POINT ESTIMATION, CONFIDENCE SETS, AND BOOTSTRAPPING IN SOME STATISTICAL MODELS

A Thesis
Submitted to the Faculty of Graduate Studies and Research
In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy
In
Mathematics
University of Regina

By
Andrei I. Volodin
Regina, Saskatchewan
July 2002

Abstract

This thesis considers at least four independent statistical problems in various statistical models, perhaps an unusual feature of a thesis in this field. Statistical problems considered here consist of classical and robust point estimation, confidence sets and bootstrapping.

Chapter 2 discovers the asymptotic expansion of the coverage probability of a confidence set centered at the James – Stein estimator and its positive part modification. The result obtained here is based on a multivariate normal population. More importantly, these results can be extended to various models, for example to multiple regression models.

Point estimation based on a new parameterization of the Birnbaum - Saunders lifetime distribution is investigated in Chapter 3. The maximum likelihood estimator and new estimators based on the method of moments and regression - quantile (least squares) method are developed. Asymptotic statistical properties of the proposed estimators are also developed. A Monte Carlo simulation study is conducted to appraise the performance of the proposed strategies for given sample sizes.

In Chapter 4, the method of weighted likelihood is applied to problems of robust estimation of parameters. The large-sample properties of the proposed estimator, along with simulation results, are provided and discussed for an exponential model.

Chapter 5 obtains convergence rates in the form of a Baum – Katz/ Hsu – Robbins/ Spitzer type result for the bootstrapped means from a sequence of random

elements and random variables that are not necessarily independent or identically distributed.

An improvement and generalization of the classical Kolmogorov exponential inequality in the case of negative dependent random variables is presented in Chapter 6.

The final chapter offers some conclusions and interesting open problems for further research.

Acknowledgements

From the bottom of my heart, I wish to extend my sincere gratitude to my thesis supervisor Professor S.Ejaz Ahmed, for introducing me to the wonderful subjects of shrinkage and robust estimation and inspired my interest in these areas. He guided me to successful completion of this thesis, and provided invaluable supervision.

I would like to express my thanks to my committee members: Dr. W. Abdul-Kader, Dr. A. Adatia and Professor R.J. Tomkins for their constructive comments.

I am greatly appreciative of the financial support from the Faculty of Graduate Studies and Research at the University of Regina.

Finally, I wish to thank Mrs. N. Griffiths and Professor B. Gilligan for their useful comments on the presentation of this thesis.

To My Father

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This dissertation has the following main themes: Asymptotic expansion of confidence sets of James – Stein estimators (Chapter 2), estimation of Birnbaum – Saunders distribution parameters (Chapter 3), the weighted likelihood procedure and robust estimation of parameters (Chapter 4), the bootstrap of mean (Chapter 5) and the Kolmogorov exponential inequality (Chapter 6).

Our goal in Chapter 2 is to find the asymptotic expansion of the coverage probability of a confidence set centered at the James – Stein estimator and its positive part modification.

The goal of Chapter 3 is to present maximum likelihood estimators for the new parameterization of Birnbaum – Saunders distribution and to find new estimators based on the moment method and the least squares (regression - quantile) method. Asymptotic mean square error of maximum likelihood estimators and moment method estimators are found. True values of the mean square error of all three estimators are calculated by the method of statistical simulations. Tables are produced, where biases, mean square errors and asymptotic mean square errors for maximum likelihood estimators and moment method estimators are presented. A conclusion about the characteristic changes in mean square error values is given for different values of parameters of the distribution.

In Chapter 4 we suggest applying the method of weighted likelihood for robust estimation of parameters by assigning zero weights to observations with small likelihood. For the example of an exponential distribution we establish that, with the help of controls on the size of rejected observations, the weighted likelihood method gives α-trimmed mean type estimators. We investigate robustness properties of the weighted likelihood estimator in comparison with the usual maximum likelihood estimator.

The main focus of Chapter 5 is to obtain the convergence rates in the form of Baum – Katz/ Hsu – Robbins/ Spitzer type result for bootstrapped means from a sequence of random elements and random variables which are not necessarily independent or identically distributed. It is important to note that the strong laws of large numbers are of practical use in establishing the strong asymptotic validity of the bootstrapped mean for random variables. The consistency of bootstrap estimators has received a lot of attention in recent years due to a growing demand for the procedure, both theoretically and practically.

We start Chapter 5 with a result on the equivalence between the convergence to zero in L^1, completely, almost surely and in probability, of a weighted sum of rowwise independent Banach space valued random elements. Next we apply this result to the field of bootstrapped means for random elements and random variables. More specifically, we obtain strong consistency for the bootstrapped mean, assuming the corresponding weak consistency (concerning the array). We impose neither conditions about the marginal or joint distributions of the random elements forming the sample nor geometric conditions on the Banach space where the random elements take values.

In Chapter 6 we present an improvement and generalization of the classical Kolmogorov exponential inequality in the case of negative dependent random variables.

Now we provide some preliminary discussion and a more detailed overview of the dissertation including brief introductions to each of the (next five) chapters.

Chapter 2. This chapter is the main strength of the thesis and showcases the asymptotic expansion of the coverage probability of a confidence sets centered at James-Stein estimators and its positive-part modification. Parametric estimation theory has been eminently developed for the last four decades in two main subjects: The large sample properties of the maximum likelihood estimator (MLE) and the shrinkage estimation problem, now referred to as the Stein-rule problem for the inadmissibility of the MLE in a small sample. As such, we are primarily interested in the estimation of a parameter when there may be some *uncertainty* about the *constraints* to modify usual estimators such as maximum likelihood or unbiased estimators for increasing their efficiencies. The research led by the remarkable observation of Stein (1956), and following Stein's result, James and Stein (1961) exhibited an estimator that under squared error loss dominates the usual estimator and thus demonstrates its inadmissibility. This result means that the unbiased rule may have an inferior risk when compared to other biased estimators. This area has received considerable attention since then and a remarkably large number of theoretical results has been produced. Indeed, Stein-rule estimation has been evaluated as an effective procedure in small samples from a practical point of view while the theoretical research has progressed in an exemplary fashion. Useful discussions on some of the implications of the pretest and shrinkage estimators in parametric theory are provided by Ahmed (2001), Kubokawa (1998) and Stigler (1990) among others. It may be worth mentioning that this is one of the two areas Professor Efron predicted for continuing research for the early 21st century in *RSS News of January 1995*.

More specifically, the problem of improving upon the usual point estimator of a multivariate normal mean has received enormous attention in the literature during the past 40 years. Without any embellishment, one can safely say that the finding of Stein is one of the most important and most discussed results in statistical decision theory. Later, many statisticians proposed other classes of improved estimators. Some of the proposed point estimators dominate the James-Stein estimator. The celebrated result of James-Stein has added a new dimension to research work in statistical inference during the past four decades. Improved estimators were derived for many distributions, many loss functions, and many point estimation problems (cf., for example Ahmed (1998a, 1998b, 1999a) and etc.). Nonetheless, the companion problem, that of the confidence set, has received comparatively little attention. This is partially due to the increased difficulty of mathematical acquiescence, and also because many of the techniques developed for point estimation do not readily carry over to the confidence set problem. In this chapter, we present the asymptotic expansion of the coverage probability of confidence sets centered at the James-Stein estimator and its positive part modification. We analytically demonstrate that substantial improvement in coverage probability over that of MLE is possible.

Statement of the Problem

If X is one observation from a p-variate normal distribution with mean vector θ and an identity covariance matrix, the confidence set

$$CS_X^0 = \{\theta : \|\theta - X\| \leq C\},$$

is a sphere centered at X and it has probability $1 - \alpha$ of covering the true value of θ if C^2 satisfies $K_p(C^2) = P\{\chi_p^2 \leq C^2\} = 1 - \alpha$. The set CS_X^0 enjoys many optimal properties; for example, it is unbiased and translation invariant. It is also minimax, that is, among all procedures with coverage probability at least $1 - \alpha$, the set CS_X^0 minimizes the maximum expected volume (cf., for example Cramér (1946) or Lehmann (1983)).

A natural question that arises is whether CS_X^0 is a unique minimax set estimator, or do other rules exist? Then since the coverage probability of CS_X^0 is constant for all θ, there would be room to increase coverage probability without increasing the volume of the set. This question was first posed by Stein (1962), who developed heuristic arguments that showed that an improved set can be developed. However, it was Joshi (1967) who proved that the set

$$CS_{\delta^J} = \{\theta : \|\theta - \delta^J(X)\| \le C\},$$

where $\delta^J(X) = (1 - a/(b + X'X))X$, has higher coverage probability than CS_X^0, subject to a and b being sufficiently small and large, respectively, and $p \ge 3$. This result indicated that large gains in coverage probability are possible. Hwang and Casella (1982) extended the method of Joshi and derived exact formulas for confidence sets centered at the James-Stein or its positive-part estimator. They established numerically that for $p \ge 4$,

$$CS_{\delta^+} = \{\theta : \|\theta - \delta^+(X)\| \le C\},$$

where $\delta^+(X) = (1-a/(b+X'X))^+X$, and $+$ denotes positive part, has higher coverage probability than that of CS_X^0, for a specified range of values of a. Since the volume of CS_{δ^+} is the same as that of CS_X^0, it follows that CS_{δ^+} is a minimax set estimator of θ. To this end, Hwang and Casella (1982) did not produce asymptotic results.

Consequently, in Chapter 2 we attempt to prove the high coverage probability based Stein-rule confidence sets asymptotically, rather than analytically.

Chapter 3. A continuous random variable X has a Birnbaum – Sounders distribution (BS-distribution) if X has the cumulative distribution function

$$F_{\alpha,\beta}(x) = 1 - \Phi\left(\alpha\left(\sqrt{\frac{\beta}{x}} - \sqrt{\frac{x}{\beta}}\right)\right), \quad x > 0, \alpha > 0, \beta > 0,$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and α and β are regarded as the location and scale parameter of the distribution, respectively.

This distribution is named B-S because of the pioneering papers by Birnbaum and Saunders (1969a) and (1969b). Birnbaum and Saunders (1969a) presented this new two-parameter family of life length distributions which is derived from a model for fatigue. The BS-distribution is widely used as a lifetime distribution in the various models of reliability theory in the case when a failure of the object under consideration appears to be due to the development of fatigue cracks (Desmond (1986), Ahmad (1988)). This derivation follows from considerations of renewal theory for the number of cycles needed to force a fatigue crack extension to exceed a critical value. Some closure properties of this family are given and some comparisons made with other families, such as lognormal, which have been previously used in fatigue studies. Further, Birnbaum and Saunders (1969b) presented a comprehensive review, both theoretical and practical, of the fitting of this family of distributions to the solution of the problem of crack development. In this paper the maximum likelihood estimation for the scale parameter β is expressed as the solution of the corresponding maximum likelihood equations, and the maximum likelihood estimation of the parameter α is expressed in terms of the estimation of β. Two iterative methods are given for solving the maximum likelihood equation, and conditions are laid out under which the iterations will converge. A strong argument is stressed for the use of the geometric mean of the harmonic and arithmetic means, as a replacement for the maximum likelihood estimation of β or at least as the initial estimate for the iterations. Birnbaum and Saunders (1969b) provided an extensive set of numerical computations consisting of 21 sets of constructed data for various n, α and β. Furthermore, three real data sets relating to actual fatigue testing were also given and analyzed in their paper.

Recently substantial research effort was devoted to estimations of the parameters of this distribution. Desmond (1986) considered estimation of the parameters for censored data. Ahmad (1988) proposed the modernization of Birnbaum and Saunders (1969b) estimation of the scale parameter β (which overestimates the median life) by the jackknife method to eliminate first-order bias. This estimate has the same limiting behavior as that of Birnbaum and Saunders (1969b). Further, Ahmad (1988) reported about an extensive simulation study. Rieck (1995) derived asymptotically optimal linear estimations for symmetrically type II censored samples. We refer to the monograph by Bogdanoff and Kozin (1985) for motivating examples of probabilistic models of cumulative damage. A more recent view on the problem of fatigue crack damages based on stochastic differential equations is suggested by Singpurwalla (1985).

Birnbaum and Saunders (1969a) considered a probability model of a fatigue crack development under cyclic loading in framework of renewal theory. However, a more general model of such phenomena can be described by recurrence equations that have a similar form to ones that produce the lognormal distribution. We refer to Cramér (1946), Parzen (1967) and Desmond (1986) for a description of the recurrence equations method in connection with the lognormal distribution. Indeed, the latter approach gives a richer picture of the physical phenomena of fatigue cracks and, moreover, the distribution of the size of a crack by a fixed moment of time can be arbitrary, while in the BS-model it has to be normal.

In Chapter 3 we derive maximum likelihood estimators for the new parameterization and find new estimators based on the method of moments and the least squares approach. Asymptotic mean square errors of maximum likelihood estimators and method of moments estimation are derived and compared computationally. In addition, numerical values of the mean square errors of all three estimators are calculated by the Monte Carlo simulations. The bias and mean square error of the proposed estimators with respect to one another have been appraised on the simulated data.

Chapter 4. In this chapter we propose the method of weighted likelihood for robust estimation of exponential distribution parameters. We establish that the suggested weighted likelihood method provides $\alpha-$trimmed mean type estimators. Further, we investigate the robustness properties of the weighted likelihood estimator (WLE) in comparison with the usual maximum likelihood estimator (MLE). We present an application of the classical likelihood that was introduced by Hu (1994, 1997) as *the relevance weighted likelihood* to problems of robust estimation of parameters. The weighted likelihood method was introduced as a generalization of the local likelihood method. For further discussion of the local likelihood method in the context of nonparametric regression, see Tibshirani and Hastie (1987), Staniswalis (1989), Fan, Heckman and Wand (1995) and others. In contrast to the local likelihood, the weighted likelihood method can be global, as demonstrated by one of the applications in Hu and Zidek (2001) where the James-Stein estimator is found to be a maximum weighted likelihood estimator with weights estimated from the data.

The theory of weighted likelihood enables a bias-precision trade off to be made without relying on a Bayesian approach. The latter permits the bias-variance trade off to be made in a conceptually straightforward manner. However, the reliance on empirical Bayes methods softens the demands for realistic prior modeling in complex problems. The weighted likelihood theory offers a simple alternative to the empirical Bayesian approach for many complex problems. At the same time, it links within a single formal framework, a diverse collection of statistical domains, such as weighted least squares, nonparametric regression, meta-analysis and shrinkage estimation. All the while the weighted likelihood principle comes with an underlying general theory that extends Wald's theory for maximum likelihood estimators, as it is shown in Hu and Zidek (2001).

We propose applying the method of weighted likelihood to the problem of the robust estimation of the parameter of an exponential distribution by assigning zero weights to observations with small likelihood. Interestingly, the weighted likelihood method yields α-trimmed mean type estimators of the parameter of interest. The statistical asymptotic properties of the WLE is developed and a simulation study is conducted to appraise the behavior of the proposed estimators for moderate samples. Further, a comparative study with the usual maximum likelihood estimator is also provided.

Chapter 5. As was mentioned before, we start Chapter 5 with a result on the equivalence between the convergence to zero in L^1, completely, almost surely and in probability, for a weighted sum of rowwise independent Banach space valued random elements. Hsu and Robbins (1947) introduced the concept of complete convergence of a sequence $\{X_n, n \geq 1\}$ of random variables. $\{X_n, n \geq 1\}$ is said to converge completely to a constant c if, for each $\epsilon > 0$,

$$\sum_{n=1}^{\infty} P[|X_n - c| > \epsilon] < \infty.$$

This concept is extended to a sequence $\{X_n, n \geq 1\}$ of random elements taking values in a real separable Banach space $(\mathcal{B}, ||.||)$. We say that $\{X_n\}$ converges completely to zero if, for each $\epsilon > 0$,

$$\sum_{n=1}^{\infty} P[||X_n|| > \epsilon] < \infty.$$

Chung's (1947) classical strong law of large numbers (SLLN) states that if $\{X_n, n \geq 1\}$ is a sequence of independent random variables with $EX_n = 0$ for all $n \geq 1$ and $\sum_{n=1}^{\infty} \frac{E\psi(|X_n|)}{\psi(n)} < \infty$ when ψ is a positive, even and continuous function such that

$$\frac{\psi(|t|)}{|t|} \uparrow \quad \text{and} \quad \frac{\psi(|t|)}{|t|^2} \downarrow \quad \text{as} \quad |t| \uparrow,$$

then $\frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow 0$ almost surely.

Hu and Taylor (1997) proved a Chung type SLLN for an array of rowwise independent random variables $\{X_{ni}, 1 \leq i \leq n, n \geq 1\}$ weighted by a sequence $\{a_n, n \geq 1\}$ of real numbers with $0 < a_n \uparrow \infty$. One of the conditions required by the authors to prove that $\frac{1}{a_n}\sum_{i=1}^{n} X_{ni} \longrightarrow 0$ almost surely, is that

$$\sum_{n=1}^{\infty}\sum_{i=1}^{n} \frac{E\psi(|X_{ni}|)}{\psi(a_n)} < \infty,$$

when ψ is a positive, even and continuous function such that

$$\frac{\psi(|t|)}{|t|^p} \uparrow \quad \text{and} \quad \frac{\psi(|t|)}{|t|^{p+1}} \downarrow \quad \text{as} \quad |t| \uparrow$$

for some integer $p \geq 2$.

Many classical theorems hold for $\mathcal{B}$-valued random elements under the assumption that the weak law of large numbers (WLLN) holds (see Kuelbs and Zinn (1979), de Acosta (1981), Choi and Sung (1988) and (1989), Wang, Rao and Yang (1993), Kuczmaszewska and Szynal (1994) and Sung (1997)).

Sung (2000) obtained a Hu and Taylor's (1997) result in a general Banach space under the assumption that WLLN holds, using a positive and even function ψ verifying the following condition, which is weaker than used in Hu and Taylor (1997):

$$\frac{\psi(|t|)}{|t|} \uparrow \quad \text{and} \quad \frac{\psi(|t|)}{|t|^{p}} \downarrow \quad \text{as} \quad |t| \uparrow$$

for some $p \geq 1$.

Ordóñez Cabrera and Sung (2002) extend Sung's (1997) result for a weighted sum $S_n = \sum_{i=1}^{k_n} a_{ni}X_{ni}$ of rowwise independent $\mathcal{B}$-valued random elements, where $\{k_n, n \geq 1\}$ is a sequence of positive integers, and $\{a_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ is an array of real constants. They use a function $\psi : (0, +\infty) \longrightarrow (0, +\infty)$ such that, for some $p \geq 1$, there exist constants $C, D > 0$ such that

$$u \geq v \implies C\frac{u}{v} \leq \frac{\psi(u)}{\psi(v)} \leq D\frac{u^p}{v^p}$$

This condition is weaker than the one used in Sung (1997).

Next, we apply the results on complete convergence of weighted sums of random elements and random variables to the problem of consistency of the bootstrap procedure. For expository purposes, we first wish to give a brief description of results related to independent identically distributed (i.i.d.) random variables. Assume that $\{X, X_n;\ n \geq 1\}$ is a sequence of i.i.d. random variables, such that X is nondegenerate, $E(X^2) < \infty$ and defined on some complete probability space $(\Omega, \mathcal{F}, P)$. We now outline the bootstrap procedure for random variables. For $\omega \in \Omega$ and $n \geq 1$, let $P_n(\omega) = \dfrac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}$ denote the empirical measure. For $n \geq 1$, let $\{\hat{X}^{\omega}_{n,j};\ 1 \leq j \leq k_n\}$ be i.i.d. random variables with law $P_n(\omega)$, where k_n is a positive integer. Let $\bar{X}_n(\omega)$ denote the sample mean of $\{X_i(\omega);\ 1 \leq i \leq n\},\ n \geq 1$, that is, $\bar{X}_n(\omega) = \dfrac{1}{n}\sum_{i=1}^{n} X_i(\omega)$.

Bickel and Freedman (1981) and Singh (1981) showed the following weak convergence of distributions for $m(n) = n,\ n \geq 1$ and almost every $\omega \in \Omega$:,

$$\mathcal{L}\left(n^{1/2}(n^{-1}\hat{S}^{\omega}_n - \bar{X}_n(\omega))\right) \to_w N(0, \sigma^2). \tag{1.1}$$

Here $\hat{S}^{\omega}_n = \sum_{j=1}^{n} \hat{X}^{\omega}_{n,j},\ n \geq 1$ and $\sigma^2 = E(X - E(X))^2$. Note that by the Glivenko-Cantelli Theorem, $P_n(\omega)$ is close to $\mathcal{L}(X)$ and, by the Lévy central limit theorem,

$$\mathcal{L}\left(n^{1/2}(n^{-1}S_n - E(X))\right) \to_w N(0, \sigma^2).$$

It follows that if $E(X^2) < \infty$, then the statistic $n^{1/2}(n^{-1}S_n - E(X))$ is close in distribution to the bootstrap statistic $n^{1/2}(n^{-1}\hat{S}^{\omega}_n - \bar{X}_n(\omega))$ for large n ω-almost surely (a.s.). This is, very roughly, the idea of the bootstrap. See Efron (1979), where this nice idea is made explicit and where it is substantiated with several important examples. Giné and Zinn (1989) proved that the existence of the second moment is necessary for there to exist positive scalars $a_n \uparrow \infty$, centering $c_n(\omega)$, and a random probability measure $\nu(\omega)$ nondegenerate with positive probability, such that

$\mathcal{L}(a_n^{-1}\hat{S}_n^\omega - c_n(\omega)) \to_w \nu(\omega)$ for almost every $\omega \in \Omega$. The limit law (1.1) tells us just the right rate at which to magnify the difference $n^{-1}\hat{S}_n^\omega - \bar{X}_n(\omega)$, which is tending a.s. to zero, in order to obtain convergence in distribution to a nondegenerate law ω-a.s. We note from (1 .1) that, for almost every $\omega \in \Omega$,

$$\frac{n^{1/2}}{x_n}(n^{-1}\hat{S}_n^\omega - \bar{X}_n(\omega)) \to 0 \text{ in probability as } n \to \infty$$

for any sequence of constants $\{x_n\}$ with $x_n \uparrow \infty$. On the other hand, strong laws of large numbers were proved by Athreya (1983) and Csörgő (1992) for the bootstrap mean. Arenal-Gutiérrez, Matran and Cuesta-Albertos (1996) analyzed the results of Athreya (1983) and Csörgő (1992). Then, by taking into account the different growth rates for the resampling size $m(n)$, they gave new and simple proofs of those results. They also provided examples that show that the sizes of resampling required by their results to ensure a.s. convergence are not far from optimal.

In the results of Chapter 5 we do not make any assumptions regarding the marginal or joint distributions of the random elements or variables taken from the sample. In this case, the main result of Hu and Taylor (1997) can be seen as a special case of the result given in Theorem 5.1. The results of Section 5.4 are published in Ahmed and Volodin (2001).

Chapter 6. In the paper Bosorgnia, Patterson and Taylor (1996) it is mentioned that in many stochastic models, the assumption that random variables are independent is not plausible. Increases in some random variables are often related to decreases in other random variables so an assumption of negative dependence is more appropriate than an assumption of independence. Lehmann (1966) investigated various conceptions of positive and negative dependence in the bivariate case.

Some interesting exponential bounds were established in Tomkins (1978). Teicher (1979) was working on similar problems to improve the Kolmogorov exponential bounds at about the same time. However, these papers are dealing with independent random variables, not negatively dependent ones.

One of the most interesting and useful examples of negative dependent random variables arises in the situation of a sample from a finite population without replacement. Hence we can apply our result for the so-called *dependent bootstrap*, that is the sample drawn without replacement from the collection of items made up of copies of sample observations. Smith and Taylor (2000) obtained consistency of the bootstrap mean. With the help of the present paper we can prove a law of iterated logarithm type results for dependent bootstrap analogous to the results of Ahmed, Li, Rosalsky and Volodin (2001) for the independent bootstrap.

We think that the dependent bootstrap is only one, of course very interesting, application of the notion of negative dependence. Another one is to apply it to limit theorems. In Chapter 5 we present an improvement and generalization of the classical Kolmogorov exponential inequality in the case of negative dependent random variables. The results of Chapter 5 are accepted for publication in Volodin (2002) in the *Pakistan Journal of Statistics* (Special issue in honor of Professor S.E. Ahmed).

To facilitate the presentation, we introduce the following notation. For two functions $u(\tau)$ and $v(\tau)$ the notation $u(\tau) \asymp v(\tau)$ as $\tau \to \infty$ means

$$0 < \liminf_{\tau\to\infty} \frac{u(\tau)}{v(\tau)} \leq \limsup_{\tau\to\infty} \frac{u(\tau)}{v(\tau)} < \infty$$

and the notation $u(\tau) \sim v(\tau)$ as $\tau \to \infty$ means

$$\lim_{\tau\to\infty} \frac{u(\tau)}{v(\tau)} = 1.$$

Moreover, we use the usual notation $u \approx v$ in the case that u is approximately equal to v.

Chapter 2

Asymptotic Expansion of the Coverage Probability of Stein-rule Estimators

2.1 Introduction

Consider a p-variate vector $(X_1, \ldots, X_p)$ having p-variate normal distribution with identity covariance matrix and consider the problem of estimating the mean vector $\theta = (\theta_1, \ldots, \theta_p)'$ by an estimator δ. A natural estimator of θ is the sample mean vector $\bar{X}$ and it is a maximum likelihood estimator. Recall that $\bar{X}$ is uniformly a minimum variance unbiased and minimax estimator (cf., for example, Cramér (1946) or Lehmann (1983)). For the admissibility of $\bar{X}$, Stein (1956) presented the result that $\bar{X}$ is inadmissible for $p > 2$, while for $p = 1$ and 2 it is admissible. James and Stein (1961) unearthed an explicit form of an estimator

$$\delta_j(\bar{X}) = \left(1 - \frac{a}{n \sum_{i=1}^{p} \bar{X}_i^{\,2}}\right) \bar{X}_j, \qquad a = p - 2 \geq 1.$$

Useful discussions on some of the implications of Stein-rule estimation are given in Ahmed (2001a), Kubokawa (1998) and Stigler (1990) among others. Ahmed and his co-investigators (2002, 2001b, 2000a, 2000b, 1999a, 1999b, 1999c, 1999d, 1998a, 1998b) have demonstrated the superiority of shrinkage estimators over the usual estimator $\bar{X}$ for a wide class of statistical models. For a Bayesian perspective we refer to Efron and Morris (1972) and Robbins (1983).

Note that when $n\sum_{i=1}^{p}\bar{X}_i^{\,2} < a$, the James-Stein estimator yields an over-shrinkage and changes the sign of each $\bar{X}_i$. Consequently, we consider a superior alternative $\delta_j^+(\bar{X})$ to this estimator by considering its positive part only. The positive-part estimator not only mitigates the over-shrinking problem but is also superior to $\delta_j(\bar{X})$ in the entire parameter space. The estimator under consideration prevents a changing of the sign of the maximum likelihood estimator. Hence

$$\delta_j^+(\bar{X}) = \left(1 - \frac{a}{n\sum_{i=1}^{p}\bar{X}_i^{\,2}}\right)^+ \bar{X}_j.$$

In passing, we remark that Hwang and Casella (1982) considered the case of $n = 1$.

2.2 Preliminaries and Coverage Probability

In this section, we describe a simple method of calculating the coverage probability based on Monte-Carlo simulations. Consider standard normal random variables $Z_j = \sqrt{n}(\bar{X}_j - \theta_j), 1 \le j \le p$ and let

$$\tau = \sqrt{n}(\sum_{j=1}^{p}\theta_j^2)^{1/2},\ Y = \sum_{j=1}^{p} Z_j^2,\ X = (\sum_{j=1}^{p}\theta_j^2)^{-1/2}\sum_{j=1}^{p}\theta_j Z_j.$$

Introduce the events

$$\begin{aligned} B &= \{n\sum_{j=1}^{p}(\delta_j(\bar{X}) - \theta_j)^2 < C^2\} \\ &= \{Y(Y + 2X\tau + \tau^2) - 2aX + a^2 - 2aY < C^2(Y + 2X\tau + \tau^2)\} \end{aligned}$$

where C^2 satisfies $K_p(C^2) = P\{\chi_p^2 \le C^2\} = 1 - \alpha$. Define

$$A = \{n\sum_{j=1}^{p}\bar{X}_j^{\,2} > a\} = \{Y + 2\tau X + \tau^2 > a\}.$$

We are interested in the coverage probability

$$Q_p(\tau) = P\{n\sum_{j=1}^{p}(\delta_j^+(\bar{X}) - \theta_j)^2 < C^2\} = P(A^c)I(\tau < C) + P(A\bigcap B).$$

The probabilities of events A and B are defined by the joint distribution of the random variables X and Y.

Proposition 2.1. *The joint distribution of the X and Y is equal to the distribution of the random variables Z and $W+Z^2$, where Z has the standard normal distribution and Z is independent of W which has χ^2_{p-1}-distribution.That is, the joint density function of X and Y equals*

$$f(x,\, y) = \frac{1}{\sqrt{2\pi}} \frac{1}{2^{\frac{p-1}{2}} \Gamma(\frac{p-1}{2})} (y-x^2)^{\frac{p-1}{2}-1} \exp\left\{-\frac{y}{2}\right\},$$

if $y - x^2 > 0$,and $f(x,\, y) = 0$ otherwise.

Proof. Put $\sigma = \left(\sum_{j=1}^p \theta_j^2\right)^{1/2}$. We would like to find the joint Fourier transform of X and Y. We have

$$\varphi(t_1,\, t_2) = E \exp\left\{it_1 \sum_{j=1}^p \theta_j Z_j/\sigma + it_2 \sum_{j=1}^p Z_j^2\right\} =$$

$$\frac{1}{(2\pi)^{p/2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\sum_{j=1}^p (x_j^2 - 2it_1\theta_j x_j/\sigma - it_2 x_j^2)\right\}.$$

The quadratic form under the exponent can be simplified as

$$-\frac{1}{2}(1-2it_2)\sum_{j=1}^p \left(x_j - \frac{it_1\theta_j}{\sigma(1-2it_2)}\right)^2 - \frac{t_1^2}{2(1-2it_2)}.$$

Since we are integrating an analytic function of a complex variable, we can consider i as a parameter and directly use the value of the Laplace - Poisson integral because the p-variate integral can be represented as a product of such integrals. As a result we obtain the following Fourier transform:

$$\varphi(t_1,\, t_2) = \exp\left\{-\frac{t_1^2}{2(1-2it_2)}\right\}(1-2it_2)^{-p/2}.$$

Now consider the inverse Fourier transform. We note that, as a function of $t_1, \varphi(t_1, t_2)$ is the Fourier transform of $N(0, 1-2it_2)$. Hence the transformation with respect to the variable t_1 gives

$$\frac{(1-2it_2)^{1/2}}{\sqrt{2\pi}} \exp\left\{-\frac{x^2(1-2it_2)}{2}\right\}(1-2it_2)^{-p/2}$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \exp\{it_2x^2\}(1-2it_2)^{-(p-1)/2}.$$

For the transformation with respect to the variable t_2 note that

$$\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \exp\{it_2x^2\}(1-2it_2)^{-(p-1)/2}$$

is the characteristic function of the random variable $\chi^2_{p-1} + x^2$. The joint inverse Fourier transform of this multiple gives

$$\frac{1}{2^{(p-1)/2}\Gamma(\frac{p-1}{2})}(y-x^2)^{\frac{p-1}{2}-1} \exp\left\{-\frac{y-x^2}{2}\right\}$$

if $y - x^2 > 0$.

Interestingly, Proposition 2.1 provides a simple method of calculating the coverage probability based on Monte-Carlo simulations. We will conduct a Monte-Carlo simulation study of this coverage probability in the following section.

2.3 Monte-Carlo Simulations

For the simulation study we have generated $N = 10^5$ Monte-Carlo simulations of sets of p independent standard normally distributed random numbers. The sum of their squares is Y and the first normal number is X. For the estimation of the coverage probability we take the sum of the relative frequencies of events A^c and $A \bigcap B$.

For the significance level .9 for the confidence set centered at $\bar{X}$ we produced Table 2.1 below. In this table, for each $p = 3, 5, 11$, in the first column we have the simulated coverage probability produced by our method, and the second column reports the values from Hwang and Casella (1982 Table 2, p.875).

Table 2.1: Coverage probability for positive-part James-Stein estimator.

p / τ	3		5		11	
0	.9567	.9565	.9875	.9879	.9995	.9994
2	.9442	.9458	.9798	.9809	.9988	.9989
4	.9078	.9062	.9354	.9343	.9945	.9949
6	.9029	.9026	.9167	.9162	.9663	.9661
8	.9011	.9014	.9092	.9093	.9435	.9443
10	.9016	.9009	.9048	.9060	.9294	.9307
15	.9002	.9004	.9026	.9027	.9123	.9147
20	.8988	.9002	.8994	.9015	.9079	.9085
25	.9003	.9001	.8999	.9010	.9046	.9055
50	.8982	.9000	.8995	.9002	.9007	.9014
100	.9008	.9000	.8997	.9001	.8996	.9004
500	.9002	.9000	.8993	.9000	.8978	.9000
1000	.8990	.9000	.8995	.9000	.8980	.9000
∞	.8980	.9000	.8992	.9000	.8976	.9000

By inspecting the coverage probabilities in the above table we find these probabilities are very close to each other for various configurations of p and τ. We note that the standard "two sigma" error in the estimation of the probability of success r in N Bernoulli trials is $\Delta = 2\sqrt{\frac{r(1-r)}{N}}$. In our case $N = 10^5, r > .9$ so the standard error in estimation of the coverage probability Q_p, obtained by the Monte-Carlo method will be $\Delta \leq .002$. Hence, the difference between the control and our numbers in the table cannot be not more than Δ. However, in practice this may not always be precise.

2.4 Asymptotic Coverage Probability

This section showcases our main results on the asymptotic coverage probability and numerical computations. We consider the two cases, that is $\tau \to \infty$ and $\tau \to 0$, for the coverage probability.

Proposition 2.2. *If $\tau \to \infty$, then the probability of the event A^c tends to 0 with exponential rate, that is, $P(A^c) \asymp \tau \exp\{-\tau^{-2}\}$.*

Proof. In terms of random variables $Z \sim N(0,\,1)$ and $W \sim \chi^2_{p-1}$ the event

$$A^c = \{W + Z^2 + 2\tau Z + \tau^2 \,<\, a\},$$

takes the form

$$\{(1 + \varepsilon Z)^2 + \varepsilon^2 W < \varepsilon^2 a\}$$

as $\varepsilon = \tau^{-1} \to 0$.

Introduce a normal random variable U with $N(1,\,\varepsilon)$-distribution and a random variable V having gamma distribution with location parameter $(p-1)/2$ and a scale parameter $2\varepsilon^2$. Then the event A^c takes the form $U^2 + V < \varepsilon^2 a$, with probability

$P(A^c)$

$$= \left[\sqrt{2\pi}2^{\frac{p-1}{2}}\varepsilon^{p-1}\Gamma\left(\frac{p-1}{2}\right)\right]^{-1} \int\limits_{-\varepsilon a}^{\varepsilon a} \exp\left\{-\frac{(u-1)^2}{2\varepsilon^2}\right\} du \int\limits_{0}^{\varepsilon^2 a - u^2} v^{\frac{p-1}{2}-1} \mathrm{e}^{-\frac{v}{2\varepsilon^2}}\, dv.$$

Substitute $v/\varepsilon^2 = y$. Then $P(A^c)$

$$= \left[\sqrt{2\pi}\varepsilon 2^{\frac{p-1}{2}}\Gamma\left(\frac{p-1}{2}\right)\right]^{-1} \int\limits_{-\varepsilon a}^{\varepsilon a} \exp\left\{-\frac{(u-1)^2}{2\varepsilon^2}\right\} du \int\limits_{0}^{a - u^2/\varepsilon^2} y^{\frac{p-1}{2}-1} \mathrm{e}^{-\frac{y}{2}}\, dy.$$

If in the upper limit of the second integral we put $u = 0$, then we will obtain the upper bound, that is

$$P(A^c) \le K_{p-1}(a)\left[\,\Phi(\sqrt{a} - \varepsilon^{-1}) - \Phi(-\sqrt{a} - \varepsilon^{-1})\right].$$

Now it is sufficient to use the well-known asymptotic formula

$$1 - \Phi(x) \sim \frac{1}{x\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}.$$

to obtain the desired result.

Proposition 2.3. *If* $\tau \to 0$, *then* $P(A^c) = K_p(a) + O(\tau^2)$.

Proof. Since $A^c = \{(X, Y) : X < (a - Y - \tau^2)/2\tau\}$, and the density $f(x, y)$ is not zero only in the region $y > x^2$, we have

$$Q_A(\tau) = P(A^c) = \int_{\tau-\sqrt{a}}^{\tau+\sqrt{a}} dx \int_{x^2}^{a-\tau^2-2\tau x} f(x, y)dy.$$

The main term of the asymptotic expansion of $Q_A(\tau)$ near $\tau \to 0$ is

$$Q_A(0) = \int_{-\sqrt{a}}^{\sqrt{a}} dx \int_{x^2}^{a} f(x, y)dy = K_p(a).$$

In order to calculate the coefficient of the first degree of τ, we evaluate the derivative $Q'_A(\tau)$. Note

$$Q'_A(\tau) = \int_{(\tau+\sqrt{a})^2}^{a-\tau^2-2\tau(\tau+\sqrt{a})} f(\tau + \sqrt{a}, y)dy$$

$$- \int_{(\tau-\sqrt{a})^2}^{a-\tau^2-2\tau(\tau-\sqrt{a})} f(\tau - \sqrt{a}, y)dy - 2\int_{\tau-\sqrt{a}}^{\tau+\sqrt{a}} f(x, a - \tau^2 - 2\tau x)dx.$$

Hence

$$Q'_A(0) = 2c_p \exp\{-a/2\} \int_{-\sqrt{a}}^{\sqrt{a}} x(a - x^2)^{\frac{p-1}{2}-1}dx = 0.$$

Now we study the asymptotic value of the probability of the event B. Indeed, these asymptotic studies are interesting by themselves. For the case $\tau \to 0$, they define the behavior of the coverage probability of the confidence set centered by the usual James

– Stein estimator but not the one centered by the positive part modification. On the other hand, by Proposition 2.2 the asymptotic value of $P(B)$ as $\tau \to \infty$ gives the asymptotic value of the coverage probability in both cases, for centering by the usual James – Stein estimator as well as for centering by its positive part modification.

We begin with a few general formulas. The probability of the event B is

$$Q_B = \int_0^{a+C^2} dy \int_{A_1}^{\sqrt{y}} f(x,\, y)dx + \int_{a+C^2}^{\infty} dy \int_{-\sqrt{y}}^{A_2} f(x,\, y)dx, \tag{2.1}$$

where

$$A_1 = A_1(y) = \min\left\{\max\left\{-\sqrt{y},\, h(y)\right\},\, \sqrt{y}\right\},$$

$$A_2 = A_2(y) = \max\left\{\min\left\{\sqrt{y},\, h(y)\right\},\, -\sqrt{y}\right\}, \text{ and}$$
$$h(y) = \frac{(y-a)^2 - C^2\tau^2 + y(\tau^2 - C^2)}{2\tau(C^2 + a - y)}.$$

A_1 and A_2 are defined by solutions of the equation $h(y) = \pm\sqrt{y}$, and thus we need the asymptotic behavior in τ here.

Theorem 2.1. *If $\tau \to \infty$, then $Q_p(\tau) = P(B) + O(\tau^{-3}) = K_p(C^2) + O(\tau^{-3})$ if $p > 3$, and, for $p = 3$,*

$$Q_3(\tau) = K_3(C^2) + \frac{C(2C^2-1)}{\tau^2\sqrt{2\pi}} \mathrm{e}^{-\frac{C^2}{2}} + O(\tau^{-3}) \tag{2.2}$$

Proof. Let $\varepsilon = \tau^{-1} \to 0$ and represent the equation $h(y) = \pm\sqrt{y}$ in the following more convenient form for the asymptotic analysis:

$$h(y) = \varepsilon\frac{C^2 + a - y}{2} + \frac{a(\varepsilon^{-1} - \varepsilon C^2)}{2(C^2 + a - y)} - \frac{1}{2\varepsilon} - \frac{\varepsilon C^2}{2} = \pm\sqrt{y}.$$

We can introduce the variable $z^2 = y$. Then we will have the following fourth degree algebraic equation

$$z^4 + \frac{2}{\varepsilon}az^3 + (\varepsilon^{-2} - C^2 - 2a)z^2 - \frac{2}{\varepsilon}(C + a)z + (a^2 - \frac{C^2}{\varepsilon^2}) = 0.$$

We emphasize that one can solve this equation directly and find all its roots. More interestingly, the following procedure seems to be simpler.

We plot both sides of the original equation in the coordinate system $(y,\, h)$. The equation has three roots, as we can see on the plot. To show that we have exactly three roots we suggest the following procedure. Preclude terms with ε, which correspond to the second order asymptotic. This will not strongly influence the plot, and coordinates of roots will be nearly the same. The hyperbola

$$h = \frac{a}{2(C^2 + a - y)\varepsilon} - \frac{1}{2\varepsilon}$$

has asymptotes $y = C^2 + a$ and $h = -1/(2\varepsilon)$ and expands with respect to these asymptotes in the left upper and right lower corners. In the upper corner it first intercepts the parabola $h = -\sqrt{y}$ giving the first root y_1 of our equation. After it intercepts the parabola $h = \sqrt{y}$, giving the second root y_2. The lower branch of the hyperbola intersects the parabola $h = -\sqrt{y}$ giving the third and last root y_3.

By the same plot we can see how to find the limits in (2.1):

$$Q_B = \int\limits_0^{y_1} dy \int\limits_{-\sqrt{y}}^{\sqrt{y}} f(x,\, y)dx + \int\limits_{y_1}^{y_2} dy \int\limits_{h(y)}^{\sqrt{y}} f(x,\, y)dx$$

$$+ \int\limits_{y_3}^{\infty} dy \int\limits_{-\sqrt{y}}^{h(y)} f(x,\, y)dx + O(\varepsilon^3) \tag{2.3}$$

Here the intervals $(y_2,\, C^2 + a)$ and $(C^2 + a,\, y_3)$ correspond to the limits of integration $(\sqrt{y},\, \sqrt{y})$ and $(-\sqrt{y},\, -\sqrt{y})$. These give the value zero for the integrals.

Now we will produce an asymptotic analysis of the roots. Note that this solution is not simple to find. First we find the zero approximations, and after this, guess the asymptotic expansion (especially for the root y_3). After this we need to calculate coefficients for powers of ε by substituting the suggested expansion in our equation.

That is, in the expanded equation we collect the coefficients of the same powers of ε. For the sake of brevity, we give the answer, since it is simple to check by substitution of these expansions of the roots into the equation. The roots are

$$y_1 = C^2 - \varepsilon g + \varepsilon^2 b + O\left(\varepsilon^3\right),\ y_2 = C^2 + \varepsilon g + \varepsilon^2 b + O\left(\varepsilon^3\right),\ y_3 = \frac{1}{4\varepsilon^2} + 2a + O(\varepsilon),$$

where $g = 2Ca,\ b = 2a(a - 2C^2)$

The last integral in (2.3) decreases faster than any positive power of ε, since it is less than

$$\int\limits_{y_3}^{\infty} dy \int\limits_{-\sqrt{y}}^{\sqrt{y}} f(x,\, y)dx = 1 - K_p(y_3) \sim 1 - K(1/(4\varepsilon^2)) \asymp \varepsilon^{-p} \exp\{-1/(8\varepsilon^2)\}.$$

Hence,

$$Q_B = \int\limits_{0}^{y_1(\varepsilon)} dy \int\limits_{-\sqrt{y}}^{\sqrt{y}} f(x,\, y)dx + \int\limits_{y_1(\varepsilon)}^{y_2(\varepsilon)} dy \int\limits_{h(y,\varepsilon)}^{\sqrt{y}} f(x,\, y)dx + O(\varepsilon^3),$$

where the function

$$h(y,\, \varepsilon) = \frac{\varepsilon}{2} \frac{(C^2 + a - y)^2 + a(\varepsilon^{-2} - C^2) - (\varepsilon^{-2} + C^2)(C^2 + a - y)}{C^2 + a - y}.$$

For $\varepsilon \to 0$ we have the expansion $Q_B(\varepsilon) = Q_B(0) + Q_B'(0)\varepsilon + Q_B''(0)\varepsilon^2 + O(\varepsilon^3)$. We now evaluate the derivatives of the function $Q_B(\varepsilon)$. For the sake of brevity, we do not write an argument ε for the functions $y_1(\varepsilon)$, $y_2(\varepsilon)$ and their derivatives. We have

$$Q_B'(\varepsilon) = y_1' \int\limits_{-\sqrt{y_1}}^{\sqrt{y_1}} f(x,\, y_1)dx + y_2' \int\limits_{h(y_2(\varepsilon),\varepsilon)}^{\sqrt{y_2}} f(x,\, y_2)dx - y_1' \int\limits_{h(y_1(\varepsilon),\varepsilon)}^{\sqrt{y_1}} f(x,\, y_1)dx,$$

$$Q_B''(\varepsilon) = y_1'' \int\limits_{-\sqrt{y_1}}^{\sqrt{y_1}} f(x,\, y_1)dx$$

$$+y_1' \left[\frac{y_1'}{2\sqrt{y_1}} f(\sqrt{y_1},\, y_1) + \frac{y_1'}{2\sqrt{y_1}} f(-\sqrt{y_1},\, y_1) + y_1' \int\limits_{-\sqrt{y_1}}^{\sqrt{y_1}} f_y'(x,\, y_1)dx \right]$$

$$+y_2'' \int\limits_{h(y_2,\,\varepsilon)}^{\sqrt{y_2}} f(x,\,y_2)dx$$

$$+y_2' \left[\frac{y_2'}{2\sqrt{y_2}} f(\sqrt{y_2},\,y_2) - h_\varepsilon'(y_2,\,\varepsilon) f(h(y_2,\,\varepsilon),\,y_2) + y_2' \int\limits_{h(y_2,\varepsilon)}^{\sqrt{y_2}} f_y'(x,\,y_2)dx \right]$$

$$-y_1'' \int\limits_{h(y_1,\,\varepsilon)}^{\sqrt{y_1}} f(x,\,y_1)dx$$

$$-y_1' \left[\frac{y_1'}{2\sqrt{y_1}} f(\sqrt{y_1}, y_1) - h_\varepsilon'(y_1, \varepsilon) f(h(y_1, \varepsilon), y_1) + y_1' \int\limits_{h(y_1,\varepsilon)}^{\sqrt{y_1}} f_y'(x, y_1)dx \right].$$

Now we evaluate the values of the functions and their derivatives at $\varepsilon = 0$.

$$y_1(0) = y_2(0) = C^2,\ y_1'(0) = -d,\ y_2'(0) = a,\ y_1''(0) = y_2''(0) = 2b.$$

Further,

$$h(y_{1,2},\,\varepsilon) = \frac{\varepsilon a^2 - 2C^2 a\varepsilon \pm g + b\varepsilon + O(\varepsilon^2)}{2(a \pm g\varepsilon - b\varepsilon^2)},$$

$$h_\varepsilon'(y_{1,2},\,\varepsilon) = \frac{a^2 - 2C^2 a + ba + g^2 + O(\varepsilon)}{2a^2}.$$

Hence $h(y_{1,2}(0),\,0) = \mp C$, and $h_\varepsilon'(y_{1,2}(0), 0) = (3a - 2C^2)/2$.

Further, recall that

$$f(x,\,y) = c_p(y - x^2)^{\frac{p-1}{2}-1} \exp\{-\frac{y}{2}\},\ y - x^2 > 0.$$

Hence for all x we have $f(\pm x, x^2) = 0$ for $p > 3$, and

$$f(\pm x,\,x^2) = \frac{1}{2\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}$$

for $p = 3$.

Substituting the obtained expressions into $Q_B'(0)$ and $Q_B''(0)$, we obtain

$$Q_B'(0) = -g \int_{-C}^{C} f(x,\,C^2)dx + g \int_{C}^{C} f(x,\,C^2)dx + g \int_{-C}^{C} f(x,\,C^2)dx = 0,$$

and if $p > 3$, then

$$Q''_B(0) = 2b\int_{-C}^{C} f(x,\, C^2)dx - g\left[\frac{-g}{2C}f(C, C^2) + \frac{-g}{2C}f(-C,\, C^2)\right.$$

$$\left. -g\int_{-C}^{C} f'_y(x, C^2)dx\right]$$

$$+2b\int_{C}^{C} f(x,\, C^2)dx + g\left[\frac{g}{2C}f(C, C^2) - \frac{3a-2C^2}{2}f(C, C^2)\right.$$

$$\left. +g\int_{C}^{C} f'_y(x, C^2)dx\right]$$

$$-2b\int_{-C}^{C} f(x, C^2)dx + g\left[\frac{-g}{2C}f(C, C^2) - \frac{3a-2C^2}{2}f(-C, C^2)\right.$$

$$\left. -g\int_{-C}^{C} f'_y(x, C^2)dx\right]$$

$$= 2b\int_{-C}^{C} f(x, C^2)dx + g^2\int_{-C}^{C} f'_y(x, C^2)dx - 2b\int_{-C}^{C} f(x, C^2)dx$$

$$-g^2\int_{-C}^{C} f'_y(x, C^2)dx = 0.$$

But if $p = 3$, then $f(\pm C,\, C^2) \neq 0$, and

$$Q''_B(0) = \left[\frac{g^2}{C} - g(3a - 2C^2)\right]\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{C^2}{2}} = \frac{Ca(2C^2 - a)}{\sqrt{2\pi}}\mathrm{e}^{-\frac{C^2}{2}}.$$

Finally, we note that $Q''_B(0)$ is positive for $C^2 = K_p^{-1}(1-\alpha)$ with $\alpha < 0.5$.

In the following theorem we consider the second order asymptotic as $\tau \to 0$ for the probability of the event B. We will prove that the coefficient for τ to the power one equals zero.

Theorem 2.2. *If $\tau \to 0$, then*

$$Q_B(\tau) = K_p\left(a + \frac{C^2}{2} + \sqrt{aC^2 + \frac{C^4}{4}}\right) - K_p\left(a + \frac{C^2}{2} - \sqrt{aC^2 + \frac{C^4}{4}}\right)$$

$$+O(\tau^2) \tag{2.4}$$

Proof. In the limits of the integrals (2.1) we don't consider terms near τ^2. Hence

$$h(y) = \frac{(y-a)^2 - C^2 y}{2\tau(C^2 + a - y)}.$$

In order to find the limits of integration it is necessary to solve the equation $h(y) = \pm\sqrt{y}$.

This equation has four roots and again, to show this we should look at the plot of the functions $h(y)$ and $\pm\sqrt{y}$ focusing only on the region $y > 0$. This is seen as follows. In the region $0 < y < C^2 + a$ the curve $h(y)$ starts from a positive point $h(0) = a^2/(2\tau(C^2 + a))$ on the h-axis. Then it decreases monotonically and intersects the parabola $h = \pm\sqrt{y}$ in two points y_1 and y_2 (the roots of the equation that we are interested in). The straight line $h = C^2 + a$ is an asymptote of $h(y)$ as $y \to C^2 + a - 0$. In the region $y > C^2 + a$, if $y \to C^2 + a(+0)$, the curve $h(y)$ tends to $+\infty$ becoming closer to this line. Further, when y increases, the curve $h(y)$ goes down to $-\infty$ monotonically and intersects the parabola $h = \pm\sqrt{y}$ in two points y_3 and y_4.

Consequently, we have four roots and $P(B)$ can be written as

$$Q_B = P(B) = \int_{y_1}^{y_2} dy \int_{h(y)}^{\sqrt{y}} f(x,y)dx \; + \int_{y_2}^{y_3} dy \int_{-\sqrt{y}}^{\sqrt{y}} f(x,\, y)dx$$

$$+ \int_{y_3}^{y_4} dy \int_{-\sqrt{y}}^{h(y)} f(x,\, y)dx \qquad (2.5)$$

Now we will find asymptotic values of all four roots of the equation $(y-a)^2 - C^2 = \pm 2\tau\sqrt{y}(C^2 + a - y)$. Noting that all roots have the same form $y_i = y_0 + a\tau$, and if in the original equation we keep the terms with τ^2, then in general we can find the coefficients of $\tau^2,\ \tau^3$ so to speak. The zero approximation is common for the first two roots and for the last two roots:

$$y_0^{\mp} = a + \frac{C^2}{2} \mp \sqrt{C^2 a + \frac{C^4}{4}}$$

The roots differ only by the addition or subtraction of one constant

$$g^{\mp} = \frac{2\sqrt{y_0^{\mp}}(C^2 + a)}{2(y_0^{\mp} - a - C^2)},$$

multiplied by τ. For that reason, with the accuracy $O(\tau^{-2})$ we have the following asymptotic approximation:

$$y_1 = y_0^- - g^-\tau,\; y_2 = y_0^- + g^-\tau,\; y_3 = y_0^+ - g^+\tau,\; y_4 = y_0^+ + g^+\tau$$

After the substitution of these roots into the limits of integration with Taylor's expansion about the points $y_0^{\pm}$ we obtain (2.5).

2.5 Final Result

Now we can formulate and prove our final result.

Theorem 2.3. *If* $\tau \to 0$, *then*

$$Q_p(\tau) = K_p \left(a + \frac{C^2}{2} + \sqrt{\frac{C^4}{4} + C^2 a} \right) + O(\tau^2) \tag{2.6}$$

Proof. The coverage probability (the confidence coefficient) is equal to $Q_p = Q_A + Q_2$, where $Q_A = P(A^c) = P(Y + 2\tau X + \tau^2 < a))$, $Q_2 = P(A \bigcap B)$. The event $A \bigcap B$ is defined by the region of integration (2.4) from which we have to delete the part corresponding to the event A^c. That is, for the calculation of $Q_2(\tau)$, we need to delete from the region defined by the limits in the integral (2.4) the part defined by the inequality $x < (a - y)/2\tau$. We do not consider the term with τ^2 here. We are interested only in the linear term in the expansion.

Again we consider a plot. The line $h = (a - y)/2\tau$ in the region $y \geq 0$ starts from the point $a/2\tau$ when $y = 0$. After this it goes down and, most importantly, does not intersect the curve

$$h(y,\,\tau)=\frac{(y-a)^2-C^2y}{2\tau(C^2+a-y)}.$$

As a matter of fact, it goes between the branches of this curve that lie in the regions

$$y<C^2+a \text{ and } y>C^2+a.$$

However, the line intersects the parabola $h=\sqrt{y}$ at the point $y=a-2\tau\sqrt{a}$, and the parabola $h=-\sqrt{y}$ at the point $y=a+2\tau\sqrt{a}$.

Further, if we define

$$y_0=a+\frac{C^2}{2}+\sqrt{\frac{C^4}{4}+C^2a},\quad g=\frac{2\sqrt{y_0}(y_0-a-C^2)}{C^2+2(y_0-a-C^2)},$$

then the limits of integration will be:

$$Q_2=\int\limits_{a-b\tau}^{a+b\tau}dy\int\limits_{(a-y)/2\tau}^{\sqrt{y}}f(x,\,y)dx$$

$$+\int\limits_{a+b\tau}^{y_0-g\tau}dy\int\limits_{-\sqrt{y}}^{\sqrt{y}}f(x,\,y)dx+\int\limits_{y_0-a\tau}^{y_0+g\tau}dy\int\limits_{-\sqrt{y}}^{h(y,\tau)}f(x,\,y)dx+O(\tau^2),$$

where $b=2\sqrt{a}$.

Hence, the region of iteration contains only the roots y_3 and y_4.

In the first integral consider the substitution $(y-a)/\tau=v$ and in the third $(y-y_0)/\tau=v$. As a result we will have

$$Q_2=\tau\int\limits_{-b}^{b}dv\int\limits_{-v/2}^{\sqrt{v\tau+a}}f(x,\,v\tau+a)dx$$

$$+\int\limits_{a+b\tau}^{y_0-g\tau}dy\int\limits_{-\sqrt{y}}^{\sqrt{y}}f(x,\,y)dx+\tau\int\limits_{-g}^{g}dv\int\limits_{-\sqrt{v\tau+y_0}}^{h(v\tau+y_0,\,\tau)}f(x,\,v\tau+y_0)dx+O(\tau^2).$$

Direct computations show that $h(v\tau+y_0,\,\tau)=-v/2$. Since near the first and third integral, we have factors of τ, then in the integrals we need to put $\tau=0$ and all other terms will be $O(\tau^2)$. Hence Q_2

$$= \tau \int_{-b}^{b} dv \int_{-v/2}^{\sqrt{a}} f(x,a)dx + \int_{a+b\tau}^{y_0-g\tau} dy \int_{-\sqrt{y}}^{\sqrt{y}} f(x,y)dx +$$

$$\tau \int_{-g}^{g} dv \int_{-\sqrt{y_0}}^{-v/2} f(x,y_0)dx + O(\tau^2)$$

$$= \tau \int_{-b}^{b} dv \int_{-v/2}^{\sqrt{a}} f(x,a)dx + \int_{a}^{y_0} dy \int_{-\sqrt{y}}^{\sqrt{y}} f(x,y)dx - \tau g \int_{-\sqrt{y_0}}^{\sqrt{y_0}} f(x,y_0)dx$$

$$-b\tau \int_{-\sqrt{a}}^{\sqrt{a}} f(x,a)dx + \tau \int_{-g}^{g} dv \int_{-\sqrt{y_0}}^{-v/2} f(x,\, y_0)dx + O(\tau^2)$$

$$= K_p(y_0) - K_p(a) + \tau[\int_{-b}^{b} dv \int_{-v/2}^{\sqrt{a}} f(x,\, a)dx - g \int_{-\sqrt{y_0}}^{\sqrt{y_0}} f(x,\, y_0)dx$$

$$-b \int_{-\sqrt{a}}^{\sqrt{a}} f(x,\, a)dx + \int_{-g}^{g} dv \int_{-\sqrt{y_0}}^{-v/2} f(x,\, y_0)dx] + O(\tau^2).$$

Now calculate the integral in the square brackets. Start with the integral

$$\int_{-\sqrt{c}}^{\sqrt{c}} f(x,\, c)dx = c_p \mathrm{e}^{-c/2} \int_{-\sqrt{c}}^{\sqrt{c}} (c - x^2)^{\frac{p-1}{2}-1} dx$$

$$= c_p \mathrm{e}^{-c/2} c^{\frac{p}{2}-1} \int_0^1 (1-\tau)^{\frac{p-1}{2}-1} \tau^{\frac{1}{2}-1} dx$$

$$= \frac{\mathrm{e}^{-\frac{c}{2}} c^{\frac{p}{2}-1}}{\sqrt{2\pi} 2^{\frac{p-1}{2}} \Gamma(\frac{p-1}{2})} \mathrm{B}\left(\frac{p-1}{2},\, \frac{1}{2}\right) = \frac{\mathrm{e}^{-\frac{c}{2}} c^{\frac{p}{2}-1}}{2^{\frac{p}{2}} \Gamma(\frac{p}{2})}.$$

The double integrals in the square brackets can be evaluated by changing the order of integration. The first double integral (recall $b = 2\sqrt{a}$) equals:

$$\int_{-b}^{b} dv \int_{-v/2}^{\sqrt{a}} f(x,\, a)dx = \int_{-\sqrt{a}}^{\sqrt{a}} f(x,\, a)dx \int_{-2x}^{2\sqrt{a}} dv$$

$$= 2c_p \mathrm{e}^{-a/2} \int\limits_{-\sqrt{a}}^{\sqrt{a}} (x + \sqrt{a})(a - x^2)^{\frac{p-1}{2}-1} dx$$

$$= 2c_p \mathrm{e}^{-a/2} a^{\frac{p-1}{2}} \mathrm{B}\left(\frac{p-1}{2}, \frac{1}{2}\right) = \frac{a^{\frac{p-1}{2}} \mathrm{e}^{-\frac{a}{2}}}{2^{\frac{p}{2}-1}\Gamma\left(\frac{p}{2}\right)}.$$

Finally, we still have to evaluate the last integral:

$$\int\limits_{-g}^{g} dv \int\limits_{-\sqrt{y_0}}^{-v/2} f(x, y_0) dx = \int\limits_{-\sqrt{y_0}}^{-g/2} f(x, y_0) dx \int\limits_{-g}^{-g} dv + \int\limits_{-g/2}^{g/2} f(x, y_0) dx \int\limits_{-g}^{-2x} dv$$

$$= 2a \int\limits_{-\sqrt{y_0}}^{-g/2} f(x, y_0) dx + \int\limits_{-g/2}^{-g/2} (g - 2x) f(x, y_0) dx$$

$$= c_p \mathrm{e}^{-\frac{y_0}{2}} y_0^{\frac{p-1}{2}-1} \left[2g \int\limits_{-\sqrt{y_0}}^{-g/2} \left(1 - \left(\frac{x}{\sqrt{y_0}}\right)^2\right)^{\frac{p-1}{2}-1} dx \right.$$

$$\left. + \int\limits_{-g/2}^{g/2} (g - 2x) \left(1 - \left(\frac{x}{\sqrt{y_0}}\right)^2\right)^{\frac{p-1}{2}-1} dx \right]$$

$$= c_p \mathrm{e}^{-\frac{y_0}{2}} y_0^{\frac{p-1}{2}-1} \left[2g \int\limits_{-\sqrt{y_0}}^{-g/2} \left(1 - \left(\frac{x}{\sqrt{y_0}}\right)^2\right)^{\frac{p-1}{2}-1} dx \right.$$

$$\left. +2g \int\limits_{-g/2}^{0} \left(1 - \left(\frac{x}{\sqrt{y_0}}\right)^2\right)^{\frac{p-1}{2}-1} dx \right]$$

$$= c_p \mathrm{e}^{-\frac{y_0}{2}} y_0^{\frac{p-1}{2}-1} 2g \int\limits_{0}^{\sqrt{y_0}} \left(1 - \left(\frac{x}{\sqrt{y_0}}\right)^2\right)^{\frac{p-1}{2}-1} dx$$

$$= 2g c_p \mathrm{e}^{-\frac{y_0}{2}} y_0^{\frac{p-1}{2}-1} \int\limits_{0}^{1} t^{-\frac{1}{2}} (1 - t)^{\frac{p-1}{2}-1} dx = g \frac{\mathrm{e}^{-\frac{y_0}{2}} y_0^{\frac{p-1}{2}-1}}{2^{\frac{p}{2}} \Gamma\left(\frac{p}{2}\right)}.$$

In the following section, numerical work is carried out to appraise our asymptotic work for various values of τ and p.

2.6 A Numerical Study

Some numerical results based on formulas (2.2) from Theorem 2.1 give us the following Table 2.2. The significance level is .9

Table 2.2 reports the coverage probability for given p and large values of τ. In the first column we have the coverage probability based on formula (2.2) from Theorem 2.1 and the second reads the values from Table 2, p.875 Hwang and Casella (1982) for the different values of τ.

Table 2.2: The coverage probability for large values of τ

τ \ p	3		5		11	
15	.9510	.9004	.9000	.9027	.9000	.9147
20	.9287	.9002	.9000	.9015	.9000	.9085
25	.9183	.9001	.9000	.9010	.9000	.9055
50	.9046	.9000	.9000	.9002	.9000	.9014
100	.9011	.9000	.9000	.9001	.9000	.9004
500	.9000	.9000	.9000	.9000	.9000	.9000
1000	.9000	.9000	.9000	.9000	.9000	.9000

For a given p, the difference between the values of two columns is not significant, particularly for $p > 3$ and large values of τ.

However, the most interesting case is when τ is small. The numerical computations of the formula (2.6) from Theorem 2.3 are given in the Table 2.3. Here we do not have values from Hwang and Casella (1982) for comparison. Therefore, we conducted a simulation study to obtain these. The significance level is .9, and for given p, in the first column we provide the simulated coverage probability. The second column represents the asymptotic coverage probability obtained by the formula (2.6) from Theorem 2.3.

Table 2.3: The coverage probability for small values of τ

τ \ p	3		5		11	
0	.8966	.9000	.8991	.9000	.8988	.9000
.1	.8995	.9000	.9007	.9000	.8976	.9000
.2	.8988	.9000	.8993	.9000	.8986	.9000
.3	.8980	.9000	.8986	.9000	.8966	.9000
.4	.8981	.9000	.8975	.9000	.8936	.9000
.5	.8968	.9000	.8947	.9000	.8910	.9000
.6	.8965	.9000	.8914	.9000	.8862	.9000
.7	.8942	.9000	.8916	.9000	.8847	.9000
.8	.8942	.9000	.8859	.9000	.8802	.9000
.9	.8936	.9000	.8832	.9000	.8754	.9000

The numerical result of the above table very well supports our analytical findings.

Chapter 3

Point Estimation of Birnbaum – Saunders Lifetime Distribution

3.1 Introduction

A continuous random variable X has a Birnbaum – Saunders distribution (BS-distribution), if X has the cumulative distribution function

$$F_{\alpha,\beta}(x) = 1 - \Phi\left\{\alpha\left(\sqrt{\frac{\beta}{x}} - \sqrt{\frac{x}{\beta}}\right)\right\}, \quad x > 0,\ \alpha > 0,\ \beta > 0, \tag{3.1}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and α and β are regarded as the location and scale parameter of the distribution, respectively.

The use of recurrence relations is given in Birnbaum and Saunders (1969a). Interestingly, these recurrence relations lead to a new parameterization of the model. For example, on a metallic sample, which has the form of a rectangular plate of thickness h and is fixed on two sides, suppose there is a cyclic loading, which results in the development of a crack. Let X_k be the size of the crack at time $k = 1, 2, \ldots$, that is, after the kth loading cycle. The following recurrence equations are derived (Desmond

(1986)):

$$X_{k+1} = X_k + Y_k g(X_k),\ k = 0, 1, \ldots, \quad Y_0 = X_0 = 0,\ g(0) \neq 0. \tag{3.2}$$

These equations connect the crack size in previous and next moments of time by a positive continuous function g and a sequence of random variables Y_1, $Y_2, \ldots$, that take care of variations in values of loading and some other physical factors that influence the development of the crack. Assume that the random variables Y_k are nonnegative, independent and identically distributed. Further, we assume the existence of the second moment of Y_k. We are interested in the time $U\,(= 1, 2, ...)$, at which the crack achieves the critical value h. By the recurrence equations (3.2) we obtain

$$\sum_{k=0}^{U-1} Y_k = \sum_{k=0}^{U-1} \frac{X_{k+1} - X_k}{g(X_k)} \approx \int_0^{X_U} \frac{dx}{g(x)}.$$

Here we assume, of course, that the increments $X_{k+1} - X_k$ are sufficiently small.

For each sufficiently large value t of the variable U, the random variable on the left hand side of the equality, can be approximated by a normal distribution with mean tm and variance $t\sigma^2$, where $m = E(Y_1),\ \sigma^2 = Var(Y_1)$. Hence, for large t,

$$P(U > t) = P\left(\int_0^{X_t} \frac{dx}{g(x)} < \int_0^{h} \frac{dx}{g(x)}\right) \approx P\left(\sum_{k=0}^{U-1} Y_k < a(h)\right)$$
$$\approx \Phi\left(\frac{a(h) - mt}{\sigma\sqrt{t}}\right),$$

where

$$a(h) = \int_0^h \frac{dx}{g(x)}$$

is a strictly increasing function of the upper limit $h\,(> 0)$, since $g(x) \geq 0$.

In this chapter we introduce a re-parameterization technique. We obtain a natural re-parameterization by letting $\lambda = a(h)/\sigma,\ \mu = m/\sigma$. Importantly, this re-parameterization fits the physics of studying phenomena since the proposed parameters λ and μ correspond to the thickness of the sample and nominal treatment loading on the sample, respectively. The cumulative distribution function of U is

$$F_{\mu,\lambda}(x) = 1 - \Phi\left(\frac{\lambda}{\sqrt{x}} - \mu\sqrt{x}\right), \quad x > 0, \lambda > 0, \mu > 0.$$

Finally, we find the relations between the usual parameters α, β and new parameters μ and λ as follows:

$$\begin{array}{rcl} \mu & = & \frac{\alpha}{\sqrt{\beta}} \\ \lambda & = & \alpha\sqrt{\beta} \end{array} \quad \text{and} \quad \begin{array}{rcl} \alpha & = & \sqrt{\mu\lambda} \\ \beta & = & \frac{\lambda}{\mu} \end{array} \tag{3.3}$$

Birnbaum and Saunders (1969b) found the expected values and mean square errors (MSE) of some statistics of BS-distributions. We can use the above relations to rewrite expected values and MSE of the statistics in terms of μ and λ.

A plan of this chapter is as follows. In Sections 3.2 -3.3 we consider two estimation strategies, namely maximum likelihood estimations (MLE), and the method of moments estimations (MME). Further, expressions for asymptotic MSE of the proposed estimators are derived analytically and some computational aspects are discussed. The regression-quantile (least square) estimation (RQE) technique is presented in Section 3.4. In Section 3.5, we provide MSE analysis of the estimators and summarize the findings.

3.2 The Method of Maximum Likelihood

The probability density function of the BS- distribution after the re-parameterization is as follows:

$$f(x\,;\,\mu,\,\lambda) = \frac{1}{2\sqrt{2\pi}}\left(\frac{\lambda}{x\sqrt{x}} + \frac{\mu}{\sqrt{x}}\right)\exp\left\{-\frac{1}{2}\left(\frac{\lambda}{\sqrt{x}} - \mu\sqrt{x}\right)^2\right\}, \quad x > 0.$$

The observed likelihood function is of the form

$$L(\mu,\,\lambda) = \sum_{k=1}^{n} \ln f(X_k,\,\mu,\,\lambda)$$
$$\asymp \sum_{k=1}^{n} \ln\left(\frac{\lambda}{X_k\sqrt{X_k}} + \frac{\mu}{\sqrt{X_k}}\right) - \frac{1}{2}\sum_{k=1}^{n}\left(\frac{\lambda}{\sqrt{X_k}} - \mu\sqrt{X_k}\right)^2.$$

We obtain the system of maximum likelihood equations by evaluating derivatives with respect to μ and λ.

$$\frac{\partial L(\mu,\lambda)}{\partial \mu} = \sum_{k=1}^{n} \frac{X_k}{\lambda+\mu X_k} + \sum_{k=1}^{n} (\lambda - \mu X_k)\,,$$
$$\frac{\partial L(\mu,\lambda)}{\partial \lambda} = \sum_{k=1}^{n} \frac{1}{\lambda+\mu X_k} + \sum_{k=1}^{n} \left(\frac{\lambda}{X_k} - \mu\right).$$

Hence the MLE $\hat{\mu}^{(MLE)}$ and $\hat{\lambda}^{(MLE)}$ of μ and λ can be obtained by simultaneously solving $\frac{\partial L(\mu,\lambda)}{\partial \mu} = 0$ and $\frac{\partial L(\mu,\lambda)}{\partial \lambda} = 0$. Further, we get

$$\mu = \frac{\lambda}{n}\sum_{k=1}^{n}\frac{1}{X_k} - \frac{1}{n}\sum_{k=1}^{n}\frac{1}{\lambda+\mu X_k} \;\; (= f_1(\mu,\,\lambda)),$$

$$\lambda = \frac{\mu}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n}\frac{X_k}{\lambda+\mu X_k} \;\; (= f_2(\mu,\,\lambda)). \tag{3.4}$$

It seems to be natural to use the iteration method for a solution of the system which is given by the above two equations. However, for a parametric space $\mu > 0,\ \lambda > 0$ there is no initial point that can ensure the convergence of the iteration process. Thus, in this situation the iteration process may diverge. A necessary (but not sufficient) condition for convergence of the iteration process is the inequality $\|\, A(\mu,\,\lambda) \,\| < 1$, where the operator $A(\mu,\,\lambda) = (f_1,\,f_2)$. In the L_2-metric the norm of operator A is equal to the largest eigenvalue of the Jacobian matrix

$$G = G(\mu,\,\lambda) = \begin{pmatrix} \frac{\partial f_1}{\partial \mu} & \frac{\partial f_1}{\partial \lambda} \\ \frac{\partial f_2}{\partial \mu} & \frac{\partial f_2}{\partial \lambda} \end{pmatrix}.$$

The equation for eigenvalues is

$$\left(\frac{1}{n}\sum_{k=1}^{n}\frac{X_k}{(\lambda+\mu X_k)^2}-a^2\right)^2$$
$$-\frac{1}{n}\left[\left(\sum_{k=1}^{n}\frac{1}{X_k}+\sum_{k=1}^{n}\frac{1}{(\lambda+\mu X_k)^2}\right)\left(\sum_{k=1}^{n}X_k+\sum_{k=1}^{n}\frac{X_k^2}{(\lambda+\mu X_k)^2}\right)\right]=0,$$

and its solutions are

$$a_1(\mu,\lambda)=\frac{1}{n}\sum_{k=1}^{n}\frac{X_k}{(\lambda+\mu X_k)^2}$$
$$+\frac{1}{\sqrt{n}}\left(\sum_{k=1}^{n}\frac{1}{X_k}+\sum_{k=1}^{n}\frac{1}{(\lambda+\mu X_k)^2}\right)\left(\sum_{k=1}^{n}X_k+\sum_{k=1}^{n}\frac{X_k^2}{(\lambda+\mu X_k)^2}\right);$$
$$a_2(\mu,\lambda)=\frac{1}{n}\sum_{k=1}^{n}\frac{X_k}{(\lambda+\mu X_k)^2}$$
$$-\frac{1}{\sqrt{n}}\left(\sum_{k=1}^{n}\frac{1}{X_k}+\sum_{k=1}^{n}\frac{1}{(\lambda+\mu X_k)^2}\right)\left(\sum_{k=1}^{n}X_k+\sum_{k=1}^{n}\frac{X_k^2}{(\lambda+\mu X_k)^2}\right).$$

Note that $a_1 > a_2$. So $\| A \| = a_1(\mu,\lambda)$. It is safe to conclude that $a_1(\mu,\lambda) > 1$ for all $\mu > 0$ and $\lambda > 0$. Hence, most likely there do not exist the initial values of parameters that will ensure the convergence of the iteration process. To this end, we will use an alternative method to obtain a solution of the maximum likelihood equations in Section 3.5.

We now derive the asymptotic information matrix

$$\mathbf{I}=\begin{pmatrix} I_{11}(\mu,\lambda) & I_{12}(\mu,\lambda) \\ I_{21}(\mu,\lambda) & I_{22}(\mu,\lambda) \end{pmatrix}.$$

First, let $C=\frac{1}{2^{3/2}\sqrt{\pi}}$ and

$$K(x,\mu,\lambda)=\left(\frac{\lambda}{x^{3/2}}+\mu\sqrt{x}\right)\exp\left\{-\frac{1}{2}\left(\frac{\lambda}{\sqrt{x}}-\mu\sqrt{x}\right)^2\right\}.$$

Hence,

$$I_{11}(\mu,\lambda)=E\left(\frac{\partial\ln f(X\,;\,\mu,\,\lambda)}{\partial\mu}\right)^2$$

$$= C\int_0^\infty \left(\frac{x}{\lambda+\mu x} + \lambda - \mu x\right)^2 K(x,\mu,\lambda)dx, \text{ and}$$

$$I_{22}(\mu,\lambda) = E\left(\frac{\partial \ln f(X\,;\,\mu,\,\lambda)}{\partial\lambda}\right)^2$$

$$= C\int_0^\infty \left(\frac{1}{\lambda+\mu x} + \mu - \frac{\lambda}{x}\right)^2 K(x,\mu,\lambda)dx,$$

$$I_{12}(\mu,\lambda) = I_{21}(\mu,\lambda) = E\left(\frac{\partial \ln f(X\,;\,\mu,\,\lambda)}{\partial\mu}\frac{\partial \ln(X\,|\,\mu,\,\lambda)^2}{\partial\lambda}\right)$$

$$= C\int_0^\infty \left(\frac{x}{\lambda+\mu x} + \lambda - \mu x\right)\left(\frac{1}{\lambda+\mu x} + \mu - \frac{\lambda}{x}\right) K(x,\mu,\lambda)dx.$$

Seemingly, these integrals may not be evaluated in closed form. However the following transformations are useful for numerical evaluation. Let $\lambda/\sqrt{x} - \mu\sqrt{x} = t$. Then

$$I_{11}(\mu,\lambda) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left[\frac{2\lambda\mu + t^2 + t\sqrt{t^2+4\lambda\mu}}{\mu(4\lambda\mu + t^2 + t\sqrt{t^2+4\lambda\mu})} + \lambda\right.$$

$$\left. - \quad \frac{2\lambda\mu + t^2 - t\sqrt{t^2+4\lambda\mu}}{2\mu}\right]^2 \mathrm{e}^{-\frac{t^2}{2}}dt,$$

$$I_{22}(\mu,\lambda) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left[\frac{2\mu}{\mu(4\lambda\mu + t^2 + t\sqrt{t^2+4\lambda\mu})} + \mu\right.$$

$$\left. - \quad \frac{2\lambda\mu^2}{2\lambda\mu + t^2 + t\sqrt{t^2+4\lambda\mu}}\right]^2 \mathrm{e}^{-\frac{t^2}{2}}dt.$$

The asymptotic MSE of MLE can be computed by solving these integrals. The computation is done by Simpson formula

$$\int_{-A}^{A} f(x)dx = \frac{h}{6}\sum_{i=0}^{m-1}\left(f(x_i) + 4f(x_{i+0.5}) + f(x_{i+1})\right),$$

where $x_0,\, x_1, \ldots, x_m$ are the points of division of an interval $[-A,\, A]$, and $x_{i+0.5} = (x_{i+1} + x_i)/2$.

The cancelation of tails in the integrals with infinite limits, that is, the choice of A, is carried out by the method of sequential approximations. We compute sequentially the values of integrals for the points $A = 10, 11, \ldots$ up to the point A where the changes in the values of the integrals become less than 10^{-3}. This gives us sufficient accuracy. Furthermore, a simulation study is conducted to appraise the properties of MLE in a practical setting. Tables 3.1-3.3 and 3.10 present the value of the biases and asymptotic and simulated MSE of the estimators.

MLE is of course a consistent estimator of a parameter, cf. Lehman (1983), Chapter 6, Sections 1-6, p.427-435. Moreover, it is asymptotically ($n \to \infty$) efficient, but since seemingly the integrals from Fisher's Information matrix may not be evaluated in closed form, we cannot produce the lower bound from the Rao-Cramér inequality. One practice one often uses the estimated Fisher's Information matrix, so this is only a theoretical problem.

In the following section we consider the method of moment estimation for the parameters of interest.

3.3 The Method of Moments

In this section we will present one of the oldest methods for deriving the point estimators of distribution parameters, the method of moments. However, this method is based on the assumption that the sample moments should provide good estimates of the corresponding population moments. Then because the population moments will be functions of population parameters, we will equate corresponding population and sample moments and solve for the desired parameters. To do so, consider the following statistics of BS- distribution for the estimation of the values of μ and λ by the moment method:

$$T_1 = \frac{1}{n}\sum_{k=1}^{n} X_k, \quad T_2 = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{X_k}, \quad T = T_1 T_2. \tag{3.5}$$

Birnbaum and Saunders (1969b) found the expected values and MSE of these statistics, which can be used to rewrite the expected values and MSE of the statistics in terms of μ and λ with the help of relations between the usual and the new parameters (cf. Section 3.1, (3.3)):

$$E_1 = E(T_1) = \frac{\lambda\mu + 1/2}{\mu^2}, \quad E_2 = E(T_2) = \frac{\lambda\mu + 1/2}{\lambda^2},$$

$$Var(T_1) = \frac{\lambda\mu + 5/4}{n\mu^4}, Var(T_2) = \frac{\lambda\mu + 5/4}{n\lambda^4},$$

$$E(T) = \frac{1}{n} + \frac{n-1}{n}\left(\frac{\lambda\mu + 1/2}{\lambda\mu}\right)^2. \tag{3.6}$$

Hence, by equating these expectations and their sample values, we find the moment method estimators (MME)

$$\hat{\mu}_n^{(MME)} = \sqrt{\frac{\sqrt{T}}{2T_1(\sqrt{T}-1)}}, \quad \hat{\lambda}_n^{(MME)} = \sqrt{\frac{T_1}{2\sqrt{T}(\sqrt{T}-1)}}.$$

The following theorem gives the asymptotic MSE of MME.

Theorem 3.1. *The MSE of* $\hat{\mu}_n^{(MME)}$ *and* $\hat{\lambda}_n^{(MME)}$ *are given by*

$$MSE(\hat{\mu}_n^{(MME)}) = \frac{(\lambda\mu + 5/4)\left(\mu^2(\lambda\mu+1)^2 + \lambda^2\mu^4\right)}{n(2\lambda\mu+1)^2} - \frac{2\lambda\mu^3(\lambda\mu+1/4)(\lambda\mu+1)}{n(2\lambda\mu+1)^2}$$

$$+O\left(\frac{1}{n^2}\right),$$

$$MSE(\hat{\lambda}_n^{(MME)}) = \frac{(\lambda\mu + 5/4)\left(\lambda^2(\lambda\mu+1)^2 + \lambda^4\mu^2\right)}{n(2\lambda\mu+1)^2} - \frac{2\lambda^3\mu(\lambda\mu+1/4)(\lambda\mu+1)}{n(2\lambda\mu+1)^2}$$

$$+O\left(\frac{1}{n^2}\right) \tag{3.7}$$

as $n \to \infty$.

Proof. We follow the asymptotic formula derivation of moments of a sample moments function, the method outlined in Cramér (1946), §27.7, p.352-358. To this end, we introduce functions $M(T_1,\, T_2) = \hat{\mu}_n$ and $L(T_1, T_2) = \hat{\lambda}_n$. Now, rewrite the functions M and L in the following canonical forms

$$M(x,y) = 2^{-1/2} x^{-1/4} y^{1/4} \left(x^{1/2} y^{1/2} - 1\right)^{1/2},$$

$$L(x,y) = 2^{-1/2} x^{1/4} y^{-1/4} \left(x^{1/2} y^{1/2} - 1\right)^{1/2}.$$

Since the statistics T_1 and T_2 will have all moments, the following asymptotic representation can be written

$$\begin{aligned}
&Var(\hat{\mu}_n^{(MME)}) \\
&= Var(T_1)\left(\frac{\partial M(E_1,E_2)}{\partial x}\right)^2 + 2cov(T_1,T_2)\frac{\partial M(E_1,E_2)}{\partial x}\frac{\partial M(E_1,E_2)}{\partial y} \\
&+ Var(T_2)\left(\frac{\partial M(E_1,E_2)}{\partial y}\right)^2 + O(n^{-2}), \\
&Var(\hat{\lambda}_n^{(MME)}) \\
&= Var(T_1)\left(\frac{\partial L(E_1,E_2)}{\partial x}\right)^2 + 2cov(T_1,T_2)\frac{\partial L(E_1,E_2)}{\partial x}\frac{\partial L(E_1,E_2)}{\partial y} \\
&+ Var(T_2)\left(\frac{\partial L(E_1,E_2)}{\partial y}\right)^2 + O(n^{-2}).
\end{aligned}$$

Direct evaluation of derivatives gives that

$$\frac{\partial M(E_1,\, E_2)}{\partial x} = \frac{\mu^3(\lambda\mu+1)}{2\lambda\mu+1}, \quad \frac{\partial M(E_1,\, E_2)}{\partial y} = \frac{\lambda^3\mu^2}{2\lambda\mu+1},$$

$$\frac{\partial L(E_1,\, E_2)}{\partial x} = \frac{\lambda^2\mu^3}{2\lambda\mu+1}, \quad \frac{\partial L(E_1,\, E_2)}{\partial y} = \frac{\lambda^3(\lambda\mu+1)}{2\lambda\mu+1}.$$

Finally, using formulas (3.6) for the moments of statistics T_1, T_2 and T we obtain the required asymptotic expansions.

The asymptotic MSE of MME can be computed by the formulas from Theorem 3.1. Further, a simulation study is performed to obtain simulated MSE of the estimators for given sample size. The results are reported in Tables 3.4 – 3.6 and 3.11.

An MME is of course a consistent estimator of a parameter. This follows from the law of large numbers and since the estimator as a function of sample moments has locally continuous inverse.

In the following section we consider another method for the estimation of parameters of the BS-distribution, the so-called regression-quantile or least square method.

3.4 The Regression - Quantile (Least Square) Method

The regression - quantile method is based on the minimization of the quadratic measure of the difference between the empirical distribution function $F_n(x)$ and the theoretical cumulative distribution function

$$F(x) = 1 - \Phi(\lambda/\sqrt{x} - \mu\sqrt{x}).$$

If $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ are order statistics of $X_1, X_2, \ldots, X_n$, then by definition the empirical distribution function is given by $F_n(X_{(k)}) = k/n,\ k = 1, \ldots, n.$

Now, we consider the following asymptotic equality:

$$\Phi^{-1}\left(1 - \frac{k}{n}\right) \approx \frac{\lambda}{\sqrt{X_{(k)}}} - \mu\sqrt{X_{(k)}}, \quad k = 1, \ldots, n-1,$$

which can be used for the parameter estimation. Hence, estimations of parameters are obtained by finding the minimum of the function

$$G(\lambda,\, \mu) = \sum_{k=1}^{n}\left(\frac{\lambda}{\sqrt{X_{(k)}}} - \mu\sqrt{X_{(k)}} - t_k\right)^2,$$

where $t_k = \Phi^{-1}(1 - k/n)$ for $k = 1, \ldots, n-1$. Since $\Phi^{-1}(\,0\,) = -\infty,\ t_n$ is chosen by the condition of further minimization of the function G. It is interesting to note that the results of statistical simulations of estimators show that the optimal choice of t_n is close to $t_{n-1} - 1$ for nearly all μ, λ and n.

We can rewrite statistics T_1 and T_2 in (3.5) in the following form:

$$T_1 = \frac{1}{n}\sum_{k=1}^{n} X_{(k)}, \quad T_2 = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{X_{(k)}}.$$

Further,

$$T_3 = \frac{1}{n}\sum_{k=1}^{n} t_k\sqrt{X_{(k)}}, \quad T_4 = \frac{1}{n}\sum_{k=1}^{n} \frac{t_k}{\sqrt{X_{(k)}}}$$

Hence, the regression - quantile estimations (RQE), are written as

$$\tilde{\mu}_n = \frac{T_2T_3 - T_4}{1 - T_1T_2}, \qquad \tilde{\lambda}_n = \frac{T_3 - T_1T_4}{1 - T_1T_2}.$$

Seemingly, it is not possible to evaluate asymptotic MSE of these estimators in closed form. The evaluation of asymptotic MSE as $n \to \infty$ seems to be mathematically intractable. Thus, we confine our investigation of the behavior of the MSE to Monte Carlo simulations.

RQE is a consistent estimator of a parameter. This follows from the Glivenko-Cantelli theorem (cf., for example, Stout (1974), Chapter 3, Section 3.2, p.124) which states that an empirical distribution function is a strongly (even uniformly) consistent estimator of a true distribution function.

In the next section, we consider numerical methods to appraise the performance of the proposed estimators.

3.5 Some Computed Analysis

In this section we will study the statistical properties of the proposed estimators by numerical methods. For MLE and MME we consider direct computations of asymptotic MSE formulas and by Monte Carlo simulations. On the other hand, the behavior of the RQE is investigated only via a simulation study.

Because the iteration process may diverge, for a solution of the system of maximum likelihood equations (cf. Section 3.2, (3.4)) in the region $\mu > 0$, $\lambda > 0$ we choose a rectangle, which is divided into 100 congruent rectangles, and we find a point $(\mu., \lambda.)$, for which the sum of squares of differences of the left hand and right hand sides of the equations (3.4) of the maximum likelihood system is obtained. The point $(\mu., \lambda.)$ is surrounded by a rectangle of smaller size, which is also divided into 100 parts and the process is repeated until the required accuracy 10^{-3} for MLE is achieved.

On the other hand, the asymptotic MSE of MME is computed by the formulas (3.7) obtained in Theorem 3.1.

Some computed biases and MSE of the MLE and MME are reported in Tables 1,2,3,10 and Tables 4,5,6,11, respectively.

In an effort to calculate the simulated bias and MSE of all proposed estimators we need a sample from the BS - distribution. First we generate a sample $Y_1, \ldots, Y_n$, of given size n from the standard normal distribution. If Y has the standard normal distribution, then the root

$$X = \left(\frac{-Y + \sqrt{Y^2 + 4\mu\lambda}}{2\mu} \right)^2$$

of the equation $\lambda/\sqrt{X} = Y + \mu\sqrt{X}$ will have the BS - distribution with parameters μ and λ. Mention that the second root $X = \left(\frac{-Y - \sqrt{Y^2+4\mu\lambda}}{2\mu} \right)^2$ of the equation will not produce the BS -distribution. A sample, $X_1, \ldots, X_n$, is thus obtained from $Y_1, \ldots, Y_n$. After we obtain 1000 replications of each estimator we compute Monte Carlo estimations of bias of the estimators and their MSE with the help of standard statistics, respectively: average of differences between the true value of the parameters and the values of simulation and an average of squares of these differences. For example, if we obtained 1000 replications of estimator of λ, that is, $\lambda_1^*, \lambda_2^*, \ldots, \lambda_{1000}^*$, we will estimate the bias as

$$\frac{1}{1000}\sum_{i=1}^{1000}(\lambda_i^* - \lambda)$$

and the MSE as

$$\frac{1}{1000}\sum_{i=1}^{1000}(\lambda_i^* - \lambda)^2.$$

The simulated bias and MSE of all estimators are calculated at selected values of μ and λ and given sample size n. The results are reported in Tables 3.1-3.12.

Finally, we present the analysis of bias and MSE of the estimators based on the preceding numerical computations. It is clear from Tables 3.1, 3.2, 3.3 and 3.10 that the MLE has a systematic positive bias (underestimation of a value of the parameter), when the true value of at least one parameter is sufficiently large and the bias is an increasing function of values of the parameters. For example, for $n = 50$ and $\lambda = 5, \mu = 10$ the bias of λ for MLE is 0.003, while for $\lambda = 50, \mu = 50$ the bias of λ is 0.084.

Tables 3.4, 3.5, 3.6 and 3.11 reveal that the MME has a systematic negative bias (overestimation of a value of the parameter). The absolute value of bias is also an increasing function of values of parameters. For example, for $n = 50$ and $\lambda = 5, \mu = 10$ the bias of λ for MME is -0.129, while for $\lambda = 50, \mu = 50$ the bias of λ is -1.277.

We observed from the Tables 3.7, 3.8, 3.9 and 3.12 that the RQE has a positive bias for the parameter λ and a negative bias for μ when the values of the parameters are small. For example, for $n = 50$ and $\lambda = 0.5, \mu = 0.5$ the bias of λ is 0.009 and the bias for μ is -0.047. For larger values of parameters a systematic overestimation of true values of parameters takes place (the estimators have a negative bias). Further, the larger the values of the parameters the larger is the value of the bias. For example, for $n = 50$ and $\lambda = 5, \mu = 10$ the bias of λ is -0.117 and the bias for μ is -0.318.

It can be safely argued that the bias of the MME and RQE is higher than that for MLE. However, for $n = 500$ the simulation study shows an even larger bias for the MLE. If $\lambda = 100, \mu = 100$ the bias of λ for MLE is 0.961, for MME is -0.211 and for RQE is -0.314. This phenomenon requires additional investigation. But maybe this results from random errors common to the method of statistical simulations.

Since all three estimators are consistent and all moments exist, we can state that the bias tends to zero when the sample size $n \to \infty$.

The asymptotic MSE and simulated MSE of the MLE are given in Tables 3.1, 3.2, 3.3 and 3.10. The tables reveal that the values of the MSE increase with increasing values of the parameters. For example, for $n = 50$ and $\lambda = 0.5, \mu = 0.5$ (Table 3.2) the asymptotic MSE and the simulated MSE of λ are 0.005 and 0.002, respectively. On the other hand, for $\lambda = 50, \mu = 50$ the MSE are 0.020 and 0.232. The simulated MSE behavior is more or less similar for different sample sizes. For example, it can be seen from Table 3.2 that when $n = 50$ and $\lambda = 0.5, \mu = 0.5$ the simulated MSE of λ are 0.005 and 0.002, respectively, while for $\lambda = 50, \mu = 0.5$ the MSE are 1.942 and 0.226. Seemingly, the asymptotic MSE and the simulated MSE are comparable, except in a few instances. A difference between the asymptotic MSE and the simulated MSE is small for proportional values of μ and λ; hence our analytical work is well supported by computational analysis.

The MSE is large when the parameters are disproportional, that is, one parameter is much bigger than the other. For example, for $n = 50$ and $\lambda = 0.5, \mu = 0.5$ the MSE of λ for MLE is 0.005, while for $\lambda = 50, \mu = 0.5$ the MSE of λ is 1.942.

Tables 3.4, 3.5, 3.6,3.11 and Tables 3.7, 3.8, 3.9 and 3.12 provide more or less similar MSE analyses for the MME and RQE to that of the MLE. For all estimators, the simulated MSE tends to zero as the sample size $n \to \infty$. Moreover, for the RQE, the MSE varies inversely as sample size n, that is, for a suitable choice of $C = C(\mu, \lambda)$ the value C/n for sufficiently large sample size n satisfactorily approximates the values of the variance. Figures 1 and 2 show simulated MSE for all three estimators as functions of the sample size when the values of the parameters are $\mu = \lambda = 5$.

It is noted that the asymptotic MSE of the MME is much higher than the MSE of the MLE, particularly for large values of the parameters. For example, for $n = 50, \lambda = 50$, and $\mu = 50$ the asymptotic MSE for the MLE of λ is 0.020, while for MME it is 25.005. From our point of view, the most interesting phenomenon is the practical equivalence of the simulated MSE and the asymptotic MSE of the MME and RQE. The latter has a little bit higher variance, but this is such a small amount that it can be explained by random errors common to statistical simulation methods. Figures 1 and 2 show the simulated MSE for all three estimators as functions of the sample size when the values of the parameters are $\mu = \lambda = 5$.

Figure 3.1: Plot of MSE for MLE, MME, and RQE of λ.

Figure 3.2: Plot of MSE for MLE, MME, and RQE of μ.

Table 3.1: Bias and MSE for Maximum Likelihood Estimators, $n = 10$

λ	μ	Simul. bias		MSE			
		λ	μ	Asympt. MSE λ	Simul. MSE λ	Asympt. MSE μ	Simul. MSE μ
0.5	0.5	-0.008	-0.010	0.023	0.002	0.023	0.002
0.5	1	-0.007	+0.007	0.019	0.002	0.076	0.010
0.5	5	-0.005	-0.002	0.008	0.002	0.765	0.232
0.5	10	+0.000	+0.012	0.004	0.000	1.736	0.011
0.5	50	+0.001	+0.022	0.000	0.000	9.708	0.248
1	0.5	+0.008	-0.008	0.076	0.001	0.019	0.002
1	1	-0.003	-0.004	0.056	0.009	0.056	0.009
1	5	-0.013	-0.025	0.017	0.007	0.434	0.197
1	10	+0.001	+0.013	0.009	0.000	0.930	0.011
1	50	+0.001	-0.006	0.002	0.000	4.926	0.239
5	0.5	-0.001	-0.006	0.766	0.229	0.008	0.002
5	1	-0.032	-0.011	0.434	0.197	0.017	0.007
5	5	-0.094	-0.090	0.097	0.192	0.097	0.188
5	10	+0.003	+0.009	0.049	0.003	0.197	0.011
5	50	+0.006	+0.043	0.010	0.002	0.997	0.228
10	0.5	+0.012	+0.000	1.736	0.011	0.004	0.000
10	1	+0.015	+0.000	0.930	0.011	0.009	0.000
10	5	+0.007	+0.004	0.197	0.011	0.049	0.003
10	10	+0.005	+0.006	0.099	0.010	0.099	0.010
10	50	+0.009	+0.046	0.020	0.007	0.499	0.206
50	0.5	-0.217	+0.001	9.708	0.243	0.001	0.000
50	1	+0.002	+0.001	4.926	0.242	0.002	0.000
50	5	+0.036	+0.004	0.997	0.230	0.010	0.002
50	10	+0.049	+0.012	0.499	0.204	0.020	0.007
50	50	+0.051	+0.054	0.100	0.194	0.100	0.189

Table 3.2: Bias and MSE for Maximum Likelihood Estimators, $n = 50$

λ	μ	Simul. bias		MSE			
		λ	μ	Asympt. MSE λ	Simul. MSE λ	Asympt. MSE μ	Simul. MSE μ
0.5	0.5	-0.003	-0.003	0.005	0.002	0.005	0.002
0.5	1	-0.003	+0.001	0.004	0.002	0.015	0.007
0.5	5	-0.003	-0.002	0.002	0.001	0.153	0.125
0.5	10	+0.000	+0.014	0.001	0.000	0.347	0.010
0.5	50	+0.000	+0.023	0.000	0.000	1.942	0.225
1	0.5	-0.000	-0.003	0.015	0.007	0.004	0.002
1	1	-0.004	-0.003	0.011	0.006	0.011	0.006
1	5	-0.006	-0.017	0.004	0.005	0.087	0.109
1	10	+0.001	+0.011	0.002	0.000	0.186	0.010
1	50	+0.001	+0.036	0.000	0.000	0.985	0.214
5	0.5	-0.012	-0.003	0.153	0.126	0.002	0.001
5	1	-0.025	-0.006	0.087	0.114	0.003	0.005
5	5	-0.044	-0.042	0.019	0.141	0.019	0.137
5	10	+0.003	+0.008	0.011	0.002	0.039	0.009
5	50	+0.005	+0.059	0.002	0.002	0.199	0.184
10	0.5	+0.018	+0.000	0.010	0.347	0.001	0.000
10	1	+0.015	+0.001	0.186	0.010	0.002	0.000
10	5	+0.006	+0.003	0.039	0.009	0.010	0.002
10	10	+0.005	+0.006	0.020	0.008	0.020	0.008
10	50	+0.010	+0.055	0.004	0.007	0.100	0.187
50	0.5	+0.018	+0.000	1.942	0.226	0.000	0.000
50	1	+0.015	+0.001	0.985	0.221	0.000	0.000
50	5	+0.034	+0.004	0.199	0.175	0.002	0.002
50	10	+0.054	+0.011	0.100	0.180	0.004	0.007
50	50	+0.084	+0.085	0.020	0.232	0.020	0.227

Table 3.3: Bias and MSE for Maximum Likelihood Estimators, $n = 100$

λ	μ	Simul. bias		MSE			
		λ	μ	Asympt. MSE λ	Simul. MSE λMSE μ	Asympt. MSE μ	Simul.
0.5	0.5	-0.001	-0.001	0.002	0.001	0.002	0.001
0.5	1	-0.002	-0.000	0.002	0.001	0.008	0.005
0.5	5	-0.002	-0.006	0.001	0.001	0.077	0.092
0.5	10	+0.000	+0.017	0.000	0.000	0.174	0.010
0.5	50	+0.000	+0.055	0.000	0.000	0.971	0.022
1	0.5	-0.002	-0.001	0.008	0.005	0.002	0.001
1	1	-0.003	-0.002	0.006	0.005	0.006	0.005
1	5	-0.003	-0.006	0.002	0.003	0.043	0.083
1	10	+0.001	+0.014	0.001	0.000	0.093	0.009
1	50	+0.001	+0.034	0.000	0.000	0.493	0.191
5	0.5	-0.015	-0.002	0.076	0.091	0.001	0.001
5	1	-0.011	-0.002	0.043	0.083	0.002	0.003
5	5	-0.034	-0.032	0.010	0.108	0.010	0.110
5	10	+0.004	+0.010	0.005	0.002	0.020	0.008
5	50	+0.005	+0.054	0.001	0.002	0.100	0.184
10	0.5	+0.011	+0.001	0.174	0.975	0.000	0.000
10	1	+0.010	+0.001	0.093	0.009	0.001	0.000
10	5	+0.007	+0.004	0.020	0.008	0.005	0.002
10	10	+0.009	+0.007	0.010	0.007	0.010	0.008
10	50	+0.014	+0.075	0.002	0.007	0.050	0.189
50	0.5	+0.025	+0.001	0.971	0.210	0.000	0.000
50	1	+0.021	+0.001	0.493	0.191	0.000	0.000
50	5	+0.056	+0.050	0.010	0.175	0.001	0.002
50	10	+0.066	+0.014	0.050	0.187	0.002	0.007
50	50	+0.076	+0.078	0.010	0.235	0.010	0.240

Table 3.4: Bias and MSE for Moment Method Estimators, $n = 10$

λ	μ	Simul. bias		MSE			
		λ	μ	Asympt. MSE λ	Simul. MSE λ	Asympt. MSE μ	Simul. MSE μ
0.5	0.5	-0.107	-0.105	0.024	0.073	0.024	0.067
0.5	1	-0.945	-0.186	0.020	0.053	0.081	0.196
0.5	5	-0.079	-0.777	0.015	0.033	1.476	3.203
0.5	10	-0.076	-1.503	0.014	0.031	11.901	5.475
0.5	50	-0.073	-7.299	0.013	0.028	127.475	279.978
1	0.5	-0.189	-0.093	0.081	0.211	0.020	0.049
1	1	-0.172	-0.169	0.069	0.164	0.069	0.155
1	5	-0.152	-0.751	0.055	0.123	1.369	2.975
1	10	-0.149	-1.476	0.052	0.117	5.244	11.454
1	50	-0.146	-7.274	0.059	0.112	126.244	278.042
5	0.5	-0.788	-0.078	1.476	3.340	0.015	0.032
5	1	-0.760	-0.150	1.369	3.069	0.055	0.119
5	5	-0.734	-0.730	1.275	2.840	1.275	2.780
5	10	-0.730	-1.454	1.262	2.809	5.050	11.122
5	50	-0.727	-7.257	1.252	2.781	125.250	276.804
10	0.5	-1.519	-0.075	5.475	12.277	0.014	0.030
10	1	-1.488	-0.148	5.244	11.715	0.052	0.114
10	5	-1.460	-0.727	5.050	11.236	1.262	2.779
10	10	-1.457	-1.452	5.025	11.168	5.025	11.088
10	50	-1.453	-7.255	5.006	11.106	125.125	276.737
50	0.5	-7.339	-0.073	127.480	284.042	0.013	0.028
50	1	-7.303	-0.145	126.243	280.903	0.051	0.111
50	5	-7.269	-0.726	125.250	278.079	1.252	2.768
50	10	-7.265	-1.451	125.125	277.638	5.005	11.070
50	50	-7.259	-7.255	125.025	277.186	125.025	276.783

Table 3.5: Bias and MSE for Moment Method Estimators, $n = 50$

λ	μ	Simul. bias		MSE			
		λ	μ	Asympt. MSE λ	Simul. MSE λ	Asympt. MSE μ	Simul. MSE μ
0.5	0.5	-0.018	-0.016	0.005	0.006	0.005	0.006
0.5	1	-0.016	-0.030	0.004	0.005	0.016	0.019
0.5	5	-0.014	-0.134	0.003	0.004	0.295	0.348
0.5	10	-0.014	-0.261	0.003	0.003	1.095	1.299
0.5	50	-0.013	-1.278	0.003	0.003	25.495	30.685
1	0.5	-0.033	-0.015	0.016	0.021	0.004	0.005
1	1	-0.030	-0.028	0.014	0.017	0.014	0.016
1	5	-0.027	-0.131	0.011	0.014	0.274	0.325
1	10	-0.026	-0.258	0.011	0.013	1.049	1.254
1	50	-0.026	-1.276	0.010	0.012	25.250	30.491
5	0.5	-0.140	-0.013	0.295	0.367	0.003	0.003
5	1	-0.135	-0.261	0.274	0.339	0.011	0.013
5	5	-0.130	-0.128	0.255	0.313	0.255	0.307
5	10	-0.129	-0.255	0.252	0.309	1.010	1.220
5	50	-0.128	-1.275	0.250	0.306	25.050	30.386
10	0.5	-0.269	-0.013	1.095	1.356	0.003	0.003
10	1	-0.263	-0.026	1.049	1.294	0.011	0.013
10	5	-0.258	-0.128	1.010	1.238	0.253	0.305
10	10	-0.257	-0.255	1.005	1.229	1.005	1.217
10	50	-0.256	-1.275	1.001	1.221	25.025	30.387
50	0.5	-1.296	-0.013	25.495	31.324	0.003	0.003
50	1	-1.288	-0.026	25.249	30.945	0.010	0.012
50	5	-1.279	-0.127	25.049	30.589	0.249	0.304
50	10	-1.279	-0.255	25.025	30.531	1.001	1.215
50	50	-1.277	-1.276	25.005	30.468	25.005	30.404

Table 3.6: Bias and MSE for Moment Method Estimators, $n = 100$

λ	μ	Simul. bias		MSE			
		λ	μ	Asympt. MSE λ	Simul. MSE λ	Asympt. MSE μ	Simul. MSE μ
0.5	0.5	-0.007	-0.006	0.002	0.003	0.002	0.003
0.5	1	-0.007	-0.012	0.002	0.002	0.008	0.009
0.5	5	-0.005	-0.049	0.001	0.002	0.148	0.163
0.5	10	-0.005	-0.095	0.001	0.002	0.548	0.604
0.5	50	-0.005	-0.465	0.001	0.001	12.748	14.133
1	0.5	-0.013	-0.006	0.008	0.009	0.002	0.002
1	1	-0.012	-0.011	0.007	0.008	0.007	0.008
1	5	-0.010	-0.047	0.005	0.006	0.137	0.151
1	10	-0.010	-0.093	0.005	0.006	0.524	0.579
1	50	-0.010	-0.465	0.005	0.006	12.624	14.018
5	0.5	-0.055	-0.005	0.148	0.168	0.001	0.002
5	1	-0.052	-0.009	0.137	0.155	0.005	0.006
5	5	-0.049	-0.047	0.127	0.143	0.127	0.141
5	10	-0.048	-0.093	0.126	0.141	0.505	0.561
5	50	-0.047	-0.467	0.125	0.140	12.525	13.939
10	0.5	-0.103	-0.005	0.548	0.620	0.001	0.002
10	1	-0.100	-0.009	0.524	0.591	0.005	0.006
10	5	-0.096	-0.047	0.505	0.566	0.126	0.139
10	10	-0.095	-0.093	0.502	0.562	0.502	0.559
10	50	-0.094	-0.467	0.499	0.559	12.512	13.933
50	0.5	-0.485	-0.005	12.748	14.312	0.001	0.001
50	1	-0.479	-0.009	12.624	14.145	0.005	0.006
50	5	-0.473	-0.047	12.525	13.996	0.139	0.125
50	10	-0.472	-0.093	12.512	13.974	0.499	0.557
50	50	-0.470	-0.468	12.502	13.951	12.502	13.933

Table 3.7: Bias and MSE for Regression - Quantile Estimators, $n = 10$

		Bias		MSE	
λ	μ	λ	μ	λ	μ
0.5	0.5	+0.004	-0.200	0.048	0.115
0.5	1	-0.007	-0.319	0.038	0.296
0.5	5	-0.029	-1.009	0.029	3.885
0.5	10	-0.037	-1.750	0.028	13.599
0.5	50	-0.048	-7.102	0.027	298.306
1	0.5	-0.013	-0.159	0.152	0.074
1	1	-0.034	-0.257	0.130	0.210
1	5	-0.074	-0.875	0.111	3.400
1	10	-0.086	-1.572	0.109	12.619
1	50	-0.103	-6.741	0.110	291.790
5	0.5	-0.295	-0.101	2.880	0.039
5	1	-0.369	-0.175	2.768	0.136
5	5	-0.483	-0.710	2.726	2.983
5	10	-0.514	-1.348	2.738	11.672
5	50	-0.555	-6.264	2.764	284.279
10	0.5	-0.738	-0.088	11.074	0.034
10	1	-0.855	-0.157	10.925	0.126
10	5	-1.027	-0.674	10.954	2.918
10	10	-1.069	-1.296	10.997	11.497
10	50	-1.129	-6.151	11.095	282.945
50	0.5	-4.833	-0.071	272.615	0.030
50	1	-5.136	-0.135	273.840	0.117
50	5	-5.547	-0.626	276.407	2.843
50	10	-5.644	-1.230	277.385	11.318
50	50	-5.782	-6.009	278.701	281.186

Table 3.8: Bias and MSE for Regression - Quantile Estimators, $n = 50$

		Bias		MSE	
λ	μ	λ	μ	λ	μ
0.5	0.5	+0.009	-0.047	0.005	0.009
0.5	1	+0.004	-0.078	0.005	0.027
0.5	5	-0.006	-0.249	0.003	0.401
0.5	10	-0.007	-0.420	0.003	1.443
0.5	50	-0.011	-1.735	0.003	31.649
1	0.5	+0.008	-0.039	0.018	0.007
1	1	-0.002	-0.063	0.015	0.021
1	5	-0.016	-0.215	0.012	0.357
1	10	-0.018	-0.375	0.013	1.361
1	50	-0.024	-1.641	0.012	31.006
5	0.5	-0.055	-0.025	0.329	0.004
5	1	-0.080	-0.043	0.310	0.014
5	5	-0.114	-0.173	0.279	0.316
5	10	-0.117	-0.318	0.315	1.289
5	50	-0.133	-1.519	0.297	30.296
10	0.5	-0.160	-0.022	1.240	0.004
10	1	-0.195	-0.039	1.203	0.013
10	5	-0.233	-0.159	1.259	0.322
10	10	-0.256	-0.315	1.186	1.225
10	50	-0.271	-1.490	1.190	30.171
50	0.5	-1.136	-0.017	29.668	0.003
50	1	-1.219	-0.033	29.617	0.012
50	5	-1.278	-0.147	31.426	0.317
50	10	-1.357	-0.298	29.739	1.207
50	50	-1.393	-1.453	29.830	30.023

Table 3.9: Bias and MSE for Regression - Quantile Estimators, $n = 100$

		Bias		MSE	
λ	μ	λ	μ	λ	μ
0.5	0.5	+0.008	-0.024	0.003	0.003
0.5	1	+0.004	-0.038	0.002	0.011
0.5	5	-0.001	-0.115	0.002	0.167
0.5	10	-0.002	-0.195	0.001	0.603
0.5	50	-0.004	-0.749	0.001	13.642
1	0.5	+0.009	-0.019	0.009	0.003
1	1	+0.003	-0.030	0.008	0.009
1	5	-0.005	-0.097	0.006	0.151
1	10	-0.007	-0.170	0.005	0.568
1	50	-0.010	-0.700	0.005	13.441
5	0.5	-0.010	-0.011	0.166	0.002
5	1	-0.024	-0.019	0.154	0.006
5	5	-0.043	-0.075	0.133	0.136
5	10	-0.048	-0.140	0.132	0.538
5	50	-0.054	-0.636	0.131	13.254
10	0.5	-0.049	-0.010	0.614	0.002
10	1	-0.069	-0.017	0.545	0.006
10	5	-0.095	-0.070	0.528	0.134
10	10	-0.102	-0.133	0.526	0.533
10	50	-0.110	-0.621	0.526	13.224
50	0.5	-0.425	-0.007	14.343	0.001
50	1	-0.477	-0.014	13.200	0.005
50	5	-0.537	-0.064	13.148	0.133
50	10	-0.551	-0.124	13.149	0.529
50	50	-0.562	-0.593	14.132	14.152

Table 3.10: Bias and MSE for Maximum Likelihood Estimators, $n = 500$

λ	μ	Simul. bias		MSE			
		λ	μ	Asympt. MSE λ	Simul. MSE λ	Asympt. MSE μ	Simul. MSE μ
100	100	+0.961	+0.961	0.002	1.154	0.002	1.157

Table 3.11: Bias and MSE for Moment Method Estimators, $n = 500$

λ	μ	Simul. bias		MSE			
		λ	μ	Asympt. MSE λ	Simul. MSE λ	Asympt. MSE μ	Simul. MSE μ
100	100	-0.211	-0.210	10.001	10.358	10.001	10.348

Table 3.12: Bias and MSE for Regression - Quantile Estimators, $n = 500$

		Bias		MSE	
λ	μ	λ	μ	λ	μ
100	100	-0.314	-0.321	10.451	10.447

Chapter 4

Weighted Likelihood Method in Robust Estimation

4.1 Introduction

Let $\{F(x;\theta);\ \ \theta \in \Theta\}$ be a parametric family of cumulative distribution functions with corresponding probability density functions $f(x;\theta)$.

Any estimate $\hat{\theta}_n$, defined by a maximum problem of the form

$$\sum_{i=1}^{n} \rho(X_i;\,\theta) = \max$$

or by an implicit equation

$$\sum_{i=1}^{n} \psi(X_i;\,\theta) = 0,$$

where ρ is an arbitrary differentiable function and

$$\psi(x;\,\theta) = \frac{\partial \rho(x;\,\theta)}{\partial \theta},$$

is called an M-estimate (or maximum likelihood type estimate, cf. chapter 3 of Huber (1981)). Of course ρ need not be differentiable (eg. $\rho(x) = |x|$). Note that the choice $\rho(x;\,\theta) = \ln f(x;\,\theta)$, where f is the density function of the distribution from which the sample $X^{(n)} = (X_1, \ldots, X_n)$ is taken, gives the ordinary maximum likelihood estimate. In this chapter we consider i.i.d. sampling.

If we consider an estimation method based on the maximizing of *weighted likelihood function*

$$L(\theta \mid X^{(n)}) = \sum_{i=1}^{n} t_i \ln f(X_i;\, \theta),$$

where t_i depends on the sample: $t_i = t_i(X^{(n)})$, then we obtain a more general notion of M-estimate defined by a solution of the equation

$$\sum_{i=1}^{n} t_i \psi(X_i;\, \theta) = 0.$$

We suggest a special choice of weights t_i, connected with the theory of robust estimation and based on the maximum likelihood method with rejection of spurious observations. Let $\hat{\theta}_n = \hat{\theta}_n(X^{(n)})$ be the usual maximum likelihood estimation of parameter θ. The weight t_i that corresponds to the observation X_i is assumed to be 1, if its estimated likelihood is sufficiently large, and 0 elsewhere, that is, (the choice of C is discussed in the next paragraph)

$$t_i = \begin{cases} 1 & \text{if } f(X_i;\, \hat{\theta}_n) > C \\ 0 & \text{elsewhere.} \end{cases}$$

Consequently, we delete all *improbable* observations from the sample. Not surprisingly, it can be seen that in the case of a unimodal probability density function f we reject only extreme order statistics. However, this may not be the case for multimodal probability density functions.

We now come back to the issue of the choice of C. We suggest this not be considered as a constant. Rather assume that $C = a/\hat{\theta}_n$, where a can be treated as the same as the selection of the critical constant in the criterion of elimination of outliers.

The proposed estimator θ_n^* of the parameter θ is defined as the solution of the equation

$$\sum_{k=1}^{m} \frac{\partial \ln f(X_{i_k};\, \theta)}{\partial \theta} = 0,$$

where $X_{i_1}, \ldots, X_{i_m}$ are the remaining observations in the sample after the rejection procedure.

Define a set of *obstructing distributions* (F denotes the cumulative distribution function with density function f) as

$$\mathcal{G} = \{G_\varepsilon(x) = (1-\varepsilon)F(x;\, \theta) + \varepsilon F(x; \theta_1),\ 0 \le \varepsilon \le 1\}, \tag{4.1}$$

where $\theta_1 \in \Theta$ and $\theta_1 \neq \theta$. Under the assumption that the sample is taken from the distribution G_ε with fixed ε, we will find the limits in probability of the estimates $\hat{\theta}_n$ and θ_n^* for $n \to \infty$ and compare their biases. The values of θ for which $|\,\theta_n^* - \theta\,|$ will be less than $|\,\hat{\theta}_n - \theta\,|$, will tell us when to favour the estimate θ_n^* over the estimate $\hat{\theta}_n$. Finally, we will compare the quadratic risks of these estimates (in spirit of Huber (1981), (4.5) - (4.6), Chapter 1).

However, we need some additional assumptions on the *obstruction coefficient* ε. Assume that with the increasing of sample size n the coefficient of "obstruction" ε decreases as $\frac{1}{n}$, that is, $\varepsilon = \lambda/n$ for some $\lambda > 0$.

4.2 Proposed Weighted Likelihood Estimator

We consider an exponential distribution of the form

$$F(x;\, \theta) = 1 - \exp\{-x/\theta\}, \quad f(x;\, \theta) = \theta^{-1}\exp\{-x/\theta\}, \quad x \ge 0,\, \theta > 0.$$

Let $\theta > 0, \Delta > 0$, and $0 < \alpha < 1$ be any positive numbers. Later we will discuss the meaning and possible choice of α as a function of the sample size. However, here for the sake of brevity we assume α is a constant. Let us define $\theta_1 = \theta(1+\Delta)$. Assume that the obstructing parameter ε is a constant. Later on, when we will deal with asymptotic analysis we will assume that $\varepsilon = \frac{\lambda}{n}$. According to (4.1), the sample $(X_1, \cdots, X_n)$ is taken from the distribution of the form

$$G(x) = G_\varepsilon(x) = 1 - \exp\left\{\frac{-x}{\theta}\right\} - \varepsilon\left(\exp\left\{\frac{-x}{\theta(1+\Delta)}\right\} - \exp\left\{\frac{-x}{\theta}\right\}\right).$$

In this case

$$\hat{\theta}_n = \overline{X} = n^{-1}\sum_{i=1}^{n} X_i,$$

and, by the strong law of large numbers,

$$\hat{\theta}_n \to \theta_\varepsilon = E(\overline{X}) = ((1-\varepsilon)\theta + \varepsilon\theta(1+\Delta) = \theta(1+\varepsilon\Delta) \text{ almost surely.}$$

In an effort to consider θ_n^*, we assume that $C = a/\hat{\theta}_n$, where a can be treated as the same as in the selection of the critical constant in the criterion of elimination of outliers. Hence, a is chosen from the condition of a small probability of rejection of an observation when we choose from the *non obstructed* exponential distribution with cumulative distribution function $F(x;\theta)$, not $G(x)$. That is, a is defined by the given small probability α in the equation

$$P\left(\max_{1\le i\le n} X_i > -\hat{\theta}_n \ln a\right) \approx 1 - \prod_{i=1}^{n} P(X_i \le -\theta \ln a) = 1 - (1-a)^n = \alpha.$$

From this equation we obtain $a = 1 - (1-\alpha)^{1/n} \approx \alpha/n$. This approximation of a is reasonable since in our calculations we will replace $\hat{\theta}_n$ by the limit in probability (even almost surely) of this estimator, that is, by θ. So, we choose the value of C as

$$C = \frac{\alpha}{n\hat{\theta}_n}.$$

Hence, we reject an observation from the sample if $X_k > -\overline{X}\ln(C\overline{X}) = \overline{X}\ln(n/\alpha)$. The estimate θ_n^* converges in probability (even almost surely) to some value $\overline{\theta}_\varepsilon$ which can be calculated as the limit of the expected values truncated at the point $A = -\theta_\varepsilon \ln(C\theta_\varepsilon) = \theta_\varepsilon \ln(n/\alpha)$ of the distribution $\overline{G}(x) = G(x)/G(A),\ 0 < x \le A$. That is,

$$\overline{\theta}_\varepsilon = \lim_{n\to\infty} \frac{1}{G(A)} \int_0^A x \left[\frac{1-\varepsilon}{\theta} \exp\left\{-\frac{x}{\theta}\right\} + \frac{\varepsilon}{\theta(1+\Delta)} \exp\left\{-\frac{x}{\theta(1+\Delta)}\right\} \right] dx. \tag{4.2}$$

Indeed, these integrals can be evaluated in closed form. However the solution is cumbersome. Thus, we cannot easily compare the relative bias $(\overline{\theta}_\varepsilon - \theta)/\theta$ of θ_n^* with the relative bias $(\theta_\varepsilon - \theta)/\theta = \varepsilon\Delta$ of $\hat{\theta}_n$. In fact,

$$G(A) = 1 - \exp\left\{\frac{-A}{\theta}\right\} - \varepsilon \left(\exp\left\{\frac{-A}{\theta(1+\Delta)}\right\} - \exp\left\{\frac{-A}{\theta}\right\} \right)$$

and the integral in the numerator of (4.2) equals

$$\theta_\varepsilon - (1-\varepsilon)(A+\theta)\exp\left\{-\frac{A}{\theta}\right\} - \varepsilon(A+\theta(1+\Delta))\exp\left\{-\frac{A}{\theta(1+\Delta)}\right\}.$$

A cogent conclusion may not be possible concerning the gain in bias using these precise formulas even if we expand $\overline{\theta}_\varepsilon$ in powers of ε. Thus, we shall confine ourselves to asymptotic analysis. For this, assume that the obstructing parameter

$$\varepsilon = \frac{\lambda}{n}$$

for some $\lambda > 0$.

Recall that we reject observations with $X_k > \overline{X} \ln(n/\alpha)$. Note that

$$\overline{X} \ln(n/\alpha) \sim \theta(1+\varepsilon\Delta)\ln(n/\alpha) = A, \text{ almost surely.}$$

The probability of rejecting an observation in the obstructed model is asymptotically equal to

$$\begin{aligned} P(X_1 > A) &= (1-\varepsilon)\exp\{-A/\theta\} + \varepsilon \exp\{-A/\theta(1+\Delta)\} \\ &= (1-\varepsilon)\left(\frac{\alpha}{n}\right)^{1+\varepsilon\Delta} + \varepsilon \left(\frac{\alpha}{n}\right)^{\frac{1+\varepsilon\Delta}{1+\Delta}} = \frac{\alpha}{n} + O\left(n^{-\frac{2+\Delta}{1+\Delta}}\right). \end{aligned}$$

Hence the asymptotic distribution of θ_n^* equals the distribution of the $\alpha-$trimmed sample mean

$$\overline{Y} = \frac{1}{n(1-\alpha)} \sum_{k=1}^{n(1-\alpha)} Y_k$$

of a random sample of size $n(1-\alpha)$ from the distribution concentrated on the interval $(0,\, A)$. The probability density of this distribution is positive only on this interval and has the form

$$f_A(x;\, \theta) = \frac{1}{G_\alpha}\left[\frac{1-\varepsilon}{\theta}\exp\left\{-\frac{x}{\theta}\right\} + \frac{\varepsilon}{\theta(1+\Delta)}\exp\left\{-\frac{x}{\theta(1+\Delta)}\right\}\right],$$

where $G_\alpha = 1 - \alpha/n + o(1/n)$.

Denote by $E_\alpha(X)$ and $Var_\alpha(X)$ the mathematical expectation and variance, respectively, of a random variable X relative to the distribution with probability density function $f_A(x;\, \theta)$. With the preliminaries accounted for, the first result can now be presented.

Theorem 4.1. *The maximum likelihood estimator $\hat{\theta}_n$ under the obstructing model has the relative bias*

$$\varepsilon\Delta = \frac{\lambda\Delta}{n}$$

and the estimator θ_n^* has the relative bias

$$\varepsilon\Delta - \frac{\alpha}{n}\ln\frac{n}{\alpha} + O\left(\frac{\varepsilon \ln n}{n}\right) = \frac{\lambda\Delta}{n} - \frac{\alpha}{n}\ln\frac{n}{\alpha} + O\left(\frac{\varepsilon \ln n}{n}\right).$$

The above relations reveal that the relative bias of the proposed estimator θ_n^* in comparison with relative bias of the MLE $\hat{\theta}_n$ decreases in the value by the order $(\alpha/n)\ln(n/\alpha)$.

Proof.

The asymptotic of the bias of the estimator θ_n^* :

$$E\theta_n^* \sim E_\alpha\overline{Y} = \frac{1}{G_\alpha}\left\{\theta(1+\varepsilon\Delta)-\right.$$

$$
\begin{aligned}
&\int_A^{\infty} x\left[\frac{1-\varepsilon}{\theta}\exp\left\{-\frac{x}{\theta}\right\}+\frac{\varepsilon}{\theta(1+\Delta)}\exp\left\{-\frac{x}{\theta(1+\Delta)}\right\}\right]dx\Bigg\}\\
&=\frac{\theta}{G_\alpha}\left[1+\varepsilon\Delta-(1-\varepsilon)\left(\frac{A}{\theta}+1\right)\exp\left\{-\frac{A}{\theta}\right\}\right.\\
&\left.-\varepsilon(1+\Delta)\left(\frac{A}{\theta(1+\Delta)}+1\right)\exp\left\{-\frac{A}{\theta(1+\Delta)}\right\}\right]\\
&\sim\frac{\theta}{G_\alpha}\left[1+\varepsilon\Delta-(1-\varepsilon)\frac{\alpha}{n}\left((1+\varepsilon\Delta)\ln\frac{n}{\alpha}+1\right)\right.\\
&\left.-\varepsilon(1+\Delta)\left(\frac{\alpha}{n}\right)^{1/(1+\Delta)}\left(\frac{1+\varepsilon\Delta}{1+\Delta}\ln\frac{n}{\alpha}+1\right)\right]\\
&\sim\theta\left(1+\frac{\alpha}{n}\right)\left[1+\varepsilon\Delta-\frac{\alpha}{n}\ln\frac{n}{\alpha}-\frac{\alpha}{n}\right]\sim\theta\left(1+\varepsilon\Delta-\frac{\alpha}{n}\ln\frac{n}{\alpha}\right)\\
&=\theta\left(1+\frac{\lambda\Delta}{n}-\frac{\alpha}{n}\ln\frac{n}{\alpha}\right)
\end{aligned}
$$

with the remainder term $O\left(n^{-\frac{2+\Delta}{1+\Delta}}\ln n\right)$.

The MLE $\hat{\theta}_n$ has mathematical expectation $\theta(1+\varepsilon\Delta)$ and the weighted likelihood estimator θ_n^* has mathematical expectation

$$
E\theta_n^*=\theta\left(1+\varepsilon\Delta-\frac{\alpha}{n}\ln\frac{n}{\alpha}+O\left(\varepsilon\frac{\ln n}{n}\right)\right).
$$

Hence, the gain in bias has the order $\frac{\alpha}{n}\ln\frac{n}{\alpha}$.

Remark 4.1. From Theorem 4.1 it is apparent that the value of α must be chosen as a quantity of order $O(1/\ln n)$; otherwise we will obtain a negative bias (underestimated values of θ). Perhaps, we can suggest here to choose α from the relation $\lambda\Delta\approx\alpha\ln(n/\alpha)$. But generally the values of λ and Δ are unknown.

Finally, we derive expressions for the quadratic risks of the proposed estimator, which are based on the following two elementary integrals

$$
\int x\mathrm{e}^{ax}dx=\mathrm{e}^{ax}\left(\frac{x}{a}-\frac{1}{a^2}\right)+c,
$$

$$
\int x^2\mathrm{e}^{ax}dx=\mathrm{e}^{ax}\left(\frac{x^2}{a}-\frac{2x}{a^2}+\frac{2}{a^3}\right)+c
$$

(mainly we use the case $a = -1$).

Theorem 4.2. *The maximum likelihood estimator $\hat{\theta}_n$ has the quadratic risk*

$$R(\hat{\theta}_n) = \frac{\theta^2}{n}\left[1 + 2\varepsilon\Delta(1+\Delta)\right] \qquad (4.3)$$

and the estimator θ_n^* has the asymptotic representation of the quadratic risk

$$R(\theta_n^*) \sim \frac{\theta^2}{n(1-\alpha)}\left[1 + 2\varepsilon\Delta(1+\Delta) - \frac{\alpha}{n}\ln^2\frac{n}{\alpha}\right]. \qquad (4.4)$$

The relative efficiency of the estimator θ_n^* in comparison with maximum likelihood estimator $\hat{\theta}_n$ has the asymptotic representation

$$\mathcal{E}_n = \frac{R(\hat{\theta}_n)}{R(\theta_n^*)} \sim (1-\alpha)\left(1 + \frac{\alpha}{n}\ln^2\frac{n}{\alpha}\right).$$

Proof. The quadratic risk of $\hat{\theta}_n$ is

$$R(\hat{\theta}_n) = \frac{E(X_1-\theta)^2}{n} = \frac{E(X_1-\theta(1+\varepsilon\Delta))^2 + (\theta-\theta(1+\varepsilon\Delta))^2}{n}$$
$$= \frac{\theta^2}{n}\left(Var\left(\frac{X}{\theta}\right) + \varepsilon^2\Delta^2\right).$$

Direct computations show that the second moment with the distribution G of the normalized random variable X/θ equals $2\left(1+\varepsilon\Delta(2+\Delta)\right)$. Hence

$$Var\left(\frac{X}{\theta}\right) = 1 + 2\varepsilon\Delta(1+\Delta) - \varepsilon^2\Delta^2.$$

Finally,

$$R(\hat{\theta}_n) = \frac{\theta^2}{n}\left[1 + 2\varepsilon\Delta(1+\Delta)\right],$$

which proves formula (4.3).

Seemingly, for the estimator θ_n^*, it is rather difficult to obtain precise results. However, it is possible to obtain a first order asymptotic. Conceivably, more precise results can be obtained based on the joint distribution of the extreme terms of the sample and the sample mean.

We have

$$R(\theta_n^*) \sim E_\alpha \left[\frac{1}{n(1-\alpha)} \sum_{k=1}^{n(1-\alpha)} (Y_k - \theta) \right]^2 = \frac{1}{n(1-\alpha)} Var_\alpha(Y) + (\mu - \theta)^2,$$

where

$$\mu = E_\alpha Y = \theta \left(1 + \varepsilon\Delta - \frac{\alpha}{n} \ln \frac{n}{\alpha} \right).$$

Further, $Var_\alpha(Y) = E_\alpha(Y^2) - \mu^2$, so

$$R(\theta_n^*) \sim \frac{1}{n(1-\alpha)} \left[E_\alpha(Y^2) - \mu^2 \right] + (\mu - \theta)^2.$$

Now,

$$\begin{aligned}
E_\alpha(Y^2) &= \frac{\theta^2}{G_\alpha} \{ 2(1 + \varepsilon\Delta(2+\Delta)) \\
&- \int_{A/\theta}^{\infty} x^2 \left[(1-\varepsilon) \exp\{-x\} + \frac{\varepsilon}{(1+\Delta)} \exp\left\{ -\frac{x}{(1+\Delta)} \right\} \right] dx \Bigg\} \\
&= \frac{\theta^2}{G_\alpha} [\, 2\,(1 + \varepsilon\Delta(2+\Delta)) - (1-\varepsilon) \exp\left\{ -\frac{A}{\theta} \right\} \left(\frac{A^2}{\theta^2} + \frac{2A}{\theta} + 2 \right) \\
&- \varepsilon(1+\Delta) 2 \exp\left\{ -\frac{A}{\theta(1+\Delta)} \right\} \left(\frac{A^2}{\theta^2(1+\Delta)^2} + \frac{2A}{\theta(1+\Delta)} + 2 \right)] \\
&\sim \theta^2 \left(1 + \frac{\alpha}{n} \right) \{ 2\,(1 + \varepsilon\Delta(2+\Delta)) \\
&- (1-\varepsilon) \frac{\alpha}{n} \left[(1+\varepsilon\Delta)^2 \ln^2 \frac{n}{\alpha} + 2(1+\varepsilon\Delta) \ln \frac{n}{\alpha} + 2 \right] \\
&- \varepsilon(1+\Delta)^2 \left(\frac{\alpha}{n} \right)^{1/(1+\Delta)} \left[\frac{(1+\varepsilon\Delta)^2}{1+\Delta)^2} \ln^2 \frac{n}{\alpha} + 2 \frac{1+\varepsilon\Delta}{1+\Delta} \ln \frac{n}{\alpha} + 2 \right] \} \\
&\sim \theta^2 \left(1 + \frac{\alpha}{n} \right) \left[2\,(1 + \varepsilon\Delta(2+\Delta)) - \frac{\alpha}{n} \ln^2 \frac{n}{\alpha} - 2\frac{\alpha}{n} \ln \frac{n}{\alpha} - \frac{2\alpha}{n} \right] \\
&\sim \theta^2 \left[2\,(1 + \varepsilon\Delta(2+\Delta)) - \frac{\alpha}{n} \ln^2 \frac{n}{\alpha} - 2\frac{\alpha}{n} \ln \frac{n}{\alpha} \right]
\end{aligned}$$

with the remainder term $O\left(n^{-\frac{2+\Delta}{1+\Delta}} \ln^2 n \right)$. Hence we have established that

$$R(\theta_n^*) \sim \frac{\theta^2}{n(1-\alpha)} \left[1 + 2\varepsilon\Delta(1+\Delta) - \frac{\alpha}{n} \ln^2 \frac{n}{\alpha} \right],$$

which completes the proof of formula (4.4).

The relative efficiency of the estimator θ_n^* to the estimator $\hat{\theta}_n$ is defined as the ratio of their risks (expression (4.3) divided by expression (4.4)). For this ratio, we have the following asymptotic representation:

$$\begin{aligned}
&\mathcal{E}_n = \frac{R(\hat{\theta}_n)}{R(\theta_n^*)} \sim (1-\alpha)\frac{1+2\varepsilon\Delta(1+\Delta)}{1+2\varepsilon\Delta(1+\Delta)-\frac{\alpha}{n}\ln^2\frac{n}{\alpha}} \\
&\sim (1-\alpha)\left[1+2\varepsilon\Delta(1+\Delta)\right]\left[1-2\varepsilon\Delta(1+\Delta)+\frac{\alpha}{n}\ln^2\frac{n}{\alpha}\right] \\
&\sim (1-\alpha)\left(1+\frac{\alpha}{n}\ln^2\frac{n}{\alpha}\right).
\end{aligned}$$

Remark 4.2. The asymptotic relative efficiency of the estimator θ_n^* of the α-trimmed mean type (cf. Huber (1981) Chapter 4, pp.104-106) is given by

$$\mathcal{E} = \lim_{n\to\infty}\mathcal{E}_n = 1-\alpha$$

for any distribution. We have the same limiting behavior of the relative efficiency.

In the following section we conduct a simulation study to evaluate the performance of the proposed estimators for a given sample.

4.3 Monte-Carlo Simulations

For the purposes of simulation, we fix $\theta > 0, 0 < \varepsilon < 1$ and $\Delta > 0$. Recall that, if a random variable U has the uniform distribution on the interval $[0,1]$, then $X = -\theta\ln(U)$ will have the exponential distribution with parameter θ. We need this result for generating random numbers from the exponential distribution.

We generate random numbers from the obstructed distribution

$$G(x) = G_\varepsilon(x) = 1-\exp\left\{\frac{-x}{\theta}\right\} - \varepsilon\left(\exp\left\{\frac{-x}{\theta(1+\Delta)}\right\} - \exp\left\{\frac{-x}{\theta}\right\}\right).$$

This procedure is organized in the following way. The main idea is that with probability $1-\varepsilon$ we need to generate a random number from the exponential distribution with parameter $\theta_1 = \theta$ and with probability ε generate a random number from the exponential distribution with parameter $\theta_1 = \theta(1+\Delta)$. To achieve this, we generate a random number u from the uniform distribution on the interval $[0, 1]$ and compare u with ε. If $u > \varepsilon$, then take $\theta_1 = \theta$. Otherwise $\theta_1 = \theta(1+\Delta)$. Next, generate a random number x from the exponential distribution with the parameter θ_1. The random number x has the cumulative distribution function $G(x)$.

The Tables 4.1 - 4.6 represent the numerical values of simulated (Sim) and asymptotic (Asy) bias and risk for both estimators $\hat{\theta}$ and θ^* for various values of n, ε and α. The asymptotic bias and risk are calculated using the formulas (4.2) and (4.4) from Theorems 4.1 and 4.2. Note that the risk of the estimate $\hat{\theta}$ does not depend on α. From the tables we see that the asymptotic formulas are sufficiently accurate for the given sample sizes.

The Tables 4.7 –4.15 represent the differences between the biases $Bias(\hat{\theta}) - Bias(\theta^*)$ and the risks $R(\hat{\theta}) - R(\theta^*)$ for different values of n, Δ, ε and α. These tables show the advantage of the estimator θ^*, especially for large values of ε and α. Interestingly, the substantial gain in risk is achieved when we take $\alpha = \varepsilon$, but the value of ε is generally assumed to be unknown.

Table 4.1: Simulated and Asymptotic Bias for estimators $\hat{\theta}$ and θ^* for $n = 30$ and $\Delta = 1.0$.

	ε =.01		ε =.03		ε =.05		ε =.07		ε =.09	
	Sim	Asy	Sim	Asy	Sim	Asy	Sim	Asy	Sim	Asy
$\hat{\theta}$	-.0291	.0100	-.0013	.0300	.0679	.0500	.0842	.0700	.1242	.0900
α =.01	-.0292	.0083	-.0013	.0283	.0650	.0483	.0776	.0683	.1214	.0883
α =.03	.0209	.0055	.0223	.0256	.0323	.0455	.0377	.0655	.0846	.0855
α =.05	.0041	.0031	.0041	.0231	.0681	.0431	.0480	.0631	.0207	.0831
α =.07	.0116	.0008	.0111	.0208	.0026	.0408	.0583	.0608	.0815	.0808
α =.09	.0039	.0014	.00377	.0186	.0182	.0386	.0179	.0586	.0819	.0786

Table 4.2: Simulated and Asymptotic Risk for Estimators $\hat{\theta}$ and θ^* for $n = 30$ and $\Delta = 1.0$.

	ε =.01		ε =.03		ε =.05		ε =.07		ε =.09	
	Sim	Asy	Sim	Asy	Sim	Asy	Sim	Asy	Sim	Asy
$\hat{\theta}$	.0163	.0208	.0244	.0224	.0223	.0240	.0209	.0256	.0395	.0272
α =.01	.0180	.0207	.0182	.0223	.0360	.0239	.0278	.0256	.0412	.0282
α =.03	.0163	.0208	.0220	.0224	.0222	.0241	.0200	.0257	.0352	.0274
α =.05	.0205	.0209	.0282	.0226	.0287	.0243	.0308	.0254	.0191	.0276
α =.07	.0179	.0211	.0200	.0228	.0277	.0245	.0222	.0262	.0361	.0279
α =.09	.0202	.0213	.0116	.0230	.0183	.0248	.0190	.0266	.0248	.0283

Table 4.3: Simulated and Asymptotic Bias for Estimators $\hat{\theta}$ and θ^* for $n = 100$ and $\Delta = 1.0$.

	ε =.01		ε =.03		ε =.05		ε =.07		ε =.09	
	Sim	Asy	Sim	Asy	Sim	Asy	Sim	Asy	Sim	Asy
$\hat{\theta}$	.0214	.0100	.0290	.0300	.0558	.0500	.0719	.0700	.1052	.0900
α =.01	.0200	.0091	.0028	.0029	.0516	.0491	.0667	.0691	.1026	.0891
α =.03	.0294	.0076	.0196	.0276	.0491	.0476	.0825	.0676	.0732	.0876
α =.05	.0091	.0062	.0152	.0262	.0637	.0462	.0817	.0662	.0678	.0862
α =.07	.0181	.0049	.0167	.0249	.0373	.0449	.0373	.0659	.0752	.0849
α =.09	.0105	.003	.0013	.0237	.0376	.0437	.0330	.0637	.0531	.0836

Table 4.4: Simulated and Asymptotic Risk for Estimators $\hat{\theta}$ and θ^* for $n = 100$ and $\Delta = 1.0$.

	ε =.01		ε =.03		ε =.05		ε =.07		ε =.09	
	Sim	Asy	Sim	Asy	Sim	Asy	Sim	Asy	Sim	Asy
$\hat{\theta}$	.0111	.0104	.0110	.0112	.0161	.0120	.0151	.0128	.0274	.0136
α =.01	.0114	.0104	.0108	.0112	.0154	.0120	.0138	.0128	.0267	.0137
α =.03	.0117	.0105	.0113	.0113	.0158	.0122	.0202	.0130	.0176	.0138
α =.05	.0099	.0106	.0084	.0115	.0180	.0123	.0180	.0132	.0159	.0140
α =.07	.0116	.0108	.0170	.0116	.0134	.0125	.0173	.0133	.0192	.0142
α =.09	.0109	.0109	.0118	.0118	.0125	.0127	.0134	.0136	.0190	.0145

Table 4.5: Simulated and Asymptotic Bias for Estimators $\hat{\theta}$ and θ^* for $n = 200$ and $\Delta = 1.0$.

	ε =.01		ε =.03		ε =.05		ε =.07		ε =.09	
	Sim	Asy	Sim	Asy	Sim	Asy	Sim	Asy	Sim	Asy
$\hat{\theta}$	.0194	.0100	.0282	.0300	.0422	.0500	.0670	.0700	.0821	.0900
α =.01	.0188	.0095	.0266	.0295	.0391	.0495	.0656	.0695	.0795	.0895
α =.03	.0076	.0087	.0183	.0287	.0507	.0487	.0609	.0687	.0839	.0887
α =.05	.0059	.0079	.0289	.0279	.0479	.0409	.0567	.0679	.0774	.0879
α =.07	.0128	.0072	.0256	.0272	.0445	.0472	.0667	.0672	.0825	.0872
α =.09	.0054	.0065	.0341	.0265	.0442	.0465	.0630	.0665	.0809	.0865

Table 4.6: Simulated and Asymptotic Risk for Estimators $\hat{\theta}$ and θ^* for $n = 200$ and $\Delta = 1.0$.

	ε =.01		ε =.03		ε =.05		ε =.07		ε =.09	
	Sim	Asy	Sim	Asy	Sim	Asy	Sim	Asy	Sim	Asy
$\hat{\theta}$	.0057	.0052	.0060	.0056	.0077	.0060	.0114	.0064	.0129	.0068
α =.01	.0057	.0052	.0058	.0056	.0072	.0060	.0113	.0064	.0125	.0068
α =.03	.0062	.0053	.0054	.0057	.0083	.0061	.0160	.0070	.0056	.0058
α =.05	.0056	.0054	.0078	.0058	.0074	.0062	.0106	.0066	.0135	.0071
α =.07	.0057	.0055	.0059	.0059	.0078	.0063	.0113	.0068	.0135	.0072
α =.09	.0052	.0056	.0066	.0060	.0067	.0064	..0117	.0069	.0138	.0073

Table 4.7: Differences Between Biases $Bias(\hat{\theta}) - Bias(\theta^*)$ and Risks $R(\hat{\theta}) - R(\theta^*)$ for $n = 30$ and $\Delta = 1.0$.

	ε =.01		ε =.03		ε =.05		ε =.07		ε =.09	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
α =.01	.0000	.0000	.0034	-.0008	.0000	.0000	.0060	-.0001	.0000	.0000
α =.03	.0000	.0000	.0035	.0001	.0023	-.0014	.0051	.0051	.0249	.0035
α =.05	.0076	.0040	.0022	-.0009	.0109	.0016	.0084	.0002	.0208	.0006
α =.07	.0059	-.0006	.0089	.0018	.0301	.0022	.0126	.0005	.0184	.0095
α =.09	.0037	.0008	.0034	.0008	.0079	.0009	.0387	.0007	.0124	-.0003

Table 4.8: Differences Between Biases $Bias(\hat{\theta}) - Bias(\theta^*)$ and Risks $R(\hat{\theta}) - R(\theta^*)$ for $n = 30$ and $\Delta = 3.0$.

	ε =.01		ε =.03		ε =.05		ε =.07		ε =.09	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
α =.01	.0007	-.003	.0282	.0240	.0601	.0213	.0588	.0319	.0598	.0458
α =.03	.0084	.0108	.0572	.0191	.0870	.0413	.0391	.0208	.0822	.0490
α =.05	.0321	.0084	.0134	-.0002	.0516	.0167	.0729	.0381	.0984	.0719
α =.07	.0217	.0049	.0589	.0248	.0940	.0560	.0991	.0364	.1176	.0760
α =.09	.0279	-.0021	.0709	.0219	.0731	.0364	.0802	.0480	.1136	.0653

Table 4.9: Differences Between Biases $Bias(\hat{\theta}) - Bias(\theta^*)$ and Risks $R(\hat{\theta}) - R(\theta^*)$ for $n = 30$ and $\Delta = 5.0$.

	ε =.01		ε =.03		ε =.05		ε =.07		ε =.09	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
α =.01	.0215	.0087	.0477	.0231	.0838	.0396	.1496	.1344	.1770	.1197
α =.03	.0334	.0161	.0966	.0641	.1717	.1334	.2675	.3158	.1661	.1403
α =.05	.0258	.0089	.0898	.0432	.1254	.0718	.1686	.1293	.2009	.1569
α =.07	.0272	.0042	01369	.0899	.1719	.1257	.1913	.1679	.2497	.2106
α =.09	.0292	.0083	.0963	.0745	.1768	.0832	.2070	.1603	.2761	.2070

Table 4.10: Differences Between Biases $Bias(\hat{\theta}) - Bias(\theta^*)$ and Risks $R(\hat{\theta}) - R(\theta^*)$ for $n = 100$ and $\Delta = 1.0$.

	ε =.01		ε =.03		ε =.05		ε =.07		ε =.09	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
α =.01	.0148	.0033	.0341	.0094	.0521	.0190	.0722	.0348	.0551	.0308
α =.03	.0272	.0032	.0419	.0112	.0593	.0196	.0693	.0327	.0764	.0468
α =.05	.0177	.0014	.0560	.0064	.0592	.0230	.0654	.0307	.0930	.0505
α =.07	.0122	.0000	.0560	.0131	.0592	.0231	.0654	.0326	.0930	.0503
α =.09	.0117	-.0004	.0475	.0111	.0664	.0190	.0817	.0324	.1043	.0483

Table 4.11: Differences Between Biases $Bias(\hat{\theta}) - Bias(\theta^*)$ and Risks $R(\hat{\theta}) - R(\theta^*)$ for $n = 100$ and $\Delta = 3.0$.

	ε =.01		ε =.03		ε =.05		ε =.07		ε =.09	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
α =.01	.0148	.0033	.0341	.0094	.0521	.0190	.0722	.0348	.0551	.0308
α =.03	.0272	.0032	.0419	.0112	.0593	.0196	.0693	.0327	.0764	.0468
α =.05	.0177	.0014	.0560	.0064	.0592	.0230	.0654	.0307	.0930	.0505
α =.07	.0122	.0000	.0560	.0131	.0592	.0231	.0654	.0326	.0930	.0503
α =.09	.0117	-.0004	.0475	.0111	.0664	.0190	.0817	.0324	.1043	.0483

Table 4.12: Differences Between Biases $Bias(\hat{\theta}) - Bias(\theta^*)$ and Risks $R(\hat{\theta}) - R(\theta^*)$ for $n = 100$ and $\Delta = 5.0$.

	ε =.01		ε =.03		ε =.05		ε =.07		ε =.09	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
α =.01	.0229	.0028	.0880	.0373	.0865	.0404	.1565	.1089	.1337	.1153
α =.03	.0237	.0056	.0939	.0374	.1559	.1105	.1646	.1193	.2179	.1929
α =.05	.0404	.0065	.1052	.0405	.1584	.0796	.1760	.1146	.2240	.1928
α =.07	.0291	.0078	.0888	.0297	.1503	.0752	.1786	.1074	.2354	.2013
α =.09	.0391	.0033	.1085	.0379	.1388	.0572	.1999	.1267	.2285	.1808

Table 4.13: Differences Between Biases $Bias(\hat{\theta}) - Bias(\theta^*)$ and Risks $R(\hat{\theta}) - R(\theta^*)$ for $n = 200$ and $\Delta = 1.0$.

	ε =.01		ε =.03		ε =.05		ε =.07		ε =.09	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
α =.01	.0000	.0000	.0013	.0000	.0031	.0004	.0012	.0001	.0033	.0004
α =.03	.0016	.0000	.0048	.0003	.0048	.0011	.0075	.0015	.0070	.0009
α =.05	.0015	.0001	.0041	.0002	.0105	.0018	.0081	.0012	.0100	.0020
α =.07	.0030	-.0002	.0065	.0003	.0076	.0005	.0139	.0016	.0151	.0027
α =.09	.0031	-.0001	.0082	.0005	.0111	.0002	.0112	.0011	.0123	.0018

Table 4.14: Differences Between Biases $Bias(\hat{\theta}) - Bias(\theta^*)$ and Risks $R(\hat{\theta}) - R(\theta^*)$ for $n = 200$ and $\Delta = 3.0$.

	ε =.01		ε =.03		ε =.05		ε =.07		ε =.09	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
α =.01	.0113	.0014	.0249	.0048	.0438	.0137	.0481	.0204	.0617	.0315
α =.03	.0107	.0010	.0351	.0069	.0419	.0119	.0641	.0235	.0692	.0369
α =.05	.0117	.0012	.0462	.0094	.0510	.0144	.0664	.0238	.0785	.0387
α =.07	.0162	.0018	.0412	.0092	.0575	.0185	.0738	.0265	.0852	.0474
α =.09	.0166	.0018	.0378	.0067	.0540	.0129	.0852	.0340	.0885	.0421

Table 4.15: Differences Between Biases $Bias(\hat{\theta}) - Bias(\theta^*)$ and Risks $R(\hat{\theta}) - R(\theta^*)$ for $n = 200$ and $\Delta = 5.0$.

	ε =.01		ε =.03		ε =.05		ε =.07		ε =.09	
	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk	Bias	Risk
α =.01	.0233	.0028	.0696	.0212	.1014	.0498	.1255	.0787	.1535	.1299
α =.03	.0304	.0038	.0743	.0218	.1200	.0564	.1536	.0927	.1765	.1637
α =.05	.0358	.0058	.0903	.0218	.1363	.0641	.1677	.1017	.2034	.1596
α =.07	.0420	.0058	.0931	.0257	.1376	.0617	.1912	.1017	.2045	.1584
α =.09	.0326	.0042	.0871	.0216	.1465	.0588	.1684	.0916	.2159	.1677

Chapter 5

The Complete Convergence Rates of the Bootstrap Mean

5.1 Introduction

The proof of the main result of Section 5.1 is based on the following theorem, proved in Hu, Rosalsky, Szynal and Volodin (1999), Theorem 3.2 (for this thesis we took $c_n = 1$ for all $n \geq 1$ and weights are built into random elements).

Theorem 5.1. *Let $\{X_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ be an array of rowwise independent random elements in a real separable Banach space such that*

$$\sum_{n=1}^{\infty} \sum_{i=1}^{k_n} P\{||X_{ni}|| > \epsilon\} < \infty \text{ for all } \epsilon > 0.$$

Suppose that there exists $s > 0$ such that, for some $0 < r \leq 2$,

$$\sum_{n=1}^{\infty} \left(\sum_{i=1}^{k_n} E||X_{ni}||^r \right)^s < \infty,$$

and

$$\sum_{i=1}^{k_n} X_{ni} \xrightarrow{P} 0.$$

Then

$$\sum_{n=1}^{\infty} P\{||\sum_{i=1}^{k_n} X_{ni}|| > \epsilon\} < \infty \text{ for all } \epsilon > 0.$$

We also will need the following lemma in de Acosta (1981) (cf. also Berger (1991)):

Lemma 5.1. *Let $\{X_{nk}, 1 \leq k \leq k_n, n \geq 1\}$ be an array of rowwise independent random elements. Then for every $p \geq 1/2$, there is a positive constant A_p depending only on p such that, for all $n \geq 1$,*

$$E\Big|||\sum_{i=1}^{k_n} X_{ni}|| - E||\sum_{i=1}^{k_n} X_{ni}||\Big|^{2p} \leq A_p E\left(\sum_{k=1}^{k_n} ||X_{nk}||^2\right)^p.$$

In the last section of this chapter we will use the following theorem on complete convergence for row sums of arrays of random variables from Hu, Szynal and Volodin (1998): Remark 2 after the Theorem. It forms the basis for the results in this section.

Theorem 5.2. *Let $\{k_n, n \geq 1\}$, be a sequence of positive integers, $k_n \to \infty$ as $n \to \infty$ and $\{Y_{nk}, 1 \leq k \leq k_n, n \geq 1\}$ be an array of rowwise independent mean zero random variables. Suppose, $E|Y_{nk}|^2 < \infty, 1 \leq k \leq m(n), n \geq 1$ and $r \geq 0$.*
Moreover, assume that

A1. $\sum_{n=1}^{\infty} n^r \sum_{k=1}^{k_n} P\{|Y_{nk}| > \epsilon\} < \infty$ *for all* $\epsilon > 0$.

A2. *There exists $J \geq 1$ such that*

$$\sum_{n=1}^{\infty} n^r \left(\sum_{k=1}^{k_n} E|Y_{nk}|^2\right)^J < \infty.$$

Then

$$\sum_{n=1}^{\infty} n^r P\{|\sum_{k=1}^{k_n} Y_{nk}| > \epsilon\} < \infty \text{ for all } \epsilon > 0.$$

We now outline the bootstrap procedure in a Banach space setting. We note that an outline of the bootstrap procedure for real random variables was presented in Chapter 1. Let $\{X_n;\ n \geq 1\}$ be a sequence of (not necessarily independent or identically distributed) random elements defined on some complete probability space $(\Omega, \mathcal{F}, P)$ which take values in a real separable Banach space. For $\omega \in \Omega$ and $n \geq 1$, let $P_n(\omega) = \dfrac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}$ denote the empirical measure. For $n \geq 1$, let $\{\hat{X}^{\omega}_{n,j};\ 1 \leq j \leq k_n\}$ be i.i.d. random elements with law $P_n(\omega)$ where k_n is a positive integer. Let $\bar{X}_n(\omega)$ denote the sample mean of $\{X_i(\omega);\ 1 \leq i \leq n\},\ n \geq 1$, that is, $\bar{X}_n(\omega) = \dfrac{1}{n}\sum_{i=1}^{n} X_i(\omega)$.

5.2 Complete Convergence of Weighted Sums in Banach Spaces

Let $\psi : (0, +\infty) \longrightarrow (0, +\infty)$ be a function. Assume there exists a constant $C > 0$ such that

$$u \geq v \implies \psi(u) \geq C\psi(v) \tag{5.1}$$

and for any $\epsilon > 0$

$$\sup_{u>0} \frac{\psi(u)}{\psi(\epsilon u)} < \infty \tag{5.2}$$

By putting $u = v$ (or by using a continuity argument, if ψ is continuous), it is clear that $0 < C \leq 1$.

Conditions (5.1) and (5.2) are weaker than any of the conditions used in the papers Chung (1947), Hu and Taylor (1997), Sung (2000), and Ahmed, Hu and Volodin (2001). Consequently, the family of functions satisfying (5.1) and (5.2) is wider than the family of functions used by other authors. The following lemma presents a sufficient condition for ψ to satisfy (5.1) and (5.2).

Lemma 5.2. *Let $\psi : (0, +\infty) \longrightarrow (0, +\infty)$ be a function such that $Ax \leq \psi(x) \leq Bx$ for all $x \in (0, +\infty)$, for some constants $A, B > 0$. Then:*

$$\frac{A}{B}\frac{u}{v} \leq \frac{\psi(u)}{\psi(v)} \leq \frac{B}{A}\frac{u}{v}$$

for all $u, v > 0$.

Throughout this section, unless otherwise specified, ψ will be a function satisfying (5.1) and (5.2).

Theorem 5.3. *Let* $k_n \to \infty$ *be a sequence of positive integers. Let* $\{X_{ni}, 1 \le i \le k_n, n \ge 1\}$ *be an array of rowwise independent* $\mathcal{B}$*-valued random elements. Assume that*

$$\sum_{n=1}^{\infty} \sum_{i=1}^{k_n} P\{||X_{ni}|| > \epsilon\} < \infty \text{ for all } \epsilon > 0, \tag{5.3}$$

and

$$\sum_{n=1}^{\infty} \left(\sum_{i=1}^{k_n} E||X_{ni}||^r \right)^s < \infty \tag{5.4}$$

for some $s > 0$ *and some* $1 < r \le 2$. *Then the following statements are equivalent:*

(i) $\sum_{i=1}^{k_n} X_{ni} \longrightarrow 0$ *in* L^1.

(ii) $\sum_{i=1}^{k_n} X_{ni} \longrightarrow 0$ *completely.*

(iii) $\sum_{i=1}^{k_n} X_{ni} \longrightarrow 0$ *almost surely.*

(iv) $\sum_{i=1}^{k_n} X_{ni} \longrightarrow 0$ *in probability.*

Proof.$(ii) \Longrightarrow (iii)$, $(iii) \Longrightarrow (iv)$ and $(i) \Longrightarrow (iv)$ are immediate. $(iv) \Longrightarrow (ii)$ is stated in Theorem 5.1.

$(iv) \Longrightarrow (i)$. Assume that (iv) holds. From (5.3) and Lemma 5.1 with $p = r/2$:

$$E\left| ||\sum_{i=1}^{k_n} X_{ni}|| - E||\sum_{i=1}^{k_n} X_{ni}|| \right|^r \le A_{r/2} E(\sum_{i=1}^{k_n} ||X_{ni}||^2)^{r/2}$$

$$\le A_{r/2} \sum_{i=1}^{k_n} E||X_{ni}||^r \longrightarrow 0$$

and so

$$||\sum_{i=1}^{k_n} X_{ni}|| - E||\sum_{i=1}^{k_n} X_{ni}|| \longrightarrow 0 \text{ in probability.}$$

Consequently, $E||\sum_{i=1}^{k_n} X_{ni}|| \longrightarrow 0$, and so (i) holds.

Theorem 5.4. *Let $k_n \to \infty$ be a sequence of positive integers. Let $\{X_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ be an array of rowwise independent $\mathcal{B}$-valued random elements, and $\{a_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ an array of constants. Assume that*

$$\sum_{n=1}^{\infty}\sum_{i=1}^{k_n} \frac{E\psi(||X_{ni}||)}{\psi(|a_{ni}|^{-1})} < \infty \tag{5.5}$$

and

$$\sum_{n=1}^{\infty}\left(\sum_{i=1}^{k_n} |a_{ni}|^r E||X_{ni}||^r\right)^s < \infty \tag{5.6}$$

for some $1 \leq r \leq 2$ and some $s > 0$. Then the following statements are equivalent:

(i) $\sum_{i=1}^{k_n} a_{ni}X_{ni} \longrightarrow 0$ *in* L^1.

(ii) $\sum_{i=1}^{k_n} a_{ni}X_{ni} \longrightarrow 0$ *completely.*

(iii) $\sum_{i=1}^{k_n} a_{ni}X_{ni} \longrightarrow 0$ *almost surely.*

(iv) $\sum_{i=1}^{k_n} a_{ni}X_{ni} \longrightarrow 0$ *in probability.*

Proof. Consider $a_{ni}X_{ni}$ instead of X_{ni} in Theorem 4.3. The only thing we need to prove is that (5.5) $\Longrightarrow$ (5.1).

For each $\epsilon > 0$ by (5.1), the Markov inequality and (5.2):

$$\sum_{n=1}^{\infty}\sum_{i=1}^{k_n} P[||a_{ni}X_{ni}|| > \epsilon]$$

$$\leq \sum_{n=1}^{\infty}\sum_{i=1}^{k_n} P[\psi(||X_{ni}||) > C\psi(\epsilon|a_{ni}|^{-1})] \leq \frac{1}{C}\sum_{n=1}^{\infty}\sum_{i=1}^{k_n} \frac{E[\psi(||X_{ni}||)]}{\psi(\epsilon|a_{ni}|^{-1})}$$

$$\leq C'\sum_{n=1}^{\infty}\sum_{i=1}^{k_n} \frac{E[\psi(||X_{ni}||)]}{\psi(|a_{ni}|^{-1})}.$$

Remark 5.1. An analysis of the proof of Theorem 5.4 shows that the condition (5.2) can be simplified to

$$\sup_{n,i} \frac{\psi(|a_{ni}|^{-1})}{\psi(\epsilon|a_{ni}|^{-1})} < \infty$$

for any $\epsilon > 0$

We obtain the complete convergence of weighted sums taking values in a Banach space of type r $(1 \leq r \leq 2)$ as a corollary of this theorem.

Recall that a separable Banach space $\mathcal{B}$ is of type r, $1 \leq r \leq 2$, if, and only if, there exists a constant C_r such that

$$E||\sum_{i=1}^{n} X_i||^r \leq C_r \sum_{i=1}^{n} E||X_i||^r$$

for all independent $\mathcal{B}$-valued random elements $X_1, ..., X_n$ with mean zero and finite r-th moments.

Corollary 5.1.*Let $k_n \to \infty$ be a sequence of positive integers. Let $\{X_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ be an array of rowwise independent $\mathcal{B}$-valued random elements, and $\{a_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ an array of constants. Assume that $EX_{ni} = 0$ for all $1 \leq i \leq k_n$ and $n \geq 1$, and that $\mathcal{B}$ is of type r $(1 \leq r \leq 2)$. Assume (5.3) and (5.6) hold for this r and some $s > 0$. Then,*

$$\sum_{i=1}^{k_n} a_{ni} X_{ni} \longrightarrow 0 \text{ completely.}$$

Proof. By Theorem 5.4, it is enough to prove that $\sum_{i=1}^{k_n} a_{ni} X_{ni} \longrightarrow 0$ in L^1.

Since $\mathcal{B}$ is of type r and $E(a_{ni} X_{ni}) = 0$, we have

$$E||\sum_{i=1}^{k_n} a_{ni} X_{ni}||^r \leq C_r \sum_{i=1}^{k_n} |a_{ni}|^r E||X_{ni}||^r \longrightarrow 0$$

as a consequence of (5.6).

Convergence in L^r implies convergence in L^1, so Theorem 5.4 implies that

$$\sum_{i=1}^{k_n} a_{ni} X_{ni} \longrightarrow 0 \text{ completely.}$$

Remark 5.2. If $C = 1, k_n = n$ and $a_{ni} = \dfrac{1}{a_n}, 1 \le i \le k_n, n \ge 1$, then Theorem 5.4 and Corollary 5.1 become Theorem 2.2, Theorem 2.3 and Corollary 2.4 in Sung (2000).

By using the functions ψ satisfying the hypothesis of Lemma 5.2, we can weaken slightly the conditions (5.5) and (5.6) in Theorem 5.4.

Theorem 5.5.*Let $k_n \to \infty$ be a sequence of positive integers. Let $\{X_{ni}, 1 \le i \le k_n, n \ge 1\}$ be an array of rowwise independent $\mathcal{B}$-valued random elements, and $\{a_{ni}, 1 \le i \le k_n, n \ge 1\}$ an array of constants. Let $\psi : (0, +\infty) \longrightarrow (0, +\infty)$ be a function such that $Ax \le \psi(x) \le Bx$ for all $x \in (0, +\infty)$, for some constants $A, B > 0$. Assume that*

$$\sum_{n=1}^{\infty} \sum_{i=1}^{k_n} \frac{E\psi(||X_{ni}||)}{\psi(|a_{ni}|^{-1})} < \infty.$$

Then the following statements are equivalent:

(i) $\sum_{i=1}^{k_n} a_{ni} X_{ni} \longrightarrow 0$ *in* L^1.

(ii) $\sum_{i=1}^{k_n} a_{ni} X_{ni} \longrightarrow 0$ *completely.*

(iii) $\sum_{i=1}^{k_n} a_{ni} X_{ni} \longrightarrow 0$ *almost surely.*

(iv) $\sum_{i=1}^{k_n} a_{ni} X_{ni} \longrightarrow 0$ *in probability.*

Proof. Note that in this case

$$\frac{A}{B} ||a_{ni} X_{ni}|| \le \frac{\psi(||X_{ni}||)}{\psi(|a_{ni}|^{-1})}.$$

For $(i) \Longrightarrow (ii)$ we refer to Corollary 4.7 in Hu, Rosalsky, Szynal and Volodin (1999).

To see that $(iv) \Longrightarrow (i)$, notice that $E||\sum_{i=1}^{k_n} a_{ni}X_{ni}|| \leq \sum_{i=1}^{k_n} |a_{ni}|E||X_{ni}|| \longrightarrow 0$, so (i) holds.

Similarly, it is easy to check the following result on complete convergence in a Banach space of type $1 \leq r \leq 2$.

Corollary 5.2. *Let $k_n \to \infty$ be a sequence of positive integers. Let $\{X_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ be an array of rowwise independent $\mathcal{B}$-valued random elements, and $\{a_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ an array of constants. Let $\psi : (0, +\infty) \longrightarrow (0, +\infty)$ be a function such that $Ax \leq \psi(x) \leq Bx$ for all $x \in (0, +\infty)$, for some constants $A, B > 0$. Assume that $EX_{ni} = 0$ for all $1 \leq i \leq k_n$ and $n \geq 1$, and $\mathcal{B}$ is of type r $(1 \leq r \leq 2)$. Assume (4.4) holds. Then*

$$\sum_{i=1}^{k_n} a_{ni}X_{ni} \longrightarrow 0 \text{ completely.}$$

5.3 Consistency of Bootstrapped Means in the Banach Space Setting

In order to prove the main result of this section, we need the following lemma:

Lemma 5.3. *If $s > 0$ then, for almost every $\omega \in \Omega$,*

$$E||\hat{X}_{n,1}^{\omega} - \bar{X}_n(\omega)||^s \leq A_s \left[\frac{1}{n}\sum_{i=1}^{n} ||X_i(\omega)||^s + ||\bar{X}_n(\omega)||^s\right],$$

where $A_s = 2^{s-1}$ for $s \geq 1$ and $A_s = 1$ for $0 < s < 1$.

Proof. For almost every $\omega \in \Omega$,

$$E||\hat{X}_{n,j}^{\omega} - \bar{X}_n(\omega)||^s = \frac{1}{n}\sum_{i=1}^{n} ||X_i(\omega) - \bar{X}_n(\omega)||^s$$

$$\leq A_s \left[\frac{1}{n} \sum_{i=1}^{n} \left(||X_i(\omega)||^s + ||\bar{X}_n(\omega)||^s \right) \right],$$

by the c_r-inequalities.

Theorem 5.6. *Let $\{X_n, n \geq 1\}$ be a sequence of random elements taking values in a real separable Banach space. Let $k_n \to \infty$ be a sequence of positive integers, and let $\{a_{nk}, 1 \leq k \leq k_n, n \geq 1\}$ be an array of real constants. Let $1 \leq r \leq 2$. Define $h_n = \sum_{k=1}^{k_n} |a_{nk}|^r$, for every $n \geq 1$. Let $\{b_n, n \geq 1\}$ and $\{d_n, n \geq 1\}$ be sequences of positive constants. Suppose that:*

(1) $\sup_{n\geq 1} \frac{1}{d_n} ||\bar{X}_n|| < \infty$ a.e. and $\sup_{n\geq 1} \frac{1}{b_n} \sum_{i=1}^{n} ||X_i||^r < \infty$ a.e.,

(2) $\sum_{n=1}^{\infty} \frac{h_n b_n}{n} < \infty$ and $\sum_{n=1}^{\infty} h_n d_n^r < \infty$,

(3) The bootstrapped mean is weakly consistent with respect to $\{a_{nk}, 1 \leq k \leq k_n, n \geq 1\}$, that is, for almost every $\omega \in \Omega$

$$||\sum_{k=1}^{k_n} a_{nk} \left(\hat{X}^{\omega}_{n,k} - \bar{X}_n(\omega) \right) || \xrightarrow{P} 0.$$

Then the bootstrapped mean is strongly consistent with respect to $\{a_{nk}, 1 \leq k \leq k_n, n \geq 1\}$, that is, for almost every $\omega \in \Omega$ and all $\epsilon > 0$,

$$\sum_{n=1}^{\infty} P \left\{ ||\sum_{k=1}^{k_n} a_{nk} \left(\hat{X}^{\omega}_{n,k} - \bar{X}_n(\omega) \right) || > \epsilon \right\} < \infty.$$

Proof. If we consider the function $\psi(t) = t^r$, the condition (5.5) in Theorem 5.4 becomes $\sum_{n=1}^{\infty} \sum_{k=1}^{k_n} |a_{nk}|^r E||X_{nk}||^r < \infty$, which implies condition (5.6) with $s = 1$. Therefore, we need only check condition (5.5) for the array $\{Z_{nk} = \hat{X}^{\omega}_{n,k} - \bar{X}_n(\omega), 1 \leq k \leq k_n, n \geq 1\}$ with $\psi(t) = t^r$.

An application of Lemma 5.3 yields for almost every $\omega \in \Omega$:

$$\sum_{n=1}^{\infty} \sum_{k=1}^{k_n} |a_{nk}|^r E||Z_{nk}||^r = \sum_{n=1}^{\infty} \sum_{k=1}^{k_n} |a_{nk}|^r E||\hat{X}^{\omega}_{n,k} - \bar{X}_n(\omega)||^r$$

$$\leq 2^{r-1} \sum_{n=1}^{\infty} \sum_{k=1}^{k_n} |a_{nk}|^r \left[\frac{1}{n} \sum_{i=1}^{n} ||X_i(\omega)||^r + ||\bar{X}_n(\omega)||^r \right]$$

$$= 2^{r-1} \left(\sum_{n=1}^{\infty} (\frac{1}{n} \sum_{i=1}^{n} ||X_i(\omega)||^r) \sum_{k=1}^{k_n} |a_{nk}|^r + \sum_{n=1}^{\infty} ||\bar{X}_n(\omega)||^r \sum_{k=1}^{k_n} |a_{nk}|^r \right)$$

$$= 2^{r-1} \left(\sum_{n=1}^{\infty} \frac{h_n}{n} \sum_{i=1}^{n} ||X_i(\omega)||^r + \sum_{n=1}^{\infty} h_n ||\bar{X}_n(\omega)||^r \right)$$

$$= 2^{r-1} \left(\sum_{n=1}^{\infty} \frac{h_n b_n}{n} \left[\frac{1}{b_n} \sum_{i=1}^{n} ||X_i(\omega)||^r \right] + \sum_{n=1}^{\infty} h_n d_n^r \left[\frac{1}{d_n} ||\bar{X}_n(\omega)|| \right]^r \right) < \infty$$

in view of (1) and (2).

By applying Theorem 5.4, we have that $\sum_{k=1}^{k_n} a_{nk} Z_{nk} \longrightarrow 0$ completely, that is, the bootstrapped mean is strongly consistent with respect to $\{a_{nk}, 1 \leq k \leq k_n, n \geq 1\}$.

Remark 5.3. In the particular case $a_{nk} = \dfrac{1}{a_n}$ for all $1 \leq k \leq k_n$ and every $n \geq 1$, where $\{a_n, n \geq 1\}$ is a sequence of positive constants, we have Theorem 3 of Ahmed, Hu and Volodin (2001), with $q = r \in [1, 2]$ and $c_n = 1$ for every $n \geq 1$.

5.4 Bootstrap of the Mean for Random Variables

The main result of this section is given in Theorem 5.7 below. The main thrust and unusual feature of Theorem 5.7 is that no assumptions are required concerning marginal and joint distributions of the random variables $\{X_n\}$. Not only that, it is not assumed that these random variables are either independent or identically distributed. In general, no moment conditions are imposed on the random variables $\{X_n\}$.

Theorem 5.7. *Let $\{X_n, n \geq 1\}$ be a sequence of random variables and let $\{a_n, n \geq 1\}$ be a sequence of positive constants. Suppose there exists $J \geq 1$ such that*

$$\frac{\max_{1 \leq i \leq n} |X_i|}{k_n a_n} \to 0 \ \text{a.s. as } n \to \infty \text{ and} \tag{5.7}$$

$$\sum_{n=1}^{\infty} n^r \left(\frac{\sum_{i=1}^{n} X_i^2}{a_n^2 k_n n} \right)^J < \infty \ \text{a.s.} \tag{5.8}$$

Then the bootstrapped mean is strongly consistent, that is, for almost all $\omega \in \Omega$ and all $\epsilon > 0$,

$$\sum_{n=1}^{\infty} n^r P\left\{ |\sum_{k=1}^{k_n} \left(\hat{X}^{\omega}_{n,k} - \bar{X}_n(\omega) \right) | > \epsilon a_n k_n \right\} < \infty.$$

Proof. We need only check the conditions of Theorem 4.2 for the array

$$\left\{ Y_{nk} = \frac{\hat{X}^{\omega}_{n,k} - \bar{X}_n(\omega)}{k_n a_n}, 1 \le k \le k_n, n \ge 1 \right\}.$$

These conditions may be rewritten as follows.

Condition 1: $\sum_{n=1}^{\infty} n^r \sum_{k=1}^{k_n} P\{|\hat{X}^{\omega}_{n,k} - \bar{X}_n(\omega)| > \epsilon k_n a_n\} < \infty$ for all $\epsilon > 0$.

Condition 2: There exists $J \ge 1$ such that

$$\sum_{n=1}^{\infty} n^r \left(\sum_{k=1}^{k_n} \frac{E|\hat{X}^{\omega}_{n,k} - \bar{X}_n(\omega)|^2}{(a_n k_n n)^2} \right)^J < \infty.$$

Further, for the first condition,

$$|\hat{X}^{\omega}_{n,k} - \bar{X}_n(\omega)| \le 2 \max_{1 \le i \le n} |X_i(\omega)|,$$

and for the second one, since

$$E|\hat{X}^{\omega}_{n,k} - \bar{X}_n(\omega)|^2 = \frac{1}{n} \sum_{i=1}^{n} |X_i(\omega) - \bar{X}_n(\omega)|^2$$
$$\le \frac{4}{n} \sum_{i=1}^{n} X_i(\omega)^2,$$

we have

$$\sum_{n=1}^{\infty} n^r \left(\sum_{k=1}^{k_n} \frac{E|\hat{X}^{\omega}_{n,k} - \bar{X}_n(\omega)|^2}{(a_n k_n n)^2} \right)^J$$
$$\le C \sum_{n=1}^{\infty} n^r \left(\frac{\sum_{i=1}^{n} X_i^2}{a_n^2 k_n n} \right)^J < \infty \text{ a.s.}$$

where C is a positive constant.

Now we wish to make the following two remarks.

Remark 5.4. If $k_n a_n \to \infty$ monotonically, then (5.7) is equivalent to the structurally simpler condition

$$\lim_{n\to\infty} \frac{X_n}{k_n a_n} = 0 \text{ a.s.} \tag{5.9}$$

Indeed, let $k_n a_n \to \infty$ monotonically and (5.7) holds. Then, for arbitrary $n \geq k \geq 2$,

$$\begin{aligned} \frac{\max_{1\leq i\leq n}|X_i|}{k_n a_n} &\leq \frac{\max_{1\leq i\leq k-1}|X_i|}{k_n a_n} + \frac{\max_{k\leq i\leq n}|X_i|}{k_n a_n} \\ &= \frac{\max_{1\leq i\leq k-1}|X_i|}{k_n a_n} + \max_{k\leq i\leq n}\frac{|X_i|}{k_n a_n} \\ &\leq \frac{\max_{1\leq i\leq k-1}|X_i|}{k_n a_n} + \sup_{i\geq k}\frac{|X_i|}{k_n a_n} \to 0, \end{aligned}$$

as first $n \to \infty$ and then $k \to \infty$. It is noted that the reverse implication is evident.

Remark 5.5. It appears that the comparison of Theorem 5.7 of this section to Theorem 2.1 of Li, Rosalsky and Ahmed (1999) may not be possible. One simple argument is that the difference in the assumptions of both theorems does not provide for comparative analysis. Li, Rosalsky and Ahmed (1999) assumed the convergence of partial sums. On the other hand we use only boundness of partial sums.

Finally, we wish to provide another generalization of the main result of Hu and Taylor (1997). We give the proof for pairwise i.i.d. random variables. Recall that Hu and Taylor (1997) only considered the i.i.d. case in their publication. In addition, the result in this investigation is sharper in the sense that it establishes convergence rates which ware not given in Hu and Taylor (1997). In their paper only a.s. convergence results were stated. Furthermore, our proof may be simpler than that of Hu and Taylor (1997).

Corollary 5.3. *Let $\{X_n, n \geq 1\}$ be pairwise independent identically distributed random variables with $E|X|^{1+\delta} < \infty$ for some $\delta > 0$ and $EX = \mu$. Then for all $\epsilon > 0$, for almost all $\omega \in \Omega$ and any real number r:*

$$\sum_{n=1}^{\infty} n^r P\left\{ |\sum_{k=1}^{n} \left(\hat{X}_{n,k}^{\omega} - \mu \right) | > \epsilon n \right\} < \infty.$$

Proof. Note that it is sufficient to show the result for $r \geq -1$, since for $r < -1$ the result is obvious. Let $a_n = 1$ and $k_n = n$.

In order to prove (5.9), we can write for arbitrary $\epsilon > 0$

$$\sum_{n=1}^{\infty} P\{|X_n| > \epsilon n\} \leq CE|X_1| < \infty.$$

Thus, by the Borel-Cantelli lemma,

$$\lim_{n\to\infty} X_n/n = 0, \text{ a.s.}$$

To prove (5.8), note that $\{X_n^2, n \geq 1\}$ are also pairwise independent identically distributed random variables. Recall that $E|X|^{1+\delta} < \infty$ and, by an application of Petrov (1996) results, it can be shown that

$$\frac{1}{n^{2/(1+\alpha)}} \sum_{i=1}^{n} X_i^2 \to 0, \text{ a.s., where } 0 < \alpha < \delta.$$

Now let $J > \frac{(1+\alpha)(1+r)}{2\alpha}$. Then

$$\sum_{n=1}^{\infty} n^r \left(\frac{1}{n^2} \sum_{i=1}^{n} X_i^2 \right)^J \leq \left(\sup_{n\geq 1} \frac{1}{n^{\frac{2}{1+\alpha}}} \sum_{i=1}^{n} X_i^2 \right)^J \sum_{n=1}^{\infty} \frac{1}{n^{\frac{2\alpha J}{1+\alpha} - r}} < \infty, \text{ a.s.}$$

Hence, by Theorem 5.7, for almost all $\omega \in \Omega$ and all $\epsilon > 0$

$$\sum_{n=1}^{\infty} n^r P\left\{ |\sum_{k=1}^{n} \left(\hat{X}_{n,k}^{\omega} - \bar{X}_n(\omega) \right) | > \epsilon n \right\} < \infty.$$

Further, by an application of Etemadi's (1981) strong law of large numbers, we have that $\bar{X}_n \to \mu$ a.s.. Hence, for almost all $\omega \in \Omega$ and all $\epsilon > 0$ there exists $N = N(\epsilon, \omega)$ such that, for all $n \geq N$, we have $\bar{X}_n - \mu < \epsilon/2$. Then

$$\begin{aligned}
& \sum_{n=1}^{\infty} n^r P\left\{ |\sum_{k=1}^{n} \left(\hat{X}_{n,k}^{\omega} - \mu \right) | > \epsilon n \right\} \\
\leq & \sum_{n=1}^{\infty} n^r P\left\{ |\sum_{k=1}^{n} \left(|\hat{X}_{n,k}^{\omega} - \bar{X}_n(\omega)| + |\bar{X}_n(\omega) - \mu| \right) > \epsilon n \right\} \\
\leq & \sum_{n=1}^{\infty} n^r P\left\{ |\sum_{k=1}^{n} \left(|\hat{X}_{n,k}^{\omega} - \bar{X}_n(\omega)| \right) > \epsilon n/2 \right\} < \infty.
\end{aligned}$$

Chapter 6

An Improvement of Kolmogorov Exponential Inequality for Negatively Dependent Random Variables

6.1 Introduction

Definition 6.1. The random variables $X_1, \cdots, X_n$ are said to be *negatively dependent* if we have

$$P\left\{\cap_{j=1}^{n}(X_j \leq x_j)\right\} \leq \Pi_{j=1}^{n} P\left\{X_j \leq x_j\right\},$$

and

$$P\left\{\cap_{j=1}^{n}(X_j \geq x_j)\right\} \leq \Pi_{j=1}^{n} P\left\{X_j \geq x_j\right\}$$

for all real $x_1, \cdots, x_n$.

6.2 Preliminary Results

The following two lemmas are used to obtain the main result in the next section. The first lemma is a simple corollary of the observation that if $X_1, \cdots, X_n$ is a sequence of negatively dependent random variables, then $e^{X_1}, \cdots, e^{X_n}$ are also negatively dependent. The same argument is used in Bosorgnia, Patterson and Taylor (1996) p.1167.

Lemma 6.1. *If $X_1, \cdots, X_n$ is a sequence of negatively dependent random variables, then*

$$E \exp\left\{\sum_{k=1}^{n} X_k\right\} \leq \Pi_{k=1}^{n} E \exp\{X_k\}.$$

The second lemma is only a technical result that will help us to improve a constant in the Kolmogorov exponential inequality.

Lemma 6.2. *Let $a > 0$ and $0 < \alpha \leq \frac{a^3}{2(e^a - 1 - a - a^2/2)}$. Then*

$$e^x - 1 - x - \frac{x^2}{2} \leq \frac{x^3}{2\alpha}$$

for all $0 \leq x \leq a$.

Proof. Consider the function

$$f(x, \alpha) = \ln\left(1 + x + \frac{x^2}{2} + \frac{x^3}{2\alpha}\right) - x.$$

We need to prove that $f(x, \alpha) \geq 0$ for all $0 < \alpha \leq \frac{a^3}{2(e^a - 1 - a - a^2/2)}$ and $0 \leq x \leq a$.

Take the derivative

$$\frac{\partial f}{\partial x} = -\frac{x^2\left(x - (3 - \alpha)\right)}{2\alpha\left(1 + x + x^2/2 + x^3/(2\alpha)\right)}.$$

Hence f is increasing with respect to x on the interval $(0, 3 - \alpha)$ and decreasing on the interval $(3 - \alpha, a)$.

Note that $f(0, \alpha) = 0$ and $f(a, \alpha) \geq 0$ since $\alpha \leq \frac{a^3}{2(e^a - 1 - a - a^2/2)}$.

6.3 Main Result

Now we can formulate and prove the main result of this section.

Theorem 6.1. *Let $X_1, \cdots, X_n$ be a sequence of negatively dependent random variables with zero means and finite variances. Let $s_n^2 = \sum_{k=1}^{n} EX_k^2$ and assume that $|X_k| \leq Cs_n$ almost surely for each $1 \leq k \leq n$ and $n \geq 1$. Then for each $a > 0$ and $n \geq 1$, the assumptions $\epsilon C \leq a$ and $0 < \alpha \leq \frac{a^3}{2(e^a-1-a-a^2/2)}$ imply that*

$$P\{S_n/s_n > \epsilon\} \leq \exp\left\{-\frac{\epsilon^2}{2}\left(1 - \frac{\epsilon C}{\alpha}\right)\right\},$$

where $S_n = \sum_{k=1}^{n} X_k$, as usual.

Proof. We will follow the proof of the classical Kolmogorov exponential inequality, cf. Stout (1974), p.263. Fix $n \geq 1$ and $a > 0$. Suppose $x = \epsilon C \leq a$. For each $1 \leq k \leq n$, since all series are absolutely convergent we can write

$$\begin{aligned}
E\exp\{\epsilon X_k/s_k\} &= 1 + \frac{\epsilon^2 EX_k^2}{2!s_n^2} + \frac{\epsilon^3 EX_k^3}{3!s_n^3} + \cdots \\
&\leq 1 + \frac{\epsilon^2 EX_k^2}{2s_n^2}\left(1 + \frac{\epsilon C}{3} + \frac{\epsilon^2 C^2}{3\cdot 4} + \cdots\right) \\
&\leq 1 + \frac{\epsilon^2 EX_k^2}{2s_n^2}\left(1 + \frac{x}{3} + \frac{x^2}{3\cdot 4} + \cdots\right) \\
&= 1 + \frac{\epsilon^2 EX_k^2}{2s_n^2}\left(e^x - x - \frac{x^2}{2}\right) \\
&\leq 1 + \frac{\epsilon^2 EX_k^2}{2s_n^2}\left(1 + \frac{x}{\alpha}\right) \text{ by Lemma 6.2} \\
&\leq \exp\left\{\frac{\epsilon^2 EX_k^2}{2s_n^2}\left(1 + \frac{x}{\alpha}\right)\right\}
\end{aligned}$$

since $1 + t \leq e^t$ for all t. By Lemma 6.1,

$$E\exp\{\epsilon S_n/s_n\} \leq \exp\left\{\frac{\epsilon^2}{2}\left(1 + \frac{\epsilon C}{\alpha}\right)\right\}.$$

Thus

$$P\{S_n/s_n > \epsilon\} \leq \exp\{-\epsilon^2\} E \exp\{\epsilon S_n/s_n\}$$
$$\leq \exp\left\{-\frac{\epsilon^2}{2}\left(1 - \frac{\epsilon C}{\alpha}\right)\right\}.$$

Remark 6.1. Even for $a = 1$ our Theorem 6.1 gives better constant $\alpha = \frac{1}{2e-5} = 2.2906 > 2$, while in the classical Kolmogorov's inequality we have $\alpha = 2$ (cf. Stout, (1974) p.263).

Remark 6.2. If $a \to 0$ then $\alpha = \frac{a^3}{2(e^a-1-a-a^2/2)} \to 3$. As $a \to \infty$ then $\alpha \to 2$. We need $a \to 0$ for the proof of the law of iterated logarithm.

Remark 6.3. Another interesting advantage of Theorem 6.1 is that we can consider any positive a, while $a = 1$ in the usual Kolmogorov inequality. In our inequality, the upper bound involves a fixed (given) C and fixed ϵ and a variable α. Now, α is a function of a, for $a \geq \epsilon C$. Note that the left-hand side of the inequality doesn't involve a anywhere, whereas the right-hand side is, in effect, a function of a. So the best possible inequality occurs when a is chosen so that $\alpha(a)$ is maximized on the interval $[\epsilon, \infty]$. However, α is a decreasing function of a, so the best value of the upper bound occurs when a is as small as possible; i.e., when $a = \epsilon C$.

In short, then, technically we have a family of inequalities – one for each value of $a \geq \epsilon C$. However, the special case where $\alpha = \alpha(\epsilon C)$ implies the validity of the inequality for all larger values of α – so there is really only one inequality, for one specific value of α.

Remark 6.4 It is interesting to obtain another exponential inequalities (Bernstein's, Prohorov's, Hoeffding's, Bennett's, Fuk-Nagaev's and etc.) for negatively dependent random variables.

Chapter 7

Conclusion and Outlook for Future Research

We conclude this dissertation by showcasing some open problems and providing some general comments.

The asymptotic analysis of the confidence coefficient of the interval estimate of the mean vector with center at the James-Stain estimator is developed in Chapter 2. Most importantly, the method and its technique lead to results that are quite general in nature. A novel aspect of the derivation is that we were able to find limits of integration in the double integral and this tells us what the confidence coefficient is. Consequently, formal calculations for asymptotic analysis are reduced to asymptotic of roots of a sufficiently simple algebraic equation. Of course, the method we developed can be successfully used for other statistical models in which we can find the James-Stein phenomena. An interesting area will be the application of the results we obtained to multiple regression problems.

The results of Chapter 3, that is, the investigations of point properties of different methods of estimation of Birnbaum - Sounders distribution parameters ensure us that maximum likelihood estimators have substantial advantages in comparison with other estimators, disregarding the difficulties connected with their computations. This result is important since we estimate the parameters that reflect a physical nature of an object, which is analyzed from a statistical point of view by this distribution. In connection with this, one could continue investigations of asymptotic behavior of the Fisher information matrix when the parameter λ tends to infinity. Moreover, we considered a class of new estimators, so-called regression-quartile (least square) estimators. This method is based on the regression analysis of sample quartiles.

In Chapter 4 we suggested the application of weighted likelihood to robust estimation. We think that this problem is methodologically important and could be of paramount interest in statistical practice. It demonstrates the nature of robustness of such estimators as the α-trimmed mean and explains the point of the matter of outliers and, namely, strongly outstanding observations, and supports the idea of the necessity of their removal from the sample and further investigation. For the sake of brevity, we considered the exponential model. It will be exceedingly interesting to continue the investigation of robust properties of weighted likelihood estimators in general for obstructed models or other gross error models. It is a necessity to investigate the behavior of the breakdown point and analyze Hampel's influence function for 0 - 1 weights that depend on sample values. That is, we plan to investigate all robustness properties mentioned in Huber's (1981) book. Finally, it is interesting to consider such types of estimators for different obstructing models, especially for normal distributions with obstruction, for both the mean and the variance.

One of the most interesting and useful examples of negative dependent random variables arrises in the situation of the sample from a finite population without replacement. Hence we can apply the results of Chapters 5 and 6 to the so-called *dependent bootstrap*, that is, a sample drawn without replacement from a collection of items made up of copies of sample observations. We think that the dependent bootstrap is only one, of course very interesting, application of the notion of negative dependence. Another one is to apply it to the limit theorems.

REFERENCES

de Acosta, A. (1981). Inequalities for B-valued random vectors with applications to the strong law of large numbers. *Ann. Probab.* **9**, 157-161.

Ahmad, I. A. (1988). Jackknife estimation for a family of life distribution. *J.Statist. Comput. and Simul.* **29**, 211-223.

Ahmed, S. E. (1998a). Improved pretest nonparametric estimation in a multivariate regression model. *Communications in Statistics: Theory and Methods* **27**, 2391-2421.

Ahmed, S. E. (1998b). Large-Sample Estimation Strategies for Eigenvalues of a Wishart Matrix. *Metrika* **47**, 35-45.

Ahmed, S. E. (1999a). Simultaneous estimation of survivor functions in exponential lifetime models. *Journal of Statistical Computation and Simulation* **63**, 235-261.

Ahmed, S. E. (2000a). Construction of improved estimators of multinomial proportions. *Communications in Statistics: Theory and Methods* **29**, 1273-1291.

Ahmed, S. E. (2001). Shrinkage estimation of regression coefficients from censored data with multiple observations. In: Empirical Bayes and Likelihood Inference (Montreal, QC, 1997), 103–120, *Lecture Notes in Statist.*, **148**, Springer, New York.

Ahmed, S. E. and Basu, A. K. (2000b). Least squares, preliminary test and Stein-type estimation in general vector $AR(p)$ models. *Statistica Neerlandica* **54**, 47-66.

Ahmed S. E., Gupta A. K., Khan, S. M., and Nicol C. J. (2001b). Simultaneous estimation of several intraclass correlation coefficients. *Ann. Inst. Statist. Math.* **53**, 354-369.

Ahmed, S.E., Hu, T.C. and Volodin, A.I. (2001). On the rate of convergence of bootstrapped means in a Banach space. *Internat. J. Math. & Math. Sci.* **25**, 629-635.

Ahmed, S. E. and Khan, S. M. (2002). Using Several Data To Structure Efficient Estimation of Intraclass Correlation Coefficients. Measurement and Multivariate Analysis, Springer-Verlag: Tokyo.

Ahmed,S.E., Li, D., Rosalsky, A., and Volodin, A. (2001). Almost sure Lim Sup behavior of bootstrapped means with applications to pairwise i.i.d. sequences and stationary ergodic sequences, *Journal of Statistical Planning and Inference* **98**, 1-14.

Ahmed, S. E. and Saleh, E. (1999b). Improved nonparametric estimation of location vector in a multivariate regression Model, *Nonparametric Statistics* **11**, 51-78.

Ahmed, S. E. and Ullah, B. (1999c). Improved biased estimation in an ANOVA model.*Linear Algebra and its Application* **289**, 3-24.

Ahmed, S. E. and Ullah, B. (1999d). To pool or not to pool: The multivariate data. *Sankhya* **61**, Series B, 266-288.

Ahmed, S. E. and Volodin A.(2001). The convergence rates of the bootstrap mean. Data Analysis from Statistical Foundations. A Festschrift in honour of the 75th birthday of D.A.S. Fraser. A.K.Md.E.Saleh (Editor). Nova Science Publications, Inc. Huntington, New York, 207-213.

Arenal-Gutiérrez, E., Matrán, C. and Cuesta-Albertos, J.A. (1996). On the unconditional strong law of large numbers for the bootstrap mean. *Statist. Probab. Lett.* **27**, 49-60.

Athreya, K.B. (1983). Strong law for the bootstrap. *Statist. Probab. Lett.* **1**, 147-150.

Berger, E. (1991). Majorization, exponential inequalities and almost sure behavior of vector-valued random variables. *Ann. Probab.* **19** 1206-1226.

Bickel, P.J. and Freedman D.A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9**, 1196-1217.

Birnbaum, Z. W., Sounders, S. C. (1969a). A new family of the life distribution, *J.Appl. Probab.* **6**, 319-327.

Birnbaum, Z. W., Sounders, S. C. (1969b). Estimation for a family of life distribution with application to fatigue. *J.Appl. Probab.* **6**, 328-337.

Bogdanoff, D., Kozin F. (1985). Probabilistic models of cumulative damage. John Wiley and Sons, New-York - Chichester.

Bosorgnia, A., Patterson, R.F. and Taylor, R.L. (1997). Chung type strong laws for arrays of random elements and bootstrapping. *Stochastic Anal. Appl.* **15**, 651-669.

Bozorgnia, A., Patterson, R.F. and Taylor, R.L. (1996). Limit theorems for dependent random variables. Proceedings of WCNA'92, Tampa, Florida, 1639 - 1650.

Choi, B.D. and Sung, S.H. (1988). On Chung's strong law of large numbers in general Banach spaces. *Bull. Austral. Math. Soc.* **37**, 93-100.

Choi, B.D. and Sung, S.H. (1989). On Teicher's strong law of large numbers in general Banach spaces. *Probab. Math. Statist.* **10**, 137-142.

Chung, K.L. (1947). Note on some strong laws of large numbers. *Amer. J. Math.* **69**, 189-192.

Cramér G. (1946). Mathematical methods of statistics. Princeton Mathematical Series, vol. 9. Princeton University Press, N.J.

Csörgő, S. (1992). On the law of large numbers for the bootstrap mean. *Statist. Probab. Lett.* **14**, 1-7.

Desmond A. (1986). On the relationship between two fatigue-life models. *IEEE Trans. in Reliability* **35**, 167-169.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7**, 1-26.

Efron, B., Morris, C. (1972). Empirical Bayes on vector observations: an extension of Stein's method. *Biometrika* **59**, 335-347.

Etemadi, N. (1981). An elementary proof of the strong law of large numbers. *Z. Wahrsch. Verw. Gebiete* **55**, 119-122.

Fan, J., Heckman, N.E., Wand, W.P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *J. Amer. Statist. Assos.* **90**, 826-838.

Giné, E. and Zinn, J. (1989). Necessary conditions for the bootstrap of the mean. *Ann. Statist.* **17**, 684-691.

Hsu, P.L. and Robbins, H. (1947). Complete convergence and the law of large numbers. *Proc. Nat. Acad. Sci. U.S.A.* **33**, 25-31.

Hu, F. (1994). Relevance weighted smoothing and a new bootstrap method. Technical report, Department of Statistics, University of British Columbia.

Hu, F. (1997). Asymptotic properties of relevance weighted likelihood estimations. *Canad. J. Statist.* **25**, 45-60.

Hu, F., Zidek, J.V. (2001). The relevance weighted likelihood with applications. In: Ahmed, S.E., Reid, N. (Eds.) Empirical Bayes and Likelihood Inference, Springer Verlag, New York, 211-235.

Hu, T.-C., Rosalsky, A., Szynal, D., Volodin, A. (1999). On complete convergence for arrays of rowwise independent random elements in Banach spaces. *Stochastic Anal. Appl.* **17**, 963-992.

Hu, T.-C., Szynal, D. and Volodin, A.I. (1998). A note on complete convergence for arrays. *Statist. Probab. Lett.* **38**, 27-31.

Hu, T.C. and Taylor, R.L. (1997). On the strong law for arrays and for the bootstrap mean and variance. *Internat. J. Math. & Math. Sci.* **20**, 375-382.

Huber, P.J. (1981). Robust Statistics. John Wiley & Sons, New York, Chichester, Brisbane, Toronto.

Hwang, J.T. and Casella, G. (1982). Minimax confidence sets for the mean of a multivariate normal distribution. *Ann. Statist.* **10**, 868-881.

James, W. and Stein, C. (1961) Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1**, 361 -380.

Joshi, V.M. (1967). Inadmissibility of the usual confidence sets for the mean of a multivariate normal population. *Ann. Math. Statist.* **38**, 1868-1875.

Kubokawa, T. (1998). The Stein phenomenon in simultaneous estimation: a review. In: Applied Statistical Science, III, 143–173, Nova Sci. Publ., Commack, NY.

Kuczmaszewska, A. and Szynal, D. (1994). On complete convergence in a Banach space. *Internat. J. Math. & Math. Sci.* **17**, 1-14.

Kuelbs, J. and Zinn, J. (1979). Some stability results for vector valued random variables. *Ann. Probab.* **7**, 75-84.

Lehmann, E. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37**, 1137 - 1153.

Lehmann, E. (1983). Theory of Point Estimation, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. A Wiley Publication in Mathematical Statistics. John Wiley & Sons, Inc., New York.

Li, D., Rosalsky, A. and Ahmed, S.E. (1999). Complete convergence of bootstrapped means and moments of the supremum of normed bootstrapped sums. *Stochastic Analysis and Applications* **17**, 799-814.

Ordóñez Cabrera, M. and Sung, S.H. On complete convergence of weighted sums of random elements. To appear in *Stochastic Anal. Appl.*

Parzen E. (1967). On models for the probability of fatigue failure of a structure. *Time Series Analysis Papers*, Holden Day, San Francisco, 532-548.

Petrov, V.V. (1996). On the strong law of large numbers. *Statist. Probab. Lett.* **26**, 377-380.

Rieck J. R. (1995). Parameter estimation for the Birnbaum – Sounders distribution based on symmetrically censored samples. *Commun.Statist. Theory and Meth.* **24**, 1721-1736.

Robbins, H. (1983). Some thoughts on empirical Bayes estimation. *Ann. Statist.* **11**, no. 3, 713-723.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap, *Ann. Statist.* **9**, 1187-1195.

Singpurwalla, N.D. (1995). Survival in dynamic environment. *Statistical Sci.* **10**, 86-103.

Smith, W.S. and Taylor, R.L. (2000). Consistency of dependent bootstrap estimators. UGA Technical Report 2000-6.

Staniswalis, J.G. (1989). The kernel estimate of a regression function in likelihood-based models. *J. Amer. Statist. Assoc.* **84**, 276-283.

Stein, C. (1956). Inadmissibility of the usual estimator of the mean of a multivariate normal distribution. In *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1**, 197-206.

Stein, C. (1962). Confidence sets for the mean of a multivariate normal distribution. *J.Roy. Statist. Soc., Ser. B* **24**, 265-296.

Stigler, S.M. (1990). The 1988 Neyman Memorial Lecture: a Galtonian perspective on shrinkage estimators. *Statist. Sci.* 5, no. 1, 147-155.

Stout, W. F. (1974). Almost Sure Convergence. Academic Press, New York.

Sung, S.H. (1997). Complete convergence for weighted sums of arrays of rowwise independent B-valued random variables. *Stochastic Anal. Appl.* **15**, 255-267.

Sung, S.H. (2000). Complete convergence for sums of arrays of random elements. *Internat. J. Math. & Math. Sci.* **23**, 789-794.

Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation *J. Amer. Statist. Assoc.* **82**, 559-567.

Teicher, H. (1979). Generalized exponential bounds, iterated logarithm and strong laws *Z. Wahr. verw. Geb.* **48**, 293-307.

Tomkins, R.J. (1978). On the law of iterated logarithm *Ann. Probab.* **6**, 162-168.

Volodin, A. (2002). On the Kolmogorov exponential inequality for negatively dependent random variables. To appear in *Pakistan Journal of Statistics.* Special issue in honor of Professor S.E. Ahmed.

Wang, X.C., Rao, M.B. and Yang, X. (1993). Convergence rates on strong laws of large numbers for arrays of rowwise independent elements, *Stochastic Anal. Appl.* **11**, 115-132.