration. n is the sample size of the simulation data set, $1-\alpha$ is the confidence level and p is the content.

Logistic			With Calibration		Without Calibration	
n	$1-\alpha$	p	Coverage	Mean (sd)	Coverage	Mean (sd)
20	0.95	0.90	0.922	8.242 (1.603)	0.921	8.213 (1.609)
20	0.95	0.95	0.916	10.064 (1.956)	0.914	10.027 (1.965)
20	0.90	0.90	0.881	7.798 (1.517)	0.891	7.805 (1.529)
20	0.90	0.95	0.870	9.574 (1.860)	0.881	9.549 (1.871)
50	0.95	0.90	0.946	7.316 (0.891)	0.946	7.446 (0.926)
50	0.95	0.95	0.941	9.074 (1.105)	0.933	9.141 (1.137)
50	0.90	0.90	0.888	6.981 (0.849)	0.905	7.182 (0.893)
50	0.90	0.95	0.887	8.672 (1.055)	0.893	8.832 (1.099)
100	0.95	0.90	0.949	6.821 (0.568)	0.971	7.011 (0.603)
100	0.95	0.95	0.950	8.480 (0.706)	0.966	8.635 (0.742)
100	0.90	0.90	0.900	6.605 (0.549)	0.942	6.823 (0.587)
100	0.90	0.95	0.899	8.212 (0.682)	0.934	8.415 (0.724)
200	0.95	0.90	0.945	6.528 (0.403)	0.983	6.701 (0.409)
200	0.95	0.95	0.947	8.119 (0.502)	0.976	8.275 (0.505)
200	0.90	0.90	0.888	6.388 (0.393)	0.963	6.567 (0.401)
200	0.90	0.95	0.888	7.941 (0.488)	0.948	8.119 (0.496)
300	0.95	0.90	0.946	6.394 (0.316)	0.981	6.569 (0.324)
300	0.95	0.95	0.945	7.953 (0.393)	0.976	8.123 (0.400)
300	0.90	0.90	0.894	6.285 (0.309)	0.967	6.460 (0.318)
300	0.90	0.95	0.891	7.811 (0.385)	0.961	7.995 (0.394)

3. gauging overall quality of medical care by comparing the remission times of future patients (once those are observed) to the established tolerance limits.

We proceed to calculate $(0.95, 0.95)$ two-sided Weibull tolerance intervals for these data with and without calibration. The calibrated confidence level is about 0.84. The results are presented in Table 2.4 and also overlaid on the histogram in Figure 2.2. As we can see, the more conservative tolerance interval without calibration would lead to an upper limit that is nearly three months larger than the calibrated tolerance interval, while both have lower limits that are similar. Thus, the calibrated $(0.95, 0.95)$ tolerance interval informs us that with 95% confidence, at least 95% of all bladder cancer remissions times will fall between 0.194 and 39.016 months. Moreover, the fact that the calibrated $(0.95, 0.95)$ tolerance interval has a lower upper limit by nearly three months could have important implications for assessing the remission times, especially given the critical role of time when dealing with cancer.

Table 2.4: $(0.95, 0.95)$ two-sided Weibull tolerance intervals for the bladder cancer data, with and without calibration. The unit for these limits is in months.

	Lower Limit	Upper Limit
Without Calibration	0.165	41.981
With Calibration	0.194	39.016

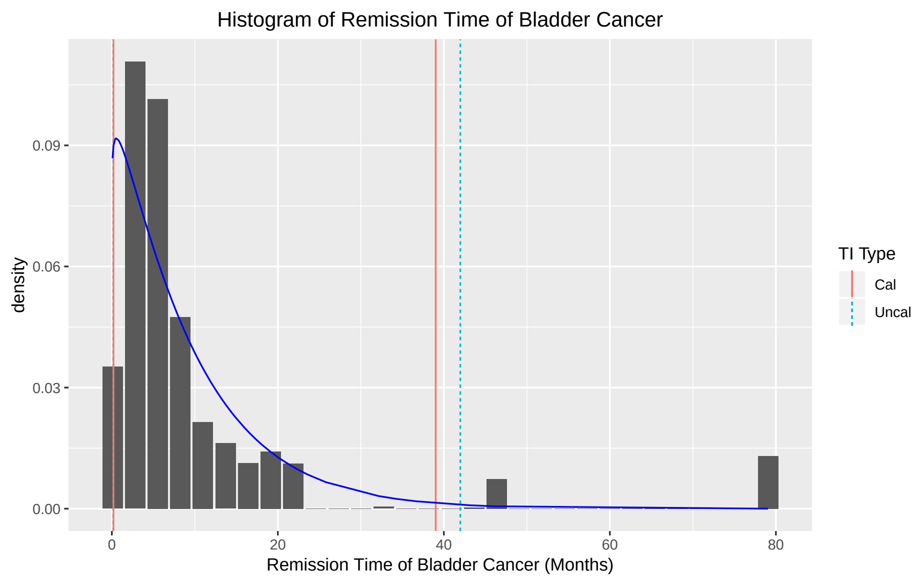


Figure 2.2: Histogram of the bladder cancer data along with the estimated Weibull density curve and vertical lines for the limits of the calibrated and uncalibrated $(0.95, 0.95)$ tolerance intervals.

Example 2: Urine Albumin-to-Creatinine Ratio (UACR) Measurements

Kidney function laboratory tests are used to evaluate how well the kidneys are working and to help identify kidney diseases. One such test measures the urine albumin-to-creatinine ratio (UACR), also called urine microalbumin, which is typically checked and monitored immediately after a patient is diagnosed with diabetes. Specifically, it helps diagnose the increased excretion of urinary albumin, which is called albuminuria.

The National Kidney Foundation [41] has published the reference ranges given in Table 2.5 for the three categories of albuminuria in chronic kidney disease. These reference ranges were established for the adult population. However, the detection and management of kidney disease in children and adolescents is a major initiative of the National Kidney Foundation.[42] Thus, reference ranges for this population could be insightful, especially for the treatment and management of juvenile diabetes. Reference ranges have been calculated using a number of statistical methods.[43] Perhaps the most common approaches have been to simply use specified percentiles from a reference sample or calculate prediction intervals. However, tolerance intervals have been advocated as a more sound approach given that they provide information about the entire population.[44, 45, 46]

Table 2.5: Albuminuria categories in chronic kidney disease for adults based on UACR measurements.

Category	UACR (mg/g)	Terms
A1	< 30	Normal to mildly increased
A2	$30 - 300$	Moderately increased
A3	> 300	Severely increased

The reference population for our analysis is created using survey participants from NHANES 1999–2014 who met the following criteria: between 12 to 17 years old, not pregnant, blood pressure $<$ 120/80 mmHg, without diabetes, no prescription medications used within the previous 30 days, and a Z-score for weight-to-height ratio ≤ 2. This yields a reference sample of size $n = 5255$. We analyze these data using an exponential distribution. The MLE of the rate parameter for this distribution is 3.186. A histogram of these data with the estimated exponential density curve overlaid is given in Figure 2.3.

We proceed to calculate $(0.95, 0.95)$ two-sided exponential tolerance intervals for these data with and without calibration. The calibrated confidence level is about 0.73. The results are presented in top-half of Table 2.6 and also overlaid on the histogram in Figure 2.3. For this dataset, the tolerance interval without calibration is only slightly more conservative relative to the calibrated tolerance interval. Regardless, the calibrated $(0.95, 0.95)$ tolerance

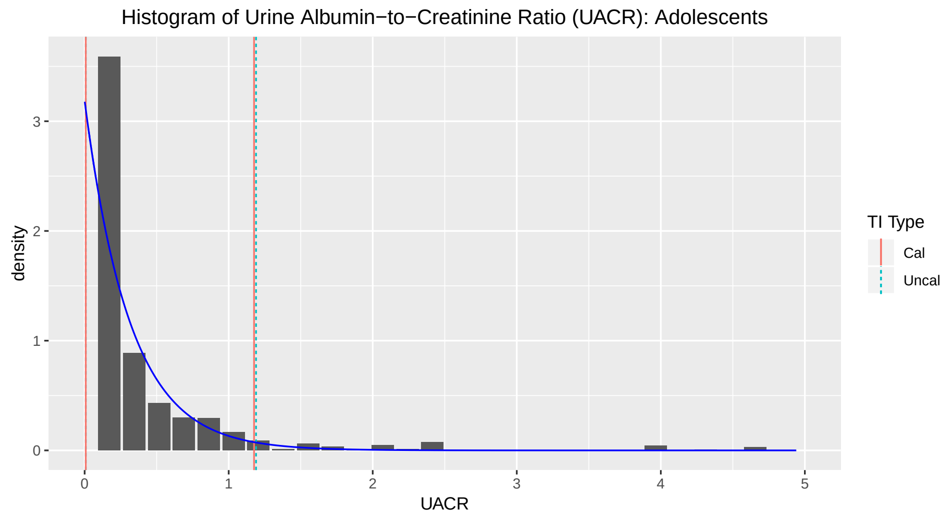


Figure 2.3: Histogram of the UACR data for the adolescent population along with the estimated exponential density curve and vertical lines for the limits of the calibrated and uncalibrated $(0.95, 0.95)$ tolerance intervals. Note that we truncated the histogram at 5 to emphasize the majority of the dataset. There are 48 observations larger than 5 not displayed on this figure. The maximum of those observations not shown is 36.88.

interval has the interpretation that with 95% confidence, at least 95% of all healthy adolescents have UACR values between 0.0078 mg/g and 1.1760 mg/g. This could help, for example, establish reference ranges for adolescents that are analogous to the A1 category for adults in Table 2.5. In fact, since we have calculated a two-sided tolerance interval, we could propose splitting A1 into two categories. Our interval could be used to characterize patients as being borderline normal to moderately increased, whereas anything that falls below our lower limit could be classified strictly as normal. However, if a direct analogue to A1 for adolescents is sought, then we could proceed to calculate a $(0.95, 0.95)$ one-sided upper exponential tolerance limit, which would be 0.7358 mg/g. As we noted, the one-sided exponential tolerance limit is exact, thus, no calibration would be necessary for this limit.

For completeness, we also created an adult reference population from NHANES 1999-2014 using the same criteria as for the adolescent reference population. For this reference population, the subjects must be over 17 years old. The MLE of the rate parameter for this

Table 2.6: $(0.95, 0.95)$ two-sided exponential tolerance intervals for the UACR data, with and without calibration. The top-half of the table gives the results for the adolescent population and the bottom-half of the table gives the results for the adult population.

	Lower Limit	Upper Limit
Adolescent Population		
Without Calibration	0.0077	1.1900
With Calibration	0.0078	1.1760
Adult Population		
Without Calibration	0.0031	0.4742
With Calibration	0.0032	0.4696

The unit for these limits is in mg/g.

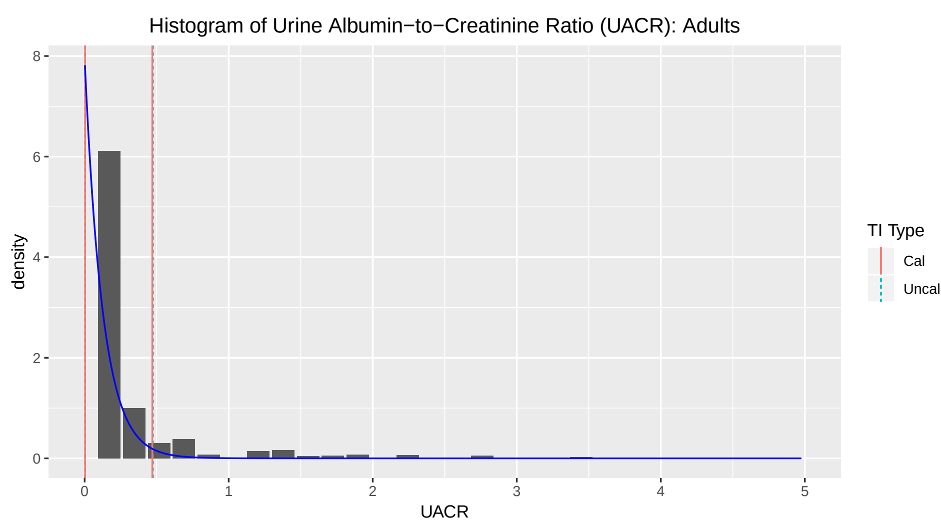


Figure 2.4: Histogram of the UACR data for the adult reference population along with the estimated exponential density curve and vertical lines for the limits of the calibrated and uncalibrated $(0.95, 0.95)$ tolerance intervals. Note that we truncated the histogram at 5 to emphasize the majority of the dataset. There are 13 observations larger than 5 not displayed on this figure. The maximum of those observations not shown is 28.17.

distribution is 0.1259, which is noticeably lower than that from the adolescent population. The uncalibrated and calibrated $(0.95, 0.95)$ tolerance intervals are reported in the bottom-half of Table 2.6. Since the MLE is lower for this population, we obtain a lower upper tolerance limit as expected. In fact, it is noticeably lower than that for the adolescent population, which would further suggest that modifying the reference ranges in Table 2.5 for categorizing chronic kidney disease could be justified.

Example 3: Breast Cancer Survival Times

Ramos et al [47] and Tahir et al [48] developed generalizations to the log-logistic distribution, which they used to analyze survival times (in months) of $n = 121$ patients with breast cancer observed from a large hospital between the years of 1929 and 1938. These analyses also showed that the classic log-logistic provides a good fit. A histogram of these data with the estimated log-logistic density curve overlaid is given in Figure 2.5. The MLEs of the scale and shape parameters for this distribution are 35.177 and 1.856, respectively. Construction of a two-sided tolerance interval for the survival times can provide potential insight into the general time frame that breast cancer patients survived around these years of operation of the hospital. Comparisons can then be made with subsequent hospital records of survival times in later years to indicate when these have (ideally) become longer.

We proceed to calculate $(0.90, 0.90)$ two-sided log-logistic tolerance intervals for these data with and without calibration. The calibrated confidence level is about 0.77. The results are presented in Table 2.7 and also overlaid on the histogram in Figure 2.5. As we can see, just like with the previous examples, the more conservative tolerance interval without calibration would lead to an upper limit that is nearly 14 months larger than the calibrated tolerance interval, while both have lower limits that are similar. Thus, the calibrated $(0.90, 0.90)$ tolerance interval informs us that with 90% confidence, at least 90% of all breast cancer survival times at this hospital will fall between 5.996 and 206.379 months.

Table 2.7: $(0.90, 0.90)$ two-sided log-logistic tolerance intervals for the breast cancer data, with and without calibration. The unit for these limits is in months.

	Lower Limit	Upper Limit
Without Calibration	5.614	220.415
With Calibration	5.996	206.379

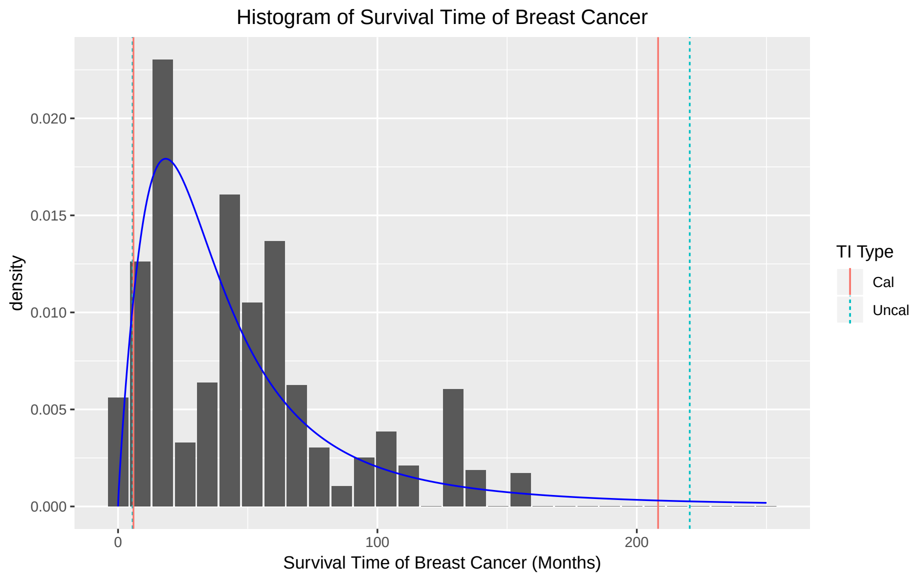


Figure 2.5: Histogram of the breast cancer survival time data along with the estimated log-logistic density curve and vertical lines for the limits of the calibrated and uncalibrated $(0.95, 0.95)$ tolerance intervals.

2.6 Conclusion and Discussion

Exact one-sided tolerance limits and exact two-sided tolerance intervals are not available for many non-normal distributions. For a given distribution, approximate one-sided tolerance limits have varying degrees of performance in terms of their coverage properties, with many such procedures being conservative relative to the nominal level. Moreover, approximate two-sided tolerance intervals constructed based on applying the Bonferroni correction to an approximate one-sided procedure will usually compound the degree of conservatism.

Our work demonstrated the efficacy of using a single-layer bootstrap calibration to achieve coverage probabilities closer to the nominal level for parametric tolerance interval procedures. The bootstrap calibration has almost exclusively been used on confidence interval procedures, with some recent applications to very specific tolerance interval problems. Our work contributed a more general treatment of the bootstrap calibration to parametric tolerance intervals. We provided extensive numerical studies of applying the bootstrap calibration to some common non-normal tolerance interval procedures. The results clearly underscored that the calibrated tolerance intervals have better coverage probabilities and overall smaller average interval widths compared to uncalibrated intervals. While we only highlighted the improvements under three different non-normal settings, improvements should be expected in any parametric setting since our calibration routine implements a parametric bootstrap, which is known to provide better accuracy for inference procedures over those based on (asymptotic) approximations.[49] We further highlighted the potential impact of using the calibrated tolerance intervals that were calculated for three real medical datasets.

This work only highlighted the use of bootstrap calibration on tolerance intervals for univariate distributions. This approach could be extended to tolerance regions for multivariate parametric distributions, however, tolerance region procedures are only developed for a limited number of such distributions, like the multivariate normal.[50, 51, 52] Additional applications of the bootstrap calibration could be coupled with novel developments to construct approximate tolerance regions for non-normal multivariate tolerance regions.

Chapter 3 Confidence Interval of the Mean and Upper Tolerance Limit for Zero-Inflated Gamma Data

3.1 Introduction

Continuous outcomes mixed with a a high proportion of zeros are frequently encountered in practical situations such as biomedical, economical, and ecological studies.[53, 54, 55, 56, 57, 58] The excessive zeros need to be considered in data analysis. A question arises regarding how those zeros are generated. Similarly to count data, there are two possible scenarios: *true zeros*, when the zeros are true observations; *censoring*, when the zeros are censored values below the detection limit. In the medical expenses data,[55, 56] zero values mean that the patient does not have any medical bill, so the zero is truly observed. Also in the fisheries data,[57] zero values occur when there is no fish in the area. On the other hand, zero values reported from a drug assay arise when the measured concentration is below the detection limit, and the drug concentration is not necessarily zero. In this work, we will limit our discussion to semicontinuous data, in which all zeros observed are true zeros. Our motivation comes from the following examples. In fisheries research, a problem of interest is to estimate the mean density of a species when the population is skewed and contains a relatively high proportion of zeros.[57] In exposure/pollution assessment, an upper confidence limit of the mean or an upper tolerance limit (also upper confidence limit of the the upper percentile) of the exposure distribution can be used to determine if the exposure levels are within the federal standards.[58] Zero-inflated data can be fit using zero-inflated lognormal (delta-lognormal) distribution. Hasan and Krishnamoorthy [59] provided inference for mean and percentile of such zero-inflated lognormal distribution using fiducial inference. Note that lognormal distribution and gamma distribution can be used quite effectively for analyzing skewed non-negative data sets. Zero-inflated continuous data could be also modeled as zero-inflated gamma (ZIG) distribution. However, the inference of ZIG distribution has not been well studied in the literature. First, we will briefly introduce the definition for ZIG distribution. The ZIG distribution is a generalized form of the gamma distribution in which a proportion of π of the observations may be zeros and the positive values follow a gamma distribution with shape parameter α and rate parameter β. The distribution function for a ZIG is given by

$$G(x;\pi,\alpha,\beta) = \pi + (1-\pi)\mathbb{I}\{x>0\}F(x;\alpha,\beta)$$

where $F(x;\alpha,\beta)$ is the distribution function for gamma distribution Gamma(α,β). We are interested in the inference for mean and the upper tolerance limit. The population mean thus is given by

$$M = \frac{(1-\pi)\alpha}{\beta}.$$

The pth quantile, denoted by q_p of the ZIG distribution is given by $G(q_p;\pi,\alpha,\beta) = p$, thus q_p is the root of the equation

$$F(q_p;\alpha,\beta) = \frac{p-\pi}{1-\pi}.$$

Hence by solving this equation we have

$$q_p = F^{-1}(\frac{p-\pi}{1-\pi};\alpha,\beta),$$

where $F^{-1}(\cdot;\alpha,\beta)$ is the quantile function for gamma distribution Gamma(α,β).

Tian [60] and Li et al [61] have applied the method of *generalized variable* approach, which was first introduced by Weerahandi,[11] to study the mean of zero-inflated lognormal data. Hannig et al [12] revealed the connection between generalized variable method and fiducial inference. The confidence intervals for the mean proposed by Tian and Li were not consistently close to normal coverage probabilities among different simulation scenarios, thus Hasan and Krishnamoorthy [59] modified their approach by changing the fiducial distribution used for zero proportion π, which yielded better result. By using the idea of *generalized fiducial inference* (GFI),[6] Chen et al [62] also explored the fiducial distributions for gamma distribution parameters and constructed confidence intervals for some important quantities such as mean and quantile. Wang et al [63] also used fiducial inference to propose a new structural equation to generate inference for gamma distribution that outperformed Chen's method without making the assumption that scale parameter is large. In this work, we shall use the fiducial approach to investigate the confidence interval for mean and upper tolerance limit of ZIG.

This chapter is organized as follows. In Section 3.2, we present the fiducial distributions for binomial and gamma parameters. In Section 3.3, we introduce the algorithm on how to calculate the fiducial confidence interval for the mean and the upper tolerance limit of the ZIG. In Section 3.4, we give an extensive simulation to demonstrate the satisfactory performance of our method. In Section 3.5, we illustrate the utility of our method using two real examples. In the last Section 3.6, we make a conclusion of our method.

3.2 Fiducial Disitributions

The aim of GFI is to define a distribution for parameters of interest that contains all the information from data. Therefore, inference for the parameters can be made through this distribution. Such distribution is called fiducial distribution, which can be interpreted as the posterior distribution without assuming prior distribution.[64] The random variable having the fiducial distribution is denoted as *generalized fiducial quantity* (GFQ). The basic idea of the GFI is to switch the role of the parameters and the data. We will explain the philosophy of GFI briefly. Suppose data Y (possibly discrete) is generated through the structural equation $Y = G(\xi, U)$ where ξ is a vector of parameters and U is some random variable with a known distribution independent of the parameter ξ. The structural equation can be regarded as data generation process that noise process U and the signal ξ will produce observed data Y. Hence, the distribution of Y can be decided via structural equation given a fixed parameter ξ and the distribution U. After the data Y is observed, we can switch the position of data and parameters by solving the structural equation conditioning on that the solution to that equation exists, then we can get $\xi = Q(Y, U)$. For more detail regarding GFI, Hannig provided more thorough discussion in his paper.[6]

For the fiducial distributions of multiple parameters, a two-stage method can be applied based on the minimal sufficient statistics. Let the parameters of interest be $\boldsymbol{\xi} = (\xi_1, \xi_2)$. Assume that the following two conditions hold:

1. If ξ_2 is known, there is a statistic $\mathcal{S}_1 = \mathcal{S}_1(\xi_2)$ that has an invertible pivotal relationship with ξ_1.

2. A statistic $\mathcal{S}_2$ exists that $\mathcal{S}_2$ and ξ_2 have an invertible pivotal relationship.

Then we can first obtain the fiducial distribution of ξ_2 and then obtain the fiducial distribution of ξ_1 given that ξ_2 is known.

ZIG can be considered as a two-part model. Suppose sample size from ZIG is n, and there are n_0 zeros. The number of zero n_0 follows a binomial distribution Binomial(n, π), and the rest of the $n - n_0$ positive values come from gamma distribution Gamma(α, β). Thus the fiducial distribution of (π, α, β) can be obtained separately for fiducial distribution of binomial parameter π and fiducial distribution of gamma parameters (α, β) since they are independent.

Binomial parameter

Let $X \sim$ Binomial(n, π) where m is the number of trials and π is the success rate. As noted in Hannig's paper,[6] a GFQ of p is

$$\mathcal{R}_\pi(x) = U_{s:n} + D(U_{s+1:n} - U_{s:n}),$$

where D is any random variable with support contained in [0,1], n is the number of trials, s is the observed number of successes, $U_1, \cdots, U_n$ are i.i.d. $\mathcal{U}(0,1)$ and $U_{s:n}$ is sth the order statistic among $U_1, \cdots, U_n$.

As we can see from the equation of GFQ for p, due to the selection of random variable D, GFQ for p will not be unique. Hannig recommended the choice D follows 50-50 mixture of 0 and 1 through extensive simulation study. Hence, the GFQ of p will be written as a 50-50 mixture of Beta$(x+1, n-x)$ and Beta$(x, n-x+1)$ distributions.

Gamma parameters

Let $X_1, \cdots, X_n$ be a random sample from a gamma distribution Gamma(α, β) with the density function

$$f(x;\alpha,\beta) = \frac{\alpha^\beta}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x > 0,$$

where α is the shape parameter and β is the rate parameter. Let $\bar{X}$ be the algorithm mean $\frac{1}{n}\sum_{i=1}^n X_i$, $\tilde{X}$ be the geometric mean $(\prod_{i=1}^n X_i)^{\frac{1}{n}}$ and $T = log(\frac{\tilde{X}}{\bar{X}})$. Then $(T, \bar{X})$ are sufficient and complete statistics. Note that T only depends on the shape parameter α.

Wang et al [63] proposed a structural equation for shape parameter α based on Cornish-Fisher expansion. Let $\kappa_i(\alpha)$ is the ith cumulant of T, λ is confidence level, the equation of $Q(\alpha, \lambda)$ is defined as

$$\begin{aligned} Q(\alpha,\lambda) =& z_\lambda + \frac{1}{6}\kappa_3'(\alpha)(z_\lambda^2 - 1) + \frac{1}{24}\kappa_4'(\alpha)(z_\lambda^3 - 3z_\lambda) - \frac{1}{36}[\kappa_3'(\alpha)]^2(2z_\lambda^3 - 5z_\lambda) + \\ & \frac{1}{120}\kappa_5'(\alpha)(z_\lambda^4 - 6z_\lambda^2 + 3) - \frac{1}{24}\kappa_3'(\alpha)\kappa_4'(\alpha)(z_\lambda^4 - 5z_\lambda^2 + 2) + \\ & \frac{1}{324}[\kappa_3'(\alpha)]^3(12z_\lambda^4 - 53z_\lambda^2 + 17), \end{aligned}$$

where $\kappa_i'(\alpha) = \kappa_i(\alpha)/[\kappa_2(\alpha)]^{\frac{1}{2}}$, $\kappa_1(\alpha) = \log(n) + \psi(\alpha) - \psi(n\alpha)$, $\kappa_i(\alpha) = \frac{1}{n^{i-1}}\psi^{(i-1)}(\alpha) - \psi^{(i-1)}(n\alpha)$, $\psi(\alpha) = \frac{d\log(\Gamma(\alpha))}{d\alpha}$ and $\psi^{(k)}(\alpha)$ is the kth derivative of $\psi(\alpha)$ and z_λ is the λ percentile of the standard normal distribution $\mathrm{N}(0,1)$. According to Cornish-Fisher expansion, the λ approximate percentile of T is given by

$$t_\lambda(\alpha) = \kappa_1(\alpha) + [\kappa_2(\alpha)]^{1/2} Q(\alpha, \lambda).$$

We can then use such approximation to construct a structural equation for the shape parameter α. Let $F(t|\alpha)$ be the CDF of the statistic T, then $F(T|\alpha)$ will follow the uniform distribution $\mathcal{U}(0,1)$. For a given $U \sim \mathcal{U}(0,1)$, we have

$$T = \kappa_1(\alpha) + [\kappa_2(\alpha)]^{\frac{1}{2}} Q(\alpha, U).$$

Based on this equation, we can get $\alpha = g(T,U)$. It should be noted that T is treated as a known quantity after the sample is obtained. Given a realization $U^* \sim \mathcal{U}(0,1)$, a realization α^* of the fiducial distribution of α can be obtained numerically as $g(T,U^*)$. As for the rate parameter β. Notice that $2n\beta\bar{X} \sim \chi^2(2n\alpha)$, T and $\bar{X}$ are independent. Given a realization $U^* \sim \mathcal{U}(0,1)$ and a realization $V^* \sim \chi^2(2ng(T,U^*))$, a realization β^* of the fiducial distribution of β can be obtained numerically as $\frac{V^*}{2n\bar{X}}$.

3.3 Inference for Mean and Upper Quantile

Hence, we can calculate the fiducial confidence interval for the mean and the upper tolerance limit of the ZIG. Firstly, we give more detail about how to simulate the GFQ for π. As we mention in the discussion of binomial parameter, it should be a 50-50 mixture of Beta$(x+1, n-x)$ and Beta$(x, n-x+1)$ distributions. Since x is a integer from $\{0, 1, \cdots, n\}$, x and $n-x$ should be positive for Beta distribution and success probability π is between $[0,1]$. Hence, the GFQ of π is defined as

$$\mathcal{R}_\pi(x) = \begin{cases} (1-D)\text{Beta}(1, n-1) & x = 0 \\ D\text{Beta}(x, n-x+1) + (1-D)\text{Beta}(x+1, n-x) & x = 1, \cdots, n-1 \\ D\text{Beta}(n, 1) + (1-D) & x = n \end{cases}$$

where $D \sim$ Bernoulli($\frac{1}{2}$). Given a realization $(\pi^*, \alpha^*, \beta^*)$ of the fiducial distribution of (π, α, β), the realization of the mean and the upper p-percentile of the ZIG will be

$$M^* = \frac{(1-\pi^*)\alpha^*}{\beta^*}$$
$$q_p^* = F^{-1}(\frac{p-\pi^*}{1-\pi^*}; \alpha^*, \beta^*).$$

Let M_λ^* be the λ percentile of the GFQ of M and $q_{p,\lambda}^*$ be the λ percentile of the FPQ of q_p, then $[M_{\lambda/2}^*, M_{1-\lambda/2}^*]$ is a $1-\lambda$ fiducial confidence interval for the mean M and $q_{p,1-\lambda}^*$ is the upper tolerance limit with content level p and confidence level $1-\lambda$. We are able to calculate those quantities using Monte-Carlo simulations through the algorithm below.

Algorithm 2 Fiducial confidence interval for the mean of ZIG and upper tolerance limit

1. Given sample $(X_1, \cdots, X_n)$, compute $T, \bar{X}$ and the number of zeros n_0.
2. Generate a realization π^* from $\mathcal{R}_\pi(n_0)$.
3. Generate a realization $U^* \sim \mathcal{U}(0,1)$, then obtain a realization $\alpha^* = g(T, U^*)$.
4. Generate a realization$V^* \sim \chi^2(2ng(T, U^*))$, then obtain a realization $\beta^* = \frac{V^*}{2n\bar{X}}$.
5. Compute M^* and q_p^* given the realization $(\pi^*, \alpha^*, \beta^*)$.
6. Repeat Steps 2-5 B times, there are B sets of M and q_p.
7. The fiducial confidence interval for the mean will be the $\lambda/2$ quantile and $1-\lambda/2$ quantile of the sample of M^* and the upper tolerance limit will be the $1-\lambda$ quantile of the sample of q_p^*.

3.4 Simulation Study

The performance of the proposed method to construct the confidence interval for mean and upper tolerance limit is assessed through a Monte Carlo simulation. The rate parameter β was fixed to be $\beta = 1$. The shape parameter was chosen from $\{0.5, 1, 3\}$. The zero proportion of the data was set to be $\{0.2, 0.4, 0.6\}$. The number of sample size was set to be $\{15, 25, 50, 100\}$. The number of fiducial distribution sample B was set to be $B = 10000$. $M = 10000$ Monte Carlo samples were used for the assessment. Estimated coverage probabilities and mean width of 95% confidence intervals for the mean and estimated coverage probabilities for $(0.95, 0.95)$ and $(0.90, 0.95)$ upper tolerance limits were calculated.

The simulation result for various simulation scenarios is given in Table 3.1. As we can see, for the different simulation settings, the coverage probabilities of 95% confidence intervals and $(0.95, 0.95)$ and $(0.90, 0.95)$ upper tolerance limits are very close to the nominal level. It should be noted that for $n = 15$ and $\pi = 0.6$, there was a high chance that the simulated datasets only contained one or two positive values, which caused problem for the estimation of Gamma distribution parameters. Thus, we excluded those cases when we calculated the coverage probability.

Table 3.1: Estimated coverage probabilities and mean width of 95% confidence intervals for the mean and estimated coverage probabilities for $(0.95, 0.95)$ and $(0.90, 0.95)$ upper tolerance limits. n is the sample size of the simulation data set, π is the proportion of the zeros, α is the shape parameter, β is the rate parameter, nrepl is the number of Monte-Carlo simulation, and nfid is the number of fiducial sample.

n	π	α	β	nrepl	nfid	cp_mean	width_mean	cp_tl_90	cp_tl_95
15	0.2	0.5	1	10000	10000	0.953	1.063	0.940	0.941
25	0.2	0.5	1	10000	10000	0.950	0.654	0.944	0.944
50	0.2	0.5	1	10000	10000	0.949	0.410	0.941	0.943
100	0.2	0.5	1	10000	10000	0.949	0.273	0.945	0.945
15	0.4	0.5	1	10000	10000	0.958	1.675	0.939	0.940
25	0.4	0.5	1	10000	10000	0.951	0.662	0.941	0.942
50	0.4	0.5	1	10000	10000	0.948	0.387	0.938	0.941
100	0.4	0.5	1	10000	10000	0.952	0.251	0.941	0.942
15	0.6	0.5	1	9945	10000	0.954	9.336	0.932	0.938
25	0.6	0.5	1	10000	10000	0.955	0.946	0.936	0.939
50	0.6	0.5	1	10000	10000	0.953	0.360	0.940	0.941
100	0.6	0.5	1	10000	10000	0.953	0.221	0.943	0.944
15	0.2	1	1	10000	10000	0.955	1.265	0.938	0.939
25	0.2	1	1	10000	10000	0.951	0.871	0.941	0.942
50	0.2	1	1	10000	10000	0.951	0.575	0.942	0.943
100	0.2	1	1	10000	10000	0.949	0.395	0.946	0.947
15	0.4	1	1	10000	10000	0.959	1.460	0.941	0.942
25	0.4	1	1	10000	10000	0.952	0.852	0.940	0.941
50	0.4	1	1	10000	10000	0.952	0.547	0.940	0.944
100	0.4	1	1	10000	10000	0.950	0.371	0.943	0.943
15	0.6	1	1	9940	10000	0.961	4.813	0.931	0.936
25	0.6	1	1	10000	10000	0.956	0.929	0.931	0.937
50	0.6	1	1	10000	10000	0.955	0.502	0.939	0.940
100	0.6	1	1	10000	10000	0.950	0.331	0.941	0.942
15	0.2	3	1	10000	10000	0.958	2.198	0.940	0.943
25	0.2	3	1	10000	10000	0.956	1.622	0.942	0.941
50	0.2	3	1	10000	10000	0.952	1.113	0.945	0.946
100	0.2	3	1	10000	10000	0.954	0.778	0.947	0.948
15	0.4	3	1	10000	10000	0.958	2.293	0.937	0.939
25	0.4	3	1	10000	10000	0.956	1.650	0.940	0.943
50	0.4	3	1	10000	10000	0.949	1.127	0.941	0.944
100	0.4	3	1	10000	10000	0.952	0.787	0.941	0.944
15	0.6	3	1	9962	10000	0.963	3.881	0.930	0.932
25	0.6	3	1	10000	10000	0.958	1.596	0.935	0.937
50	0.6	3	1	10000	10000	0.952	1.055	0.941	0.943
100	0.6	3	1	10000	10000	0.949	0.729	0.944	0.948

3.5 Applications

Air Contaminants Data

The example is about the measurement of worker exposure to air contaminants.[58] The dataset consists of 15 measurements of the concentration of airborne chlorine (in parts per million), where 6 of them were 0. Non-zero values have been modeled by lognormal distribution.[58, 59, 60] Here, we fit the non-zero observations using gamma distribution. The fitted result is given in the following figure.

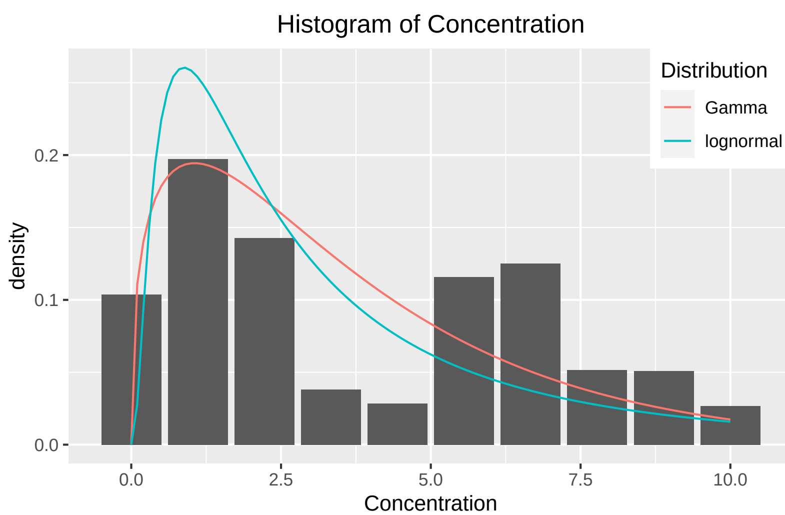


Figure 3.1: Concentration of airborne chlorine fit with the zero-inflated gamma distribution, the histogram shows the positive part of the data fit with gamma distribution

Tolerance limit has been widely used for environmental data for setting a limit. The federal standard threshold limit value (TLV) is 1.0 ppm for chlorine. The goal is to determine if the data could show that the company was in compliance with the standard. Here we calculated the $(0.95, 0.90)$ upper tolerance limit was 12.983, which was above the standard TLV 1.0. Thus, our conclusion was consistent with the Tian [60] and Hasan and Krishnamoorthy,[59] which was that the company was not in compliance with the standard.

Diagnostic Test Charges Data

We used the example of diagnostic test charges among older adults with depression. There were 40 patients, and 10 of them had no diagnostic tests during the study period. Zhou and Tu,[56] Hasan and Krishnamoorthy,[59] and Tian [60] analysed positive part of the data using lognormal distribution.

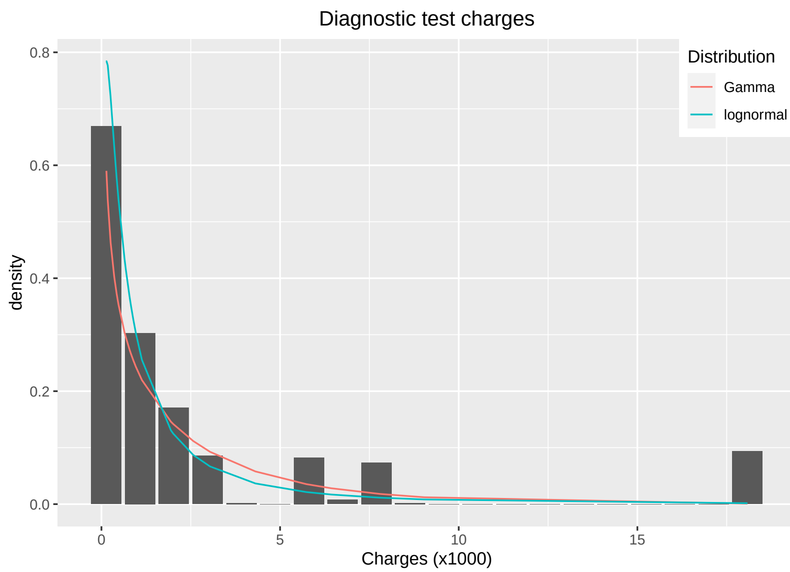


Figure 3.2: Histogram of the diagnostic test charges along with the estimated lognormal and gamma density curve

We are interested in estimating the mean patient's charges. Below is the 95% CI for the mean. The result for zero-inflated lognormal is based on fiducial method proposed by Hasan and Krishnamoorthy.[59]

Table 3.2: The 95% CIs for the mean based on different distribution.

Distribution	95% CI for the mean
Zero-inflated Gamma	(1120.7, 3060.7)
Zero-inflated Lognormal	(987.8, 4654.2)

3.6 Conclusion and Discussion

In this work, fiducial confidence interval for the mean of zero-inflated gamma was proposed by using the fiducial distribution for binomial parameter [6] and the fiducial distribution for gamma parameters.[63] The proposed fiducial confidence interval enjoys close to nominal coverage probability for different combinations of binomial parameter and gamma parameters. The fiducial based method is easy to implement and possesses good small sample property. We also proposed the calculation of upper tolerance limit, which is also the upper confidence limit of the upper quantile. The coverage probability of the proposed upper tolerance limit approaches to the nominal level as the sample size increases. The result is very satisfactory overall for different simulation scenarios. Our method provides some inference for the semicontinuous data when the positive part of the data is fit better using gamma distribution than lognormal distribution.

Chapter 4 Generalized Fiducial Inference on the Mean of Zero-Inflated Poisson and Poisson Hurdle Models

4.1 Introduction

The Poisson distribution is arguably one of the most commonly used models for count data. As such, a large number of inferential tools are available for Poisson-based models, such as for the ratio of two Poisson rates,[65] Poisson regression models,[66] and Poisson point processes.[67] Assuming the Poisson as an underlying distribution for parametric modeling can be a fairly strong assumption since one must be willing to posit that their data are equi-dispersed. In practice, count data almost ubiquitously demonstrate over-dispersion, which can be attributed to, for example, (spatio-)temporal dependency, unexplained heterogeneity, and/or excess zeros.[68]

The problem of excess zeros was first addressed by Mullahy, who proposed a two-part model that permits a more flexible data-generating process: zeros are from a binomial distribution while positive values are from a truncated distribution. Such a model can accommodate under- and over-dispersion. The model using a zero-truncated Poisson is often called the Poisson hurdle (PH) model.[13] Later, Lambert [14] extended this phenomenon of excess zeros to the count regression setting, but also framed the problem differently with respect to *how* the zeros were generated. Specifically, a certain number of zeros are expected to be generated according to the assumed count distribution (*random zeros*) while the excess zeros are assumed to be generated from a separate, degenerate process (*structural zeros*). This framework results in a zero-inflated model, which is a two-component mixture model with one component for the assumed count distribution and the second component a degenerate distribution at zero. In Lambert's work, [14] the development was in the context of zero-inflated Poisson (ZIP) regression models. Regardless, both PH and ZIP models accommodate the notion of excess zeros in a Poisson setting, but how the zeros are generated is treated differently under the two models. Moreover, both models tend to have comparable performance regarding goodness-of-fit measures, which underscores how the application should provide the guidance in determining the way the zeros are generated.

The more complex data setting posed by zero-inflation opens the door to additional inference considerations, many coupled with their own challenges. For example, there is a bevy of score tests developed for testing the presence of zero-inflation in various count data settings.[69, 70, 71, 72, 73]. Bhattacharya et al [74] used a general Bayesian setup for detecting if zero-inflation is present in the data, however, it is challenging to justify the selection of the prior distribution. Score-based tests are also available for testing the presence of overdispersion, which can be caused by zero-inflation.[75, 76, 77] With the exception of large-sample-based approaches for constructing confidence intervals on regression parameters in zero-inflated regression and hurdle regression models, there is no panacea for constructing reliable, accurate confidence intervals for other parameters in their non-regression counterparts, such as the population mean of univariate ZIP and PH distributions.

Deriving confidence intervals for a more complex data setting, like the presence of excess zeros, is challenging in the frequentist setting. Typically, one resorts to normal-based theory, but finite sample properties can be highly unreliable. Bayesian approaches suffer from the challenge to justify the selection of the prior distribution, just like we noted with Bhattacharya's work earlier.[74] Alternatively, one can consider fiducial inference as proposed by Fisher.[78] Fiducial inference struggled to gain popularity among statisticians because of perceived deficiencies in the general approach. However, later works have developed more sophisticated procedures coupled with rigorous theory to mitigate such criticisms, all while reflecting the core tenets of the fiducial paradigm. For example, Weerahandi introduced generalized confidence intervals (GCIs) constructed by generalized pivotal quantities (GPQs),[11] Hannig et al [12] further established the connection between GCI and the fiducial argument of Fisher. For the purposes of our study, we turn to generalized fiducial distributions as they often lead to attractive solution with asymptotically correct frequentist coverage levels. Moreover, many simulation studies have shown that generalized fiducial solutions have very good small sample properties.[6, 8] There is also some work on fiducial approaches for discrete distributions. Mathew and Young [79] developed fiducial tolerance intervals for functions of discrete random variables, while Hannig et al [80] presented an extensive summary about computing the generalized fiducial distribution for parameters of some common discrete distributions. In this chapter, we shall consider using the fiducial inference for the mean of ZIP and PH distributions.

This chapter is organized as follows. In Section 4.2, we give a brief sketch of generalized fiducial inference, with emphasis on the discrete data setting. In Section 4.3, we derive the respective fiducial distributions of the ZIP mean and PH mean. In Section 4.4, we present a numerical study to illustrate the good coverage probabilities of GCIs for the ZIP mean and PH mean constructed using fiducial inference. An analysis of urinary tract infection data is presented in Section 4.5. In Section 4.6, we make some concluding remarks.

4.2 Generalized Fiducial Inference

The aim of generalized fiducial inference is to define a distribution for parameters of interest that contains all of the information from data. Therefore, inference on the parameters can be made through this distribution. The tenet of generalized fiducial inference is to switch the role of the parameters and the data. We now briefly explain the philosophy of generalized fiducial inference.

Suppose that data Y are generated through the structural equation $Y = G(\xi, U)$, where ξ is a vector of parameters and U is some random variable with a known distribution independent of the parameter ξ. The structural equation can be regarded as a data generation process where the noise process U and the signal ξ will produce observed data Y. Hence, the distribution of Y can be determined via the structural equation given a fixed parameter ξ and the distribution U. After the data Y are observed, we can switch the position of the data and parameters by solving the structural equation conditioned on that the solution to that equation exists. Thus, we can get $\xi = Q(Y, U)$. For more details regarding this setup, we refer to Hannig's paper.[6]

Generalized Fiducial Inference on Discrete Data

Let Y now be a discrete random variable with the distribution function $F(\cdot|\theta)$. We know that if $U \sim \mathcal{U}(0,1)$, data following the distribution $F(\cdot|\theta)$ can be generated through $Y = F^{-}(U|\theta)$, where $F^{-}(a|\theta) = \inf\{y : a \leq F(y|\theta)\}$ is the inverse function. According to the philosophy of generalized fiducial inference, we need to solve the data generating equation to get the parameter as a function of the data and a known random distribution. Assume for each fixed y, the distribution is a nonincreasing function of θ. It follows that $Q_y^+(u) = \sup\{\theta : F(y|\theta) = u\}$ and $Q_y^-(u) = \inf\{\theta : F(y_-|\theta) = u\}$ exist and satisfy $F(y|Q_y^+(u)) = F(y_-|Q_y^-(u)) = u$. It follows that the closure of the inverse image is $\bar{Q}_y(u) = [Q_y^-(u), Q_y^+(u)]$. Hannig chose a 50-50 mixture of the upper and lower bound as the generalized fiducial distribution for the parameter.[80]

For the fiducial distributions of multiple parameters, a two-stage method can be applied based on the minimal sufficient statistics. Let the parameters of interest be $\boldsymbol{\xi} = (\xi_1, \xi_2)$. Assume that the following two conditions hold:

1. If ξ_2 is known, there is a statistic $\mathcal{S}_1 = \mathcal{S}_1(\xi_2)$ that has an invertible pivotal relationship with ξ_1.

2. A statistic $\mathcal{S}_2$ exists that $\mathcal{S}_2$ and ξ_2 have an invertible pivotal relationship.

Then we can first obtain the fiducial distribution of ξ_2 and then obtain the fiducial distribution of ξ_1 given that ξ_2 is known.

4.3 Fiducial Distributions for Poisson Data with Excess Zeros

Fiducial Distribution of ZIP Mean

The ZIP distribution has probability mass function

$$p(x|\pi,\lambda) = \pi I_{\{0\}}(x) + (1-\pi)\frac{\lambda^x e^{-\lambda}}{x!} I_{\{\mathbb{N}^+\}}(x), \tag{4.1}$$

where $I_{\{A\}}(z)$ is the indicator function that z belongs to the set A and $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$. The following proposition establishes the minimal sufficient statistic for a ZIP distribution:

Proposition 4.3.1. *Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$ be a random sample from a ZIP distribution. Denote the sum of the random sample as S and the number of zeros of the random sample as K, where $S = \sum_{i=1}^n X_i$ and $K = \sum_{i=1}^n I_{\{0\}}(X_i)$. Consequently the minimal sufficient statistic is (S, K).*

Proof. First we need to prove (S, K) is sufficient. The joint density of $(X_1, \ldots, X_n)$ is

$$\begin{aligned}
p(x_1,\ldots,x_n|\pi,\lambda) &= \prod_{i=1}^n \left\{ \pi I_{\{0\}}(x_i) + (1-\pi)\frac{\lambda_i^x e^{-\lambda}}{x_i!} I_{\{\mathbb{N}\}}(x_i) \right\} \\
&= \prod_{i=1}^n \left\{ [\pi + (1-\pi)e^{-\lambda}] I_{\{0\}}(x_i) + (1-\pi)\frac{\lambda_i^x e^{-\lambda}}{x_i!} I_{\{\mathbb{N}^+\}}(x_i) \right\} \\
&= \prod_{i=1}^n \left\{ [\pi + (1-\pi)e^{-\lambda}]^{I_{\{0\}}(x_i)} [(1-\pi)\frac{\lambda_i^x e^{-\lambda}}{x_i!}]^{(1-I_{\{0\}}(x_i))} \right\} \\
&= \left(\frac{\pi + (1-\pi)e^{-\lambda}}{(1-\pi)e^{\lambda}} \right)^{\sum_{i=1}^n I_{\{0\}}(x_i)} [(1-\pi)e^{-\lambda}]^n \lambda^{\sum_{i=1}^n x_i (1-I_{\{0\}} x_i)} \\
&\qquad \prod_{i=1}^n \left(\frac{1}{x_i!} \right)^{(1-I_{\{0\}}(x_i))} \\
&= \left(\frac{\pi + (1-\pi)e^{-\lambda}}{(1-\pi)e^{\lambda}} \right)^{\sum_{i=1}^n I_{\{0\}}(x_i)} [(1-\pi)e^{-\lambda}]^n \lambda^{\sum_{i=1}^n x_i} \\
&\qquad \prod_{i=1}^n \left(\frac{1}{x_i!} \right)^{(1-I_{\{0\}}(x_i))},
\end{aligned}$$

where $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$. According to the factorization theorem, (S, K) is sufficient.

Now we want to show (S, K) is minimal sufficient. Assume that we have another sample $\boldsymbol{Y} = (Y_1, \dots, Y_n)^{\mathrm{T}}$. The ratio of the two density functions is

$$\frac{p(x_1, \dots, x_n|\pi, \lambda)}{p(y_1, \dots, y_n|\pi, \lambda)} = \left(\frac{\pi + (1-\pi)e^{-\lambda}}{(1-\pi)e^{\lambda}}\right)^{\sum_{i=1}^n I_{\{0\}}(x_i) - \sum_{i=1}^n I_{\{0\}}(y_i)} \lambda^{\sum_{i=1}^n x_i - \sum_{i=1}^n y_i} f(\boldsymbol{x}, \boldsymbol{y}),$$

where f is a function that does not depend on the parameters. The ratio is free of (π, λ) if and only if $\left(\sum_{i=1}^n x_i, \sum_{i=1}^n I_{\{0\}}(x_i)\right) = \left(\sum_{i=1}^n y_i, \sum_{i=1}^n I_{\{0\}}(y_i)\right)$. Hence (S, K) is minimal sufficient. □

It immediately follows that $K \sim \text{Binomial}(n, \pi + (1-\pi)e^{-\lambda})$, and $(S \mid K = k)$ has the same distribution as $\sum_{i=1}^{n-k} Y_i$, where Y_i are independent Poisson(λ) random variables conditioned on the event $\{Y_i \geq 1\}$. We also need the following proposition regarding sums of zero-truncated Poisson distributions:

Proposition 4.3.2. *Let $Y_1, Y_2, \dots$ be independent Poisson(λ) random variables conditioned on the event $\{Y_i \geq 1\}$. Then*

$$\begin{aligned} P\left(\sum_{j=1}^m Y_j = k\right) &= \frac{\lambda^k}{k!(e^{\lambda}-1)^m} \sum_{j=0}^m \binom{m}{j} (-1)^{m-j} j^k \\ &= \frac{\lambda^k m! S(k, m)}{k!(e^{\lambda}-1)^m} I_{\{m, m+1, \dots\}}(k), \end{aligned}$$

where $S(k, m) = \frac{1}{m!} \sum_{j=0}^m \binom{m}{j} (-1)^{m-j} j^k$ is the Stirling number of the second kind.

Proof. The proof follows by mathematical induction, and can be found.[81] □

Denote the distribution function of the sum of m zero-truncated Poisson(λ) by $F_1(k|m, \lambda)$, the distribution of Poisson with mean parameter λ by $F_P(k|\lambda)$, and the distribution function of Binomial(n, p) random variables by $F_B(k|n, p)$. It follows that

$$F_1(k \mid m, \lambda) = P(\sum_{j=1}^m Y_j \leq k) = \sum_{j=1}^m (-1)^{m-j} \binom{m}{j} \frac{e^{\lambda j}}{(e^{\lambda}-1)^m} F_P(k|\lambda j).$$

We will use the inverse distribution functions as a data generating equation:

$$K = F_B^{-1}(U_1|n, \pi + (1-\pi)e^{-\lambda}) \text{ and } S = F_1^{-1}(U_2|n-K, \lambda),$$

where U_1, U_2 are independent $\mathcal{U}(0,1)$. When $K = n$, the value of S is set as 0.

After observing $K = k$ and $S = s$, and inverting the data generating equation, we see that

$$B_{k,n-k+1}(U_1^*) \leq \pi + (1-\pi)e^{-\lambda} \leq B_{k+1,n-k}(U_1^*), \text{ and } H_{n-k,s-1}(U_2^*) \leq \lambda \leq H_{n-k,s}(U_2^*),$$

where $B_{a,b}(u)$ is the quantile function of the Beta(a,b) distribution evaluated at u and $H_{m,s}(u)$ is the solution (in λ) of the equation $F_1(s \mid m, \lambda) = u$. Thus, the sample from the fiducial distribution is obtained by sampling (U_1^*, U_2^*), and using the above inequalities to solve for π and λ. Consequently, when the parameter of interest is $\mu = (1-\pi)\lambda$, the mean of the ZIP distribution, we have

$$\frac{H_{n-k,s-1}(U_2^*)(1 - B_{k+1,n-k}(U_1^*))}{1 - e^{-H_{n-k,s-1}(U_2^*)}} \leq \mu \leq \frac{H_{n-k,s}(U_2^*)(1 - B_{k,n-k+1}(U_1^*))}{1 - e^{-H_{n-k,s}(U_2^*)}},$$

if $k < n$. When $k = n$ then $\mu = 0$.

Finally, we need to select a representative region for the fiduciual sample. Following Hannig's recommendation,[80] we choose a 50-50 mixture of the upper and lower bound in the case of $k < n$. In the case of $k = n$, we select 0.

Fiducial Distribution of PH Mean

The derivation of the fiducial distribution for the PH model follows that for the ZIP model *mutatis mutandis*. The PH distribution has probability mass function:

$$p(x|\pi, \lambda) = \pi I_{\{0\}}(x) + (1-\pi)\frac{\lambda^x e^{-\lambda}}{x!(1-e^{-\lambda})} I_{\mathbb{N}^+}(x).$$

Note that after reparameterization, a ZIP distribution characterized by the binomial parameter μ_1 and the Poisson parameter λ_1 can be expressed as a PH distribution that is characterized by the binomial parameter $\mu_2 = \mu_1 + (1 - e^{-\lambda_1})$ and the truncated Poisson parameter $\lambda_2 = \lambda_1$. Hence, the likelihoods of the two models are equivalent. The selection of the model should be based on how zeros are generated. Note that this equivalency does not hold in the ZIP regression and PH regression settings. In those settings, the likelihoods are based on a conditional distribution (i.e., Y given some covariates) for determining the estimates of the regression parameters. While the final likelihoods will typically be similar, they will not be equal.

The following proposition establishes the minimal sufficient statistic for a PH distribution:

Proposition 4.3.3. *Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$ be a random sample from a PH distribution. Denote the sum of the random sample as S and the number of zeros of the random sample as K, where $S = \sum_{i=1}^n X_i$ and $K = \sum_{i=1}^n I_{\{0\}}(X_i)$. Consequently the minimal sufficient statistic is (S, K).*

Proof. The proof is very similar to the proof of Proposition 4.3.1. First we need to prove (S, K) is sufficient. The joint density of $(X_1, \ldots, X_n)$ is

$$\begin{aligned} p(x_1, \ldots, x_n | \pi, \lambda) &= \prod_{i=1}^n \left\{ \pi I_{\{0\}}(x_i) + (1-\pi) \frac{\lambda_i^x e^{-\lambda}}{x_i!} I_{\{1,\ldots\}}(x_i) \right\} \\ &= \prod_{i=1}^n \left\{ \pi^{I_{\{0\}}(x_i)} [(1-\pi) \frac{\lambda_i^x e^{-\lambda}}{x_i!}]^{(1-I_{\{0\}}(x_i))} \right\} \\ &= (\frac{\pi}{(1-\pi)e^{\lambda}})^{\sum_{i=1}^n I_{\{0\}}(x_i)} [(1-\pi)e^{-\lambda}]^n \lambda^{\sum_{i=1}^n x_i (1 - I_{\{0\}} x_i)} \\ &\qquad \prod_{i=1}^n (\frac{1}{x_i!})^{(1-I_{\{0\}}(x_i))} \\ &= (\frac{\pi}{(1-\pi)e^{\lambda}})^{\sum_{i=1}^n I_{\{0\}}(x_i)} [(1-\pi)e^{-\lambda}]^n \lambda^{\sum_{i=1}^n x_i} \prod_{i=1}^n (\frac{1}{x_i!})^{(1-I_{\{0\}}(x_i))}. \end{aligned}$$

According to the factorization theorem, (S, K) is sufficient.

Now we want to show (S, K) is minimal sufficient. Assume we have another sample $(Y_1, \ldots, Y_n)$, the ratio of the two density function is

$$\frac{p(x_1, \ldots, x_n|\pi, \lambda)}{p(y_1, \ldots, y_n|\pi, \lambda)} = \big(\frac{\pi}{(1-\pi)e^{\lambda}}\big)^{\sum_{i=1}^n I_{\{0\}}(x_i) - \sum_{i=1}^n I_{\{0\}}(y_i)} \lambda^{\sum_{i=1}^n x_i - \sum_{i=1}^n y_i} f(\boldsymbol{x}, \boldsymbol{y}),$$

where f is a function that does not depend on the parameters.

The ratio is free of (π, λ) if and only if

$$(\sum_{i=1}^n x_i, \sum_{i=1}^n I_{\{0\}}(x_i)) = (\sum_{i=1}^n y_i, \sum_{i=1}^n I_{\{0\}}(y_i)).$$

Hence (S, K) is minimal sufficient. □

Immediately, we see that $K \sim \text{Binomial}(n, \pi)$, and $(S \mid K = k)$ has the same distribution as $\sum_{i=1}^{n-k} Y_i$, where Y_i are independent Poisson(λ) random variables conditioned on the event $\{Y_i \geq 1\}$. We will then use the inverse distribution functions as a data generating equation

$$K = F_B^{-1}(U_1|n, \pi) \text{ and } S = F_1^{-1}(U_2|n - K, \lambda),$$

where U_1, U_2 are independent $\mathcal{U}(0, 1)$. When $K = n$ the value of S is again set as 0.

After observing $K = k$ and $S = s$, and inverting the data generating equation, we see that

$$B_{k,n-k+1}(U_1^*) \leq \pi \leq B_{k+1,n-k}(U_1^*) \text{ and } H_{n-k,s-1}(U_2^*) \leq \lambda \leq H_{n-k,s}(U_2^*),$$

where $B_{a,b}(u)$ and $H_{m,s}(u)$ are as defined for the ZIP setting. Thus, the sample from the fiducial distribution is obtained by sampling (U_1^*, U_2^*) and using the above inequalities to solve for π and λ. Consequently, when the parameter of interest is $\mu = \frac{(1-\pi)\lambda}{1-e^{-\lambda}}$, the mean of the PH distribution, we have

$$\frac{H_{n-k,s-1}(U_2^*)(1-B_{k+1,n-k}(U_1^*))}{1-e^{-H_{n-k,s-1}(U_2^*)}} \leq \mu \leq \frac{H_{n-k,s}(U_2^*)(1-B_{k,n-k+1}(U_1^*))}{1-e^{-H_{n-k,s}(U_2^*)}},$$

if $k < n$. When $k = n$ then $\mu = 0$. Thus, it turns out that the fiducial distribution of the mean of the ZIP and the mean of the PH are the same.

Finally, just as in the ZIP setting, the selection of a representative region for the fiducial sample is to choose a 50-50 mixture of the upper and lower bound in the case of $k < n$. In the case of $k = n$ we select 0.

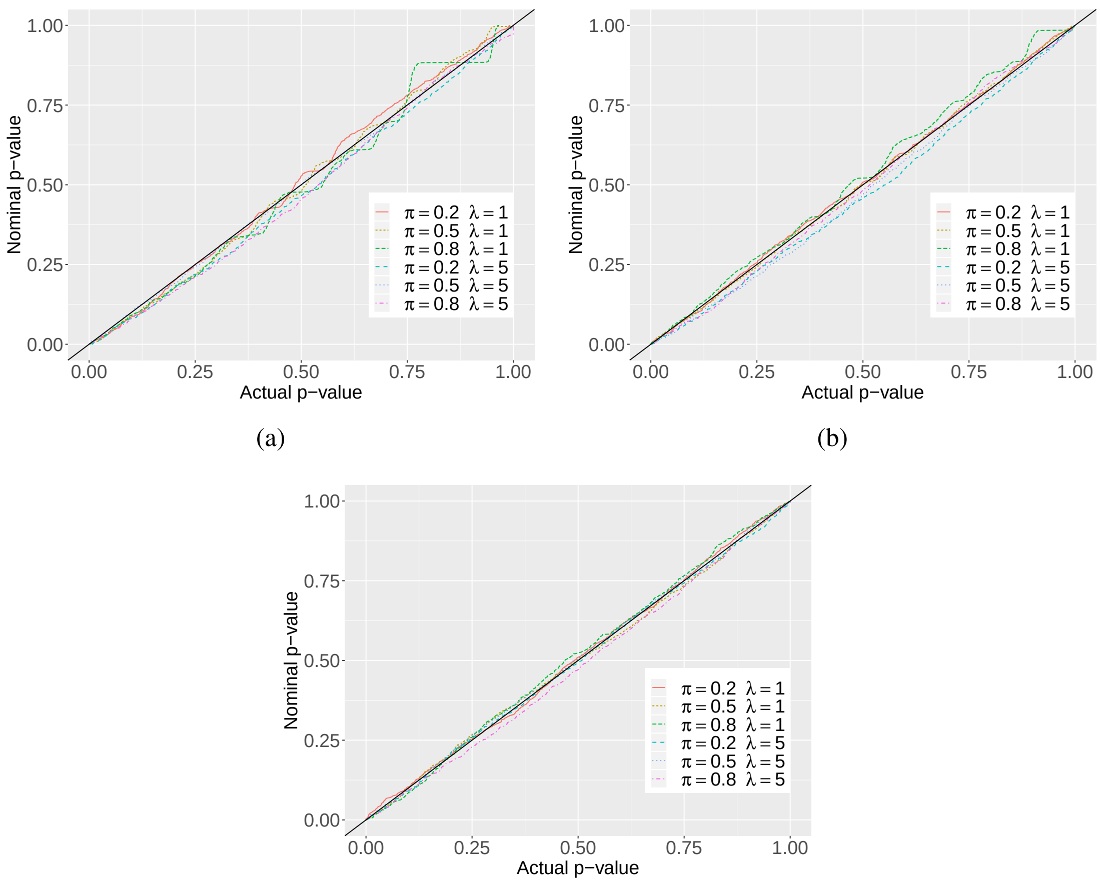
4.4 Simulation Study

We next assess the performance of the GCI just presented through an extensive simulation study. The sample sizes used to assess the finite sample performance of the GCI include $n \in \{15, 30, 100\}$. For the parameters, the mixture proportion π is selected from $\{0.2, 0.5, 0.8\}$ and the mean λ of the Poisson distribution is selected from $\{1, 5\}$. The simulation settings for the PH distribution are the same as for the ZIP distribution: sample sizes $n \in \{15, 30, 100\}$, mixture proportions $\pi \in \{0.2, 0.5, 0.8\}$, and mean of the Poisson distribution $\lambda \in \{1, 5\}$. Note that when $\pi = 0$ in the ZIP setting or $\pi = e^{-\lambda}$ in the PH setting, the data are actually simulated from the Poisson distribution Poisson(λ). Moreover, we demonstrate the performance of our approach when there is no under-/over-dispersion in the data. Specifically, the same values of n and λ are considered, but no mixing proportion is present (or equivalently $\pi = 0$). The number of Monte Carlo samples for our simulations is set to 1000 and the number of fiducial samples used is 500. For each simulation scenario, we estimated the probability $Q(X) = P(\mathcal{R}_M(X) < M|X)$, where M is the mean of the distribution. If the generalized fiducial inference were exact, then $Q(X)$ should follow a standard uniform distribution, which could be examined through $Q - Q$ plots. The coverage probabilities and the average widths of the GCIs are also reported.

The first set of simulation results is for the ZIP distributions. The results are given in Table 4.1. As we can see, for the different simulation scenarios, the coverage probabilities are pretty close to nominal except for the sample size $n = 15$ and $\lambda = 1$. Under such a setting, the simulated samples are almost all zeros, thus compromising the inference. The $Q - Q$ plots for the different sample sizes are given in Figure 4.1. As the sample size n increases, the agreement between the actual p-value and the nominal p-value improves.

Table 4.1: Estimated coverage probabilities and average widths for the GCIs of the different ZIP and PH distributions used in our simulation study.

n	λ	π	**ZIP**		**PH**	
			Cov. Prob.	Avg. Width	Cov. Prob.	Avg. Width
15	1	0.2	0.966	1.054	0.956	1.031
		0.5	0.980	0.983	0.957	1.056
		0.8	0.975	0.839	0.946	0.925
	5	0.2	0.959	2.849	0.941	2.841
		0.5	0.956	2.916	0.957	2.922
		0.8	0.941	2.342	0.936	2.355
30	1	0.2	0.956	0.717	0.949	0.702
		0.5	0.952	0.653	0.947	0.725
		0.8	0.959	0.533	0.973	0.587
	5	0.2	0.948	2.012	0.942	2.000
		0.5	0.949	2.082	0.947	2.090
		0.8	0.948	1.627	0.969	1.643
100	1	0.2	0.947	0.385	0.947	0.378
		0.5	0.950	0.342	0.948	0.385
		0.8	0.957	0.248	0.955	0.294
	5	0.2	0.940	1.098	0.949	1.103
		0.5	0.951	1.151	0.943	1.150
		0.8	0.948	0.872	0.942	0.878



(a) (b) (c)

Figure 4.1: Q-Q plots for ZIP models with sample size of (a) $n = 15$, (b) $n = 30$, and (c) $n = 100$.

The second set of simulation results is for the PH distributions. The results are also given in Table 4.1. Again, for different simulation scenarios, the coverage probabilities are close to nominal. The $Q - Q$ plots for different sample sizes are given in Figure 4.2. The same asymptotic behavior identified in the ZIP setting is also observed from the present simulation results; specifically, as the sample size n increases, the agreement between the actual p-value and the nominal p-value improves.

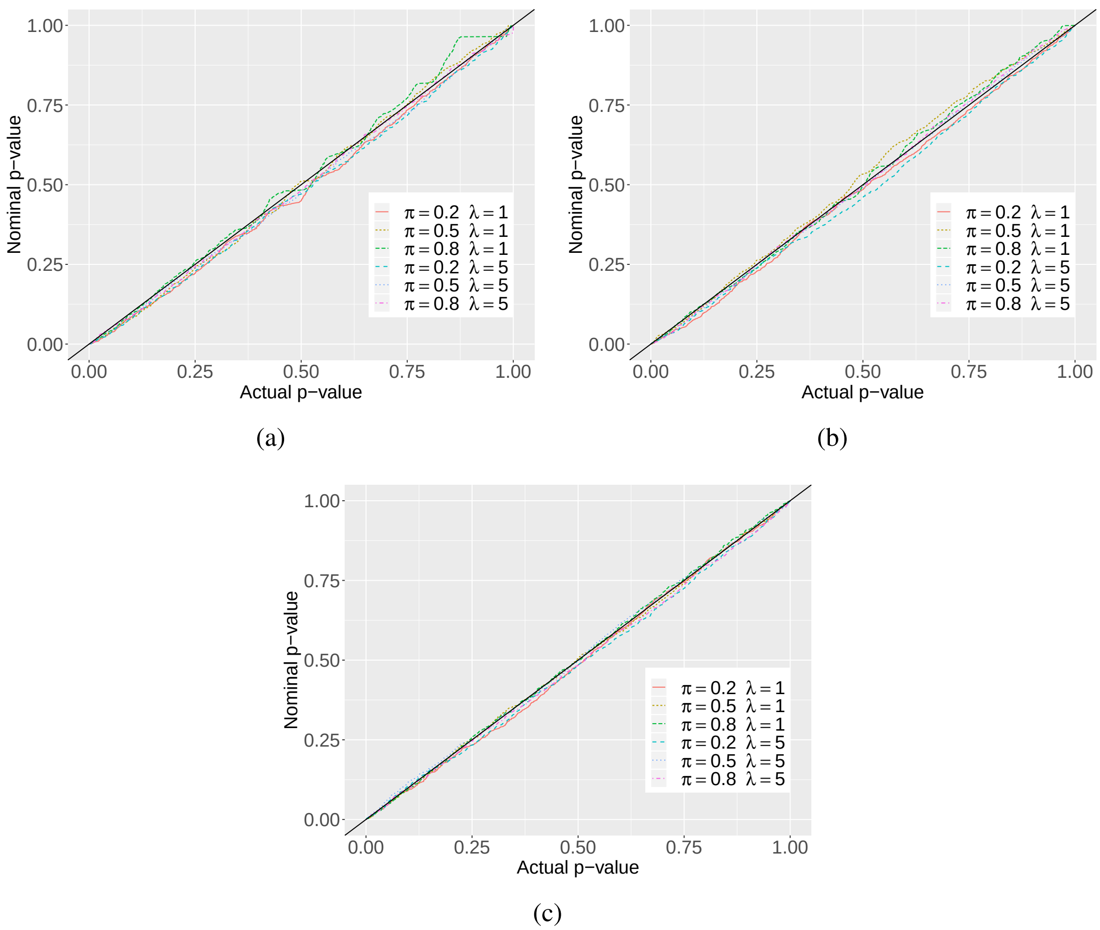
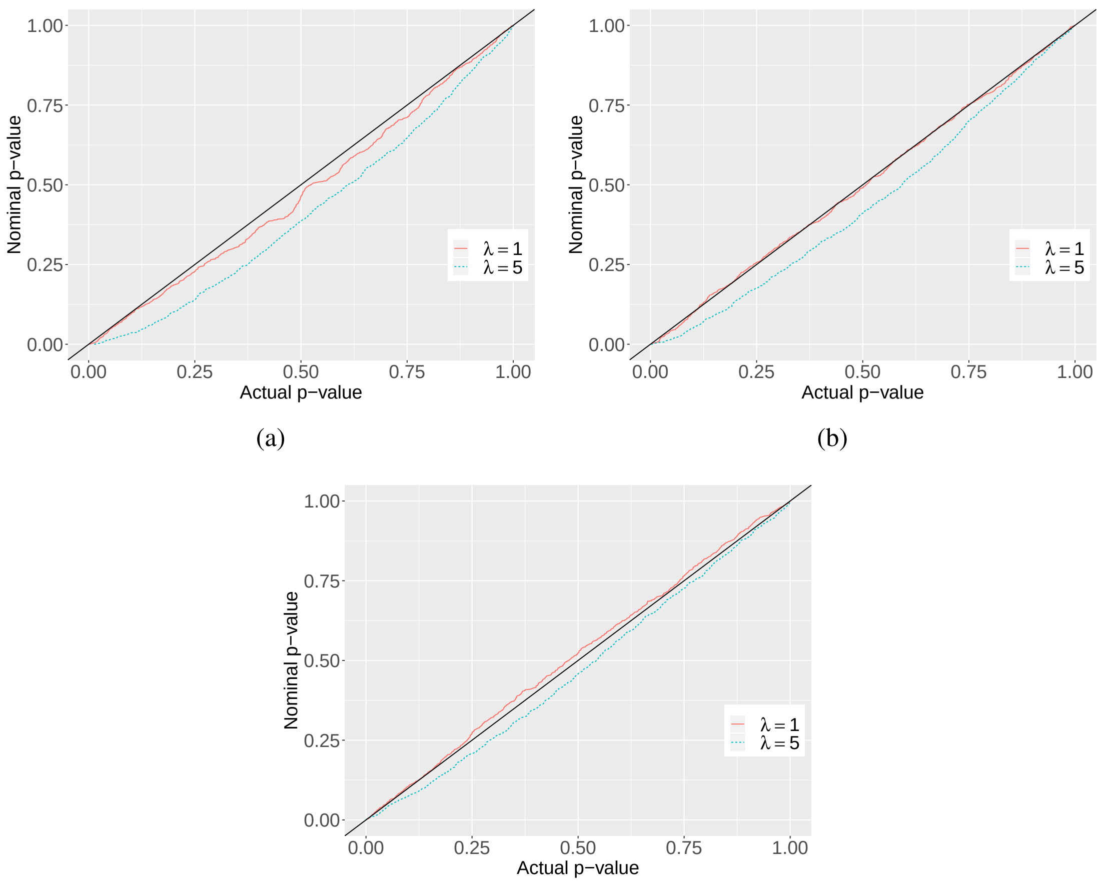


Figure 4.2: Q-Q plots for PH models with sample size of (a) $n = 15$, (b) $n = 30$, and (c) $n = 100$.

The last set of simulation results we consider is for the Poisson distribution. The results are given in Table 4.2. As noted earlier, the Poisson is just as a special case for the ZIP and PH distributions such that there are no excessive zeros. Again, for different simulation scenarios, the coverage probabilities are close to nominal. This illustrates that regardless if zero-inflation is present in the data, our method is still appropriate for constructing a confidence interval of the mean. The $Q - Q$ plots for different sample sizes are given in Figure 4.3. Only moderate discrepancies are noticeable when the sample size is small ($n = 15$) or moderate ($n = 30$); however, the tail behavior appears to be very good. Since we want to construct a 95% confidence interval, it is not a concern as long as the tails are accurate, which is confirmed by our results in Table 4.2. The asymptotic behavior is also observed from the simulation results: as the sample size n increases, the agreement between the actual p-value and the nominal p-value improves.

Table 4.2: Estimated coverage probabilities and average widths for the GCIs of different Poisson distributions used in our simulation study.

n	λ	Cov. Prob.	Avg. Width
15	1	0.957	1.053
	5	0.949	2.434
30	1	0.951	0.729
	5	0.961	1.671
100	1	0.951	0.395
	5	0.949	0.886



(a) (b) (c)

Figure 4.3: Q-Q plots for Poisson models with sample size of (a) $n = 15$, (b) $n = 30$, and (c) $n = 100$.

4.5 Application: Urinary Tract Infection Data

We construct GCIs for a ZIP distribution fit to data on urinary tract infections (UTIs). These data came from $n = 98$ HIV-infected men who were treated by the Department of Internal Medicine at the Utrecht University Hospital in the Netherlands. The frequency of times those patients had a UTI was recorded as X. The frequency table is given in Table 4.3. The data were analyzed using a score test to detect if zero-inflation exists.[69] Later Bhattacharya et al [74] and Bayarri et al [82] applied Bayesian testing to test for zero-inflation. All of these analyses favor a ZIP distribution. Moreover, the use of a zero-inflated distribution is appropriate because the zeros are likely arising from two subgroups of men: one group that are otherwise healthy aside from having HIV (structural zeros), and one group that has a history of other issues with their urinary system (e.g., kidney stones) and, thus, could be at higher risk of eventually developing a UTI (random zeros).

The fiducial sample is set to be $10,000$. The 95% GCI for the average number of UTIs that the patients have is $(0.160, 0.435)$. Even though one can easily calculate the sample mean from these data ($\bar{x} = 0.266$) and infer that, for example, the average number is less than 1, the approach we have presented now affords us with the additional insights that accompany confidence interval interpretations, such as the reliability of our estimate of the mean and how far the spread of that interval falls away from a particular value of interest.

Table 4.3: The frequency table of the number of UTIs recorded in the patients admitted at the Department of Internal Medicine at the Utrecht University Hospital.

X	0	1	2	3	Total
Frequency	81	9	7	1	98

4.6 Conclusion and Discussion

In this chapter, generalized fiducial inference on ZIP and PH distributions was studied for the first time and applied to a healthcare dataset. The practical contribution of this method is that one can now easily calculate and report a confidence interval along with an estimate of the mean if using either a ZIP or PH model. The theoretical advantage of this method is that it achieves good small sample properties except for when the zero proportion π is large and the Poisson parameter λ is small. Also, it does not depend on the selection of priors like Bayesian inference, but it only relies on the data generation equation. A simulation study demonstrated that, for the confidence interval of the mean of ZIP and PH distributions, the generalized fiducial inference works very well for various scenarios. Since the Poisson distribution can be considered as a special case of ZIP or PH distribution, the simulation also shows our method for ZIP and PH distributions can accommodate constructing the confidence interval of the mean of a Poisson distribution. Thus, if the goal is only to construct a confidence interval for the mean of the count data, our approach can be applied directly since it will not be necessary to detect for zero-inflation or decide if the data is under/over-dispersed.

We note that there is some computational limitation of the proposed method since it involves finding the root of a sum of factorials. When the sample size is large or the Poisson mean parameter is large, the computational effort could become prohibitive. Uniformly valid approximation exists for Stirling numbers of the second kind, [83] which can alleviate some of the computational burden, but this can translate to worse results for coverage probabilities. Such simulation results are not shown here. We also note that generalized fiducial inference can be very similar to Bayesian inference when a fiducial distribution is obtained. We highlighted earlier that Bhattacharya et al [74] used Bayesian inference to test for the presence of zero-inflation. Future research will be focused on extending the use generalized fiducial inference for selecting the model between Poisson distribution and ZIP/PH models. Moreover, there are broader inference considerations when fitting ZIP/PH regression models, such as joint confidence intervals on the regression parameters and simultaneous confidence intervals over the values of the covariate space. These are further extensions worth considering in the generalized fiducial framework.

Chapter 5 Approximate Statistical Limits for Birnbaum-Saunders Distribution Based on Generalized Fiducial Inference

5.1 Introduction

Birnbaum–Saunders (BS) distribution was first proposed by Birnbaum and Saunders [15] as a failure time distribution for fatigue failure caused under cyclic loading. Desmond [16] derived the distribution in a more general way based on a biological model. BS distribution has been shown to have some connection with other commonly used distributions: BS distribution can be obtained by making a monotone transformation on the standard normal distribution, and it can also be developed using an equal mixture of the inverse Gaussian distribution and its reciprocal. Extensive work has been done on the statistical inference of the BS distribution because of its wide range of applications. The maximum likelihood estimators (MLE) of the shape and scale parameters based on a complete sample were derived by Birnbaum and Saunders.[15] Engelhardt et al [84] obtained the asymptotic distributions and the asymptotic confidence intervals for the parameters. Ng et al [85] proposed a bias reduction method to reduce the bias of the MLEs and the modified moment estimators. Wu and Wong [86] utilized higher order likelihood inference to get better interval estimations for the parameters of BS distribution. Wang derived the generalized confidence intervals (GCIs) for the shape parameter and explored some important reliability quantities such as mean, quantiles and reliability function.[87] Li and Xu further developed fiducial inference for the parameters of BS distribution.[88] It should be noted the Weerahandi [11] introduced GCIs constructed by generalized pivotal quantities (GPQs), and Hannig et al [12] further established the connection between GCI and the fiducial argument of Fisher. Balakrishnan et al [89] have a detailed summary about some statistical inference properties and applications of BS distribution in their review paper.

There has been a lot of research focusing on developing the point and interval estimation of the parameters for the BS distribution, while for other statistical intervals like tolerance interval and prediction interval, which are useful in reliability applications, are not well developed. For instance, the fatigue life of aluminum coupons can be often modeled by BS distribution. If we only have a single future fatigue life of the aluminum coupon, it can be compared with the upper prediction limit computed from the historical measurements to check whether the aluminum coupon has defect. In the manufacture setting, we are more interested in drawing an inference about a proportion of a distribution than a parameter of a distribution only. Tolerance interval, an interval which contains a specified proportion of a sampled population with some confidence level, can be established to set the limit. Developing a tolerance interval for the fatigue life of aluminum coupons can be used for the engineer to check if the aluminum coupons are designed properly. There are only a limited literature discussing tolerance interval and prediction interval. Chang and Tang constructed the tolerance interval for BS distribution based on the confidence limits of the parameters because of the monotone relationship between the parameters and the pth quantile of the BS distribution.[90] However, such tolerance limits were very conservative. Parametric bootstrap has been widely used to make inference for complex and difficult distribution settings. Lu and Chang [91] studied the approximate prediction intervals for future realizations of BS distribution based on the bootstrap method. However, the coverage probabilities of the prediction intervals constructed by parametric bootstrap method tended to be liberal when sample size was small. Generalized fiducial inference has been proved to have correct asymptotic correct coverage and good small sample property for making interval estimation of BS distribution.[87, 88] Likelihood based inference has also been applied the BS distribution to improve the interval estimation for parameters.[86] Although the higher-order likelihood asymptotic procedure has similar performance relative to generalized fiducial inference in the small sample setting, higher-order likelihood asymptotic procedure is more complicated and it is difficult to apply it on other quantities of interest such as mean and quantile. In contrast, the use of generalized fiducial method is more straightforward. Wang et al introduced how to obtain prediction intervals using generalized fiducial inference, which have approximate nominal coverage probabilities in the frequentist settings.[92] Here, we will obtain the approximate upper tolerance limit and upper prediction limit of a single observation based on generalized fiducial method.

The remainder of the chapter is organized as follows. In Section 5.2, we give a formal definition of BS distribution, the prediction interval and the tolerance interval. In Section 5.3, we develop the upper tolerance limit and upper prediction limit for a single future observation using generalized fiducial inference. In Section 5.4, the performance of two approximate limits calculated by the generalized fiducial inference is assessed by Monte Carlo simulation. In Section 5.5, two real examples are provided to illustrate the two methods. We end our chapter with some conclusion remarks.

5.2 Some Definitions

Definition and Properties of BS Distribution

The two-parameter BS distribution with density

$$f(t;\alpha,\beta)=\frac{1}{2\sqrt{2\pi}\alpha\beta}\left[\left(\frac{\beta}{t}\right)^{1/2}+\left(\frac{\beta}{t}\right)^{3/2}\right]\exp\left[-\frac{1}{2\alpha^2}\left(\frac{t}{\beta}+\frac{\beta}{t}-2\right)\right]\quad t,\alpha,\beta>0.$$

We can denote BS distribution shape α and scale β as BS(α,β). BS is related to normal distribution as follows. If $T\sim$ BS(α,β), let

$$Z=\frac{1}{\alpha}\left[\left(\frac{T}{\beta}\right)^{1/2}-\left(\frac{T}{\beta}\right)^{-1/2}\right],$$

it can be easily to show that Z follows standard normal distribution. T can also be represented as a function of Z:

$$T=\frac{\beta}{4}\left[\alpha Z+\sqrt{\alpha^2Z^2+4}\right]^2,$$

Equation 5.2 can be used effectively to generate BS random variables from standard normal distribution. Since T is an increasing function fo Z, $100p$th percentile of BS distribution is also the function of $100p$th percentile of standard normal distribution, denoted by

$$W_p = \frac{\beta}{4}\left[\alpha z_p + \sqrt{(\alpha^2 z_p^2 + 4)}\right]^2,$$

where z_p is the standard normal $100p$th percentile.

It has been shown by Engelhardt et al [84] that the joint distribution of $(\hat{\alpha}, \hat{\beta})$ is bivariate normal, denoted by

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \sim N\left[\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} \frac{\alpha^2}{2n} & 0 \\ 0 & \frac{\beta^2}{n(1/4+\alpha^{-2}+I(\alpha))} \end{pmatrix}\right]$$

where

$$I(\alpha) = 2\int_0^\infty \left[\left(2 + \frac{\alpha^2 x^2}{2} + \alpha x\sqrt{1 + \frac{\alpha^2 x^2}{4}}\right)^{-1} - 1/2\right]^2 d\Phi(x).$$

We can get the asymptotic confidence intervals for the parameters based on normal theory, and obtain other parameters of interest using delta method. However, such asymptotic confidence interval obtained can display poor coverage in small sample settings since it can include values that do not belong to the parameter space with positive probability.[86]

Prediction Interval

Let T be a random variable of BS distribution. $\boldsymbol{T} = (T_1, T_2, \ldots, T_n)^{\mathrm{T}}$ be a random sample from the same BS distribution. A future measurement T_{n+1} is from the same BS distribution. A $(1-\gamma)$ two-sided prediction interval $(PL(\boldsymbol{T}), PU(\boldsymbol{T}))$ for a single measurement satisfies the condition

$$P_T(PL(\boldsymbol{T}) \leq T_{n+1} \leq PU(\boldsymbol{T})) = 1 - \gamma.$$

When $PL(\boldsymbol{T}) = -\infty$, $(1-\gamma)$ upper prediction limit is $(-\infty, PU(\boldsymbol{T})]$. Similarly, when $PU(\boldsymbol{T}) = \infty$, $(1-\gamma)$ lower prediction limit is $[PL(\boldsymbol{T}), +\infty)$.

Tolerance Interval

When the number of future measurements is either large or unknown, the tolerance interval is of more interest than the prediction interval. Let T be a random variable of BS distribution. $\boldsymbol{T} = (T_1, T_2, \ldots, T_n)^{\mathrm{T}}$ be a random sample from the same BS distribution. A $(p, 1-\gamma)$ two-sided tolerance interval $(TL(\boldsymbol{T}), TU(\boldsymbol{T}))$ satisfies the condition

$$P_{\boldsymbol{T}}\{P_T\big(TL(\boldsymbol{T}) \le T \le TU(\boldsymbol{T})|\boldsymbol{T}\big) \ge p\} = 1-\gamma,$$

where p is the the tolerance limit's *content* and $(1-\gamma)$ is the *confidence level*.

When $TL(\boldsymbol{T}) = -\infty$, $(p, 1-\gamma)$ one-sided upper tolerance limit is $(-\infty, TU(\boldsymbol{T})]$. Similarly, when $TU(\boldsymbol{T}) = \infty$, $(p, 1-\gamma)$ one-sided lower tolerance limit is $[TL(\boldsymbol{T}), +\infty)$.

The upper tolerance limit can be considered as the upper confidence limit of the $100p$th percentile W_p which satisfy the follow condition:

$$P_{\boldsymbol{T}}\{W_p \le U(\boldsymbol{T})\} = 1-\gamma.$$

Similarly, the lower tolerance limit can be considered as the lower confidence limit of the $(1-p)100$th percentile.

Generalized fiducial inference

The aim of fiducial inference is to define a distribution for parameters of interest that contains the all the information from data. Therefore, inference for the parameters can be made through this distribution. The basic idea of the fiducial inference is to switch the role of the parameters and the data. We will explain the philosophy of fiducial inference briefly.

Suppose that data Y are generated through the structural equation $Y = G(\xi, U)$, where ξ is a vector of parameters and U is some random variable with a known distribution independent of the parameter ξ. The structural equation can be regarded as a data generation process where the noise process U and the signal ξ will produce observed data Y. Hence, the distribution of Y can be determined via the structural equation given a fixed parameter ξ and the distribution U. After the data Y are observed, we can switch the position of the data and parameters by solving the structural equation conditioned on that the solution to that equation exists. Thus, we can get $\xi = Q(Y, U)$. For more detail regarding to generalized fiducial inference, Hannig has more thorough discussion in his paper.[6]

5.3 The Proposed Methods

GCIs of the BS parameters have been studied by Wang [87] using generalized variable approach [11]. Li and Xu [88] made generalized fiducial inference for BS parameters. The fiducial intervals of the parameters coincide with the generalized confidence interval, but the former is easier to implement. Hannig et al [12] pointed out the fiducial pivotal quantity is a subclass generalized pivotal quantity used to construct generalized confidence intervals. We will also propose the approximate upper tolerance limit and upper prediction limit of a single observation based on generalized fiducial inference.

Let $\boldsymbol{T} = (T_1, T_2, \ldots, T_n)^{\mathrm{T}}$ be a random sample from BS(α, β) with observations data $(t_1, t_2, \cdots, t_n)$. Let

$$X_i = \frac{1}{\alpha}\left[\left(\frac{T_i}{\beta}\right)^{1/2} - \left(\frac{T_i}{\beta}\right)^{-1/2}\right],$$

where $i = 1, 2, \cdots, n$ and $X_i \overset{i.i.d}{\sim} \mathcal{N}(0, 1)$. Denote $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S^2 = \frac{1}{n-1}(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2)$, then $\bar{X}$ and S^2 are independent. We have

$$Y = \frac{\bar{X}}{S/\sqrt{n}} \sim t(n-1),$$

where $t(n-1)$ is the student t distribution with degree of freedom $n-1$. As we can see that Y only depends on the data $\boldsymbol{T}$ and the parameter β since the parameter α is canceled between the nominator and denominator. Let Y^* be a realization from $t(n-1)$, then a realization β^* of the fiducial sample of β can be obtained by solving the following equation given the data $\boldsymbol{T}$ is known

$$Y^* = g(\boldsymbol{T}, \beta). \tag{5.1}$$

Denote $S_1 = \frac{1}{n}\sum_{i=1}^{n} T_i, S_2 = \frac{1}{n}\sum_{i=1}^{n} T_i^{-1}, S_3 = \frac{1}{n}\sum_{i=1}^{n} T_i^{1/2}, S_4 = \frac{1}{n}\sum_{i=1}^{n} T_i^{-1/2}$, after doing some algebra by solving equation 5.1, we have

$$\begin{aligned}[(n-1)S_4^2 - (S_2 - S_4^2)(Y^*)^2]\beta^2 - 2[(n-1)S_3S_4 + (S_3S_4 - 1)(Y^*)^2)]\beta + \\ [(n-1)S_3^2 - (S_1 - S_3^2)(Y^*)^2] = 0. \end{aligned} \tag{5.2}$$

The equation 5.2 has two solutions β_1 and β_2. Noted that the realization $-Y^*$ yields the same equation as equation 5.2. Sun [93] has proved that Y is a strictly monotone decreasing function regarding to β when $\beta > 0$. Hence, we let the realization β^* of β be

$$\beta^* = \begin{cases} \min\{\beta_1, \beta_2\}, & \text{if } Y^* \geq 0 \\ \max\{\beta_1, \beta_2\}, & \text{if } Y^* < 0 \end{cases}$$

If $\beta^* \leq 0$, then regenerate a realization Y^* until $\beta^* > 0$.

Since $\sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$, let V^* be a realization from $\chi^2(n)$, the realization α^* of α

$$\alpha^* = \sqrt{\frac{nS_1/\beta^* + nS_2\beta^* - 2n}{V^*}}.$$

The realization W_p^* of W_p is just to replace the α and β with α^* and β^*.The generalized fiducial method for upper tolerance limit is summarized as follows:

Let $\boldsymbol{T} = (T_1, T_2, \ldots, T_n)^{\mathrm{T}}$ be a random sample from BS(α, β). Suppose that T_{n+1} is a future measurement from the same BS distribution. Let $F_{\boldsymbol{\theta}}(t)$ be the CDF of the continuous random variable Y with parameter $\boldsymbol{\theta}$, then the statistic $U = F_{\boldsymbol{\theta}}(Y)$ will follow

Algorithm 3 Upper tolerance limit

1. Given a BS random sample $(t_1, t_2, \cdots, t_n)$, calculate S_1, S_2, S_3 and S_4
2. Generate a realization Y^* from $t(n-1)$ and get a realization β^* until $\beta^* > 0$
3. Generate a realization V^* from $\chi^2(1)$ and get a realization α^*
4. Calculate the realization $W_p^* = \frac{\beta^*}{4}\left[\alpha^* z_p + \sqrt{\alpha^{*2} z_p^2 + 4}\right]^2$
5. Repeat Steps 2-4 B times, then we have B fiducial samples of W_p, denoted by $W_{p1}^*, W_{p2}^*, \cdots, W_{pB}^*$, the $(p, 1-\gamma)$ upper tolerance limit will be the upper $1-\lambda$ quantile of the fiducial samples of W_p

the uniform distribution $\mathcal{U}(0,1)$, once we know the quantile function of Y, we can generate sample for Y using $F_{\boldsymbol{\theta}}^{-1}(U)$. (Casella and Berger, 2002) The construction of the prediction interval will also base on this idea. The future observation can be generated from the BS distribution with a set of realizations of BS parameters, and it can be formulated as $\tilde{T} = F_{\alpha^*,\beta^*}^{-1}(U^*) = \frac{\beta^*}{4}\left[\alpha^* Z_{U^*} + \sqrt{\alpha^{*2} Z_{U^*}^2 + 4}\right]^2$, where $F_{\alpha,\beta}^{-1}$ is the quantile function of the BS function with parameters α and β, Z_p is the pth quantile of the standard normal distribution, (α^*, β^*) is a realization from the fiducial distribution of (α, β) and U^* is a realization from the standard uniform distribution $\mathcal{U}(0,1)$. By utilizing these realizations, the prediction interval can be easily constructed. The $1-\gamma$ prediction interval will be the $\gamma/2$ quantile and $1-\gamma/2$ quantile of the fiducial samples. The generalized fiducial method for prediction interval is summarized as follows:

Algorithm 4 Prediction interval for a single future measurement T_{n+1}

1. Given a BS random sample $(t_1, t_2, \cdots, t_n)$, calculate S_1, S_2, S_3 and S_4
2. Generate a realization Y^* from $t(n-1)$ and get a realization β^* until $\beta^* > 0$
3. Generate a realization V^* from $\chi^2(1)$ and get a realization α^*
4. Generate a realization U^* from $\mathcal{U}(0,1)$
5. Calculate the realization $T^*_{n+1} = \frac{\beta^*}{4}\left[\alpha^* Z_{U^*} + \sqrt{\alpha^{*2} Z^2_{U^*} + 4}\right]^2$
6. Repeat Steps 2-4 B times, then we have B fiducial samples of T_{n+1}, denoted by $T^*_{n+1,1}, T^*_{n+1,2}, \cdots, T^*_{n+1,B}$, the $1-\gamma$ upper prediction interval will be $1-\gamma$ quantile of the fiducial samples of T_{n+1}

5.4 Simulation Study

To study the accuracy of the proposed methods, we performed the following Monte Carlo simulation to study the coverage probabilities the proposed upper prediction limit and upper tolerance limit for the BS distribution under different simulation scenarios. The sample size n was chosen from $\{10, 30, 100\}$ and the shape parameter α was set to be $\{0.1, 0.5, 2\}$. Without loss of generality, the scale parameter β was kept fixed at 1. The density curves of the different BS distributions used for simulation were given in the following figure.

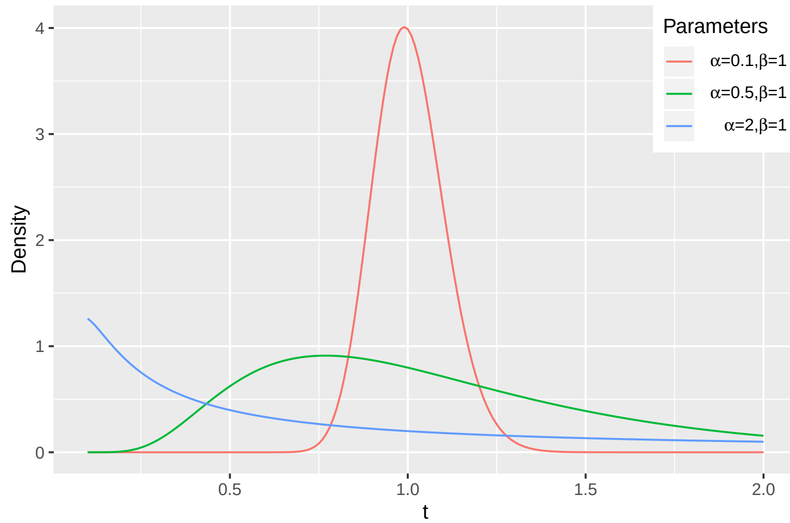


Figure 5.1: The density curves of the different BS distributions used for simulation.

Simulating BS distribution per se was not straightforward, so here BS distribution was simulated using the its relationship with standard normal distribution based on Equation 5.2.

All the results were based on 10000 Monte Carlo runs. We calculated the coverage probabilities of prediction intervals with confidence levels $\{90\%, 95\%\}$ and $(p, 1-\alpha)$ upper tolerance limit with p and α from $\{90\%, 95\%\}$.

It is observed that the coverage percentages of the proposed upper prediction limits and upper tolerance limits are close to the nominal coverage probabilities in all cases.

Table 5.1: Estimated coverage probabilities of the upper tolerance limits and upper prediction limits. utl_95_cp is the coverage probability of (0.95, 0.95) upper tolerance limit, utl_90_cp is the coverage probability of (0.90, 0.95) upper tolerance limit, pi_95_cp is the coverage probability of 95% upper prediction limit, pi_90_cp is the coverage probability of 90% upper prediction limit.

α	n	utl_95_cp	utl_90_cp	pi_95_cp	pi_90_cp
0.1	10	0.946	0.947	0.953	0.904
	30	0.949	0.949	0.951	0.900
	100	0.948	0.948	0.949	0.899
0.5	10	0.949	0.949	0.947	0.895
	30	0.950	0.950	0.952	0.901
	100	0.951	0.950	0.950	0.897
2	10	0.947	0.947	0.955	0.905
	30	0.950	0.950	0.950	0.902
	100	0.949	0.949	0.948	0.897

5.5 Applications

Aluminum Coupon Data

The data set is given by Birnbaum and Saunders [15] about the fatigue life of 6061-T6 aluminum coupons cut parallel to the direction of rolling and oscillated at 18 cycles per second. It consists of 101 observations with maximum stress per cycle 31,000 psi. The data are presented in Table 5.2.

Table 5.2: Fatigue lifetime data of 6061-T6 aluminum coupons.

70	90	96	97	99	100	103	104	104	105	107	108	108
108	109	109	112	112	113	114	114	114	116	119	120	120
120	121	121	123	124	124	124	124	124	128	128	129	129
130	130	130	131	131	131	131	131	132	132	132	133	134
134	134	134	134	136	136	137	138	138	138	139	139	141
141	142	142	142	142	142	142	144	144	145	146	148	148
149	151	151	152	155	156	157	157	157	157	158	159	162
163	163	164	166	166	168	170	174	196	212			

The $90\%/95\%$ upper prediction limits and $(0.95, 0.95)/(0.90, 0.95)$ upper tolerance limits are given in Table 5.3.

If we care about a single future observation of the fatigue life of the aluminum coupon,

Table 5.3: Upper tolerance limits and upper prediction limits in Example 1.

(0.95, 0.95) Upper tolerance limit	183.33
(0.90, 0.95) Upper tolerance limit	171.25
95% Upper Prediction limit	176.66
90% Upper Prediction limit	164.70

we should calculate the 95% upper prediction limit, which means we have 95% confidence that a single future fatigue life of 6051-T6 aluminum coupon will be below 176.66. A $(0.95, 0.95)$ upper tolerance limit means that we have 95% confidence that 95% of the fatigue life of 6061-T6 aluminum coupons will below 183.33.

Bearings Data

This example is from McCool [94] on the fatigue life in hours of 10 bearings of a certain type. The data are given in Table 5.4. Xu and Tang [95] indicated that BS distribution fitted the data very well. Wang [92] and Li and Xu [88] applied fiducial inference for parameters and other quantities of BS distribution.

Table 5.4: Fatigue lifetime data of bearings of a certain type.

152.7	172.0	172.5	173.3	193.0	204.7	216.5	234.9	262.6	422.6

The $90\%/95\%$ upper prediction limits and $(0.95, 0.95)/(0.90, 0.95)$ upper tolerance limits are given in Table 5.5.

Table 5.5: Upper tolerance limits and upper prediction limits in Example 2.

(0.95, 0.95) Upper tolerance limit	437.64
(0.90, 0.95) Upper tolerance limit	375.13
95% Upper Prediction limit	348.08
90% Upper Prediction limit	311.68

If we care about a single future observation of the fatigue life of the bearing, we should calculate the 95% upper prediction limit, which means we have 95% confidence that a single future fatigue life of bearing will be below 348.08. A $(0.95, 0.95)$ upper tolerance limit means that we have 95% confidence that 95% of the fatigue life of the bearing of a certain type will below 437.64.

5.6 Conclusion and Discussion

BS distribution is widely used in fatigue life modeling, it is critically important to calculate the upper tolerance limit and upper prediction limit of the BS distribution to determine a change in the manufacturing process. Since W_p is a function of α and β, we can certainly use delta method to derive the upper tolerance limit. However, such tolerance limit obtained can display poor coverage in small sample situation since it can include values that do not belong to the parameter space with positive probability. A general technique of frequentist prediction intervals is to find and compute a pivotal quantity of the historical measurements. However, it is not easy to construct pivotal quantity for BS distribution. Generalized fiducial inference often has good small properties and relatively easy to implement. Hence, we construct the prediction and tolerance limits for BS distribution based on generalized fiducial inference. We conducted extensive simulation study under different scenarios, the simulation results revealed the convincing performance of the proposed method: the coverage probabilities of the upper tolerance limits and upper prediction limits are consistently close to the nominal levels. The proposed method works very well when the sample size is small, which has been shown in many applications for generalized fiducial inference. Two real examples are used to demonstrate the effectiveness of the proposed method. Our current method only works when the data is complete. However, it is not uncommon that lifetime data could be censored. Traditionally, Type-I and Type-II were the commonly used censoring schemes. However, neither of these censoring methods can remove experimental units during the experiment. Type-I and Type-II progressive censoring are proposed to address the problem of removing the experimental unit. Progressive censoring is a useful method to reduce cost and obtain additional reliability information,[96] There have been some studies on progressively censored BS distribution using likelihood based method.[97, 98, 99] Future work could be extended to using generalized fiducial inference to construct statistical limits for BS distribution when the data are collected under progressively type-II censoring.

Chapter 6 Summary and Future Work

6.1 Summary

In this dissertation, we consider the problems of obtaining the statistical intervals (e.g., confidence, prediction, or tolerance) for several distributions under complex settings. Compared to the construction of confidence interval and prediction interval, the construction of tolerance interval presents more challenges because two levels *confidence* level and *content* level are needed. The two-sided tolerance interval for non-normal distributions has not been studied. Zero-inflated data are commonly found in, among other areas, pharmaceutical and quality control applications. Zero-inflated Poisson and zero-inflated gamma distribution have been widely used for modeling count data and continuous data with excess zeros, respectively. Birnbaum-Saunders distribution has been commonly applied in modeling the fatigue data. The statistical intervals for situations mentioned above have not been well studied. Here we employ bootstrap calibration and generalized fiducial inference to help address the problems encountered in the construction of statistical intervals.

In Chapter 2, we present a focused treatment on using a single-layer bootstrap calibration to improve the coverage probabilities of two-sided parametric tolerance intervals. Simulation results clearly demonstrate the improved coverage probabilities towards the nominal level over the uncalibrated setting. Applications to medical data for various parametric distributions also highlight the utility of constructing these calibrated tolerance intervals. In Chapter 3, the problems of constructing a confidence interval for the mean and calculating an upper tolerance limit of a zero-inflated gamma population are considered using generalized fiducial inference. Our simulation studies indicate that the proposed method is very satisfactory in terms of coverage properties and precision. Two applications also highlight the utility of the proposed method. In Chapter 4, the generalized fiducial inference is used to construct confidence interval for the means of zero-inflated Poisson and Poisson Hurdle. The proposed methods are assessed by the intensive simulation study. An illustrative example demonstrates utility of the inference methods. In Chapter 5, the upper tolerance limit and prediction limit are constructed for the Birnbaum-Saunders distribution also using generalized fiducial inference. The proposed methods are assessed by a simulation study, and two illustrative examples are used to demonstrate the inference methods.

6.2 Future Work

In Chapter 2, we emphasized how to use bootstrap calibration to improve the coverage probabilities for two-sided parametric tolerance intervals for univariate non-normal distributions. On the other hand, tolerance region procedures have only been developed for a limited number of such distributions, such as the multivariate normal distribution.[50, 51, 52] This work can be extended to improve the coverage probabilities of the tolerance region for non-normal multivariate distributions using bootstrap calibration. Given the computational burden of bootstrap calibration even in the univariate distribution setting, it will likely be a concern for multivariate distributions.

In Chapters 3 and 4 we explored the use of statistical intervals for zero-inflated data. Our research focused on the construction of statistical interval for zero-inflated distributions using generalized fiducial inference. Moreover, there are broader inference considerations when fitting zero-inflated regression models, such as joint confidence intervals on the regression parameters and simultaneous confidence intervals over the values of the covariate space. These are further extensions worth considering in the generalized fiducial framework.

In Chapter 5, we developed statistical limits for Birnbaum-Saunders distribution without any censoring. However, it is not uncommon that lifetime data could be censored. Traditionally, Type-I and Type-II were the commonly used censoring schemes. However, neither of these censoring methods can remove experimental units during the experiment. Type-I and Type-II progressive censoring are proposed to address the problem of removing the experimental unit. There have been some studies on progressively censored Birnbaum-Saunders distribution using likelihood based method.[97, 98, 99] Future work could be extended to using generalized fiducial inference to construct statistical limits for Birnbaum-Saunders distribution when the data are collected under progressively type-II censoring.

Bibliography

[1] W. Loh. Calibrating Confidence Coefficients. *J Am Stat Assoc*, 82(397):155–162, 1987.

[2] W. Loh. Bootstrap Calibration for Confidence Interval Construction and Selection. *Stat Sin*, 1(2):477–491, 1991.

[3] T. J. DiCiccio, M. A. Martin, and G. A. Young. Fast and Accurate Approximate Double Bootstrap Confidence Intervals. *Biometrika*, 79(2):285–295, 1992.

[4] F. W. Scholz. On Exactness of the Parametric Double Bootstrap. *Stat Sin*, 4(2):477–492, 1994.

[5] D. McCarthy, K. Zhang, L. D. Brown, R. Berk, A. Buja, E. I. George, and L. Zhao. Calibrated Percentile Double Bootstrap for Robust Linear Regression Inference. *Stat Sin*, 28(4):2565–2589, 2018.

[6] J. Hannig. On Generalized Fiducial Inference. *Statistica Sinica*, 19(2):491–544, 2009.

[7] R. A. Fisher. Inverse probability. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 26, pages 528–535. Cambridge University Press, 1930.

[8] L. E, J. Hannig, and H. Iyer. Fiducial Intervals for Variance Components in an Unbalanced Two-Component Normal Mixed Linear Model. *Journal of the American Statistical Association*, 103(482):854–865, 2008.

[9] S. B. Vardeman. What about the other intervals? *The American Statistician*, 46(3):193–197, 1992.

[10] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, Boca Raton, FL, 1993.

[11] S. Weerahandi. Generalized confidence intervals. In *Exact statistical methods for data analysis*. Springer, New York, NY, 1995.

[12] J. Hannig, H. Iyer, and P. Patterson. Fiducial generalized confidence intervals. *Journal of the American Statistical Association*, 101(473):254–269, 2006.

[13] J. Mullahy. Specification and Testing of Some Modified Count Data Models. *Journal of Econometrics*, 33(3):341–365, 1986.

[14] D. Lambert. Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics*, 34(1):1–14, 1992.

[15] Z. W. Birnbaum and S. C. Saunders. Estimation for a family of life distributions with applications to fatigue. *Journal of Applied probability*, 6(2):328–347, 1969.

[16] A. Desmond. Stochastic models of failure in random environments. *Canadian Journal of Statistics*, 13(3):171–183, 1985.

[17] S. Zhai, T. Mathew, and Y. Huang. Comparison of Drug Dissolution Profiles: A Proposal Based on Tolerance Limits. *Statist Med*, 35(29):5464–5476, 2016.

[18] X. Dong, Y. Tsong, M. Shen, and J. Zhong. Using Tolerance Intervals for Assessment of Pharmaceutical Quality. *J Biopharm Stat*, 25(2):317–327, 2015.

[19] A. Fedorov, W. M. Wells, R. Kikinis, C. M. Tempany, and M. G. Vangel. Application of Tolerance Limits to the Characterization of Image Registration Performance. *IEEE Trans Med Imaging*, 33(7):1541–1550, 2014.

[20] D. S. Young, C. M. Gordon, S. Zhu, and B. D. Olin. Sample Size Determination Strategies for Normal Tolerance Intervals Using Historical Data. *Qual Eng*, 28(3):337–351, 2016.

[21] D. J. Lizotte and A. Tahmasebi. Prediction and Tolerance Intervals for Dynamic Treatment Regimes. *Stat Methods Med Res*, 26(4):1611–1629, 2017.

[22] Environmental Protection Agency. *Data Quality Assessment: Statistical Methods for Practitioners*. U.S. Environmental Protection Agency, Washington, DC, USA, 2006.

[23] International Atomic Energy Agency. *Safety Report Series No. 52: Best Estimate Safety Analysis for Nuclear Plants: Uncertainty Evaluation*. IAEA Publishing Section, Vienna, Austria, 2008.

[24] International Organization for Standardization. *ISO 16269-6: Statistical Interpretation of Data – Part 6: Determination of Statistical Tolerance Intervals*. International Organization for Standardization, Geneva, Switzerland, 2014.

[25] K. Krishnamoorthy and T. Mathew. *Statistical Tolerance Regions: Theory, Applications, and Computation*. John Wiley & Sons, Hoboken, NJ, 2009.

[26] S. S. Wilks. Determination of Sample Sizes for Setting Tolerance Limits. *Ann Math Statist*, 12(1):91–96, 1941.

[27] S. S. Wilks. Statistical Prediction with Special Reference to the Problem of Tolerance Limits. *Ann Math Statist*, 13(4):400–409, 1942.

[28] A. Wald. An Extension of Wilks' Method for Setting Tolerance Limits. *Ann Math Statist*, 14(1):45–55, 1943.

[29] W. A. Jensen. Approximations of Tolerance Intervals for Normally Distributed Data. *Qual Reliability Eng International*, 25(5):571–580, 2009.

[30] D. S Young. Normal Tolerance Intervals in the tolerance Package. *R J*, 8(2):200–212, 2016.

[31] D. Hoffman. One-Sided Tolerance Limits for Balanced and Unbalanced Random Effects Models. *Technometrics*, 52(3):303–312, 2010.

[32] L. Zhang, T. Mathew, H. Yang, K. Krishnamoorthy, and I. Cho. Tolerance Limits for a Ratio of Normal Random Variables. *J Biopharm Stat*, 20(1):172–184, 2009.

[33] M. Flouri, S. Zhai, T. Mathew, and I. Bebu. Tolerance Limits and Tolerance Intervals for Ratios of Normal Random Variables Using a Bootstrap Calibration. *Biometrical Journal*, 59(3):550–566, 2017.

[34] D. S. Young. tolerance: An R Package for Estimating Tolerance Intervals. *J Stat Softw*, 36(5):1–39, 2010.

[35] L. J. Bain and M. Engelhardt. Simple Approximate Distributional Results for Confidence and Tolerance Limits for the Weibull Distribution Based on Maximum Likelihood Estimators. *Technometrics*, 23(1):15–20, 1981.

[36] W. C. Guenther. On the Use of Best Tests to Obtain Best β-Content Tolerance Intervals. *Stat Neerl*, 25(4):191–202, 1971.

[37] W. R. Blischke and D. N. P. Murthy. *Reliability: Modeling, Prediction, and Optimization*, volume 767. John Wiley & Sons, 2011.

[38] L. Bain. *Statistical analysis of reliability and life-testing models: theory and methods*. Routledge, 2017.

[39] L. T. Fernholz and J. A. Gillespie. Content-Corrected Tolerance Limits Based on the Bootstrap. *Tehnometrics*, 43(2):147–155, 2001.

[40] E. T. Lee and J. Wang. *Statistical Methods for Survival Data Analysis*, volume 476. John Wiley & Sons, 2003.

[41] National Kidney Foundation. ACR. `https://www.kidney.org/kidneydisease/siemens_hcp_acr`. Accessed: 2019-09-26.

[42] R. J. Hogg, S. Furth, K. V. Lemley, R. Portman, G. J. Schwartz, J. Coresh, E. Balk, J. Lau, A. Levin, A. T. Kausz, G. Eknoyan, and A. S. Levey. National Kidney Foundation's Kidney Disease Outcomes Quality Initiative Clinical Practice Guidelines for Chronic Kidney Disease in Children and Adolescents: Evaluation, Classification, and Stratification. *Pediatrics*, 111(6):1416–1421, 2003.

[43] E. K. Harris and J. C. Boyd. *Statistical Bases of Reference Values in Laboratory Medicine*. Marcel Dekker, Inc., New York, NY, 1995.

[44] E. B. Brown, H. K. Iyer, and C. M. Wang. Tolerance Intervals for Assessing Individual Bioequivalence. *Statist Med*, 16(7):803–820, 1997.

[45] X. Dong and T. Mathew. Central Tolerance Regions and Reference Regions for Multivariate Normal Populations. *J Multivar Anal*, 134:50–60, 2015.

[46] L. N. Rathnayake and P. K. Choudhary. Tolerance Bands for Functional Data. *Biometrics*, 72(2):503–512, 2016.

[47] M. W. A. Ramos, G. M. Cordeiro, P. R. D. Marinho, Cí. R. B. Dias, and G. G. Hamedani. The Zografos-Balakrishnan Log-Logistic Distribution: Properties and Applications. *J Stat Theory Appl*, 12(3):225–244, 2013.

[48] M. H. Tahir, M. Mansoor, M. Zubair, and G. Hamedani. McDonald Log-Logistic Distribution With an Application to Breast Cancer Data. *J Stat Theory Appl*, 13(1):65–82, 2014.

[49] A. C. Davison, D. V. Hinkley, and G. A. Young. Recent Developments in Bootstrap Methodology. *Stat Sci*, 18(2):141–157, 2003.

[50] S. John. A Tolerance Region for Multivariate Normal Distribution. *Sankhya Ser A*, 25(4):363–368, 1963.

[51] V. Chew. Confidence, Prediction, and Tolerance Regions for the Multivariate Normal Distribution. *J Am Stat Assoc*, 61(315):605–617, 1966.

[52] K. Krishnamoorthy and S. Mondal. Improved Tolerance Factors for Multivariate Normal Distributions. *Commun Stat Simul Comput*, 35(2):461–478, 2006.

[53] S. Taylor and K. Pollard. Hypothesis tests for point-mass mixture data with application to omics data with many zero values. *Statistical Applications in Genetics and Molecular Biology*, 8(1):1–43, 2009.

[54] A. M. C. Vieira, J. P. Hinde, and C. G. B. Demétrio. Zero-inflated proportion data models applied to a biological control assay. *Journal of Applied Statistics*, 27(3):373–389, 2000.

[55] W. Tu and X. Zhou. A wald test comparing medical costs based on log-normal distributions with zero valued costs. *Statistics in medicine*, 18(20):2749–2761, 1999.

[56] X. Zhou and W. Tu. Confidence intervals for the mean of diagnostic test charge data containing zeros. *Biometrics*, 56(4):1118–1125, 2000.

[57] David Fletcher. Confidence intervals for the mean of the delta-lognormal distribution. *Environmental and Ecological Statistics*, 15(2):175–189, 2008.

[58] WJ Owen and TA DeRouen. Estimation of the mean for lognormal data containing zeroes and left-censored values, with applications to the measurement of worker exposure to air contaminants. *Biometrics*, pages 707–719, 1980.

[59] M. S. Hasan and K. Krishnamoorthy. Confidence intervals for the mean and a percentile based on zero-inflated lognormal data. *Journal of Statistical Computation and Simulation*, 88(8):1499–1514, 2018.

[60] L. Tian. Inferences on the mean of zero-inflated lognormal data: the generalized variable approach. *Statistics in medicine*, 24(20):3223–3232, 2005.

[61] X. Li, X. Zhou, and L. Tian. Interval estimation for the mean of lognormal data with excess zeros. *Statistics & Probability Letters*, 83(11):2447–2453, 2013.

[62] P. Chen and Z. Ye. Approximate statistical limits for a gamma distribution. *Journal of Quality Technology*, 49(1):64–77, 2017.

[63] B. X. Wang and F. Wu. Inference on the gamma distribution. *Technometrics*, 60(2):235–244, 2018.

[64] B. Efron. R. a. fisher in the 21st century. *Statistical Science*, 13(2):95–122, 1998.

[65] K. Gu, H. K. Ng, M. L. Tang, and W. R. Schucany. Testing the Ratio of Two Poisson Rates. *Biometrical Journal*, 50(2):283–298, 2008.

[66] A. C. Cameron and P. K. Trivedi. Regression-Based Tests for Overdispersion in the Poisson Model. *Journal of Econometrics*, 46(3):347–364, 1990.

[67] K. Itô. *Poisson Point Processes and Their Applications to Markov Processes*. Springer, New York, NY, 2015.

[68] A. C. Cameron and Trivedi P. K. *Regression Analysis of Count Data*. Cambridge, New York, NY, 2^{nd} edition, 2013.

[69] J. van den Broek. A Score Test for Zero Inflation in a Poisson Distribution. *Biometrics*, 51(2):738–743, 1995.

[70] N. Janaskul and J. P. Hinde. Score Tests for Zero-Inflated Poisson Models. *Computational Statistics and Data Analysis*, 40(1):75–96, 2002.

[71] N. Janaskul and J. P. Hinde. Score Tests for Extra-Zero Models in Zero-Inflated Negative Binomial Models. *Communications in Statistics - Simulation and Computation*, 38(1):92–108, 2008.

[72] G. Cao, W. Hsu, and D. Todem. A Score-Type Test for Heterogeneity in Zero-Inflated Models in a Stratified Population. *Statistics in Medicine*, 33(12):2103–2114, 2014.

[73] D. Todem, W. Hsu, and J. P. Fine. A Quasi-Score Statistic for Homogeneity Testing Against Covariate-Varying Heterogeneity. *Scandinavian Journal of Statistics*, 45(3):465–481, 2018.

[74] A. Bhattacharya, B. S. Clarke, and G. S. Datta. A Bayesian test for excess zeros in a zero-inflated power series distribution. In N. Balakrishnan, E. A. Peña, and M. J. Silvapulle, editors, *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, pages 89–104. Institute of Mathematical Statistics, Beachwood, OH, 2008.

[75] M. Ridout, J. Hinde, and C. G. B. Demétrio. A Score Test for Testing a Zero-Inflated Poisson Regression Model Against Zero-Inflated Negative Binomial Alternatives. *Biometrics*, 57(1):219–223, 2001.

[76] D. B. Hall and K. S. Berenhaut. Score Tests for Heterogeneity and Overdispersion in Zero-Inflated Poisson and Binomial Regression Models. *The Canadian Journal of Statistics*, 30(3):415–430, 2002.

[77] D. Deng and S. R. Paul. Score Tests for Zero-Inflation and Over-Dispersion in Generalized Linear Models. *Statistica Sinica*, 15(1):257–276, 2005.

[78] R. A. Fisher. The Fiducial Argument in Statistical Inference. *Annals of Eugenics*, 6(4):391–398, 1935.

[79] T. Mathew and D. S. Young. Fiducial-Based Tolerance Intervals for Some Discrete Distributions. *Computational Statistics and Data Analysis*, 61:38–49, 2013.

[80] J. Hannig, H. Iyer, R. C. S. Lai, and T. C. M. Lee. Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association*, 111(515):1346–1361, 2016.

[81] L. Springael and I. Van Nieuwenhuyse. On the Sum of Independent Zero-Truncated Poisson Random Variables. *Research paper 2006-011, June*, pages 1–15, 2006.

[82] M. J. Bayarri, J. O. Berger, and G. S. Datta. Objective Bayes Testing of Poisson Versus Inflated Poisson Models. In *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, volume 3 of *B. Clarke and S. Ghosal (eds.) IMS Collections*, pages 105–121. Institute of Mathematical Statistics, Beachwood, OH, 2008.

[83] N. M. Temme. Asymptotic Estimates of Stirling Numbers. *Studies in Applied Mathematics*, 89(3):233–243, 1993.

[84] M. Engelhardt, L. J. Bain, and F. T. Wright. Inferences on the parameters of the birnbaum-saunders fatigue life distribution based on maximum likelihood estimation. *Technometrics*, 23(3):251–256, 1981.

[85] H. K. T. Ng, D. Kundu, and N. Balakrishnan. Modified moment estimation for the two-parameter birnbaum–saunders distribution. *Computational Statistics & Data Analysis*, 43(3):283–298, 2003.

[86] J. Wu and A. C. M. Wong. Improved interval estimation for the two-parameter birnbaum–saunders distribution. *Computational statistics & data analysis*, 47(4):809–821, 2004.

[87] B. X. Wang. Generalized interval estimation for the birnbaum–saunders distribution. *Computational Statistics & Data Analysis*, 56(12):4320–4326, 2012.

[88] Y. Li and A. Xu. Fiducial inference for birnbaum-saunders distribution. *Journal of Statistical Computation and Simulation*, 86(9):1673–1685, 2016.

[89] N. Balakrishnan and D. Kundu. Birnbaum-saunders distribution: A review of models, analysis, and applications. *Applied Stochastic Models in Business and Industry*, 35(1):4–49, 2019.

[90] D. S.. Chang and C. T. Loon. Percentile bounds and tolerance limits for the birnbaum-saunders distribution. *Communications in Statistics-Theory and Methods*, 23(10):2853–2863, 1994.

[91] M. Lu and D. S. Chang. Bootstrap prediction intervals for the birnbaum-saunders distribution. *Microelectronics Reliability*, 37(8):1213–1216, 1997.

[92] C. Wang, J. Hannig, and H. K. Iyer. Fiducial prediction intervals. *Journal of Statistical Planning and Inference*, 142(7):1980–1990, 2012.

[93] Z. L. Sun. The confidence intervals for the scale parameter of the birnbaum–saunders fatigue life distribution. *Acta Armamentarii*, 30(11):1558–1561, 2009.

[94] N. R. Mann, N. D. Singpurwalla, and R. E. Schafer. *Methods for statistical analysis of reliability and life data*. Wiley, 1974.

[95] A. Xu and Y. Tang. Reference analysis for birnbaum–saunders distribution. *Computational statistics & data analysis*, 54(1):185–192, 2010.

[96] N Balakrishnan and R. Aggarwala. *Progressive censoring: theory, methods, and applications*. Springer Science & Business Media, 2000.

[97] A. J. Lemonte and S. L. P. Ferrari. Testing hypotheses in the birnbaum–saunders distribution under type-ii censored samples. *Computational Statistics & Data Analysis*, 55(7):2388–2399, 2011.

[98] B. Pradhan and D. Kundu. Inference and optimal censoring schemes for progressively censored birnbaum–saunders distribution. *Journal of Statistical Planning and Inference*, 143(6):1098–1108, 2013.

[99] M. Barros, R. Ospina, and A. Tsuyuguchi. Goodness-of-fit tests for the birnbaum-saunders distribution with censored reliability data. *IEEE Transactions on Reliability*, 63(2):543–554, 2014.

Vita

Yixuan Zou

Education

- **M.S. in Statistics** University of Kentucky, Lexington, KY, 2014-2016
- **B.S. in Statistics** Wuhan University, Wuhan, China, 2010-2014

Experience

- **Teaching Assistant** University of Kentucky Department of Statistics, Lexington, KY August 2019-May 2020
- **Biostatistician Intern** Emmes, Rockville, MD, June 2019-August 2020
- **Teaching Assistant** University of Kentucky Department of Statistics, Lexington, KY August 2018-May 2019
- **Senior Data Management Analyst** University of Kentucky College of Pharmacy, Lexington, KY June 2016-August 2018
- **Teaching Assistant** University of Kentucky Department of Statistics, Lexington, KY July 2014-May 2016

Publication

- Clofazimine plasma exposure is associated with diarrheal status. Arnold SL, Conrad TM, **Zou Y**, Tam PI, Barrett LK, Chen CY, et al. *Clin Pharmacol Ther. Submitted.*
- Improving Coverage Probabilities for Parametric Tolerance Intervals via Bootstrap Calibration. **Zou Y**, Young DS. *Stat Med. Published.*
- Using Medical Claims Database to Develop a Population Disease Progression Model for Leuprorelin-treated Subjects with Hormone-Sensitive Prostate Cancer. **Zou Y**, Tang F, Talbert JC, Wang P, Ng CM. *PLoS One. 2020 Mar 24;15(3):e0230571.*
- Population Pharmacokinetics of Exendin-(9-39) and Clinical Dose Selection in Patients with Congenital Hyperinsulinism. Ng CM, Tang F, Seeholzer S, **Zou Y**, De León, DD. *Br J Clin Pharmacol.* 2018 Mar;84(3):520-532.