ABSTRACT

WILSON, DAVID HUNT. Signed Scale Measures: An Introduction and Application. (Under the direction of Dennis Boos and Jacqueline Hughes-Oliver.)

The role of the Interquartile Range in constructing a boxplot provides the rationale for considering the halves of the "box" in a boxplot. The two halves of the boxplot are viewed as measures of distance from a measure of location. This viewpoint is the genesis for considering a new class of parameters called signed scale parameters.

A conceptual framework for signed scale parameters is introduced and four classes of signed scale parameters are discussed in detail. The small sample and asymptotic behaviors for several signed scale estimators are examined for nine distributions.

A new boxplot construction rule that uses one pair of signed scale estimators is introduced. The common, skew adjusted, and signed boxplots are compared with respect to their propensity to label observations as "outliers" and their ability to provide skewness information.

SIGNED SCALE MEASURES: AN INTRODUCTION AND APPLICATION

by

David Hunt Wilson

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

DEPARTMENT OF STATISTICS

Raleigh
2002

APPROVED BY:

DENNIS BOOS Co-chair of Advisory Committee	JACQUELINE HUGHES-OLIVER Co-chair of Advisory Committee
LEONARD STEFANSKI	JACQUELIN DIETZ

For My Parents

Biography

David Hunt Wilson was born September 28, 1969 in Columbia, South Carolina. He received his B.S. degree in actuarial science from the University of North Carolina at Chapel Hill in 1991. He started his graduate work in Statistics at Duke University in 1991 but transferred to the University of Kentucky in 1993 where he received an M.S. degree in Statistics in 1995. He then joined the graduate program in Statistics at North Carolina State University in January 1996. Upon completion of his doctoral degree, David will become a full-time Senior Statistician at the Research Triangle Institute.

Acknowledgement

I want to thank Dr. Dennis Boos and Dr. Jacqueline M. Hughes-Oliver for their mentoring, motivation, and patience while I researched and prepared my dissertation. I would also like to thank the other members of my advisory committee, Dr. Leonard Stefanski and Dr. Jackie Dietz, for their help with problems I encountered while performing my research. I would like to express my gratitude to my parents, Tom and Sandra Wilson, for their support and encouragement during my long trek through graduate school.

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

The discipline of Statistics is often used to make comparisons and to examine data. An investigator may be interested in comparing the characteristics of different populations or comparing the behavior of one statistic relative to another. One particular method of investigation and comparison is known as the boxplot and was introduced by Tukey (1977). The boxplot is a graphical method of summarizing data and is used to quickly gain information about location, spread, and skewness.

Various modifications of Tukey's boxplot have been introduced over the years, yet most of them rely on one particular notion of scale, the interquartile range (IQR), which is the difference between the third and first quartiles. While the IQR is relatively easy to calculate, especially for small sample sizes, it is most often a less efficient measure of scale than other scale estimators such as the standard deviation and mean absolute deviation. However, one way in which the boxplot provides information about skewness is via the distance of the median from the first quartile and the distance of the median from the third quartile. If the median is closer to the first quartile then the data (or distribution) is said to be skewed to the right. Conversely, if the median is closer to the third quartile then the distribution is said to be skewed to the left.

The IQR is unique among scale estimators in that its value can be represented graphically by marking the location of the first and third quartiles. It is not clear how to replace the IQR with a more efficient scale estimator and be able to, uniquely, graphically represent the estimator by marking two points. It is helpful to view the IQR from a slightly different perspective. Let the distance between the third quartile and the median be referred to as IQR^+ and let the distance between the median and the first quartile be referred to as IQR^-. Using this notation, the IQR can be expressed as $\text{IQR}=\text{IQR}^+ + \text{IQR}^-$. Viewing IQR^+ and IQR^- as distances from a measure of location, an approach yielding a more efficient boxplot will consist of replacing IQR^+ and IQR^- with alternative estimators. The quantities IQR^+ and IQR^- are referred to as signed scale estimators (or parameters when describing a population) since IQR^{+} only uses values greater $(+)$ than a measure of location and IQR^{-} only uses values less than $(-)$ that measure of location. This dissertation will introduce the notion of signed scale measures and provide an example illustrating a use for signed scale measures in the area of exploratory data analysis.

The remainder of this chapter will define signed scale parameters and list properties they should satisfy. The notion of signed scale estimators follows and several classes of signed scale estimators are introduced. Chapter 2 will examine the asymptotic and small sample properties of several signed scale estimators under the Uniform(0,1), Normal(0,1), Laplace(0,1), t_5, Logistic(0,1), Exponential(1), χ_4^2, Extreme Value(0,1), and Lognormal(0,1) distributions. Lastly, Chapter 3 uses the notion of signed scale estimators to introduce a new type of boxplot. The new boxplot will be compared and contrasted with two other boxplot types with respect to its ability to detect unusual observations and provide information about skewness, both for small samples and asymptotically.

1.1 Signed Scale Measures and Estimators of Scale

1.1.1 Scale Parameters and Estimation

Let X be a random variable with distribution function F_X, and let $S(F)$ be a functional defined for distribution functions F. Following the work of Ferguson (1967, p.164), $S(F_X)$ is a scale parameter for the distribution of the random variable X if it satisfies the following properties:

1. Positivity: $S(F_X) \geq 0$.

2. Translation Invariance: $S(F_X)$ is also a scale parameter for the distribution of the random variable $a + X$, where a is a finite constant.

3. Equivariance: $|b|\, S(F_X)$ is a scale parameter for the distribution of the random variable bX, where b is a finite constant.

A common problem in statistics is the estimation of scale parameters through various techniques, such as maximum likelihood and the method of moments. The estimators derived through these techniques should satisfy properties similar to those of the parameters being estimated. Specifically, let $\mathbf{x}_n = (x_1, \ldots, x_n)^T$ be a random sample from a distribution F_X with scale parameter $S(F_X)$, and let the sample distribution function be denoted $F_{\mathbf{x}_n}$. A statistic $S(F_{\mathbf{x}_n})$ will be viewed as a scale estimator if it satisfies three properties:

1. Positivity: $S(F_{\mathbf{x}_n}) \geq 0$ with equality only when $x_1 = x_2 = \ldots = x_n$.

2. Translation Invariance: $S(F_{a+\mathbf{x}_n}) = S(F_{\mathbf{x}_n})$, where a is a finite constant, and $F_{a+\mathbf{x}_n}$ denotes the sample distribution function for the sample with i^{th}element $a + x_i$.

3. Equivariance: $S(F_{b\mathbf{x}_n}) = |b| S(F_{\mathbf{x}_n})$, where b is a finite constant, and $F_{b\mathbf{x}_n}$ denotes the sample distribution function for the sample with i^{th}element bx_i.

The scale estimators considered in this paper satisfy these properties. In general, any estimator mentioned that satisfies these properties will be referred to as a scale estimator.

While the definition of scale parameter is convenient, and many estimation techniques exist that yield estimators with desirable properties, there is no guarantee that distributions from different classes will have scale parameters that have the same interpretation or are directly comparable. Moreover, scale estimators that have desirable properties under one distribution may have undesirable properties under another.

1.1.2 Signed Scale Parameters

The concept of signed scale parameters is an extension of the concept of scale parameters. However, each scale parameter is associated with two signed scale parameters, a "positive" and a "negative" parameter. The terms "positive" and "negative" do not denote the sign of the scale parameter, but rather whether the scale estimator uses values larger than some location measure or smaller than some location measure in its construction. Conceptually, a positive signed scale parameter reflects the variability of a distribution above some specified location measure and a negative signed scale parameter reflects the variability below the specified location measure. Typically, the specified location measure is a measure of central tendency, such as the mean or median.

To define our signed scale functionals it is simplest to use scale functionals $S(F)$ for positive random variables Y (see Bickel and Lehmann 1976, p. 1152, although we do not require stochastic ordering). We define such functionals to satisfy

1. Positivity: $S(F_Y) \geq 0$.

2. Equivariance: $S(F_{bY}) = bS(F_Y)$ for $b > 0$.

Discussion of signed scale parameters will be restricted to continuous random variables with positive density at a central measure of location $\mu(F)$. Let X be a random variable with distribution function F and with a measure of location $\mu(F)$. It is assumed that $\mu(F)$ satisfies $\mu(F_{a+bX}) = a + b\mu(F_X)$ for constants a and b. A *positive signed scale functional* T_+ measures scale or average distance above $\mu(F)$ and satisfies

1. Positivity: $T_+(F_X) \geq 0$.

2. Translation invariance: $T_+(F_{a+X}) = T_+(F_X)$ for finite constant a.

3. Equivariance: $T_+(F_{bX}) = bT_+(F_X)$ for constant $b > 0$.

Negative signed scale functionals T_- measure scale below $\mu(F)$ and satisfy the above three properties with T_- replacing T_+ above.

Generally we define our signed scale functionals via scale functionals for positive random variables as follows. Suppose that S is a scale functional for a positive random variable as defined above. Following the work of Bickel and Lehmann (1976), if $S(F)$ is a scale parameter for the distribution of a positive random variable X, then a positive signed scale parameter is a functional $T_+(F)$ defined by the relationship

$$T_+(F) = S(F_+) \tag{1.1}$$

where

$$F_+(x) = \frac{F(x + \mu(F)) - F(\mu(F))}{1 - F(\mu(F))} \text{ for } x > 0 \tag{1.2}$$

is just the distribution function of $|X - \mu(F)|$ conditional on $X > \mu(F)$. Similarly, a negative signed scale parameter is a functional $T_-(F)$ defined by the relationship

$$T_-(F) = S(F_-) \tag{1.3}$$

where

$$F_-(x) = \frac{F(\mu(F)) - F(\mu(F) - x)}{F(\mu(F))} \text{ for } x > 0 \tag{1.4}$$

is just the distribution function of $|X - \mu(F)|$ conditional on $X < \mu(F)$.

It is easy to verify that $T_+(F)$ and $T_-(F)$ as defined above from scale functionals for positive random variables satisfy the definition of signed scale functionals.

1.2 Classes of Scale and Signed Scale Parameters

Scale parameters, and thus signed scale parameters, can be classified into various classes. Four such classes of scale parameters are the p^{th} absolute central moment class, the α^{th} trimmed p^{th} absolute central moment class, the quantile class, and the A-estimator class. These classes are defined as follows.

1.2.1 p^{th} Absolute Central Moment Class

Given a random variable X with distribution function F and measure of location $\mu(F)$, the p^{th} absolute "central" moment of F is defined as

$$T_p(F) = \left[E_F|X-\mu(F)|^p\right]^{1/p} = \left[\int_{-\infty}^{+\infty}|x-\mu(F)|^p\,dF(x)\right]^{1/p}.$$

This is a slight abuse of terminology since this phrase is usually used only when the center or measure of location is the mean. The quantity $T_p(F)$ is often used as a measure of the scale of a distribution. Common values of p are 1 and 2, and $\mu(F)$ is typically the mean or median. The related scale functional for a positive random variable X is

$$S_p(F) = \left[\int_0^{+\infty}|x|^p\,dF(x)\right]^{1/p}.$$

Following the definition of a positive signed scale parameter as given by (1.1) and (1.2), the positive p^{th} absolute central moment scale parameter is

$$T_{p,+}(F) = \left\{\frac{1}{1-F(\mu(F))}\int_{\mu(F)}^{\infty}|x-\mu(F)|^p\,dF(x)\right\}^{1/p},$$

while the negative p^{th} absolute central moment scale parameter is obtained from (1.3) and (1.4) as

$$T_{p,-}(F) = \left\{\frac{1}{F(\mu(F))}\int_{-\infty}^{\mu(F)}|x-\mu(F)|^p\,dF(x)\right\}^{1/p}.$$

1.2.2 α^{th} Trimmed p^{th} Absolute Central Moment Class

Given a random variable X with distribution function F and measure of location $\mu(F)$, and a number $\alpha \in (0, .5)$, then the α^{th} trimmed p^{th} absolute central moment of F is defined as

$$T_{p,\alpha}(F) = \left[\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)}|x-\mu(F)|^p\,dF(x)\right]^{1/p}.$$

The quantity $T_{p,\alpha}(F)$ is said to be a trimmed measure of scale. Common values of p are 1 and 2, while common values of α are 0.05 and 0.1. The related scale functional for a positive random variable X is

$$S_{p,\alpha}(F) = \left[\int_0^{F^{-1}(1-\alpha)}|x|^p\,dF(x)\right]^{1/p}.$$

Applying the definition of positive signed scale parameters given by (1.1) and (1.2), the positive signed α^{th} trimmed p^{th} absolute moment scale parameter is

$$T_{p,\alpha,+}(F) = \left\{ \frac{1}{1-\alpha-F(\mu(F))} \int_{\mu(F)}^{F^{-1}(1-\alpha)} |x-\mu(F)|^p \, dF(x) \right\}^{1/p},$$

while the negative α^{th} trimmed p^{th} absolute moment scale parameter is obtained from (1.3) and (1.4) as

$$T_{p,\alpha,-}(F) = \left\{ \frac{1}{F(\mu(F))-\alpha} \int_{F^{-1}(\alpha)}^{\mu(F)} |x-\mu(F)|^p \, dF(x) \right\}^{1/p}.$$

1.2.3 Quantile Class

Let X be a random variable with distribution function F and let $\alpha \in (0, .5)$, then a quantile measure of scale is written as

$$T_\alpha(F) = F^{-1}(1-\alpha) - F^{-1}(\alpha).$$

The related scale functional for a positive random variable X is

$$S_\alpha(F) = F^{-1}(1-\alpha).$$

One of the simplest notions of scale, the IQR, can be expressed simply as $T_{0.25}(F)$. Notice that $T_\alpha(F)$ can be expanded and written as

$$T_\alpha(F) = \left(F^{-1}(1-\alpha) - F^{-1}(.5)\right) + \left(F^{-1}(.5) - F^{-1}(\alpha)\right).$$

Now letting the central measure of location be $F^{-1}(.5)$, the corresponding positive signed quantile scale measure as obtained from (1.1) and (1.2) is

$$T_{\alpha,+}(F) = F^{-1}(1-\alpha) - F^{-1}(.5),$$

and the corresponding negative signed quantile scale measure given by (1.3) and (1.4) is

$$T_{\alpha,-}(F) = F^{-1}(.5) - F^{-1}(\alpha).$$

1.2.4 A-estimator Class

The development of A-estimators arose as an extension of a process that is used to generate an estimator of scale. Consider the case of an i.i.d. sample of size n, denoted $\mathbf{x}_n = (x_1, \ldots, x_n)$ and with sample distribution function $F_{\mathbf{x}_n}$ from a Normal(0,σ^2) distribution. Then, $n \cdot var(\overline{x}) = \sigma^2$. In such a situation, the square root of n times the sample standard deviation of the location estimator $\overline{x}$ can be used as a estimator of scale.

Following the work of Lax (1985), an A-estimate of scale is defined analogously from the asymptotic variance of a location estimator. Specifically, given a scale estimate $S(F_{\mathbf{x}_n})$, distribution F, a positive constant c, and a function Ψ, the M-estimate of location is defined to be the solution $\mu(F_{\mathbf{x}_n})$ of the following equation due to Huber (1964): $\sum_{i=1}^{n} \Psi\left(\left(x_i - \mu\left(F_{\mathbf{x}_n}\right)\right)/cS\left(F_{\mathbf{x}_n}\right)\right) = 0$. Under certain regularity conditions and when F is a symmetric distribution,

$$\mu\left(F_{\mathbf{x}_n}\right) \text{ is } AN\left(\mu\left(F\right), A_{\Psi,F}/n\right)$$

where

$$A_{\Psi,F} = \left(cS\left(F\right)\right)^2 \frac{\int_{-\infty}^{\infty} \Psi^2\left(u\right) dF\left(x\right)}{\left[\int_{-\infty}^{\infty} \Psi'\left(u\right) dF\left(x\right)\right]^2}$$

is the asymptotic variance of the M-estimator of location based on the function Ψ and with data coming from distribution F where $u = (x - \mu(F))/cS(F)$.

Now, the function Ψ can be used to limit the influence of points far from the estimated center $\mu(F_{\mathbf{x}_n})$ of the sample. Huber, for example, proposed the function

$$\Psi_H\left(u\right) = \begin{cases} -b & u < -b \\ u & -b \leq u \leq b \\ b & u > b \end{cases}$$

for some $b > 0$. This function gives less weight to those observations whose values are larger in magnitude than b. In terms of A-estimators, if observations are far away from the center, then they receive less weight than observations close to the center.

So, for a random variable X with distribution function F, a constant c, a scale parameter $S(F)$, a location parameter $\mu(F)$, and a function Ψ, an A-estimator scale parameter is defined to be $\left(A_{\Psi,F}\right)^{\frac{1}{2}}$ and can be written as

$$T_{\Psi,c}\left(F\right) = \left(A_{\Psi,F}\right)^{\frac{1}{2}} = \left\{\left(cS\left(F\right)\right)^2 \frac{\int_{-\infty}^{\infty} \Psi^2\left(u\right) dF(x)}{\left[\int_{-\infty}^{\infty} \Psi'\left(u\right) dF(x)\right]^2}\right\}^{1/2}$$

where $u = \left(x - \mu\left(F\right)\right)/cS\left(F\right)$.

The related scale functional for a positive random variable X is

$$S_{\Psi,c}(F) = \left\{ (cS(F))^2 \frac{\int_0^\infty \Psi^2(u)\, dF(x)}{\left[\int_0^\infty \Psi'(u)\, dF(x)\right]^2} \right\}^{1/2}$$

where $u = x/cS(F)$.

Applying the definition of positive signed scale parameter as given by (1.1) and (1.2), the positive signed A-estimator scale parameter is

$$T_{\Psi,c,+}(F) = \left\{ (cS(F))^2 (1 - F(\mu(F))) \frac{\int_{\mu(F)}^\infty \Psi^2(u)\, dF(x)}{\left[\int_{\mu(F)}^\infty \Psi'(u)\, dF(x)\right]^2} \right\}^{1/2}.$$

Notice that this signed parameter deviates slightly from the strict definition of a positive signed scale parameter because the scale parameter $S(F)$ is based on observations below $\mu(F)$.

Similarly, the negative signed A-estimator scale parameter follows from (1.3 and (1.4) as

$$T_{\Psi,c,-}(F) = \left\{ (cS(F))^2 F(\mu(F)) \frac{\int_{-\infty}^{\mu(F)} \Psi^2(u)\, dF(x)}{\left[\int_{-\infty}^{\mu(F)} \Psi'(u)\, dF(x)\right]^2} \right\}^{1/2}.$$

Again, this parameter does not strictly follow the definition of a negative signed scale parameter due to the inclusion of $S(F)$.

1.3 Classes of Signed Scale Estimators

The formulation of the four classes of signed scale parameters in terms of functionals readily allows creation of corresponding classes of estimators by evaluating the functionals under the sample distribution $F_{\mathbf{x}_n}$. The forms of the estimators for the four classes of signed scale parameters are as follows.

1.3.1 Signed Scale Estimators Derived From p^{th} Absolute Central Moment Scale Parameters

Given a sample $\mathbf{x}_n = (x_1, \ldots, x_n)$ with sample distribution function $F_{\mathbf{x}_n}$, an estimator of the positive signed scale parameter $T_{p,+}(F)$ is

$$T_{p,+}(F_{\mathbf{x}_n}) = \left\{ \frac{1}{1 - F_{\mathbf{x}_n}(\mu(F_{\mathbf{x}_n}))} \frac{1}{n} \sum |x_i - \mu(F_{\mathbf{x}_n})|^p I(x_i \geq \mu(F_{\mathbf{x}_n})) \right\}^{1/p}$$

which can be simplified to

$$T_{p,+}\left(F_{\mathbf{x}_n}\right)=\left\{\frac{1}{\sum I\left(x_i\geq\mu\left(F_{\mathbf{x}_n}\right)\right)}\sum\left|x_i-\mu\left(F_{\mathbf{x}_n}\right)\right|^p I\left(x_i\geq\mu\left(F_{\mathbf{x}_n}\right)\right)\right\}^{1/p}.$$

Similarly, an estimator of the negative signed scale parameter $T_{-,p}\left(F\right)$ is

$$T_{p,-}\left(F_{\mathbf{x}_n}\right)=\left\{\frac{1}{F_{\mathbf{x}_n}\left(\mu\left(F_{\mathbf{x}_n}\right)\right)}\frac{1}{n}\sum\left|x_i-\mu\left(F_{\mathbf{x}_n}\right)\right|^p I\left(x_i\leq\mu\left(F_{\mathbf{x}_n}\right)\right)\right\}^{1/p}$$

which also can be simplified to

$$T_{p,-}\left(F_{\mathbf{x}_n}\right)=\left\{\frac{1}{\sum I\left(x_i\leq\mu\left(F_{\mathbf{x}_n}\right)\right)}\sum\left|x_i-\mu\left(F_{\mathbf{x}_n}\right)\right|^p I\left(x_i\leq\mu\left(F_{\mathbf{x}_n}\right)\right)\right\}^{1/p}.$$

1.3.2 Signed Scale Estimators Derived From α^{th} Trimmed p^{th} Absolute Central Moment Scale Parameters

Given a sample $\mathbf{x}_n=(x_1,\ldots,x_n)$ with sample distribution function $F_{\mathbf{x}_n}$, an estimator of the positive signed trimmed scale parameter $T_{p,\alpha,+}\left(F\right)$ is

$$T_{p,\alpha,+}\left(F_{\mathbf{x}_n}\right)=\left\{\frac{1}{1-\alpha-F_{\mathbf{x}_n}\left(\mu\left(F_{\mathbf{x}_n}\right)\right)}\frac{1}{n}\sum\left|x_i-\mu\left(F_{\mathbf{x}_n}\right)\right|^p I\left(\mu\left(F_{\mathbf{x}_n}\right)\leq x_i\leq F_{\mathbf{x}_n}^{-1}\left(1-\alpha\right)\right)\right\}^{1/p}$$

which can be reduced to

$$\left\{\frac{1}{\sum I\left(\mu\left(F_{\mathbf{x}_n}\right)\leq x_i\leq F_{\mathbf{x}_n}^{-1}\left(1-\alpha\right)\right)}\sum\left|x_i-\mu\left(F_{\mathbf{x}_n}\right)\right|^p I\left(\mu\left(F_{\mathbf{x}_n}\right)\leq x_i\leq F_{\mathbf{x}_n}^{-1}\left(1-\alpha\right)\right)\right\}^{1/p}.$$

An estimator of the negative signed trimmed scale parameter $T_{-,p,\alpha}\left(F\right)$ is

$$T_{p,\alpha,-}\left(F_{\mathbf{x}_n}\right)=\left\{\frac{1}{F_{\mathbf{x}_n}\left(\mu\left(F_{\mathbf{x}_n}\right)\right)-\alpha}\frac{1}{n}\sum\left|x_i-\mu\left(F_{\mathbf{x}_n}\right)\right|^p I\left(F_{\mathbf{x}_n}^{-1}\left(\alpha\right)\leq x_i\leq\mu\left(F_{\mathbf{x}_n}\right)\right)\right\}^{1/p}$$

which is reducible to

$$\left\{\frac{1}{\sum I\left(F_{\mathbf{x}_n}^{-1}\left(\alpha\right)\leq x_i\leq\mu\left(F_{\mathbf{x}_n}\right)\right)}\sum\left|x_i-\mu\left(F_{\mathbf{x}_n}\right)\right|^p I\left(F_{\mathbf{x}_n}^{-1}\left(\alpha\right)\leq x_i\leq\mu\left(F_{\mathbf{x}_n}\right)\right)\right\}^{1/p}.$$

1.3.3 Signed Scale Estimators Derived From Quantile Scale Parameters

Given a sample $\mathbf{x}_n = (x_1, \ldots, x_n)$ with sample distribution function $F_{\mathbf{x}_n}$, an estimator of the positive signed quantile scale measure $T_{\alpha,+}(F)$ is just

$$T_{\alpha,+}(F_{\mathbf{x}_n}) = F_{\mathbf{x}_n}^{-1}(1-\alpha) - F_{\mathbf{x}_n}^{-1}(.5),$$

and, similarly, an estimator of the negative signed quantile scale measure $T_{-,\alpha}(F)$ is just

$$T_{\alpha,-}(F_{\mathbf{x}_n}) = F_{\mathbf{x}_n}^{-1}(.5) - F_{\mathbf{x}_n}^{-1}(\alpha).$$

1.3.4 Signed Scale Estimators Derived From A-estimator Scale Parameters

Since A-estimator class scale parameters depend on a location parameter and a scale parameter, following the work of Lax (1985), all A-estimator class scale parameters considered will use the median as the location parameter and the median absolute deviation from the median (MAD) as the scale parameter. With these guidelines in mind, given a sample $\mathbf{x}_n = (x_1, \ldots, x_n)$ with sample distribution function $F_{\mathbf{x}_n}$, an estimator of the positive signed A-estimator scale parameter $T_{\Psi,c,+}(F)$ is

$$T_{\Psi,c,+}(F_{\mathbf{x}_n}) = \left\{ (cS(F_{\mathbf{x}_n}))^2 \left(1 - F_{\mathbf{x}_n}\left(F_{\mathbf{x}_n}^{-1}(.5)\right)\right) \frac{\frac{1}{n}\sum \Psi^2(u_i) I(u_i > 0)}{\left[\frac{1}{n}\sum \Psi'(u_i) I(u_i > 0)\right]^2} \right\}^{1/2}$$

where $u_i = \left(x_i - F_{\mathbf{x}_n}^{-1}(.5)\right)/cS(F_{\mathbf{x}_n})$. This can be simplified to

$$T_{\Psi,c,+}(F_{\mathbf{x}_n}) = \left\{ \frac{n}{2} (cS(F_{\mathbf{x}_n}))^2 \frac{\sum \Psi^2(u_i) I(u_i > 0)}{\left[\sum \Psi'(u_i) I(u_i > 0)\right]^2} \right\}^{1/2}.$$

In Monte Carlo studies performed by Lax (1985), the Ψ function that yielded the best relative efficiency of A-estimators of scale was the bisquare function. The bisquare function is defined as

$$\Psi(u_i) = u_i \left(1 - u_i^2\right)^2 I_{(-1<u_i<1)}$$

where $u_i = (x_i - t)/cS$ for some chosen t and S. Notice that this function is different from Huber's function in two ways. Only observations with normalized deviations between -1 and $+1$ are given a non-zero value. In other words, only those observations between $t - cS$ and $t + cS$ are given non-zero values. Additionally, Huber's function is non-decreasing, while the bisquare function is redescending.

A function $\Psi(u)$ is redescending if $\Psi(u) \to 0$ as $u \to \infty$. Accordingly, the bisquare function gives observations far away from the center little or no weight, while non-decreasing functions, such as Huber's, always give some weight to outlying observations. All further A-estimator class scale parameters and the derived estimates will use only the bisquare function. Using this function, the form of the estimator of the positive signed A-estimator scale parameter reduces to

$$T_{\Psi,c,+}(F_{\mathbf{x}_n}) = \left\{ \frac{n}{2} (cS(F_{\mathbf{x}_n}))^2 \frac{\sum u_i^2 \left(1-u_i^2\right)^4 I(0 < u_i < 1)}{\left[\sum (1-u_i^2)(1-5u_i^2) I(0 < u_i < 1)\right]^2} \right\}^{1/2}.$$

Following similar logic, an estimator of the negative signed A-estimator scale parameter is

$$T_{\Psi,c,-}(F_{\mathbf{x}_n}) = \left\{ \frac{n}{2} (cS(F_{\mathbf{x}_n}))^2 \frac{\sum u_i^2 \left(1-u_i^2\right)^4 I(-1 < u_i < 0)}{\left[\sum (1-u_i^2)(1-5u_i^2) I(-1 < u_i < 0)\right]^2} \right\}^{1/2}$$

where $u_i = \left(x_i - F_{\mathbf{x}_n}^{-1}(.5)\right)/cS(F_{\mathbf{x}_n})$.

1.4 Properties of Signed Scale Estimators

Since signed scale estimators derive from regular scale estimators, it is reasonable to require that any potential signed scale estimator have properties similar to unsigned scale estimators. While there are other properties of scale estimators that could be specified (Bickel and Lehmann 1976), given a sample $\mathbf{x}_n = (x_1, \ldots, x_n)$ and sample distribution function denoted $F_{\mathbf{x}_n}$, it will be required that signed scale estimators $T_+(F_{\mathbf{x}_n})$ and $T_-(F_{\mathbf{x}_n})$ satisfy the following three properties

1) Positivity $\quad T_+(F_{\mathbf{x}_n}) \geq 0$, $T_-(F_{\mathbf{x}_n}) \geq 0$ for any $\mathbf{x}_n$.
2) Translation Invariance $\quad T_+(F_{a+\mathbf{x}_n}) = T_+(F_{\mathbf{x}_n})$, $T_-(F_{a+\mathbf{x}_n}) = T_-(F_{\mathbf{x}_n})$ for finite constant a.
3) Equivariance $\quad T_+(F_{b\mathbf{x}_n}) = bT_+(F_{\mathbf{x}_n})$, $T_-(F_{b\mathbf{x}_n}) = bT_-(F_{\mathbf{x}_n})$ for constant $b > 0$.

Lemma 1 *Signed scale estimators satisfy the three properties.*

Proof. Let $\mathbf{x}_n = (x_1, \ldots, x_n)$ be a random sample from distribution function F_X with sample distribution function $F_{\mathbf{x}_n}$. Since $T_+(F_{\mathbf{x}_n}) = S(F_{\mathbf{x}_n,+})$ and since S is assumed to be a scale estimator, then all positive signed scale estimators satisfy the three properties. Additionally, since $T_-(F_{\mathbf{x}_n}) = S(F_{\mathbf{x}_n,-})$ then all negative signed scale estimators satisfy the three properties. ■

Lemma 2 *A-estimator class derived signed scale estimators satisfy the three properties.*

Proof. Let $\mathbf{x}_n = (x_1, \ldots, x_n)$ be a random sample from distribution F_X with sample distribution function $F_{\mathbf{x}_n}$, let a and $b > 0$ be arbitrary constants, and $u_i = \left(x_i - F_{\mathbf{x}_n}^{-1}(.5)\right)/cS\left(F_{\mathbf{x}_n}\right)$. Then,

$$T_{\Psi,c,+}\left(F_{\mathbf{x}_n}\right) = \left\{\frac{n}{2}\left(cS\left(F_{\mathbf{x}_n}\right)\right)^2 \frac{\sum u_i^2\left(1-u_i^2\right)^4 I\left(0<u_i<1\right)}{\left[\sum\left(1-u_i^2\right)\left(1-5u_i^2\right)I\left(0<u_i<1\right)\right]^2}\right\}^{1/2}$$

satisfies property 1 because the numerator is the sum of positive quantities and the denominator is squared. Now consider the perturbed sample with i^{th} component $a + bx_i$, and denote the sample distribution function of this new sample by $F_{a+b\mathbf{x}_n}$. Then,

$$T_{\Psi,c,+}\left(F_{a+bx_n}\right) = \left\{\frac{n}{2}\left(cS\left(F_{a+bx_n}\right)\right)^2 \frac{\sum \mu_i^2\left(1-\mu_i^2\right)^4 I\left(0<\mu_i<1\right)}{\left[\sum\left(1-\mu_i^2\right)\left(1-5\mu_i^2\right)I\left(0<\mu_i<1\right)\right]^2}\right\}^{1/2}$$

where $\mu_i = \left((a+bx_i) - F_{a+b\mathbf{x}_n}^{-1}(.5)\right)/cS\left(F_{a+b\mathbf{x}_n}\right)$. Because $F_{a+b\mathbf{x}_n}^{-1}(.5) = a + bF_{\mathbf{x}_n}^{-1}(.5)$, μ_i can be written as $b\left(x_i - F_{\mathbf{x}_n}^{-1}(.5)\right)/cS\left(F_{a+b\mathbf{x}_n}\right)$. Letting $median(x_i) = median(\mathbf{x}_n)$,

$$S\left(F_{\mathbf{x}_n}\right) = median\left(\left|x_i - median\left(\mathbf{x}_n\right)\right|, \quad i=1,...,n\right)$$

and

$$S\left(F_{a+b\mathbf{x}_n}\right) = median\left(\left|(a+bx_i) - median\left(a+b\mathbf{x}_n\right)\right|, \quad i=1,...,n\right)$$

which can be simplified to

$$S\left(F_{a+b\mathbf{x}_n}\right) = |b|\, median\left(\left|x_i - median\left(\mathbf{x}_n\right)\right|, \quad i=1,...,n\right).$$

Because we assume $b > 0$, this reduces to

$$S\left(F_{a+b\mathbf{x}_n}\right) = (b) median\left(\left|x_i - median\left(\mathbf{x}_n\right)\right|, \quad i=1,...,n\right) = bS\left(F_{\mathbf{x}_n}\right).$$

Therefore, μ_i can be written as

$$\mu_i = b\left(x_i - F_{\mathbf{x}_n}^{-1}(.5)\right)/cS\left(F_{a+b\mathbf{x}_n}\right) = b\left(x_i - F_{\mathbf{x}_n}^{-1}(.5)\right)/cbS\left(F_{\mathbf{x}_n}\right) = \left(x_i - F_{\mathbf{x}_n}^{-1}(.5)\right)/cS\left(F_{\mathbf{x}_n}\right).$$

Thus $T_{\Psi,c,+}\left(F_{a+b\mathbf{x}_n}\right)$ can be written as

$$\left\{\frac{n}{2}\left(cbS\left(F_{\mathbf{x}_n}\right)\right)^2 \frac{\sum u_i^2\left(1-u_i^2\right)^4 I\left(0<u_i<1\right)}{\left[\sum\left(1-u_i^2\right)\left(1-5u_i^2\right)I\left(0<u_i<1\right)\right]^2}\right\}^{1/2}$$

which equals

$$b\left\{\frac{n}{2}\left(cS\left(F_{\mathbf{x}_n}\right)\right)^2\frac{\sum u_i^2\left(1-u_i^2\right)^4 I\left(0<u_i<1\right)}{\left[\sum\left(1-u_i^2\right)\left(1-5u_i^2\right)I\left(0<u_i<1\right)\right]^2}\right\}^{1/2}=bT_{\Psi,c,+}\left(F_{\mathbf{x}_n}\right),$$

showing that properties 2 and 3 hold. By similar arguments, properties 1 through 3 hold for $T_{\Psi,c,-}\left(F_{\mathbf{x}_n}\right)$. ■

Chapter 2

THEORY OF SIGNED SCALE ESTIMATORS

2.1 Choosing Specific Signed Scale Estimators

A subset of the four classes of signed scale estimators was selected for further comparison. From the p^{th} Absolute Moment Class of estimators was chosen the pair of signed estimators Sd^{+} and Sd^{-}, where $p = 2$ and $\mu\left(F_{\mathbf{x}_n}\right) = \overline{x}$, defined as

$$\mathrm{Sd}^{+} = \left\{ \frac{1}{\sum I\left(x_i \geq \overline{x}\right)} \sum \left|x_i - \overline{x}\right|^2 I\left(x_i \geq \overline{x}\right) \right\}^{1/2}$$

and

$$Sd^{-} = \left\{ \frac{1}{\sum I\left(x_i \leq \overline{x}\right)} \sum \left|x_i - \overline{x}\right|^2 I\left(x_i \leq \overline{x}\right) \right\}^{1/2}.$$

From the same class, the pair of signed estimators Abs^{+} and Abs^{-} , where $p = 1$ and $u\left(F_{\mathbf{x}_n}\right) = F_{\mathbf{x}_n}^{-1}(.5)$, were also selected. These estimators are defined as

$$\mathrm{Abs}^{+} = \frac{1}{\sum I\left(x_i \geq F_{\mathbf{x}_n}^{-1}(.5)\right)} \sum \left|x_i - F_{\mathbf{x}_n}^{-1}(.5)\right| I\left(x_i \geq F_{\mathbf{x}_n}^{-1}(.5)\right)$$

and

$$\mathrm{Abs}^{-} = \frac{1}{\sum I\left(x_i \leq F_{\mathbf{x}_n}^{-1}(.5)\right)} \sum \left|x_i - F_{\mathbf{x}_n}^{-1}(.5)\right| I\left(x_i \leq F_{\mathbf{x}_n}^{-1}(.5)\right).$$

The pair of signed estimators Sdtrim$^+$ and Sdtrim$^-$, with $\alpha = .1$, $p = 2$, and $u(F_{\mathbf{x}_n}) = \overline{x}$ were chosen from the α^{th} Trimmed p^{th} Absolute Moment Class and are defined as

$$\text{Sdtrim}^+ = \left\{ \frac{1}{\sum I\left(\overline{x} \le x_i \le F_{\mathbf{x}_n}^{-1}(.9)\right)} \sum |x_i - \overline{x}|^2 I\left(\overline{x} \le x_i \le F_{\mathbf{x}_n}^{-1}(.9)\right) \right\}^{1/2}$$

and

$$\text{Sdtrim}^- = \left\{ \frac{1}{\sum I\left(F_{\mathbf{x}_n}^{-1}(.1) \le x_i \le \overline{x}\right)} \sum |x_i - \overline{x}|^2 I\left(F_{\mathbf{x}_n}^{-1}(.1) \le x_i \le \overline{x}\right) \right\}^{1/2}.$$

Another pair of signed estimators, Abstrim$^+$ and Abstrim$^-$, with $\alpha = .1$, $p = 1$, and $u(F_{\mathbf{x}_n}) = F_{\mathbf{x}_n}^{-1}(.5)$, were picked from the same classes and are defined as

$$\text{Abstrim}^+ = \frac{1}{\sum I\left(F_{\mathbf{x}_n}^{-1}(.5) \le x_i \le F_{\mathbf{x}_n}^{-1}(.9)\right)} \sum \left|x_i - F_{\mathbf{x}_n}^{-1}(.5)\right| I\left(F_{\mathbf{x}_n}^{-1}(.5) \le x_i \le F_{\mathbf{x}_n}^{-1}(.9)\right)$$

and

$$\text{Abstrim}^- = \frac{1}{\sum I\left(F_{\mathbf{x}_n}^{-1}(.1) \le x_i \le F_{\mathbf{x}_n}^{-1}(.5)\right)} \sum \left|x_i - F_{\mathbf{x}_n}^{-1}(.5)\right| I\left(F_{\mathbf{x}_n}^{-1}(.1) \le x_i \le F_{\mathbf{x}_n}^{-1}(.5)\right).$$

From the Quantile class of estimators comes IQR$^+$ and IQR$^-$ which are defined as

$$\text{IQR}^+ = F_{\mathbf{x}_n}^{-1}(.75) - F_{\mathbf{x}_n}^{-1}(.5)$$

and

$$\text{IQR}^- = F_{\mathbf{x}_n}^{-1}(.5) - F_{\mathbf{x}_n}^{-1}(.25).$$

The last set of estimators are derived from the A-estimator class. Since the Ψ function, center of location T, and measure of scale S are picked to be the bisquare function, median, and median absolute deviation from the median, then the entire set of estimators derived from this class differs only in the choice of the constant c. Positive and negative signed estimators derived from this class are denoted by $S_{Bi,c,+}$ and $S_{Bi,c,-}$, respectively, where Bi indicates the choice of the bisquare function for Ψ. $S_{Bi,c,+}$ and $S_{Bi,c,-}$ estimators are considered for $c = 15$ and $c = 20$. The form of these estimators is

$$S_{Bi,c,+} = \left\{ \frac{n}{2} \left(cS\left(F_{\mathbf{x}_n}\right)\right)^2 \frac{\sum u_i^2 \left(1 - u_i^2\right)^4 I\left(0 < u_i < 1\right)}{\left[\sum \left(1 - u_i^2\right)\left(1 - 5u_i^2\right) I\left(0 < u_i < 1\right)\right]^2} \right\}^{1/2}$$

and

$$S_{Bi,c,-} = \left\{ \frac{n}{2} \left(cS\left(F_{\mathbf{x}_n}\right) \right)^2 \frac{\sum u_i^2 \left(1-u_i^2\right)^4 I\left(-1 < u_i < 0\right)}{\left[\sum \left(1-u_i^2\right)\left(1-5u_i^2\right) I\left(-1 < u_i < 0\right)\right]^2} \right\}^{1/2}$$

where $S\left(F_{\mathbf{x}_n}\right)$ is the median of $\left|x_i - F_{\mathbf{x}_n}^{-1}(.5)\right|$ and $u_i = \left(x_i - F_{\mathbf{x}_n}^{-1}(.5)\right)/cS\left(F_{\mathbf{x}_n}\right)$. The following sections examine the asymptotic and small sample properties of these seven pairs of estimators.

2.2 Asymptotic Properties

2.2.1 Consistency

The signed scale estimators considered are weakly consistent under certain conditions. The following theorem by Stefanski (1997, personal communication) will be helpful in establishing the consistency of Abs$^+$, Abs$^-$, Sd$^+$, and Sd$^-$.

Theorem 1 *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. random vectors from a distribution F_X, θ_0 a finite constant, $g(x, \theta)$ a real-valued function, and $\widehat{\theta}_n$ an estimate of θ_0. Then under the following conditions*
1) There exists a neighborhood $N(\theta_0)$ of θ_0 such that $g(x,\theta)$ is a monotone function of θ for $\theta \in N(\theta_0)$ and for all x in the support of X_1.
2) $H(\theta) = E\{g(X_1, \theta)\}$ exists and is finite in a neighborhood of θ_0 and is continuous at θ_0.

$$\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i, \widehat{\theta}_n\right) \stackrel{p}{\longrightarrow} H(\theta_0) \text{ as } n \longrightarrow \infty \quad \text{as long as } \widehat{\theta}_n \stackrel{p}{\longrightarrow} \theta_0 \text{ as } n \longrightarrow \infty.$$

Proof. Assume without loss of generality that g is nondecreasing. Let $t_n = \frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i, \widehat{\theta}_n\right)$ and $\widehat{t}_n = \frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i, \theta_0\right)$. Then, for $\delta > 0$,

$$\left|t_n - \widehat{t}_n\right| = \left|t_n - \widehat{t}_n\right| I\left(\left|\widehat{\theta}_n - \theta_0\right| > \delta\right) + \left|t_n - \widehat{t}_n\right| I\left(\left|\widehat{\theta}_n - \theta_0\right| \le \delta\right) = R_{n,1} + R_{n,2}.$$

For any $\varepsilon > 0$

$$P\left(R_{n,1} > \varepsilon\right) \le P\left(\left|\widehat{\theta}_n - \theta_0\right| > \delta\right) \longrightarrow 0 \text{ as } n \longrightarrow \infty \text{ since } \widehat{\theta}_n \stackrel{p}{\longrightarrow} \theta_0.$$

Now, given $\varepsilon > 0$, choose $\delta_\varepsilon > 0$ such that $|H(\theta_0 + \delta_\varepsilon) - H(\theta_0 - \delta_\varepsilon)| < \varepsilon^2$, then

$$\begin{aligned}
R_{n,2} &= \left| \frac{1}{n} \sum_{i=1}^{n} g\left(\mathbf{X}_i, \widehat{\theta}_n\right) - \frac{1}{n} \sum_{i=1}^{n} g(\mathbf{X}_i, \theta_0) \right| I\left(\left|\widehat{\theta}_n - \theta_0\right| \le \delta_\varepsilon\right) \\
&\le \frac{1}{n} \sum_{i=1}^{n} \left| g\left(\mathbf{X}_i, \widehat{\theta}_n\right) - g(\mathbf{X}_i, \theta_0) \right| I\left(\left|\widehat{\theta}_n - \theta_0\right| \le \delta_\varepsilon\right) \\
&\le \frac{1}{n} \sum_{i=1}^{n} \left| g(\mathbf{X}_i, \theta_0 + \delta_\varepsilon) - g(\mathbf{X}_i, \theta_0 - \delta_\varepsilon) \right| I\left(\left|\widehat{\theta}_n - \theta_0\right| \le \delta_\varepsilon\right) \\
&\le \frac{1}{n} \sum_{i=1}^{n} \left[g(\mathbf{X}_i, \theta + \delta_\varepsilon) - g(\mathbf{X}_i, \theta - \delta_\varepsilon) \right]
\end{aligned}$$

Call the last expression W_n and note that it is non-negative. Then, by the Markov inequality,

$$P(R_{n,2} > \varepsilon) \le P(W_n > \varepsilon) \le \frac{E(W_n)}{\varepsilon} < \frac{\varepsilon^2}{\varepsilon} = \varepsilon.$$

Therefore, $R_{n,1}$ and $R_{n,2}$ each converge in probability to zero. Noting that t_n can be written as $\widehat{t}_n + \left(t_n - \widehat{t}_n\right)$, and that $\widehat{t}_n \xrightarrow{p} H(\theta_0)$ by the W.L.L.N., then $\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i, \widehat{\theta}\right) \xrightarrow{p} H(\theta_0)$ as $n \longrightarrow \infty$. ■

The following Theorems 2 and 3 show consistency of signed scale estimators from the p^{th} Absolute Moment Class of estimators. Theorem 2 deals with positive signed estimators and Theorem 3 deals with negative signed estimators.

Theorem 2 *Let $X_1, \ldots, X_n$ be distributed i.i.d. F such that $E(|X_1| I(X_1 > 0))^p < \infty$, $\mu(F) < \infty$, and F is continuous at $\mu(F)$. If $\mu(F_{\mathbf{x}_n}) \xrightarrow{P} \mu(F)$, then*

$$T_{p,+}(F_{\mathbf{x}_n}) = \left\{ \frac{1}{1 - F_{\mathbf{x}_n}(\mu(F_{\mathbf{x}_n}))} \frac{1}{n} \sum |x_i - \mu(F_{\mathbf{x}_n})|^p I(x_i \ge \mu(F_{\mathbf{x}_n})) \right\}^{1/p} \xrightarrow{P}$$

$$\left\{ \frac{1}{1 - F(\mu(F))} \int_{\mu(F)}^{\infty} |X_1 - \mu(F)|^p \, dF \right\}^{1/p} = T_{p,+}(F).$$

Proof. The denominator of the first ratio can be written as $\frac{1}{n}\sum_{i=1}^{n} I(x_i > \mu(F_{\mathbf{x}_n}))$. Letting $g(x, \theta) = I(x > \theta)$, note that $g(x, \theta)$ is a monotone decreasing function of θ for fixed x. Also,

$$E(g(X_1, \theta)) = P(X_1 > \theta) = (1 - F(\theta)) < \infty \ for \ all \ \theta.$$

By assumption, $H(\theta) = 1 - F(\theta)$ is continuous at $\theta = \mu(F)$. Thus the conditions of Theorem 1 hold and $1 - F_{x_n}(\mu(F_{x_n})) \xrightarrow{P} 1 - F(\mu(F))$. And since the reciprocal function is continuous,

$$\frac{1}{1 - F_{\mathbf{x}_n}\left(\mu\left(F_{\mathbf{x}_n}\right)\right)} \xrightarrow{P} \frac{1}{1 - F\left(\mu\left(F\right)\right)}.$$

Letting $g(x, \theta) = |x - \theta|^p I(x > \theta)$, note that $g(x, \theta)$ is monotone decreasing as a function of θ for fixed x. Also note that $g(x, \theta)$, when viewed as a function of θ for fixed x, is continuous for all θ. Now,

$$E\left|g\left(X_1, \theta\right)\right| = E\left(\left|X_1 - \theta\right|^p I\left(X_1 > \theta\right)\right) < \infty \quad \forall\, \theta \text{ since } E\left(\left|\, X_1\right| I\left(X_1 > 0\right)\right)^p < \infty$$

and $E\left(g\left(X_1, \theta\right)\right)$ is continuous at $\theta = \mu(F)$ since $g(x, \theta)$ is continuous in θ. Therefore, by Theorem 1,

$$\frac{1}{n}\sum\left|x_i - \mu\left(F_{\mathbf{x}_n}\right)\right|^p I\left(x_i \geq \mu\left(F_{\mathbf{x}_n}\right)\right) \xrightarrow{P} \int_{\mu(F)}^{\infty}\left|x - \mu(F)\right|^p dF(x).$$

Using these two convergence results, Slutsky's theorem, and the fact that the p^{th} root function is continuous, then

$$T_{p,+}\left(F_{\mathbf{x}_n}\right) \xrightarrow{P} T_{p,+}(F).$$

■

Corollary 1 *Abs^+ and Sd^+ are positive signed estimators from the p^{th} Absolute Moment Class of estimators where $\mu\left(F_{\mathbf{x}_n}\right) = F_{\mathbf{x}_n}^{-1}(.5)$, $p = 1$ and $\mu\left(F_{\mathbf{x}_n}\right) = \overline{x}$, $p = 2$, respectively. By Theorem 2, these estimators converge in probability if Theorem 4 holds with $p = 1/2$ and $E\left|X_1\right| < \infty$.*

Theorem 3 *Let $X_1, \ldots, X_n$ be distributed i.i.d. F such that $E\left(\left|X_1\right| I\left(X_1 < 0\right)\right)^p < \infty$ and $\mu(F) < \infty$. If $\mu\left(F_{\mathbf{x}_n}\right) \xrightarrow{P} \mu(F)$, then*

$$T_{p,-}\left(F_{\mathbf{x}_n}\right) = \left\{\frac{1}{F_{\mathbf{x}_n}\left(\mu\left(F_{\mathbf{x}_n}\right)\right)}\frac{1}{n}\sum\left|x_i - \mu\left(F_{\mathbf{x}_n}\right)\right|^p I\left(x_i \leq \mu\left(F_{\mathbf{x}_n}\right)\right)\right\}^{1/p} \xrightarrow{P}$$

$$\left\{\frac{1}{F\left(\mu(F)\right)}\int_{-\infty}^{\mu(F)}\left|x - \mu(F)\right|^p dF(x)\right\}^{1/p} = T_{p,-}(F).$$

Proof. Similar to Theorem 2. ■

Corollary 2 *Abs^- and Sd^- are negative signed estimators from the p^{th} Absolute Moment Class of estimators where $\mu\left(F_{\mathbf{x}_n}\right) = F_{\mathbf{x}_n}^{-1}(.5)$, $p = 1$ and $\mu\left(F_{\mathbf{x}_n}\right) = \overline{x}$, $p = 2$, respectively. By Theorem 3, these estimators converge in probability if Theorem 4 holds with $p = 1/2$ and $E\left|X_1\right| < \infty$.*

The following theorem provided by Serfling (1980) will be useful in showing the consistency of estimators from the Quantile class.

Theorem 4 *Let* $0 < p < 1$. *If* $F^{-1}(p) = \inf\{x : F(x) \geq p\}$ *is the unique solution* x *of* $F(x-) \leq p \leq F(x)$, *then* $F_{\mathbf{x}_n}^{-1}(p) \overset{wp1}{\rightarrow} F^{-1}(p)$.

Theorem 4, along with Slutsky's theorem, yields the following two theorems and corresponding corollaries.

Theorem 5 *If* F *satisfies Theorem 4 for* $p = 1/2$ *and* $p = 1-\alpha$, *then* $T_{\alpha,+}(F_{\mathbf{x}_n}) = F_{\mathbf{x}_n}^{-1}(1-\alpha) - F_{\mathbf{x}_n}^{-1}(.5)$ *converges with probability 1 to* $T_{\alpha,+}(F) = F^{-1}(1-\alpha) - F^{-1}(.5)$.

Corollary 3 *Assuming the conditions of Theorem 5 for* $\alpha = .25$, $IQR^{+} = T_{.25,+}(F_{\mathbf{x}_n}) = F_{\mathbf{x}_n}^{-1}(.75) - F_{\mathbf{x}_n}^{-1}(.5)$ *converges with probability 1 to* $T_{.25,+}(F) = F^{-1}(.75) - F^{-1}(.5)$.

Theorem 6 *If* F *satisfies Theorem 4 for* $p = 1/2$ *and* $p = \alpha$, *then* $T_{\alpha,-}(F_{\mathbf{x}_n}) = F_{\mathbf{x}_n}^{-1}(.5) - F_{\mathbf{x}_n}^{-1}(\alpha)$ *converges with probability 1 to* $T_{\alpha,-}(F) = F^{-1}(.5) - F^{-1}(\alpha)$.

Corollary 4 *Assuming the conditions of Theorem 6 for* $\alpha = .25$, $IQR^{-} = T_{.25,-}(F_{\mathbf{x}_n}) = F_{\mathbf{x}_n}^{-1}(.5) - F_{\mathbf{x}_n}^{-1}(.25)$ *converges with probability 1 to* $T_{.25,-}(F) = F^{-1}(.5) - F^{-1}(.25)$.

Consistency of estimators derived from the α^{th} Trimmed p^{th} Absolute Moment and the A-estimator classes follows by relying upon the notion of U-statistics with estimated parameters (Iverson and Randles 1989). For statistics of the form $U_n\left(\widehat{\lambda}_n\right) = \frac{1}{n}\sum_{i=1}^{n} h\left(X_i, \widehat{\lambda}_n\right)$ where $\widehat{\lambda}_n$ is an estimator of a population parameter λ, consistency will follow from Theorem 7 due to Iverson and Randles (1989).

Theorem 7 *Suppose that* $E\left|h(X_1;\lambda)\right| < \infty$ *and that there is a neighborhood* $K(\lambda)$ *of* λ, *such that if* $D(\lambda, d)$ *is a sphere centered at* λ *with radius* d *satisfying* $D(\lambda, d) \subset K(\lambda)$, *then*

(i) $\lim_{d\to 0} E\left[\sup_{\gamma\in D(\lambda,d)} \left|h(X_1;\gamma) - h(X_1;\lambda)\right|\right] = 0.$

Under these assumptions:

(A) If $\widehat{\lambda}_n \xrightarrow{P} \lambda$, *then* $U_n\left(\widehat{\lambda}_n\right) \xrightarrow{P} E_\lambda h(X_1;\lambda)$.

(B) If $\widehat{\lambda}_n \xrightarrow{WP1} \lambda$, *then* $U_n\left(\widehat{\lambda}_n\right) \xrightarrow{WP1} E_\lambda h(X_1;\lambda)$.

For positive signed estimators derived from the α^{th} Trimmed p^{th} Absolute Moment Class, Theorems 8 and 9 give conditions under which the estimators are consistent.

Theorem 8 *Let $X_1, \ldots, X_n$ be distributed i.i.d. F such that $E|X_1|^p < \infty$ for $p \geq 1$ and $\mu(F) < \infty$. Then*

$$\left\{\frac{1}{1-\alpha-F_{\mathbf{x}_n}\left(\mu\left(F_{\mathbf{x}_n}\right)\right)}\frac{1}{n}\sum\left|x_i-\mu\left(F_{\mathbf{x}_n}\right)\right|^p I\left(\mu\left(F_{\mathbf{x}_n}\right)\leq x_i\leq F_{\mathbf{x}_n}^{-1}(1-\alpha)\right)\right\}^{1/p}\xrightarrow{P}$$
$$\left\{\frac{1}{1-\alpha-F\left(\mu\left(F\right)\right)}\int_{\mu(F)}^{F^{-1}(1-\alpha)}\left|x-\mu\left(F\right)\right|^p dF\left(x\right)\right\}^{1/p}$$

if $\mu(F_{x_n})$ converges in probability to $\mu(F)$. If $\mu(F_{x_n})$ converges wp1 to $\mu(F)$, then the above convergence is wp1.

Proof. The quantity $\frac{1}{n}\sum\left|x_i-\mu\left(F_{\mathbf{x}_n}\right)\right|^p I\left(\mu\left(F_{\mathbf{x}_n}\right)\leq x_i\leq F_{\mathbf{x}_n}^{-1}(1-\alpha)\right)$ can be viewed as a U-Statistic with estimated parameters $F_{\mathbf{x}_n}^{-1}(1-\alpha)$ and $\mu\left(F_{\mathbf{x}_n}\right)$, kernel

$$h\left(X_1,\gamma=(\gamma_1,\gamma_2)\right)=\left|X_1-\gamma_1\right|^p I\left(\gamma_1\leq X_1\leq\gamma_2\right)\ ,$$

and $\theta(\gamma)=E_\lambda\left[h\left(X_1;\gamma\right)\right]$ where $\lambda=\left(\mu\left(F\right),F^{-1}\left(1-\alpha\right)\right)$. First,

$$E\left[\left|h\left(X_1,\gamma\right)\right|\right]=E\left[\left|X_1-\gamma_1\right|^p I\left(\gamma_1\leq X_1\leq\gamma_2\right)\right]\leq E\left[\left|X_1-\gamma_1\right|^p\right]<\infty$$

since $E|X_1|^p<\infty$. Second,

$$\left|h\left(X_1,\gamma\right)-h\left(X_1,\lambda\right)\right|=\left|\left|X_1-\gamma_1\right|^p I\left(\gamma_1\leq X_1\leq\gamma_2\right)-\left|X_1-\lambda_1\right|^p I\left(\lambda_1\leq X_1\leq\lambda_2\right)\right|$$

$$\leq\left|\left|X_1-\lambda_1\right|^p\left[I\left(\gamma_1\leq X_1\leq\gamma_2\right)-I\left(\lambda_1\leq X_1\leq\lambda_2\right)\right]\right|+\left|\left(\left|X_1-\gamma_1\right|^p-\left|X_1-\lambda_1\right|^p\right)I\left(\gamma_1\leq X_1\leq\gamma_2\right)\right|.$$

Therefore,

$$E\left[\sup_{\gamma\in D(\lambda,d)}\left|h\left(X_1,\gamma\right)-h\left(X_1,\lambda\right)\right|\right]\leq E\left[\sup_{\gamma\in D(\lambda,d)}\left|X_1-\lambda_1\right|^p\left|I\left(\gamma_1\leq X_1\leq\gamma_2\right)-I\left(\lambda_1\leq X_1\leq\lambda_2\right)\right|\right]$$

$$+E\left[\sup_{\gamma\in D(\lambda,d)}\left|\left|X_1-\gamma_1\right|^p-\left|X_1-\lambda_1\right|^p\right|I\left(\gamma_1\leq X_1\leq\gamma_2\right)\right].$$

The first term is bounded by

$$\int||x|+\lambda_1|^p[I(\lambda_1-d\leq x\leq\lambda_1)+I(\lambda_1\leq x\leq\lambda_1+d)+I(\lambda_2-d\leq x\leq\lambda_2)+I(\lambda_2\leq x\leq\lambda_2+d)]dF(x)$$

and clearly converges to 0 as $d\to 0$ by the dominated convergence theorem. The second term is bounded by $c\sup_{\gamma\in D(\lambda,d)}\left|\lambda_1-\gamma_1\right|\leq cd$ where c is the Lipschitz constant associated with $|x-a|^p$.

Therefore, $\lim_{d\to 0} \mathrm{E}\left[\sup_{\gamma\in D(\lambda,d)} |h(X_1,\gamma) - h(X_1,\lambda)|\right] = 0$ and condition 1 of Theorem 7 holds, so that

$$\frac{1}{n}\sum |x_i - u(F_{\mathbf{x}_n})|^p I\left(\mu(F_{\mathbf{x}_n}) \le x_i \le F_{\mathbf{x}_n}^{-1}(1-\alpha)\right) \xrightarrow{P} \int_{\mu(F)}^{F^{-1}(1-\alpha)} |x - u(F)|^p \, dF(x).$$

With this result, the fact that $F_{\mathbf{x}_n}(\mu(F_{\mathbf{x}_n}))$ converges in probability to $F(\mu(F))$, the fact that the reciprocal function is continuous, and the fact that the p^{th} root function is continuous, then

$$T_{+,p,\alpha}(F_{\mathbf{x}_n}) \xrightarrow{P} \left\{\frac{1}{1-\alpha-F(\mu(F))}\int_{\mu(F)}^{F^{-1}(1-\alpha)} |x-\mu(F)|^p \, dF(x)\right\}^{1/p} = T_{+,p,\alpha}(F).$$

■

Corollary 5 *Abstrim$^+$ and Sdtrim$^+$ are positive signed estimators from the α^{th} Trimmed p^{th} Absolute Moment Class of estimators where $\alpha = 0.1$, $\mu(F_{\mathbf{x}_n}) = F_{\mathbf{x}_n}^{-1}(.5)$, $p = 1$ and $\mu(F_{\mathbf{x}_n}) = \overline{x}$, $p = 2$, respectively. By Theorems 4 and 8, these estimators converge wp1.*

Theorem 9 *Let $X_1, \ldots, X_n$ be distributed i.i.d. F such that $E|X_1|^p < \infty$ for $p \ge 1$ and $\mu(F) < \infty$. Then*

$$\left\{\frac{1}{F_{\mathbf{x}_n}(\mu(F_{\mathbf{x}_n})) - \alpha}\frac{1}{n}\sum |x_i - \mu(F_{\mathbf{x}_n})|^p I\left(F_{\mathbf{x}_n}^{-1}(\alpha) \le x_i \le \mu(F_{\mathbf{x}_n})\right)\right\}^{1/p} \xrightarrow{P}$$

$$\left\{\frac{1}{F(\mu(F))-\alpha}\int_{F^{-1}(\alpha)}^{\mu(F)} |x-\mu(F)|^p \, dF(x)\right\}^{1/p}$$

if $\mu(F_{x_n})$ converges in probability to $\mu(F)$. If $\mu(F_{x_n})$ converges wp1 to $\mu(F)$, then the above convergence is wp1.

Proof. Proof similar to Theorem 8. ■

Corollary 6 *Abstrim$^-$ and Sdtrim$^-$ are negative signed estimators from the α^{th} Trimmed p^{th} Absolute Moment Class of estimators where $\alpha = 0.1$, $\mu(F_{\mathbf{x}_n}) = F_{\mathbf{x}_n}^{-1}(.5)$, $p = 1$ and $\mu(F_{\mathbf{x}_n}) = \overline{x}$, $p = 2$, respectively. Assuming the conditions of Theorem 4 with $p = .5$ and $E|X_1|^2 < \infty$, these estimators converge in probability.*

Theorems 10 and 11 show the consistency of the estimators $S_{Bi,c,+}$ and $S_{Bi,c,-}$ derived from the A-estimator class.

Theorem 10 *Let $X_1, \ldots, X_n$ be distributed i.i.d. F such that $E(X_1) < \infty$ and let $\Psi(u_i) = u_i\left(1-u_i^2\right)^2 I(-1<u_i<1)$. Then*

$$S_{Bi,c,+} = \frac{1}{\sqrt{2}} cS\left(F_{\mathbf{x}_n}\right)\sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}\Psi^2(u_i)\, I(u_i>0)}{\left[\frac{1}{n}\sum_{i=1}^{n}\Psi'(u_i)\, I(u_i>0)\right]^2}} \xrightarrow{P}$$

$$\frac{1}{\sqrt{2}} cS(F)\sqrt{\frac{\int_{F^{-1}(.5)}^{\infty}\Psi^2\left(\frac{X_1-F^{-1}(.5)}{cS(F)}\right)dF}{\left[\int_{F^{-1}(.5)}^{\infty}\Psi'\left(\frac{X_1-F^{-1}(.5)}{cS(F)}\right)dF\right]^2}}$$

if $\widehat{\lambda}_n = \left(F_{\mathbf{x}_n}^{-1}(.5), S\left(F_{\mathbf{x}_n}\right)\right)$ converges in probability to $\lambda = \left(F^{-1}(.5), S(F)\right)$. If $\widehat{\lambda}_n$ converges wp1 to λ, then the above convergence is wp1.

Proof. The quantity $(1/n)\sum_{i=1}^{n}\Psi^2(u_i)\, I(u_i>0)$ can be viewed as a U-Statistic with estimated parameters $F_{\mathbf{x}_n}(.5)$ and $S\left(F_{\mathbf{x}_n}\right)$, kernel

$$h\left(X_1, \gamma=(\gamma_1,\gamma_2)\right) = \Psi^2\left(\frac{X_1-\gamma_1}{c\gamma_2}\right) I(X_1>\gamma_1)$$

and $\theta(\gamma) = E_\lambda\left[h(X_1;\gamma)\right]$. Now, using the fact that the Ψ function is bounded in absolute value by 0.28125,

$$E\left[\left|h(X_1,\gamma)\right|\right] = E\left[\left|\Psi^2\left(\frac{X_1-\gamma_1}{c\gamma_2}\right) I(X_1>\gamma_1)\right|\right] \le E\left[\left|\Psi^2\left(\frac{X_1-\gamma_1}{c\gamma_2}\right)\right|\right] \le 0.28125^2 < \infty.$$

Now,

$$\left|h(X_1,\gamma)-h(X_1,\lambda)\right| = \left|\Psi^2\left(\frac{X_1-\gamma_1}{c\gamma_2}\right) I(X_1>\gamma_1) - \Psi^2\left(\frac{X_1-\lambda_1}{c\lambda_2}\right) I(X_1>\lambda_1)\right|$$

$$\le \left|\Psi^2\left(\frac{X_1-\gamma_1}{c\gamma_2}\right)\left[I(\gamma_1<X_1)-I(\lambda_1<X_1)\right]\right| +$$

$$\left|\left[\Psi^2\left(\frac{X_1-\gamma_1}{c\gamma_2}\right)-\Psi^2\left(\frac{X_1-\lambda_1}{c\lambda_2}\right)\right] I(\lambda_1<X_1)\right|$$

and the first term of the sum is such that

$$E\left[\sup_{\gamma\in D(\lambda,d)}\left|\Psi^2\left(\frac{X_1-\gamma_1}{c\gamma_2}\right)\left[I(\gamma_1<X_1)-I(\lambda_1<X_1)\right]\right|\right]$$

$$\le (.28125)\int\left[I(\lambda_1-d\le x\le\lambda_1)+I(\lambda_1\le x\le\lambda_1+d)\right]dF(x)\to 0$$

as $d \to 0$ by continuity of F at $\lambda_1 = F^{-1}(.5)$. The second term is such that

$$E\left[\sup_{\gamma\in D(\lambda,d)}\left|\left[\Psi^2\left(\frac{X_1-\gamma_1}{c\gamma_2}\right)-\Psi^2\left(\frac{X_1-\lambda_1}{c\lambda_2}\right)\right]I\left(\lambda_1<X_1\right)\right|\right]$$

$$\leq \frac{L_C}{c}E\left[\sup_{\gamma\in D(\lambda,d)}\left|\left(\frac{X_1-\gamma_1}{\gamma_2}\right)-\left(\frac{X_1-\lambda_1}{\lambda_2}\right)\right|\right]$$

where L_C is the Lipschitz constant associated with the function Ψ^2. Furthermore,

$$\frac{L_C}{c}E\left[\sup_{\gamma\in D(\lambda,d)}\left|\left(\frac{X_1-\gamma_1}{\gamma_2}\right)-\left(\frac{X_1-\lambda_1}{\lambda_2}\right)\right|\right]\leq\frac{L_C}{c}E\left[\sup_{\gamma\in D(\lambda,d)}\left|X_1\left(\frac{1}{\gamma_2}-\frac{1}{\lambda_2}\right)+\left(\frac{\lambda_1}{\lambda_2}-\frac{\gamma_1}{\gamma_2}\right)\right|\right]$$

$$\leq\frac{L_C}{c}E\left[\left|X_1\right|\left|\frac{1}{\lambda_2-d}-\frac{1}{\lambda_2}\right|+\left|\frac{\lambda_1}{\lambda_2}-\frac{\lambda_1+d}{\lambda_2-d}\right|\right]$$

$$\leq\frac{L_C}{c}E\left|X_1\right|\left|\frac{1}{\lambda_2-d}-\frac{1}{\lambda_2}\right|+\frac{L_C}{c}\left|\frac{\lambda_1}{\lambda_2}-\frac{\lambda_1+d}{\lambda_2-d}\right| \quad \to 0$$

as $d \to 0$.

Therefore, by Theorem 7, $S_{Bi,c,+}$ converges in probability and with probability 1. ■

Theorem 11 *Let $x_1, \ldots, x_n$ be distributed F with $E\left(X_1\right) < \infty$. If Ψ is the function $\Psi\left(u_i\right) = u_i\left(1-u_i^2\right)^2 I\left(-1<u_i<0\right)$, then,*

$$S_{Bi,c,-}=\frac{1}{\sqrt{2}}cS\left(F_{\mathbf{x}_n}\right)\sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}\Psi^2\left(u_i\right)I\left(u_i<0\right)}{\left[\frac{1}{n}\sum_{i=1}^{n}\Psi'\left(u_i\right)I\left(u_i<0\right)\right]^2}}\overset{P,WP1}{\to}cS^*\sqrt{\frac{\int_{-\infty}^{F^{-1}(.5)}\Psi^2\left(\frac{X_1-F^{-1}(.5)}{cS(F)}\right)dF}{\left[\int_{-\infty}^{F^{-1}(.5)}\Psi'\left(\frac{X_1-F^{-1}(.5)}{cS(F)}\right)dF\right]^2}}$$

depending upon whether $\widehat{\lambda}_n = \left(F_{\mathbf{x}_n}^{-1}(.5), S\left(F_{\mathbf{x}_n}\right)\right)$ converges strongly or weakly to $\lambda = \left(F^{-1}(.5), S(F)\right)$.

Proof. Proof similar to Theorem 10. ■

2.2.2 Asymptotic Normality

Given any function $\Psi(X,t)$, an associated functional T, defined on distribution functions F, is such that $T(F)$ is defined as a solution t_0 of the equation

$$\int\Psi\left(x,t_0\right)dF(x)=0.$$

The functional $T(\cdot)$ is called the M-functional corresponding to Ψ. For a sample $X_1, \ldots, X_n$ from F, the M-estimate corresponding to Ψ is a solution T_n, denoted $T\left(F_{\mathbf{x}_n}\right)$, of the equation

$$\sum_{i=1}^{n} \Psi(X_i, T_n) = \int \Psi(X, T_n)\, dF_n(X) = 0.$$

While the statistical properties of M-estimates of location are summarized quite succinctly in Serfling (1980), the statistical properties of other types of M-estimates require a more general approach. Specifically, the signed scale estimators Sd^+, Sd^-, Abs^+, Abs^-, Abstrim^+, Abstrim^-, Sdtrim^+, Sdtrim^-, $S_{Bi,c,+}$, and $S_{Bi,c,-}$, although not estimators of location, can be obtained as M-estimates corresponding to various Ψ functions. The Ψ functions that generate signed scale estimators are vector valued since the signed scale parameters contain estimates of unknown population parameters (Huber 1967). As such, the asymptotic distributions of these signed scale estimators will follow from the joint asymptotic distribution of the signed scale estimators and any population parameter estimates used in their construction. The asymptotic distributions of the signed scale estimators will be derived from the asymptotic distributions of convenient functions of the signed scale estimators. Table 2.1 lists the Ψ functions that correspond to functions of the positive signed scale estimators and their nuisance estimators, while Table 2.2 lists the Ψ functions that correspond to functions of the negative signed scale estimators and their nuisance estimators.

Table 2.3 lists the corresponding M-estimates, denoted $T(F_{\mathbf{x}_n})$, and shows the functional relationship between the positive signed scale estimators and corresponding $T(F_{\mathbf{x}_n})$, while Table 2.4 lists the corresponding M-estimates, denoted $T(F_{\mathbf{x}_n})$, and shows the functional relationship between the negative signed scale estimators and corresponding $T(F_{\mathbf{x}_n})$. Tables 2.5 and 2.6 list the values, $T(F)$, that the M-estimates are estimating, for the positive and negative signed scale estimators, respectively.

The asymptotic distributions of these signed M-estimates follow from the work of Huber (1967). Under certain regularity conditions specified by Huber, $\sqrt{n}(T_n - T(F))$ is asymptotically normal with mean 0 and covariance matrix $\Lambda^{-1} C (\Lambda')^{-1}$ where

$$C = E\left[\Psi(X, T(F))\, \Psi(X, T(F))'\right]$$

is the covariance matrix of $\Psi(X, T(F))$ and $\Lambda = \frac{d}{dt} E(\Psi(X,t))_{t=t_0}$ where t_0 denotes the parameter that is being estimated by $T(F_{\mathbf{x}_n})$. Assuming that the Huber regularity conditions hold, the asymptotic means and variances of the M-estimates in column 2 of Tables 2.3 and 2.4 are derived from these formulas. The asymptotic means and variances of the signed scale estimators in column 1 of Tables 2.3 and 2.4 are derived from the asymptotic means and variances of the M-estimates in column 2 of Tables 2.3 and 2.4.

Table 2.1: Ψ Functions Used for the Positive Signed Estimators

$\begin{pmatrix} \left(\frac{1}{1-t_2}\right)(X-t_3)^2 I(X \geq t_3) - t_1 \\ I(X \leq t_3) - t_2 \\ X - t_3 \end{pmatrix}$	for Sd$^+$
$\begin{pmatrix} t_1 - 2\lvert x - t_2 \rvert I(X \geq t_2) \\ I(X \leq t_2) - \frac{1}{2} \end{pmatrix}$	for Abs$^+$
$\begin{pmatrix} t_1 - \frac{1}{2}(ct_3)^2 \left(\left(\frac{X-t_4}{ct_3}\right) - 2\left(\frac{X-t_4}{ct_3}\right)^3 + \left(\frac{X-t_4}{ct_3}\right)^5\right)^2 I(t_4 \leq X \leq t_4 + ct_3) \\ t_2 - \left(1 - 6\left(\frac{X-t_4}{ct_3}\right)^2 + 5\left(\frac{X-t_4}{ct_3}\right)^4\right) I(t_4 < X < t_4 + ct_3) \\ I(\lvert X - t_4 \rvert \leq t_3) - \frac{1}{2} \\ I(X \leq t_4) - \frac{1}{2} \end{pmatrix}$	for S$_{Bi,c,+}$
$\begin{pmatrix} t_1 - \frac{1}{.9-t_3}(X-t_4)^2 I(t_4 \leq X \leq t_2) \\ I(X \leq t_2) - .9 \\ I(X \leq t_4) - t_3 \\ X - t_4 \end{pmatrix}$	for Sdtrim$^+$
$\begin{pmatrix} t_1 - \left(\frac{1}{.9-.5}\right)\lvert X - t_2 \rvert I(t_2 \leq X \leq t_3) \\ I(X \leq t_2) - .5 \\ I(X \leq t_3) - .9 \end{pmatrix}$	for Abstrim$^+$

Table 2.2: Ψ Functions Used for the Negative Signed Estimators

$\begin{pmatrix} \left(\frac{1}{1-t_2}\right)(X-t_3)^2 I(X \le t_3) - t_1 \\ I(X \le t_3) - t_2 \\ X - t_3 \end{pmatrix}$	for Sd$^-$
$\begin{pmatrix} t_1 - 2\|x-t_2\| I_{(X \le t_2)} \\ I_{(X \le t_2)} - \frac{1}{2} \end{pmatrix}$	for Abs$^-$
$\begin{pmatrix} t_1 - \frac{1}{2}(ct_3)^2 \left(\left(\frac{X-t_4}{ct_3}\right) - 2\left(\frac{X-t_4}{ct_3}\right)^3 + \left(\frac{X-t_4}{ct_3}\right)^5\right)^2 I(t_4 - ct_3 \le X \le t_4) \\ t_2 - \left(1 - 6\left(\frac{X-t_4}{ct_3}\right)^2 + 5\left(\frac{X-t_4}{ct_3}\right)^4\right) I(t_4 - ct_3 \le X \le t_4) \\ I(\|X-t_4\| \le t_3) - \frac{1}{2} \\ I(X \le t_4) - \frac{1}{2} \end{pmatrix}$	for $S_{Bi,c,-}$
$\begin{pmatrix} t_1 - \frac{1}{t_3 - .1}(X-t_4)^2 I(t_2 \le X \le t_4) \\ I(X \le t_2) - .1 \\ I(X \le t_4) - t_3 \\ X - t_4 \end{pmatrix}$	for Sdtrim$^-$
$\begin{pmatrix} t_1 - \left(\frac{1}{.5-.1}\right)\|X-t_3\| I(t_2 \le X \le t_3) \\ I(X \le t_2) - .1 \\ I(X \le t_3) - .5 \end{pmatrix}$	for Abstrim$^-$

Table 2.3: Relationship Between Positive Signed Scale Estimators and Convenient Functionals

Estimator	Functional
$Sd^+ = \sqrt{T_1(F_{\mathbf{x}_n})}$	$T_1(F_{\mathbf{x}_n}) = \frac{1}{1-F_{\mathbf{x}_n}(\overline{x})} \sum_{i=1}^{n} \frac{1}{n}(x_i - \overline{x})^2 I(x_i \ge \overline{x})$
Abs$^+ = T_1(F_{\mathbf{x}_n})$	$T_1(F_{\mathbf{x}_n}) = \frac{2}{n} \sum_{i=1}^{n} \left\|x_i - F_{\mathbf{x}_n}^{-1}\left(\frac{1}{2}\right)\right\| I\left(x_i \ge F_{\mathbf{x}_n}^{-1}\left(\frac{1}{2}\right)\right)$
$S_{Bi,c,+}^{+} = \sqrt{\frac{T_1(F_{\mathbf{x}_n})}{[T_2(F_{\mathbf{x}_n})]^2}}$	$T_1(F_{\mathbf{x}_n}) = \frac{1}{2}\left(cF_{y_i}^{-1}(.5)\right)^2 \frac{1}{n} \sum_{i=1}^{n} \left[u_{F_{\mathbf{x}_n}} - 2u_{F_{\mathbf{x}_n}}^3 + 5u_{F_{\mathbf{x}_n}}^4\right]^2 I(*)$ $T_2(F_{\mathbf{x}_n}) = \frac{1}{n} \sum_{i=1}^{n} \left[\left(1 - 6u_{F_{\mathbf{x}_n}}^2 + 5u_{F_{\mathbf{x}_n}}^4\right)\right] I(*)$ $* = \left(F_{\mathbf{x}_n}^{-1}(.5) \le x_i \le F_{\mathbf{x}_n}^{-1}(.5) + cF_{\mathbf{y}_n}^{-1}(.5)\right), y_i = \left\|x_i - F_{\mathbf{x}_n}^{-1}(.5)\right\|$
Sdtrim$^+ = \sqrt{T_1(F_{\mathbf{x}_n})}$	$T_1(F_{\mathbf{x}_n}) = \frac{1}{.9 - F_{\mathbf{x}_n}(\overline{x})} \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 I\left(\overline{x} \le x_i \le F_{\mathbf{x}_n}^{-1}(.9)\right)$
Abstrim$^+ = T_1(F_{\mathbf{x}_n})$	$T_1(F_{\mathbf{x}_n}) = \left(\frac{1}{.9-.5}\right) \frac{1}{n} \sum_{i=1}^{n} \left\|x_i - F_{\mathbf{x}_n}^{-1}(.5)\right\| I\left(F_{\mathbf{x}_n}^{-1}(.5) \le x_i \le F_{\mathbf{x}_n}^{-1}(.9)\right)$

Table 2.4: Relationship Between Negative Signed Scale Estimators and Convenient Functionals

Estimator	Functional
$Sd^{-}=\sqrt{T_1(F_{\mathbf{x}_n})}$	$T_1(F_{\mathbf{x}_n}) = \frac{1}{F_{\mathbf{x}_n}(\overline{x})}\sum_{i=1}^{n}\frac{1}{n}(x_i-\overline{x})^2 I(x_i \le \overline{x})$
Abs^{-}=$T_1(F_{\mathbf{x}_n})$	$T_1(F_{\mathbf{x}_n}) = \frac{2}{n}\sum_{i=1}^{n}\left\|x_i - F_{\mathbf{x}_n}^{-1}\left(\frac{1}{2}\right)\right\| I\left(x_i \le F_{\mathbf{x}_n}^{-1}\left(\frac{1}{2}\right)\right)$
$S_{Bi,c,-}^{-}=\sqrt{\frac{T_1(F_{\mathbf{x}_n})}{[T_2(F_{\mathbf{x}_n})]^2}}$	$T_1(F_{\mathbf{x}_n}) = \frac{1}{2}\left(cF_{y_i}^{-1}(.5)\right)^2 \frac{1}{n}\sum_{i=1}^{n}\left[u_{F_{\mathbf{x}_n}} - 2u_{F_{\mathbf{x}_n}}^3 + 5u_{F_{\mathbf{x}_n}}^4\right]^2 I(*)$ $T_2(F_{\mathbf{x}_n}) = \frac{1}{n}\sum_{i=1}^{n}\left[\left(1-6u_{F_{\mathbf{x}_n}}^2 + 5u_{F_{\mathbf{x}_n}}^4\right)\right] I(*)$ $* = \left(F_{\mathbf{x}_n}^{-1}(.5) - cF_{\mathbf{y}_n}^{-1}(.5) \le x_i \le F^{-1}(.5)\right),\ y_i = \left\|x_i - F_{\mathbf{x}_n}^{-1}(.5)\right\|$
Sdtrim^{-}=$\sqrt{T_1(F_{\mathbf{x}_n})}$	$T_1(F_{\mathbf{x}_n}) = \frac{1}{F_{\mathbf{x}_n}(\overline{x})-.1}\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2 I\left(F_{\mathbf{x}_n}^{-1}(.1) \le x_i \le \overline{x}\right)$
Abstrim^{-}=$T_1(F_{\mathbf{x}_n})$	$T_1(F_{\mathbf{x}_n}) = \left(\frac{1}{.5-.1}\right)\frac{1}{n}\sum_{i=1}^{n}\|x_i - F_{\mathbf{x}_n}(.5)\| I\left(F_{\mathbf{x}_n}^{-1}(.1) \le x_i \le F_{\mathbf{x}_n}^{-1}(.5)\right)$

Table 2.5: Functionals Used To Derive Asymptotic Means of Positive Signed Estimators

$T(F)$
$\left(\frac{1}{1-F(E(X))}\right)\int_{E(X)}^{\infty}(x-E(X))^2 dF$ for Sd^{+}
$2\int_{F^{-1}\left(\frac{1}{2}\right)}^{\infty}\left\|x - F^{-1}\left(\frac{1}{2}\right)\right\| dF$ for Abs^{+}
$\begin{pmatrix} \frac{1}{2}\left(cF_*^{-1}\left(\frac{1}{2}\right)\right)^2 \int_m^{m+cF_*^{-1}\left(\frac{1}{2}\right)}\left[u_F - 2(u_F)^3 + 5(u_F)^4\right]^2 dF \\ \int_m^{m+cF_*^{-1}\left(\frac{1}{2}\right)}\left[\left(1-6(u_F)^2+5(u_F)^4\right)\right] dF \\ * = \left\|X - F^{-1}\left(\frac{1}{2}\right)\right\|, m = F^{-1}\left(\frac{1}{2}\right) \end{pmatrix}$ for $S_{Bi,c,+}$
$\left(\frac{1}{.9-F(E(X))}\right)\int_{E(X)}^{F^{-1}(.9)}(x-E(X))^2 dF$ for Sdtrim^{+}
$\left(\frac{1}{.9-.5}\right)\int_{F^{-1}(.5)}^{F^{-1}(.9)}\left\|x - F^{-1}(.5)\right\| dF$ for Abstrim^{+}

Table 2.6: Functionals Used To Derive Asymptotic Means of Negative Signed Estimators

$T(F)$
$\left(\frac{1}{F(E(X))}\right)\int_{-\infty}^{E(X)}(x-E(X))^2 dF$ for Sd^{-}
$2\int_{-\infty}^{F^{-1}\left(\frac{1}{2}\right)}\left\|x - F^{-1}\left(\frac{1}{2}\right)\right\| dF$ for Abs^{-}
$\begin{pmatrix} \frac{1}{2}\left(cF_*^{-1}\left(\frac{1}{2}\right)\right)^2 \int_{m-cF_*^{-1}\left(\frac{1}{2}\right)}^{m}\left[u_F - 2(u_F)^3 + 5(u_F)^4\right]^2 dF \\ \int_{m-cF_*^{-1}\left(\frac{1}{2}\right)}^{m}\left[\left(1-6(u_F)^2+5(u_F)^4\right)\right] dF \\ * = \left\|X - F^{-1}\left(\frac{1}{2}\right)\right\|, m = F^{-1}\left(\frac{1}{2}\right) \end{pmatrix}$ for $S_{Bi,c,-}$
$\left(\frac{1}{F(E(X))-.1}\right)\int_{F^{-1}(.1)}^{E(X)}(x-E(X))^2 dF$ for Sdtrim^{-}
$\left(\frac{1}{.5-.1}\right)\int_{F^{-1}(.1)}^{F^{-1}(.5)}\left\|x - F^{-1}(.5)\right\| dF$ for Abstrim^{-}

Using these formulas, Sd^{+} is seen to be asymptotically normal with mean $\sqrt{\theta}$ and variance

$$\frac{1}{4\left(\theta\left(\alpha-1\right)^{2}\right)}\left(\begin{array}{c} B_{11}-\theta^{2}\alpha+2B_{13}\theta f\left(\mu\right)-2B_{13}\theta f\left(\mu\right)\alpha-2\theta B_{23}L_{13}\alpha-2\sigma^{2}\theta f\left(\mu\right)L_{13}\alpha+ \\ \sigma^{2}L_{13}^{2}-2B_{11}\alpha+B_{11}\alpha^{2}+\theta^{2}\alpha^{2}+2B_{13}L_{13}-4B_{13}L_{13}\alpha+2B_{13}L_{13}\alpha^{2}+ \\ 2\theta^{2}B_{23}f\left(\mu\right)+2\theta B_{23}L_{13}+\sigma^{2}\theta^{2}\left[f\left(\mu\right)\right]^{2}-2\sigma^{2}L_{13}^{2}\alpha+\sigma^{2}L_{13}^{2}\alpha^{2}+ \\ 2\sigma^{2}\theta f\left(\mu\right)L_{13} \end{array}\right)$$

where $\theta=\left(\frac{1}{1-\alpha}\right)\int_{\mu}^{\infty}\left(x-\mu\right)^{2}dF$, $\mu=E\left(X\right)$, $\alpha=F\left(\mu\right)$, $\sigma^{2}=E\left[\left(X-\mu\right)^{2}\right]$, $B_{11}=\left(\frac{1}{1-\alpha}\right)^{2}\int_{\mu}^{\infty}\left(x-\mu\right)^{4}dF-\theta^{2}$, $B_{13}=\left(\frac{1}{1-\alpha}\right)\int_{\mu}^{\infty}\left(x-\mu\right)^{3}dF$, $B_{23}=\int_{-\infty}^{\mu}xdF-\alpha\mu$, and $L_{13}=\left(\frac{1}{1-\alpha}\right)\left(f\left(\mu\right)\mu^{2}-2\int_{\mu}^{\infty}xdF+2\mu\left(1-\alpha\right)-\mu^{2}f\left(\mu\right)\right)$.

Abs^{+} is asymptotically normal with mean θ and variance

$$-\frac{1}{\left(2f\left(\mu\right)\right)^{2}}\left(-4B_{11}\left[f\left(u\right)\right]^{2}+4\theta f\left(\mu\right)-1\right)$$

where $\mu=F^{-1}\left(.5\right)$, $\theta=2\int_{\mu}^{\infty}\left|x-\mu\right|dF$, and $B_{11}=4\int_{\mu}^{\infty}\left|x-\mu\right|^{2}dF-\theta^{2}$.

Sdtrim^{+} is asymptotically normal with mean $\sqrt{c\theta}$ and variance

$$\frac{c}{400\theta}\left(\begin{array}{c} -100B_{11}+20\theta\alpha^{2}-40\theta\alpha\mu+20\theta\mu^{2}-100L_{13}\theta-9\alpha^{4}+ \\ 36\alpha^{3}\mu-54\alpha^{2}\mu^{2}+36\alpha\mu^{3}-9\mu^{4}-10L_{13}\alpha^{2}+20L_{13}\alpha\mu-10L_{13}\mu^{2}-100L_{13}^{2}\sigma^{2} \end{array}\right)$$

where $\mu=E\left(X\right)$, $\sigma^{2}=E\left[\left(X-\mu\right)^{2}\right]$, $\alpha=F^{-1}\left(.9\right)$, $\theta=\int_{\mu}^{\alpha}\left(x-\mu\right)^{2}dF$, $c=\frac{1}{.9-F(\mu)}$, $B_{11}=\int_{\mu}^{\alpha}\left(x-\mu\right)^{4}dF-\theta^{2}$, and $L_{13}=2\int_{\mu}^{\alpha}xdF+2\mu\left(F\left(\mu\right)-F\left(\alpha\right)\right)$.

Abstrim^{+} is asymptotically normal with mean 2.5θ and variance

$$\frac{1}{16\left[f\left(\mu\right)\right]^{2}}\left(\begin{array}{c} 100B_{11}\left[f\left(\mu\right)\right]^{2}+20\theta\left[f\left(\mu\right)\right]^{2}\mu-20\theta\left[f\left(\mu\right)\right]^{2}\beta-40\theta f\left(\mu\right)+9\alpha^{2}\left[f\left(\mu\right)\right]^{2}- \\ 18\mu\left[f\left(\mu\right)\right]^{2}\beta+9\beta^{2}\left[f\left(\mu\right)\right]^{2}+4f\left(\mu\right)\mu-4f\left(\mu\right)\beta+4 \end{array}\right)$$

where $\mu=F^{-1}\left(.5\right)$, $\beta=F^{-1}\left(.9\right)$, $\theta=\int_{\alpha}^{\beta}\left|x-\mu\right|dF\left(x\right)$, and $B_{11}=\int_{\alpha}^{\beta}\left|x-\mu\right|^{2}dF\left(x\right)-\theta^{2}$.

$S_{Bi,c,+}$ is also asymptotically normal. However, formulas for the asymptotic mean and variances are quite lengthy and not presented here. The formulas for the asymptotic means and variances of the negative signed estimators are similar to those of the positive signed estimators.

The asymptotic distributions of the signed scale estimators IQR^{+} and IQR^{-} follow from Theorem 12 and Theorem 13 due to Serfling (1980).

Theorem 12 *Let $0<p_1<\ldots<p_k<1$. Suppose that F has a density f in neighborhoods of $F^{-1}\left(p_1\right),\ldots,F^{-1}\left(p_k\right)$ and that f is positive and continuous at $F^{-1}\left(p_1\right),\ldots,F^{-1}\left(p_k\right)$. Then*

$\left(F_{\mathbf{x}_n}^{-1}(p_1), \ldots, F_{\mathbf{x}_n}^{-1}(p_k)\right)$ *is asymptotically normal with mean vector* $\left(F^{-1}(p_1), \ldots, F^{-1}(p_k)\right)$ *and covariances* $\frac{\sigma_{ij}}{n}$ *where*

$$\sigma_{ij} = \frac{p_i(1-p_j)}{f\left(F^{-1}(p_i)\right) f\left(F^{-1}(p_j)\right)} \qquad for \quad i \leq j$$

and $\sigma_{ij} = \sigma_{ji}$ *for* $i > j$.

Theorem 13 *In* R^k, *the random vectors* X_n *converge in distribution to the random vector* X *if and only if each linear combination of the components of* X_n *converges in distribution to the same linear combination of the components of* X.

Using Theorems 12 and 13, Theorem 14 gives the asymptotic distribution of IQR^+ and Theorem 15 gives the asymptotic distribution of IQR^-.

Theorem 14 IQR^+ *is asymptotically normal with mean* $F^{-1}(.75) - F^{-1}(.5)$ *and variance*

$$\frac{1}{4\left[f\left(F^{-1}(.5)\right)\right]^2} + \frac{3}{16\left[f\left(F^{-1}(.75)\right)\right]^2} - \frac{1}{4f\left(F^{-1}(.5)\right) f\left(F^{-1}(.75)\right)}.$$

Proof. By Theorem 12, $\left(F_{\mathbf{x}_n}^{-1}(.5), F_{\mathbf{x}_n}^{-1}(.75)\right)$ is asymptotically normal with mean

$$\mu = \left(F^{-1}(.5), F^{-1}(.75)\right)'$$

and asymptotic covariance matrix

$$\sum = \frac{1}{n}\begin{pmatrix} \frac{.25}{[f(F^{-1}(.5))]^2} & \frac{.125}{f(F^{-1}(.5))f(F^{-1}(.75))} \\ \frac{.125}{f(F^{-1}(.5))f(F^{-1}(.75))} & \frac{.1875}{[f(F^{-1}(.75))]^2} \end{pmatrix}.$$

Now, using this result along with Theorem 13, $\mathrm{IQR}^+ = F_{\mathbf{x}_n}^{-1}(.75) - F_{\mathbf{x}_n}^{-1}(.5)$ is asymptotically normal with mean

$$F^{-1}(.75) - F^{-1}(.5)$$

and asymptotic variance

$$\frac{.25}{\left[f\left(F^{-1}(.5)\right)\right]^2} + \frac{.1875}{\left[f\left(F^{-1}(.75)\right)\right]^2} - 2\left(\frac{.125}{f\left(F^{-1}(.5)\right) f\left(F^{-1}(.75)\right)}\right),$$

which simplifies to

$$\frac{1}{4\left[f\left(F^{-1}(.5)\right)\right]^2} + \frac{3}{16\left[f\left(F^{-1}(.75)\right)\right]^2} - \frac{1}{4f\left(F^{-1}(.5)\right) f\left(F^{-1}(.75)\right)}.$$

■

Theorem 15 *IQR^- is asymptotically normal with mean $F^{-1}(.5) - F^{-1}(.25)$ and variance*

$$\frac{1}{4\left[f\left(F^{-1}(.5)\right)\right]^2} + \frac{3}{16\left[f\left(F^{-1}(.75)\right)\right]^2} - \frac{1}{4f\left(F^{-1}(.5)\right)f\left(F^{-1}(.75)\right)}.$$

Proof. By Theorem 12, $\left(F_{\mathbf{x}_n}^{-1}(.25), F_{\mathbf{x}_n}^{-1}(.5)\right)$ is asymptotically normal with mean

$$\mu = \left(F^{-1}(.25), F^{-1}(.5)\right)'$$

and asymptotic covariance matrix

$$\sum = \frac{1}{n}\begin{pmatrix} \frac{.1875}{[f(F^{-1}(.25))]^2} & \frac{.125}{f(F^{-1}(.25))f(F^{-1}(.5))} \\ \frac{.125}{f(F^{-1}(.25))f(F^{-1}(.75))} & \frac{.25}{[f(F^{-1}(.5))]^2} \end{pmatrix}.$$

Now, using this result along with Theorem 13, $\text{IQR}^- = F_{\mathbf{x}_n}^{-1}(.5) - F_{\mathbf{x}_n}^{-1}(.25)$ is asymptotically normal with mean

$$F^{-1}(.5) - F^{-1}(.25)$$

and asymptotic variance

$$\frac{.25}{\left[f\left(F^{-1}(.5)\right)\right]^2} + \frac{.1875}{\left[f\left(F^{-1}(.25)\right)\right]^2} - 2\left(\frac{.125}{f\left(F^{-1}(.5)\right)f\left(F^{-1}(.25)\right)}\right),$$

which simplifies to

$$\frac{1}{4\left[f\left(F^{-1}(.5)\right)\right]^2} + \frac{3}{16\left[f\left(F^{-1}(.25)\right)\right]^2} - \frac{1}{4f\left(F^{-1}(.5)\right)f\left(F^{-1}(.25)\right)}.$$

■

Using the asymptotic results for the signed scale estimators, the asymptotic means under the nine distributions considered in this paper are presented in Table 2.7 below. It should be noted that the distributions are not standardized with respect to mean and variance because comparisons will be made after adjusting for the asymptotic means. The densities used for the Logistic(0,1) and the Laplace(0,1) distributions are

$$f(x) = \frac{\exp(-x)}{\{1 + \exp(-x)\}^2} \qquad \text{for } x \in \Re$$

and

$$f(x) = \frac{1}{2}\exp(-|x|) \qquad \text{for } x \in \Re.$$

Sd^-, $S_{Bi,15,-}$, and $S_{Bi,20,-}$ have the largest means under all distributions except for the Exponential(1) and Lognormal(0,1) distributions. Under the symmetric distributions these three estimators have means which are very close to each other, but under the Extreme Value(0,1) and χ^2_4 distributions, the means of these three estimators differ somewhat. Sd$^-$ and Sdtrim$^-$ have the largest means under the Exponential(1) and Lognormal(0,1) distributions. Abstrim$^-$ has the smallest mean under any distribution. Sdtrim$^-$ and IQR$^-$ have similar means under the symmetric distributions and are only slightly larger than Abstrim$^-$ in those cases. Abs$^-$ and IQR$^-$ have similar means under the skewed distributions and are the next smallest after Abstrim$^-$.

Table 2.7: Asymptotic Means for the Negative Signed Estimators Under Several Distributions

	IQR$^-$	Sd$^-$	Abs$^-$	Sdtrim$^-$	Abstrim$^-$	$S_{Bi,15,-}$	$S_{Bi,20,-}$
Uniform(0,1)	0.25	0.29	0.25	0.23	0.20	0.29	0.29
Normal(0,1)	0.67	1.00	0.80	0.66	0.56	1.00	1.00
Logistic(0,1)	1.10	1.81	1.39	1.10	0.92	1.77	1.79
Laplace(0,1)	0.69	1.41	1.00	0.74	0.60	1.31	1.34
t_5	0.73	1.29	0.95	0.73	0.61	1.21	1.23
Ex.Val.(0,1)	0.69	1.02	0.76	0.78	0.57	0.92	0.91
χ^2_4	1.43	2.09	1.46	1.73	1.16	1.71	1.70
Exp.(1)	0.41	0.65	0.39	0.57	0.32	0.44	0.44
Lognormal(0,1)	0.49	1.00	0.48	0.90	0.39	0.54	0.54

The asymptotic means for the Positive Signed scale estimators are listed in Table 2.8. As with the negative signed estimators, the distributions are not standardized with respect to mean and variance because comparisons will be made after adjusting for the asymptotic means. The asymptotic means for the positive signed estimators are identical to the asymptotic means of the negative signed estimators under the symmetric distributions. The asymptotic means of the positive signed estimators under the skewed distributions are larger than the corresponding asymptotic means of the negative signed estimators. Sd$^+$, $S_{Bi,15,+}$, and $S_{Bi,20,+}$ estimators have the largest means for any distribution and are relatively close under all distributions except the Lognormal(0,1) distribution where Sd$^+$ has a mean almost twice as large as the next largest mean.

Table 2.8: Asymptotic Means for the Positive Signed Estimators Under Several Distributions

	IQR$^+$	Sd$^+$	Abs$^+$	Sdtrim$^+$	Abstrim$^+$	$S_{Bi,15,+}$	$S_{Bi,20,+}$
Uniform(0,1)	0.25	0.29	0.25	0.23	0.20	0.29	0.29
Normal(0,1)	0.67	1.00	0.80	0.66	0.56	1.00	1.00
Logistic(0,1)	1.10	1.81	1.39	1.10	0.92	1.77	1.79
Laplace(0,1)	0.69	1.41	1.00	0.74	0.60	1.31	1.34
t_5	0.73	1.29	0.95	0.73	0.61	1.21	1.23
Ex.Val.(0,1)	0.88	1.56	1.18	0.81	0.75	1.51	1.54
χ_4^2	2.03	3.65	2.75	1.84	1.73	3.52	3.59
Exp.(1)	0.69	1.41	1.00	0.63	0.60	1.25	1.29
Lognormal(0,1)	0.96	3.59	1.77	0.92	0.86	1.82	1.96

The presentation of the asymptotic means is ordered by the skewness and kurtosis of the nine distributions. Table 2.9 gives the skewness and kurtosis of the nine distributions under consideration. All further results will be presented in a manner that preserves the ordering used here.

Table 2.9: Skewness and Kurtosis Ordering of Distributions

Distribution	Skewness	Kurtosis	Mean
Uniform(0,1)	0	1.8	.5
Normal(0,1)	0	3	0
Logistic(0,1)	0	4.2	0
Laplace(0,1)	0	6	0
t_5	0	9	0
Ex.Val.(0,1)	1.14	5.4	0.57721
χ_4^2	1.41	6	4
Exp.(1)	2	9	1
Lognormal(0,1)	6.18	113.94	2.7183

2.2.3 Asymptotic Efficiency

Since the signed estimators are not necessarily estimating the same quantity, comparing the asymptotic variances of the estimators within a distribution does not offer much insight into the performance of the estimators. For this reason, the notion of an asymptotic standardized variance will be employed. For a given estimator, the asymptotic standardized variance is defined to be the asymptotic variance of the estimator divided by the square of the asymptotic mean of the estimator. Specifically, for an estimator T_n with asymptotic variance σ^2 and asymptotic mean μ, the asymptotic standardized variance(ASV) is defined to be

$$ASV\left(T_n\right) = \frac{\sigma^2}{\mu^2} = \left(CV\right)^2.$$

The asymptotic standardized variance is unitless and allows comparison between different estimators. If an estimator has an asymptotic mean of 0 then the asymptotic standardized variance is not calculable. Under the nine distributions considered in this paper, the signed scale estimators have non-zero asymptotic means and thus can be compared by means of the ASV. Table 2.10 lists the asymptotic standardized variances for the negative signed estimators.

For the negative signed estimators, Sd^- has the smallest standardized variances under the Uniform(0,1), Normal(0,1), Extreme Value(0,1), χ^2_4, Exponential(1), and Lognormal(0,1) distributions. Sdtrim^- has the smallest standardized variances under the Logistic(0,1) and t_5 distributions. $S_{Bi,15,+}$ has the smallest standardized variance under the Laplace(0,1) distribution. IQR^- has the largest standardized variance under every distribution except for the t_5 distribution where Sd^- has the largest standardized variance.

Table 2.10: Asymptotic Standardized Variances of the Negative Signed Estimators

	IQR^-	Sd^-	Abs^-	Sdtrim^-	Abstrim^-	$S_{Bi,15,-}$	$S_{Bi,20,-}$
Uniform(0,1)	3.00	0.48	1.67	0.77	2.33	1.48	1.45
Normal(0,1)	3.20	0.84	1.47	1.11	2.29	1.29	1.27
Logistic(0,1)	3.31	1.37	1.62	1.26	2.36	1.52	1.55
Laplace(0,1)	4.16	2.50	2.00	1.91	2.77	1.90	1.94
t_5	3.35	3.70	1.85	1.33	2.39	1.68	1.80
Ex.Val.(0,1)	3.25	0.60	1.69	1.03	2.55	1.40	1.37
χ^2_4	3.43	0.81	2.08	1.07	2.88	1.77	1.75
Exp.(1)	4.06	1.21	3.05	1.38	3.72	2.75	2.74
Lognormal(0,1)	4.35	2.61	3.20	3.03	4.03	2.82	2.80

The asymptotic standardized variances for the positive signed estimators are given in Table 2.11. For the positive signed estimators, Sd^+ has the smallest standardized variance at the Uniform(0,1) and Normal(0,1) distributions. Sdtrim^+ has the smallest standardized variance at the Logistic(0,1) and t_5 distributions. $S_{Bi,15,+}$ has the lowest standardized variance under the Laplace(0,1), Extreme

Value(0,1), and χ^2_4 distributions. $S_{Bi,20,+}$ has the lowest standardized variances under the Exponential(1), and Lognormal(0,1) distributions. IQR$^+$ has the largest asymptotic standardized variance under all distributions except for the t_5 and Lognormal(0,1) distributions where Sd$^+$ has the largest standardized variances.

Table 2.11: Asymptotic Standardized Variances of the Positive Signed Estimators

	IQR$^+$	Sd$^+$	Abs$^+$	Sdtrim$^+$	Abstrim$^+$	$S_{Bi,15,+}$	$S_{Bi,20,+}$
Uniform(0,1)	3.00	0.48	1.67	0.77	2.33	1.48	1.45
Normal(0,1)	3.20	0.84	1.47	1.11	2.29	1.29	1.27
Logistic(0,1)	3.31	1.37	1.62	1.26	2.36	1.52	1.55
Laplace(0,1)	4.16	2.50	2.00	1.91	2.77	1.90	1.94
t_5	3.35	3.70	1.85	1.33	2.39	1.68	1.80
Ex.Val.(0,1)	3.58	2.35	1.73	1.88	2.44	1.62	1.64
χ^2_4	3.71	2.35	1.73	2.18	2.49	1.62	1.62
Exp.(1)	4.16	3.40	2.00	3.01	2.77	1.92	1.82
Lognormal(0,1)	5.23	41.26	4.37	5.56	3.69	3.02	2.84

In order to compare the various estimators, the asymptotic relative efficiency, calculated as the ratio of the asymptotic standardized variances, is given in Tables 2.12 and 2.13. A relative efficiency larger than 1 indicates that a particular signed (positive/negative) estimator is more efficient than the corresponding signed (positive/negative) IQR estimator. Table 2.12 lists the asymptotic relative efficiencies of the negative signed scale estimators while Table 2.13 lists the asymptotic relative efficiencies of the positive signed scale estimators.

Table 2.12: Asymptotic Relative Efficiency of the Negative Signed Estimators

	IQR^-	Sd^-	Abs^-	Sdtrim^-	Abstrim^-	$S_{Bi,15,-}$	$S_{Bi,20,+}$
Uniform(0,1)	1.00	6.25	1.79	3.90	1.28	2.04	2.08
Normal(0,1)	1.00	3.85	2.17	2.88	1.39	2.50	2.50
Logistic(0,1)	1.00	2.44	2.04	2.63	1.41	2.17	2.13
Laplace(0,1)	1.00	1.67	2.08	2.18	1.52	2.17	2.13
t_5	1.00	0.90	1.82	2.52	1.41	2.00	1.85
Ex.Val.(0,1)	1.00	5.42	1.92	3.16	1.28	2.33	2.38
χ^2_4	1.00	4.23	1.65	3.21	1.19	1.94	1.96
Exp.(1)	1.00	3.33	1.33	2.94	1.09	1.47	1.49
Lognormal(0,1)	1.00	1.67	1.35	1.44	1.08	1.54	1.56

Table 2.13: Asymptotic Relative Efficiency of the Positive Signed Estimators

	IQR^+	Sd^+	Abs^+	Sdtrim^+	Abstrim^+	$S_{Bi,15,+}$	$S_{Bi,20,+}$
Uniform(0,1)	1.00	6.25	1.79	3.90	1.28	2.04	2.08
Normal(0,1)	1.00	3.85	2.17	2.88	1.39	2.50	2.50
Logistic(0,1)	1.00	2.44	2.04	2.63	1.41	2.17	2.13
Laplace(0,1)	1.00	1.67	2.08	2.18	1.52	2.17	2.13
t_5	1.00	0.90	1.82	2.52	1.41	2.00	1.85
Ex.Val.(0,1)	1.00	1.52	2.08	1.90	1.47	2.22	2.17
χ^2_4	1.00	1.58	2.14	1.70	1.49	2.29	2.29
Exp.(1)	1.00	1.22	2.08	1.38	1.52	2.17	2.27
Lognormal(0,1)	1.00	0.13	1.19	0.94	1.43	1.72	1.85

The asymptotic relative efficiency of a particular negative signed estimator under a particular distribution is defined to be the asymptotic standardized variance of IQR^- divided by the asymptotic standardized variance of that estimator. If a negative signed estimator has an asymptotic relative efficiency larger than 1 then it performs better than IQR^-. As seen in Table 2.12, all negative signed estimators except for Sd^- under the t_5 distribution, have a better relative efficiency than IQR^-. Sd^- performs the best at the Uniform(0,1), Normal(0,1), Extreme Value(0,1), χ^2_4, Exponential(1), and Lognormal(0,1) distributions, and is 1.7-6.3 times more efficient than IQR^-. Sdtrim^- performs the best at the Logistic(0,1), Laplace(0,1), and t_5 distributions, and is 2.2-2.6 times more efficient than IQR^-.

The asymptotic relative efficiency of a particular positive signed estimator under a particular distribution is the asymptotic standardized variance of IQR$^+$ divided by the asymptotic standardized variance of that estimator. If a positive signed estimator has an asymptotic relative efficiency larger than 1 then that estimator outperforms IQR$^+$. For the positive signed estimators, all positive signed estimators outperform IQR$^+$, except for Sd$^+$ at the Lognormal(0,1) and t_5 distributions and Sdtrim$^+$ at the Lognormal(0,1) distribution. Sd$^+$ performs best at the Uniform(0,1) and Normal(0,1) distributions and is 3.9-6.3 times more efficient than IQR$^+$. Sdtrim$^+$ performs the best under the Logistic(0,1), Laplace(0,1), and t_5 distributions and is between 2.2-2.6 times more efficient than IQR$^+$. $S_{Bi,15,+}$ performs the best under the χ^2_4 and Extreme Value(0,1) distributions and is 2.2 times more efficient than IQR$^+$. $S_{Bi,20,+}$ has the highest efficiency under the Exponential(1) and Lognormal(0,1) distributions and is between 1.9-2.3 times more efficient than IQR$^+$.

In order to determine if any class of signed estimators performs the best over both the negative and positive estimators, Table 2.14 gives the best and second best performing signed scale estimators under each distribution.

Table 2.14: Best and Second Best Performing Signed Estimators Along With Relative Efficiencies

Distribution	Negative(1)	Negative(2)	Positive(1)	Positive(2)
Uniform(0,1)	Sd$^-$ (6.25)	Sdtrim$^-$ (3.90)	Sd$^+$ (6.25)	Sdtrim$^+$ (3.90)
Normal(0,1)	Sd$^-$ (3.85)	Sdtrim$^-$ (2.88)	Sd$^+$ (3.85)	Sdtrim$^+$ (2.88)
Logistic(0,1)	Sdtrim$^-$ (2.63)	Sd$^-$ (2.44)	Sdtrim$^+$ (2.63)	Sd$^+$ (2.44)
Laplace(0,1)	Sdtrim$^-$ (2.18)	$S_{Bi,15,-}$ (2.17)	Sdtrim$^+$ (2.18)	$S_{Bi,15,+}$ (2.17)
t_5	Sdtrim$^-$ (2.52)	$S_{Bi,15,-}$ (2.00)	Sdtrim$^+$ (2.52)	$S_{Bi,15,+}$ (2.00)
Ex.Val.(0,1)	Sd$^-$ (5.42)	Sdtrim$^-$ (3.16)	$S_{Bi,15,+}$ (2.22)	$S_{Bi,20,+}$ (2.17)
χ^2_4	Sd$^-$ (4.20)	Sdtrim$^-$ (3.21)	$S_{Bi,15,+}$ (2.29)	$S_{Bi,20,+}$ (2.29)
Exp.(1)	Sd$^-$ (3.33)	Sdtrim$^-$ (2.94)	$S_{Bi,20,+}$ (2.27)	$S_{Bi,15,+}$ (2.17)
Lognormal(0,1)	Sd$^-$ (1.67)	$S_{Bi,20,-}$ (1.56)	$S_{Bi,20,+}$ (1.85)	$S_{Bi,15,+}$ (1.72)

For the heavy tailed symmetric distributions, the t_5, Logistic(0,1), and Laplace(0,1), Sdtrim$^-$ is the best performer among all negative signed estimators while Sdtrim$^+$ is the best performer among all positive signed estimators. However, while the performance of Sdtrim$^+$ and Sdtrim$^-$ is roughly 2.4 times that of IQR$^+$ and IQR$^-$, respectively, the performance of the second best performers ($S_{Bi,15,-}$, $S_{Bi,15,+}$,Sd$^+$, and Sd$^-$) is still roughly 2.2 times that of IQR$^+$ and IQR$^-$.

Sd$^+$ and Sd$^-$ perform the best under the Uniform(0,1) and Normal(0,1) distributions, having 3.9-6.3 times the efficiency of IQR$^+$ and IQR$^-$. Sdtrim$^+$ and Sdtrim$^-$ perform a little less efficiently than Sd$^+$ and Sd$^-$ and are 2.9 -3.9 times more efficient than IQR$^+$ and IQR$^-$.

Under the skewed distributions, Sd^- is the best performer among all other negative signed estimators and is between 1.7-5.4 times more efficient than IQR^-. Sdtrim^- is slightly behind Sd^-, in terms of efficiency, under the Extreme Value(0,1), χ^2_4, and Exponential(1) distributions with an efficiency that is between 2.9-3.2 times that of IQR^-. Under the Lognormal distribution, the next closest performer to Sd^- is $S_{Bi,20,-}$, which is 1.6 times more efficient than IQR^-.

For the positive signed estimators under the skewed distributions, the best performing estimators are $S_{Bi,15,+}$ and $S_{Bi,20,+}$. $S_{Bi,15,+}$ is the best performer under the Extreme Value(0,1) and χ^2_4, while $S_{Bi,20,+}$ is the best performer under the Exponential(1) and Lognormal(0,1) distributions. However, both of these estimators have efficiencies within .13 of each other, making their performance under these distributions effectively identical.

Unfortunately, there is no one particular pair of signed scale estimators that performs the best across all of the nine distributions. The asymptotic relative efficiencies suggest three classes of performance. For the symmetric non-heavy tailed distributions, Sd^- and Sd^+ clearly outperform the other estimators. For the symmetric heavy-tailed distributions, Sdtrim^- and Sdtrim^+ outperform all other estimators though the signed A-estimators are close in performance. For the skewed distributions, there is no pair of signed estimators which dominates the others. Among all negative signed estimators, Sd^- has the largest relative efficiency under all of the skewed distributions. The two positive signed A-estimators have the highest efficiency among all other positive signed estimators under the skewed distributions. Even though the A-estimators do not dominate the other estimators under the skewed distributions, the A-estimators do dominate the IQR based estimators and the positive signed A-estimators perform much better than Sd^+. As such, A-estimators appear to be the estimators of choice under the skewed distributions.

2.3 Small Sample Comparisons

While the asymptotic performance of the signed scale estimators provides motivation for choosing "the best" pair of signed scale estimators, in practice, these estimators will be used for samples with a finite size. In particular, since the performance of these estimators under samples of size greater than about 30 will generally agree with the asymptotic performance, the small sample performance of the signed scale estimators was examined via Monte Carlo simulation. 5000 pseudo-random samples of sizes n=10, 20, 30, and 100 were generated for each distribution and the relative efficiency of each estimator was estimated. As with the asymptotic comparisons, the signed scale estimators estimate different quantities for a given sample size as well as asymptotically. Instead of comparing the signed scale estimators by their sampling variances, the estimators will be compared using finite sample standardized variances. Given a sample of size n, an estimator T_n, and an asymptotic mean of μ, the finite sample standardized variance is defined to be the variance of T_n divided by the square of the asymptotic mean of T_n. The finite sample standardized variance can be expressed as

$$SV(T_n) = \frac{var(T_n)}{\mu^2}.$$

The finite sample standardized variance is unitless and allows comparisons of different estimators under various distributions and at different sample sizes. The finite sample relative efficiency of a positive signed scale estimator will be calculated as the finite sample standardized variance of IQR^+ divided by the finite sample standardized variance of that estimator. Similarly, the finite sample relative efficiency of a negative signed scale estimator will be calculated as the finite sample standardized variance of IQR^- divided by the finite sample standardized variance of that estimator. As with the asymptotic inferences, a signed estimator with finite sample relative efficiency larger than one indicates that the estimator outperforms either IQR^+ or IQR^-, depending as whether or not the signed estimator is negative or positive.

The Monte Carlo simulation requires finite sample estimation of quantiles. Splus 4 release 3 was used for all simulation and its quantile function was used. Given a sorted vector $\mathbf{X}_n$ of length n and a number $p \in (0,1)$, Splus returns $\mathbf{X}_n[1+(n-1)p]$ if $1+(n-1)p$ is integer. Otherwise Splus interpolates between a=$\mathbf{X}_n[floor(1+(n-1)p)]$ and b=$\mathbf{X}_n[1+floor(1+(n-1)p)]$ to return $(1-c)*a+c*b$ where c=$(1+(n-1)p)mod1$.

2.3.1 Symmetric Distributions

Since the signed scale estimators estimate the same quantity when the underlying distribution is symmetric, the relative efficiency of a particular signed pair of estimators is defined to be the average of the relative efficiencies of the negative signed estimator and the positive signed estimator. Table 2.15 lists the relative efficiencies of the signed scale estimator pairs for the symmetric distributions and for samples of size 10, 20, 30, and 100.

The $\text{Sd}^{+/-}$ pair of estimators clearly performs the best under the Uniform(0,1), Normal(0,1), and Logistic(0,1) distributions for each sample size. Under the Logistic(0,1) distribution, however, the $\text{Abs}^{+/-}$, $\text{Sdtrim}^{+/-}$, $S_{Bi,15,+/-}$, and $S_{Bi,20,+/-}$ estimators perform almost as well as $\text{Sd}^{+/-}$ for each sample size. Under the Laplace(0,1) distribution the three pairs of estimators $\text{Abs}^{+/-}$, $S_{Bi,15,+/-}$, and $S_{Bi,20,+/-}$ all perform similarly and are roughly 1.6-2.0 times more efficient than $\text{IQR}^{+/-}$ at each sample size. For the heavy-tailed t_5 distribution, $\text{Abs}^{+/-}$ and $S_{Bi,20,+/-}$ perform the best when the sample size is 10. For samples of size 20 and 30, $\text{Sdtrim}^{+/-}$ performs the best with $S_{Bi,15,+/-}$ and $S_{Bi,20,+/-}$ performing almost as well.

Table 2.15: Small Sample Relative Efficiencies For Symmetric Distributions(+/− paired)

Distribution	N	IQR	Sd	Abs	Sdtrim	Abstrim	$S_{Bi,15}$	$S_{Bi,20}$
Uniform(0,1)	10	1.00	3.09	1.54	1.95	1.14	0.45	1.37
	20	1.00	4.42	1.70	2.85	1.24	1.69	1.79
	30	1.00	4.91	1.73	3.15	1.26	1.81	1.88
	100	1.00	5.79	1.79	3.65	1.29	1.97	2.03
	∞	1.00	6.25	1.79	3.90	1.28	2.04	2.08
Normal(0,1)	10	1.00	2.40	1.75	1.53	1.16	1.41	1.71
	20	1.00	3.09	2.00	2.19	1.30	2.10	2.16
	30	1.00	3.26	2.05	2.37	1.33	2.20	2.26
	100	1.00	3.55	2.08	2.65	1.35	2.33	2.38
	∞	1.00	3.85	2.17	2.88	1.39	2.50	2.50
Logistic(0,1)	10	1.00	1.83	1.63	1.36	1.13	1.51	1.60
	20	1.00	2.11	1.82	1.90	1.27	1.84	1.85
	30	1.00	2.20	1.89	2.15	1.31	1.95	1.94
	100	1.00	2.32	1.99	2.44	1.38	2.08	2.05
	∞	1.00	2.44	2.04	2.63	1.41	2.17	2.13
Laplace(0,1)	10	1.00	1.49	1.69	1.13	1.13	1.57	1.68
	20	1.00	1.70	1.89	1.57	1.28	1.88	1.94
	30	1.00	1.73	1.94	1.77	1.35	1.97	1.98
	100	1.00	1.74	2.05	2.10	1.47	2.14	2.12
	∞	1.00	1.67	2.08	2.18	1.52	2.17	2.13
t_5	10	1.00	1.04	1.42	1.17	1.11	1.35	1.42
	20	1.00	1.25	1.62	1.77	1.27	1.69	1.66
	30	1.00	1.32	1.67	1.97	1.30	1.76	1.69
	100	1.00	1.29	1.80	2.41	1.39	1.95	1.84
	∞	1.00	0.90	1.82	2.52	1.41	2.00	1.85
Av.Est.SE		NA	0.06	0.03	0.04	0.01	0.04	0.04

The last row of Table 2.15 provides an estimate of the average standard error (of the finite sample relative efficiencies) for the entries in each column, excluding the IQR column. These standard error estimates suggest that certain estimators perform similarly even though the relative efficiency estimates suggest a ranking. For example, $\text{Sd}^{+/-}$ under the Laplace(0,1) for samples of size 30 appears to outperform $\text{Sdtrim}^{+/-}$. However, these two estimators have estimates within one standard error of each other. Although not shown, estimates of the average standard error for each row were calculated. Those estimates did not indicate any significant variation by sample size, within a particular distribution.

In an attempt to determine if any particular pair of signed estimators performs the best, or close to the best, across all nine distributions, Table 2.16 lists the best and second best performing pair of signed estimators for each distribution and sample size.

$\text{Sd}^{+/-}$ performs the best under the Uniform(0,1) and Normal(0,1) distributions for all sample sizes. $\text{Sd}^{+/-}$ performs the best under the Logistic(0,1) distribution for samples of size 10, 20, and 30. For the larger sample size and asymptotically, $\text{Sdtrim}^{+/-}$ is the best performer under the Logistic(0,1) distribution. The pair of A-estimators have the best overall performance under the Laplace(0,1) distribution even though the $\text{Sdtrim}^{+/-}$ estimators perform marginally better at the larger sample sizes and asymptotically. $\text{Sdtrim}^{+/-}$ performs the best under the t_5 distribution for samples of size 20, 30 and 100 although $S_{Bi,15,+/-}$ performs almost as well as $\text{Sdtrim}^{+/-}$.

Overall, the $\text{Sd}^{+/-}$ estimators appear to be the best performers for the non-heavy tailed symmetric distributions and even for small samples in the case of the moderately heavy tailed Logistic(0,1) distribution. The $\text{Sdtrim}^{+/-}$ estimators perform the best for the samples of size 100 and asymptotically under the Laplace(0,1) distribution. The $\text{Sdtrim}^{+/-}$ estimators perform the best among all samples sizes except for samples of size 10 under the t_5 distribution. The pair of A-estimators have a performance similar to that of the $\text{Sdtrim}^{+/-}$ estimators under the Laplace(0,1) and t_5 distributions and perform better than them under some of the smaller sample sizes.

In general, $\text{Sd}^{+/-}$ is the pair of estimators that performs the best for the Uniform(0,1), Normal(0,1), and Logistic(0,1) distributions, while $S_{Bi,15,+/-}$ performs the best, or close to the best, under the Laplace(0,1) and t_5 distributions.

2.3.2 Skewed Distributions

Since the positive and negative signed scale estimator estimate different quantities when the underlying distribution is skewed, small sample comparisons must be made separately for the positive and negative signed scale estimators.

Table 2.16: Best Performing Estimators Under Symmetric Distributions(+/− paired)

Distribution	N	1^{st}	2^{nd}
Uniform(0,1)	10	$\text{Sd}^{+/-}$ (3.09)	$\text{Sdtrim}^{+/-}$ (1.95)
	20	$\text{Sd}^{+/-}$ (4.42)	$\text{Sdtrim}^{+/-}$ (2.85)
	30	$\text{Sd}^{+/-}$ (4.91)	$\text{Sdtrim}^{+/-}$ (3.15)
	100	$\text{Sd}^{+/-}$ (5.79)	$\text{Sdtrim}^{+/-}$ (3.65)
	∞	$\text{Sd}^{+/-}$ (6.25)	$\text{Sdtrim}^{+/-}$ (3.90)
Normal(0,1)	10	$\text{Sd}^{+/-}$ (2.40)	$\text{Abs}^{+/-}$ (1.75)
	20	$\text{Sd}^{+/-}$ (3.09)	$\text{Sdtrim}^{+/-}$ (2.19)
	30	$\text{Sd}^{+/-}$ (3.26)	$\text{Sdtrim}^{+/-}$ (2.37)
	100	$\text{Sd}^{+/-}$ (3.55)	$\text{Sdtrim}^{+/-}$ (2.65)
	∞	$\text{Sd}^{+/-}$ (3.85)	$\text{Sdtrim}^{+/-}$ (2.88)
Logistic(0,1)	10	$\text{Sd}^{+/-}$ (1.83)	$\text{Abs}^{+/-}$ (1.63)
	20	$\text{Sd}^{+/-}$ (2.11)	$\text{Sdtrim}^{+/-}$ (1.90)
	30	$\text{Sd}^{+/-}$ (2.20)	$\text{Sdtrim}^{+/-}$ (2.15)
	100	$\text{Sdtrim}^{+/-}$ (2.44)	$\text{Sd}^{+/-}$ (2.32)
	∞	$\text{Sdtrim}^{+/-}$ (2.63)	$\text{Sd}^{+/-}$ (2.44)
Laplace(0,1)	10	$\text{Abs}^{+/-}$ (1.69)	$S_{Bi,20,+/-}$ (1.68)
	20	$S_{Bi,20,+/-}$ (1.94)	$\text{Abs}^{+/-}$ (1.89)
	30	$S_{Bi,20,+/-}$ (1.98)	$S_{Bi,15,+/-}$ (1.97)
	100	$S_{Bi,15,+/-}$ (2.14)	$S_{Bi,20,+/-}$ (2.12)
	∞	$\text{Sdtrim}^{+/-}$ (2.18)	$S_{Bi,15,+/-}$ (2.17)
t_5	10	$S_{Bi,20,+/-}$ (1.42) ,$\text{Abs}^{+/-}$ (1.42)	$S_{Bi,15,+/-}$ (1.35)
	20	$\text{Sdtrim}^{+/-}$ (1.77)	$S_{Bi,15,+/-}$ (1.69)
	30	$\text{Sdtrim}^{+/-}$ (1.97)	$S_{Bi,15,+/-}$ (1.76)
	100	$\text{Sdtrim}^{+/-}$ (2.41)	$S_{Bi,15,+/-}$ (1.95)
	∞	$\text{Sdtrim}^{+/-}$ (2.52)	$S_{Bi,15,+/-}$ (2.00)

Negative Signed Scale Estimators

Table 2.17 lists the relative efficiencies of the negative signed estimators for all skewed distributions for samples of size 10, 20, 30, and 100.

Sd$^-$ performs the best under the Extreme Value(0,1), χ^2_4, Exponential(1), and Lognormal(0,1) distributions for all three sample sizes. Under the Extreme Value(0,1), χ^2_4, and Exponential(1) distributions, Sdtrim$^-$ performs second best to Sd$^-$, yet approaches Sd$^-$ in performance as both the sample size increases and the skewness of the distributions increases. Under the Lognormal(0,1) distribution, the second best performing estimators are $S_{Bi,15,-}$ and $S_{Bi,20,-}$ followed by Sdtrim$^-$.

As with Table 2.15, the last row of Table 2.17 lists an average of an estimate of the standard error of the entries in each column excepting the IQR$^-$ column. The Sd based estimates have the largest estimated standard errors followed by the bisquare based estimators and lastly by the Abs based estimators.

As with the symmetric distributions, Table 2.18 lists the best and second best performing negative signed estimators for each skewed distribution and sample size, in an attempt to identify an estimator that works the best across all of the skewed distributions.

Sd$^-$ performs the best over all of the skewed distributions and for each sample size. However, Sdtrim$^-$ performance appears to approach that of Sd$^-$ as the skewness of the distributions increases. Additionally, for the highly skewed Lognormal(0,1) distribution, Sd$^-$ performs almost on par with $S_{Bi,15,-}$ and $S_{Bi,20,-}$.

The performance situation for samples of size 20 is not as clear as with samples of size 10 and size 30. Abs$^+$ performs the best under the χ^2_4 and Exponential(1) distributions when the sample size is 20. $S_{Bi,20,+}$ has the best performance among all other positive signed estimators when the sample size is 20 and the distribution is either the Extreme Value(0,1) or the Lognormal(0,1). However, both Abs$^+$ and $S_{Bi,20,+}$ have almost identical performance when the sample size is 20 and the distribution is either Extreme Value(0,1) or χ^2_4. It appears that as the skewness of the distributions increases that the performance of Abs$^+$ falls behind that of $S_{Bi,20,+}$.

Positive Signed Scale Estimators

Table 2.19 lists the relative efficiencies of the positive signed estimators for all skewed distributions for samples of size 10, 20, 30, and 100.

The performance behavior of the positive signed estimators is somewhat different from the behavior of the negative signed estimators due to the fact that the skewed distributions are skewed right. As such, Abs$^+$ performs the best across all of the skewed distributions when the sample size is 10. When the sample size is 30, $S_{Bi,20,+}$ performs the best over all the skewed distributions,

Table 2.17: Finite Efficiencies For Negative Signed Estimators Under Skewed Distributions

Distribution	N	IQR^-	Sd^-	Abs^-	$Sdtrim^-$	$Abstrim^-$	$S_{Bi,15,-}$	$S_{Bi,20,-}$
Ex.Val.(0,1)	10	1.00	2.93	1.63	1.94	1.11	1.65	1.78
	20	1.00	3.76	1.79	2.56	1.20	2.04	2.09
	30	1.00	3.90	1.81	2.69	1.23	2.05	2.09
	100	1.00	4.48	1.89	3.12	1.27	2.24	2.28
	∞	1.00	5.42	1.92	3.16	1.28	2.33	2.38
χ_4^2	10	1.00	2.69	1.40	2.03	1.05	1.42	1.20
	20	1.00	3.35	1.57	2.54	1.16	1.74	1.77
	30	1.00	3.65	1.60	2.76	1.18	1.79	1.81
	100	1.00	4.17	1.67	3.12	1.20	1.94	1.97
	∞	1.00	4.20	1.63	3.18	1.18	1.92	1.94
Exp.(1)	10	1.00	2.44	1.17	2.11	0.98	1.10	1.18
	20	1.00	2.89	1.25	2.52	1.04	1.32	1.33
	30	1.00	2.96	1.30	2.58	1.07	1.38	1.39
	100	1.00	3.32	1.32	2.89	1.08	1.45	1.46
	∞	1.00	3.33	1.33	2.94	1.09	1.47	1.49
Lognormal(0,1)	10	1.00	1.40	1.19	1.20	0.96	1.26	1.28
	20	1.00	1.62	1.32	1.43	1.05	1.44	1.45
	30	1.00	1.56	1.34	1.34	1.07	1.48	1.49
	100	1.00	1.63	1.36	1.43	1.08	1.52	1.53
	∞	1.00	1.67	1.35	1.44	1.08	1.54	1.56
Av.Est.SE		NA	0.06	0.01	0.05	0.01	0.02	0.02

Table 2.18: Best Performing Negative Signed Estimators Under Skewed Distributions

Distribution	N	1^{st}	2^{nd}
Ex.Val.(0,1)	10	Sd$^-$ (2.93)	Sdtrim$^-$ (1.94)
	20	Sd$^-$ (3.76)	Sdtrim$^-$ (2.56)
	30	Sd$^-$ (3.90)	Sdtrim$^-$ (2.69)
	100	Sd$^-$ (4.48)	Sdtrim$^-$ (3.12)
	∞	Sd$^-$ (5.42)	Sdtrim$^-$ (3.16)
χ_4^2	10	Sd$^-$ (2.69)	Sdtrim$^-$ (2.03)
	20	Sd$^-$ (3.35)	Sdtrim$^-$ (2.54)
	30	Sd$^-$ (3.65)	Sdtrim$^-$ (2.76)
	100	Sd$^-$ (4.17)	Sdtrim$^-$ (3.12)
	∞	Sd$^-$ (4.20)	Sdtrim$^-$ (3.18)
Exp.(1)	10	Sd$^-$ (2.44)	Sdtrim$^-$ (2.11)
	20	Sd$^-$ (2.89)	Sdtrim$^-$ (2.52)
	30	Sd$^-$ (2.96)	Sdtrim$^-$ (2.58)
	100	Sd$^-$ (3.32)	Sdtrim$^-$ (2.89)
	∞	Sd$^-$ (3.33)	Sdtrim$^-$ (2.94)
Lognormal(0,1)	10	Sd$^-$ (1.40)	$S_{Bi,20,-}$ (1.28)
	20	Sd$^-$ (1.62)	$S_{Bi,20,-}$ (1.45)
	30	Sd$^-$ (1.56)	$S_{Bi,15,-}$ (1.49)
	100	Sd$^-$ (1.63)	$S_{Bi,20,-}$ (1.53)
	∞	Sd$^-$ (1.67)	$S_{Bi,20,-}$ (1.56)

Table 2.19: Finite Efficiencies For Positive Signed Estimators Under Skewed Distributions

Distribution	N	IQR^+	Sd^+	Abs^+	$Sdtrim^+$	$Abstrim^+$	$S_{Bi,15,+}$	$S_{Bi,20,+}$
Ex.Val.(0,1)	10	1.00	1.34	1.66	0.90	1.13	1.25	1.31
	20	1.00	1.56	1.86	1.31	1.28	1.59	1.89
	30	1.00	1.51	1.91	1.52	1.36	1.93	1.95
	100	1.00	1.61	2.08	1.76	1.44	2.19	2.18
	∞	1.00	1.52	2.08	1.90	1.47	2.22	2.17
χ_4^2	10	1.00	1.37	1.66	0.81	1.14	0.03	1.45
	20	1.00	1.47	1.88	1.17	1.31	1.73	1.86
	30	1.00	1.43	1.91	1.32	1.35	1.87	1.93
	100	1.00	1.60	2.09	1.58	1.44	2.18	2.23
	∞	1.00	1.54	2.09	1.66	1.45	2.23	2.23
Exp.(1)	10	1.00	1.10	1.61	0.65	1.12	0.25	1.19
	20	1.00	1.27	1.83	0.89	1.28	0.20	1.62
	30	1.00	1.25	1.91	1.07	1.37	1.75	1.95
	100	1.00	1.24	2.06	1.31	1.49	2.06	2.20
	∞	1.00	1.22	2.08	1.38	1.52	2.17	2.27
Lognormal(0,1)	10	1.00	0.55	1.10	0.44	1.00	0.04	0.06
	20	1.00	0.51	1.22	0.51	1.14	0.84	1.42
	30	1.00	0.46	1.19	0.64	1.26	1.41	1.63
	100	1.00	0.22	1.20	0.88	1.39	1.69	1.81
	∞	1.00	0.13	1.19	0.94	1.43	1.72	1.85
Av.Est.SE		NA	0.03	0.03	0.02	0.02	0.03	0.03

although Abs$^+$ has identical performance under the χ^2_4 distribution.

The last row of Table 2.19 provides an average of estimates of the standard errors of each entry within a column. For these skewed right distributions, the trimmed estimators have the lowest standard errors while the other estimators have similar standard errors.

The best and second best performing positive signed estimators, under the skewed distributions, are presented in Table 2.20. For small samples, Abs$^+$ outperforms all other positive signed estimators, though the A-estimators are very close in terms of efficiency. For the larger samples of size 20, 30, and 100, $S_{Bi,20,+}$ appears to be the overall best performer.

Table 2.20: Best Performing Positive Signed Estimators Under Skewed Distributions

Distribution	Sample Size	1^{st}	2^{nd}
Ex.Val.(0,1)	10	Abs$^+$ (1.66)	Sd$^+$ (1.34)
	20	$S_{Bi,20,+}$ (1.89)	Abs$^+$ (1.86)
	30	$S_{Bi,20,+}$ (1.95)	$S_{Bi,15,+}$ (1.93)
	100	$S_{Bi,15,+}$ (2.19)	$S_{Bi,20,+}$ (2.18)
	∞	$S_{Bi,15,+}$ (2.22)	$S_{Bi,20,+}$ (2.17)
χ^2_4	10	Abs$^+$ (1.66)	$S_{Bi,20,+}$ (1.45)
	20	Abs$^+$ (1.88)	$S_{Bi,20,+}$ (1.86)
	30	$S_{Bi,20,+}$ (1.93)	Abs$^+$ (1.91)
	100	$S_{Bi,20,+}$ (2.23)	$S_{Bi,15,+}$ (2.18)
	∞	$S_{Bi,15,+}$ (2.23) , $S_{Bi,20,+}$ (2.23)	Abs$^+$ (2.09)
Exp.(1)	10	Abs$^+$ (1.61)	$S_{Bi,20,+}$ (1.19)
	20	Abs$^+$ (1.83)	$S_{Bi,20,+}$ (1.62)
	30	$S_{Bi,20,+}$ (1.95)	Abs$^+$ (1.91)
	100	$S_{Bi,20,+}$ (2.20)	Abs$^+$ (2.06) ,$S_{Bi,15,+}$ (2.06)
	∞	$S_{Bi,20,+}$ (2.27)	$S_{Bi,15,+}$ (2.17)
Lognormal(0,1)	10	Abs$^+$ (1.10)	IQR$^+$ (1.00),Abstrim$^+$ (1.00)
	20	$S_{Bi,20,+}$ (1.42)	Abs$^+$ (1.22)
	30	$S_{Bi,20,+}$ (1.63)	$S_{Bi,15,+}$ (1.41)
	100	$S_{Bi,20,+}$ (1.81)	$S_{Bi,15,+}$ (1.69)
	∞	$S_{Bi,20,+}$ (1.85)	$S_{Bi,15,+}$ (1.72)

2.3.3 Summary of Simulation Results

For symmetric distributions with low kurtosis, the $\text{Sd}^{+/-}$ pair of estimators appears to perform well for small samples. When the symmetric distribution has a high kurtosis, the $S_{Bi,20,+/-}$ estimator performs quite well in comparison with the other pairs of signed estimators.

When the underlying distribution is skewed, $S_{Bi,20,+}$ performs as well as, or almost as well as, the other positive signed estimators for samples under 30. Sd^{-} performs the best among all other negative signed estimators when the underlying distribution is skewed and the sample size is less than 30. But, for highly skewed distributions, the performance of $S_{Bi,20,-}$ approaches that of Sd^{-}.

In short, for symmetric, low kurtosis, distributions, the $\text{Sd}^{+/-}$ pair are the estimators of choice. When the kurtosis of the symmetric distribution is large, the $S_{Bi,20,+/-}$ pair are the estimators of choice. For highly skewed distributions, $S_{Bi,20,+/-}$ are the estimators of choice for samples under size 30. For lightly skewed distributions, $S_{Bi,20,+}$ is the overall best performer among all other positive signed estimators, while Sd^{-} is the overall best performer among all other negative signed estimators.

2.4 Recommendations

The motivation behind Signed Scale Estimators arose from examination of boxplots and consideration of the Interquartile Range as the sum of two estimators. The previous sections dealt with the positive and negative estimators separately, and no one pair of estimators clearly dominated the others. This section will examine the asymptotic and small sample properties of the sum of the signed scale estimator pairs with the goal of identifying a particular signed estimator pair that will be used in the construction of a modified boxplot. For purposes of these comparisons, the goal is to examine the general trend of the standardized variances as a function of sample size to see how the speed of convergence of the summed estimators compares with that of the individual signed estimators. The subsequent tables will include values for samples of size 10, 30, 100, and asymptotic values as well.

Table 2.21 lists the small sample and asymptotic standardized variances of signed scale estimators under the symmetric distributions. The estimator pairs which have the smallest standardized variances overall are $\text{Sd}^{+/-}$, $\text{Sdtrim}^{+/-}$, $S_{Bi,15,+/-}$, and $S_{Bi,20,+/-}$. A couple of numbers are of special note in this table. The $S_{Bi,15,+/-}$ estimator for samples of size 10 from the Uniform(0,1) distribution has a standardized variance of 4.54, which is 3 times its asymptotic standardized variance. The asymptotic standardized variance for $\text{Sd}^{+/-}$ is 3.70 under the t_5 distribution, yet the estimate of the standardized variance is only 2.55 for samples of size 100.

Table 2.21: Standardized Variances For Positive and Negative Signed Estimators Under Symmetric Distributions

Distribution	N	IQR	Sd	Abs	Sdtrim	Abstrim	$S_{Bi,15}$	$S_{Bi,20}$
Uniform(0,1)	10	1.62	0.52	1.05	0.83	1.43	4.54	1.19
	30	2.44	0.50	1.41	0.77	1.94	1.35	1.30
	100	2.83	0.49	1.58	0.77	2.19	1.43	1.39
	∞	3.00	0.48	1.67	0.77	2.33	1.48	1.45
Normal(0,1)	10	1.96	0.82	1.12	1.28	1.68	1.39	1.15
	30	2.71	0.83	1.32	1.14	2.03	1.23	1.20
	100	3.06	0.86	1.47	1.15	2.26	1.31	1.29
	∞	3.20	0.84	1.47	1.11	2.29	1.29	1.27
Logistic(0,1)	10	2.16	1.18	1.33	1.58	1.91	1.43	1.35
	30	2.90	1.32	1.54	1.35	2.21	1.49	1.50
	100	3.12	1.35	1.57	1.28	2.26	1.50	1.53
	∞	3.31	1.37	1.62	1.26	2.36	1.52	1.55
Laplace(0,1)	10	3.05	2.05	1.80	2.69	2.69	1.94	1.81
	30	3.72	2.15	1.92	2.11	2.75	1.89	1.88
	100	4.13	2.37	2.01	1.97	2.81	1.93	1.95
	∞	4.16	2.50	2.00	1.91	2.77	1.90	1.94
t_5	10	2.30	2.21	1.63	1.96	2.08	1.71	1.63
	30	2.92	2.22	1.75	1.48	2.24	1.66	1.72
	100	3.28	2.55	1.83	1.36	2.37	1.68	1.79
	∞	3.35	3.70	1.85	1.33	2.39	1.68	1.80

The Monte Carlo sample size is 5000, and the relative standard error of the entries is approximately 0.02.

Table 2.22 lists the small sample and asymptotic standardized variances for the negative signed scale estimators under the skewed distributions. Sd^- has the smallest standardized variance in every case. One value of note in Table 2.22 is the asymptotic value of Sd^- under the Extreme Value(0,1) distribution. The limiting value is 0.60, yet even for samples of size 100 the estimated standardized variance is 0.73. This suggests that Sd^- has a relatively slow convergence under this distribution.

Table 2.22: Standardized Variances For Negative Signed Estimators Under Skewed Distributions

Distribution	N	IQR^-	Sd^-	Abs^-	Sdtrim^-	Abstrim^-	$S_{Bi,15,-}$	$S_{Bi,20,-}$
Ex.Val.(0,1)	10	2.19	0.75	1.34	1.13	1.97	1.33	1.23
	30	2.80	0.72	1.55	1.04	2.27	1.36	1.34
	100	3.26	0.73	1.72	1.04	2.56	1.46	1.43
	∞	3.25	0.60	1.69	1.03	2.55	1.40	1.37
χ^2_4	10	2.37	0.88	1.69	1.17	2.25	1.67	1.97
	30	2.97	0.81	1.86	1.08	2.52	1.66	1.64
	100	3.34	0.80	2.00	1.07	2.78	1.72	1.69
	∞	3.43	0.81	2.08	1.07	2.88	1.77	1.75
Exp.(1)	10	3.08	1.27	2.64	1.46	3.13	2.81	2.62
	30	3.58	1.21	2.76	1.39	3.33	2.59	2.57
	100	3.82	1.15	2.90	1.32	3.55	2.64	2.62
	∞	4.06	1.21	3.05	1.38	3.72	2.75	2.74
Lognormal(0,1)	10	4.06	2.90	3.42	3.38	4.24	3.22	3.18
	30	4.15	2.66	3.09	3.09	3.87	2.80	2.78
	100	4.28	2.62	3.15	3.00	3.95	2.81	2.79
	∞	4.35	2.61	3.20	3.03	4.03	2.82	2.80

The Monte Carlo sample size is 5000, and the relative standard error of the entries is approximately 0.02.

Table 2.23 lists the small sample and asymptotic standardized variances for the positive signed scale estimators under the skewed distributions. Abs^+ has the smallest standardized variance for samples of size 10, while one or the other of the two A-estimators has the smallest standardized variance in every other situation. Several values of are particular interest in this table. The $S_{Bi,15,+}$ estimator has unusually large estimates of standardized variance for samples of size 10 under the χ^2_4, Lognormal(0,1), and Exponential(1) distributions. The $S_{Bi,20,+}$ estimator also has an unusually large estimate for samples of size 10 under the Lognormal(0,1) distributions. These two estimators are most likely unstable at such small sample sizes due to the high skewness of the χ^2_4 and Lognormal(0,1) distributions. Another point of interest is the behavior of Sd$^+$ under the Lognormal(0,1) distribution. The standardized variances of Sd$^+$ under this distribution are the largest for any given sample size, except for samples of size 10. Additionally, the asymptotic standardized variance is 8 times that of any other estimator, and the standardized variance estimate for samples of size 100 is only approximately half that of the asymptotic value.

Table 2.23: Standardized Variances For Positive Signed Estimators Under Skewed Distributions

Distribution	N	IQR$^+$	Sd$^+$	Abs$^+$	Sdtrim$^+$	Abstrim$^+$	$S_{Bi,15,+}$	$S_{Bi,20,+}$
Ex.Val.(0,1)	10	2.41	1.80	1.45	2.67	2.13	1.93	1.84
	30	3.07	2.03	1.60	2.02	2.26	1.59	1.57
	100	3.47	2.16	1.67	1.98	2.40	1.59	1.59
	∞	3.58	2.35	1.73	1.88	2.44	1.62	1.64
χ^2_4	10	2.51	1.83	1.52	3.10	2.20	82.72	1.72
	30	3.10	2.17	1.62	2.35	2.30	1.66	1.60
	100	3.49	2.18	1.67	2.21	2.43	1.60	1.57
	∞	3.71	2.35	1.73	2.18	2.49	1.62	1.62
Exp.(1)	10	2.88	2.62	1.79	4.40	2.56	11.57	2.41
	30	3.62	2.89	1.90	3.37	2.65	2.06	1.86
	100	3.98	3.21	1.94	3.05	2.67	1.93	1.81
	∞	4.16	3.40	2.00	3.01	2.77	1.92	1.82
Lognormal(0,1)	10	4.44	8.05	4.03	10.19	4.44	108.53	69.08
	30	5.01	10.99	4.22	7.79	3.97	3.55	3.08
	100	5.13	23.75	4.27	5.82	3.69	3.04	2.83
	∞	5.23	41.26	4.37	5.56	3.69	3.02	2.84

The Monte Carlo sample size is 5000, and the relative standard error of the entries is approximately 0.02.

The comparison of relative efficiencies and standardized variances of the signed scale estimators suggests that either one of Sd$^{+/-}$, Sdtrim$^{+/-}$, $S_{Bi,15,+/-}$, $S_{Bi,20,+/-}$, or Abs$^{+/-}$ may outperform the others in various situations. In order to see how the sums of these individual estimators perform, Table 2.24 lists the small sample and asymptotic standardized variances of the sums of the signed scale estimators under symmetric distributions. The sum of the A-estimators and the sum of the Sd$^{+/-}$ pair have standardized variances that are extremely close, and smallest under the Uniform(0,1) and Normal(0,1) distributions. The sum of Abs$^{+/-}$ pair and the sum of the A-estimators have the overall smallest standardized variances under the Logistic(0,1), Laplace(0,1), and t_5 distributions. The summed estimators have standardized variance estimates that are uniformly smaller than the individual signed estimator standardized variances. The overall speed of convergence for the summed estimators is a bit faster than that of the individual signed estimators. There are two entries in Table 2.24 that are of particular note. The estimates of standardized variance for the $S_{Bi,15}$ summed estimator for samples of size 10 under from the Uniform(0,1) and Logistic(0,1) are extremely large compared to the corresponding asymptotic values. These values indicate that the $S_{Bi,15}$ summed estimator is highly variable in these cases.

Table 2.24: Standardized Variances For Summed Estimators Under Symmetric Distributions

Distribution	N	IQR	Sd	Abs	Sdtrim	Abstrim	$S_{Bi,15}$	$S_{Bi,20}$
Uniform(0,1)	10	0.69	0.26	0.34	0.54	0.62	604.91	0.34
	30	0.89	0.21	0.33	0.50	0.64	0.22	0.21
	100	0.96	0.21	0.34	0.50	0.66	0.21	0.21
	∞	1.00	0.20	0.33	0.50	0.67	0.20	0.20
Normal(0,1)	10	0.96	0.52	0.55	0.88	0.88	0.53	0.50
	30	1.18	0.49	0.54	0.78	0.87	0.49	0.48
	100	1.32	0.52	0.59	0.82	0.93	0.53	0.52
	∞	1.36	0.50	0.57	0.80	0.92	0.51	0.50
Logistic(0,1)	10	1.14	0.71	0.68	1.07	1.05	41.00	0.66
	30	1.33	0.76	0.69	0.96	1.00	0.69	0.69
	100	1.48	0.77	0.71	0.92	1.02	0.70	0.71
	∞	1.47	0.80	0.71	0.91	1.01	0.71	0.72
Laplace(0,1)	10	1.65	1.15	0.98	1.81	1.53	1.16	0.97
	30	1.86	1.18	0.98	1.45	1.40	1.03	1.00
	100	2.00	1.24	0.99	1.37	1.39	1.04	1.03
	∞	2.08	1.25	1.00	1.31	1.38	1.05	1.04
t_5	10	1.19	1.19	0.83	1.30	1.14	0.84	0.82
	30	1.38	1.20	0.82	1.04	1.05	0.81	0.84
	100	1.52	1.43	0.85	1.01	1.08	0.84	0.88
	∞	1.50	2.00	0.85	0.95	1.04	0.84	0.90

The Monte Carlo sample size is 5000, and the relative standard error of the entries is approximately 0.02.

Table 2.25 lists the small sample and asymptotic standardized variances of the sums of the signed scale estimators under the skewed distributions. The sum of the $\text{Abs}^{+/-}$ pair and the sum of the

A-estimators have the lowest standardized variances among all sums of signed scale estimators. The $\text{Abs}^{+/-}$ pair has the smallest standardized variances for the samples of size 10. The standardized variance of the sum of the A-estimators is less than that of the sum of $\text{Abs}^{+/-}$ under the Lognormal distribution for samples of size 30, 100, and asymptotically. There are several entries in Table 2.25 that are of interest. The $S_{Bi,15}$ and $S_{Bi,20}$ summed estimators have unusually large standardized variance estimates for samples of size 10 from the Exponential(1) and Lognormal(0,1) distributions. This behavior is similar to the behavior of the $S_{Bi,15,+}$ and $S_{Bi,20,+}$ estimators under the skewed distributions as evidenced in Table 2.23. Overall, the standardized variances of the summed estimators are smaller than the standardized estimates of the corresponding positive signed estimators. The standardized variances of the summed estimators are, almost always, smaller than the standardized variance of the corresponding negative signed estimators except for the Sd^{-} and Sdtrim^{-} estimators. These two negative signed estimators have uniformly smaller standardized variances than the corresponding summed estimators. The convergence rate of the standardized variances for the summed estimators is similar to the convergence rate of the negative and positive signed estimators.

Based on the small sample comparisons and the asymptotic comparisons, the pair of signed estimators that appears to be the most efficient, individually and summed, is the $\text{Abs}^{+/-}$ pair. The A-estimator pairs follow closely behind in terms of overall appeal and performance. As such, the $\text{Abs}^{+/-}$ pair will be considered in more detail in Chapter 3 where a new type of boxplot will be introduced.

Table 2.25: Standardized Variances For Summed Estimators Under Skewed Distributions

Distribution	N	IQR	Sd	Abs	Sdtrim	Abstrim	$S_{Bi,15}$	$S_{Bi,20}$
Ex.Val.(0,1)	10	1.18	0.94	0.73	1.29	1.09	1.02	0.88
	30	1.38	1.05	0.77	1.15	1.05	0.77	0.77
	100	1.51	1.09	0.78	1.12	1.12	0.75	0.76
	∞	1.58	1.15	0.80	1.10	1.11	0.78	0.79
χ^2_4	10	1.25	1.12	0.83	1.46	1.19	1.31	0.88
	30	1.55	1.28	0.89	1.37	1.22	0.91	0.89
	100	1.60	1.34	0.87	1.25	1.18	0.87	0.87
	∞	1.72	1.35	0.90	1.26	1.23	0.89	0.89
Exp.(1)	10	1.69	1.61	1.16	2.08	1.63	4.91	4.07
	30	2.10	1.94	1.28	2.00	1.69	1.46	1.34
	100	2.10	2.13	1.26	1.82	1.60	1.36	1.29
	∞	2.21	2.19	1.28	1.77	1.62	1.33	1.28
Lognormal(0,1)	10	2.95	6.11	2.99	5.32	3.15	940.94	18.77
	30	2.94	11.03	3.11	4.34	2.54	2.52	2.35
	100	3.03	12.65	2.94	3.45	2.42	2.32	2.24
	∞	3.03	27.33	3.02	3.25	2.36	2.27	2.21

The Monte Carlo sample size is 5000, and the relative standard error of the entries is approximately 0.02.

Chapter 3

A NEW RULE FOR GENERATING BOXPLOTS

3.1 Motivation

The boxplot is a method of displaying univariate data developed by John Tukey (1977). The main features of a boxplot are illustrated by a hypothetical boxplot shown in Figure 3.1.

The features that are readily noticeable are “the box,” defined by the values Q_1 and Q_3, that captures the notion of the middle half of the data and a line across the center of the box that indicates a central value of the data, Q_2. Lines, or whiskers, extend outward from the two ends of the box to the two most extreme values that are still within the upper and lower fences, denoted F_U and F_L respectively. Additionally, observations that are outside the fences, called “outsiders”, are presented individually so that they will attract attention and receive further scrutiny.

Although the boxplot as introduced by Tukey uses particular estimates of the first and third quartiles for Q_1 and Q_3, many current statistical software packages use various forms of the first and third quartiles as the default option when constructing boxplots (Frigge et al. 1989). Construction of the common boxplot follows the same procedure, regardless of the definition used in calculating the first and third quartiles. Construction of a common boxplot follows the method outlined in Figure 3.1 and uses the first and third quartiles for the values Q_1 and Q_3. Based on work by Hoaglin et al. (1986), values of k are typically 1.5 or 3.0. Following their conclusion that a value of 3.0 is too large, a value of 1.5 will be used for all common boxplots.

A drawback to the common boxplot, aside from the many definitions of quartiles that exist, is that the lengths of the upper and lower fences are identical since the lengths are multiples of

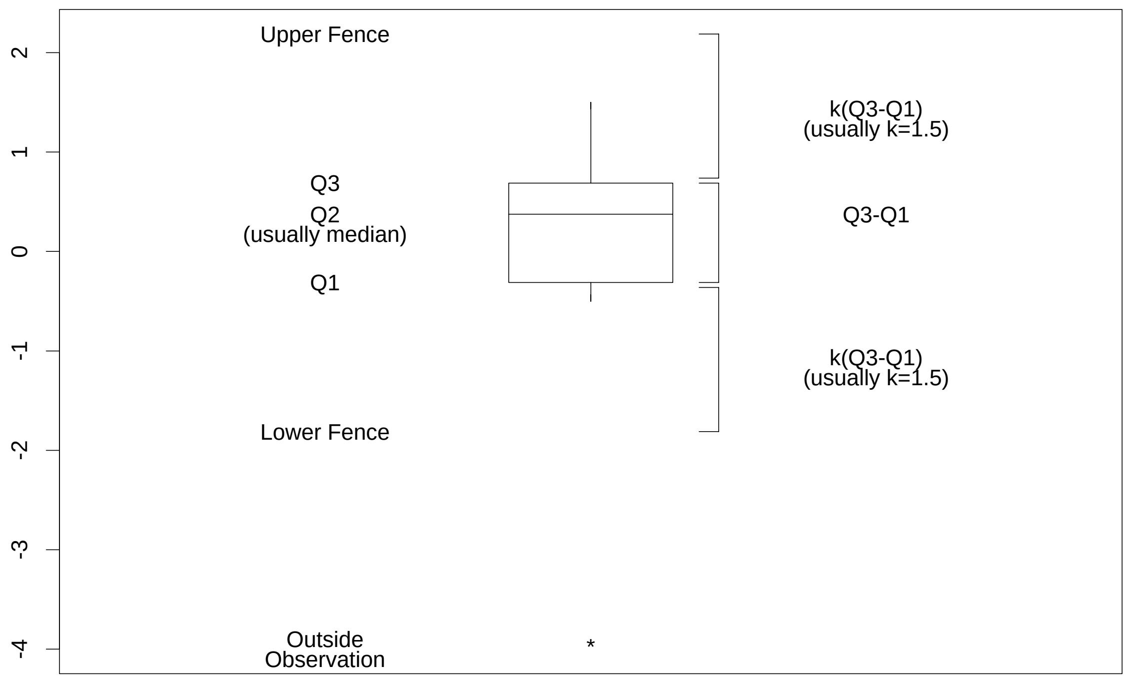

Figure 3.1: Construction of a Hypothetical Boxplot.

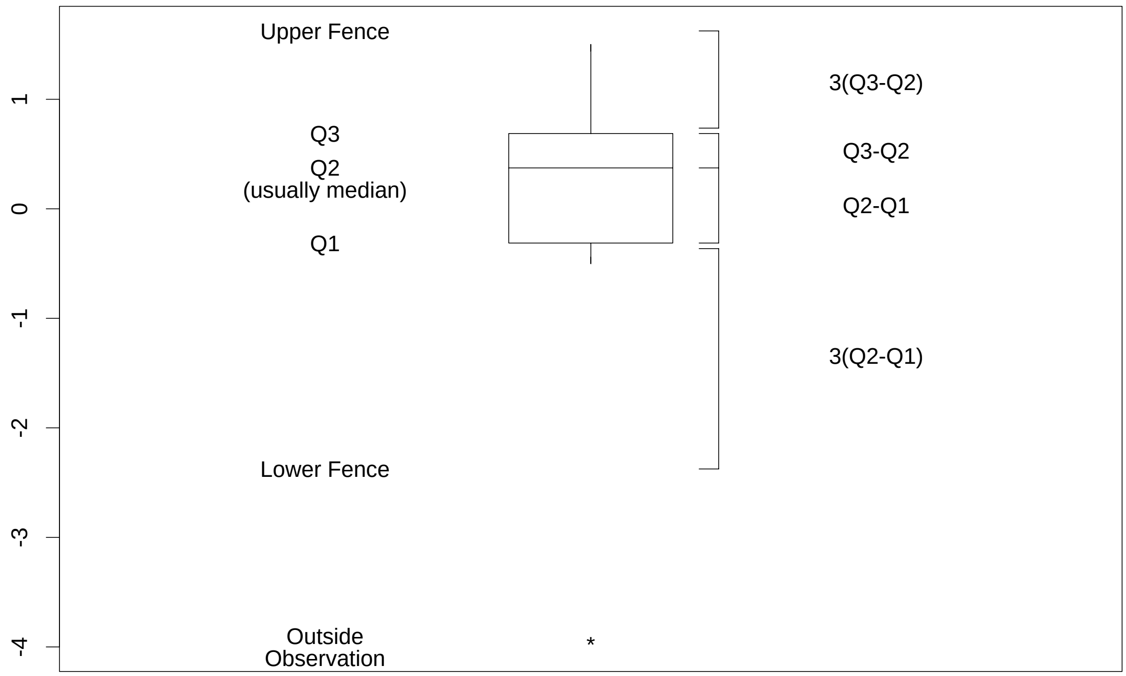

Figure 3.2: Construction of a Skew Adjusted Boxplot.

the interquartile range. Since the fences are used to distinguish observations that are outside from those that are inside, using fences of the same length for skewed distributions may overestimate the number of observations that are outside on the skewed side and underestimate the number of observations that are outside on the other side. An alternative rule for constructing fences in an attempt to capture the presence of skewness was presented by Kimber (1990). Figure 3.2 illustrates the construction of a boxplot, referred to as a skew adjusted boxplot, that uses Kimber's rule for constructing fences.

The difference between the skew adjusted boxplot and the common boxplot lies in the construction of the upper and lower fences. For the skew adjusted boxplot, the lower fence is defined to be $Median - 4 * (Median - Q_1)$ and the upper fence is defined to be $Median + 4 * (Q_3 - Median)$. The lower and upper fences have lengths of $3 * (Median - Q_1)$ and $3 * (Q_3 - Median)$, respectively. For symmetric distributions, the quantities $Median - Q_1$ and $Q_3 - Median$ are identical and the upper and lower fences have lengths of

$$3 * (Median - Q_1) = 3 * (.5 * IQR) = 1.5 * IQR\ ,$$

which agree with the lengths of the fences in the common boxplot. For non-symmetric data, however, the lower and upper fences will most likely be of differing lengths. Just as with the common boxplot, observations beyond the lower and upper fences are said to be outside.

One way in which a boxplot provides information about skewness is via the relative lengths of the two parts of its box. The relative lengths of the two parts of the box can be quantified by the ratio Q,

$$\begin{aligned} Q &= \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} \\ &= \frac{IQR^+ - IQR^-}{IQR^+ + IQR^-}. \end{aligned}$$

As seen previously, IQR^+ and IQR^- are inefficient signed estimators, as compared with signed estimators from Chapter 2 such as Abs^+ and Abs^-. Since the common and skew adjusted boxplots rely upon these inefficient estimators, information provided by the boxplots about the structure of the underlying data may well be obscured. Specifically, since the construction of the box part of the common and skew adjusted boxplots rely upon IQR^+ and IQR^-, then any skewness information derived from the shape of the box may be hidden or muted due to the inefficiency of the underlying estimators. Of course, the common and skew adjusted boxplots use IQR^+ and IQR^- to determine the lengths of their fences which, in turn, influence the lengths of the whiskers. The inefficiency of these estimators may cause information about skewness derived from the whiskers to also be hidden.

Additionally, since boxplots are often used to identify unusual or interesting data, the method by which upper and lower fences are constructed must be carefully chosen. The common boxplot uses a rule which leads to upper and lower fences equidistant from the ends of the box. This rule, while easy to implement, will tend to label too many observations as unusual when the data come from a skewed distribution. On the other hand, the skew adjusted boxplot uses a rule which attempts to use the skewness of the data in the construction of the upper and lower fences. This rule identifies fewer observations as outsiders than the common boxplot rule, when the data come from a skewed distribution. However, the skew adjusted boxplots' upper and lower fences are based on multiples of IQR^+ and IQR^-, which have been shown to be inefficient as compared with other signed scale estimators.

Section 2 of the chapter will provide a new boxplot construction rule, compare the visual information provided between all three boxplot rules under three examples, and examine the asymptotic values of the quantities used in the boxplot construction under various distributions. The tendency of the various boxplots to label observations as outside will be examined for small samples via Monte Carlo simulation and the results will be presented in Section 3. Lastly, Section 4 will examine the efficiency of skewness measures derived from the common boxplot and the new signed boxplots via Monte Carlo simulation.

3.2 Signed Scale Boxplot Rule

Both the common boxplot and the skew adjusted boxplot use quartiles in their construction. An alternative to both of these types of boxplots uses signed scale estimators instead of quartiles. Based on the asymptotic results in Chapter 2, the upper and lower signed scale estimators Abs^+ and Abs^- outperform IQR^+ and IQR^- under a variety of distributions. This pair of signed scale estimators will be used in place of the quartiles to construct another type of boxplot. Figure 3.3 shows the construction of what will be referred to as a signed scale boxplot.

Specifically, let $Q_1 = Median - .845 * Abs^-$ and let $Q_3 = Median + .845{*}\text{Abs}^+$. The lower fence is defined to be $F_L = Median - 4 * (.845 * Abs^-)$ and the upper fence is defined to be $F_U = Median + 4 * (.845 * Abs^+)$. The lower and upper fences have lengths of $3 * (.845 * Abs^-)$ and $3 * (.845 * Abs^+)$, respectively. For symmetric data, the quantities $.845 * Abs^+$ and $.845 * Abs^-$ are identical and lead to the lower and upper fences having the same length. For non-symmetric data, the lower and upper fences will be of different lengths. The choice of the value .845 is such that, as the sample size increases for data coming from a Normal(0,1) distribution, the shapes of the three boxplots converge to the same shape. Tables 3.1 and 3.2 give the limiting values for Q_1, Q_2, Q_3, lower fence, and upper fence for various distributions; details are given in Appendix A.2. It should be noted that all three boxplot construction rules create fences outside the domain of the Uniform(0,1) distribution.

For the (right) skewed distributions, the following relationships exist for limiting values,

$$F_L\,(common) \leq F_L\,(skewed) \leq F_L\,(signed)$$

and

$$F_U\,(common) \leq F_U\,(skewed) \leq F_U\,(signed)\,.$$

Letting Q_1, Q_2, and Q_3 represent the first, second, and third quartiles of F, then $Q_3 - Q_2 \geq$

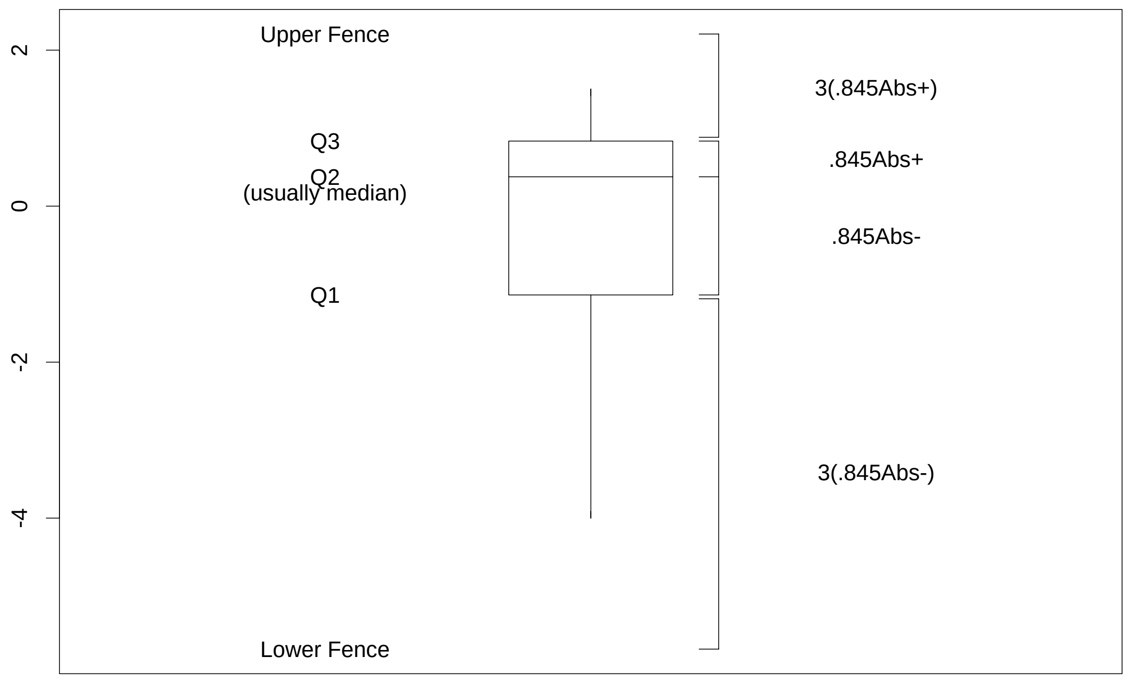

Figure 3.3: Construction of a Signed Boxplot.

$Q_2 - Q_1$. Since $Q_3 - Q_1 = (Q_3 - Q_2) + (Q_2 - Q_1)$ then $Q_3 - Q_1 \leq 2(Q_3 - Q_2)$ and thus

$$F_U\,(common) = Q_3 + 1.5(Q_3 - Q_1) \leq Q_3 + 1.5(2(Q_3 - Q_2) = Q_3 + 3(Q_3 - Q_2) = F_U\,(skewed)$$

Noting that $F_U\,(skewed)$ can be written as $Q_2 + 4(Q_3 - Q_2)$ and that

$$F_U\,(signed) = Q_2 + 4((.845)T_{1,+}\,(F))$$

where $T_{1,+}\,(F)$ is the limiting value of Abs^+ as defined in Section 1.2.1 on page 5 with $\mu\,(F) = F^{-1}\,(.5)$, then $F_U\,(skewed) \leq F_U\,(signed)$ if $(.845)T_{1,+}\,(F) \geq Q_3 - Q_2$. Using the limiting values of Abs^+ given in Table 2.8 and the limiting values of Q_3 and Q_2 given in Table 3.1 for the four right-skewed distributions considered here, we obtain $F_U\,(skewed) \leq F_U\,(signed)$. A similar analysis shows that the relationships between the lower fences also holds.

By choice of the constant .845 in constructing the signed boxplot, the lower and upper fences are identical for all three boxplots under the normal distribution. The common and skew adjusted boxplots agree for every symmetric distribution while, except for the Uniform(0,1) distribution, the signed boxplot has the following relationship with them,

$$F_L\,(signed) \leq F_L\,(common) = F_L\,(skew)$$

and

$$F_U\,(signed) \geq F_U\,(common) = F_U\,(skew)\,.$$

Letting Q_1, Q_2, and Q_3 represent the first, second, and third quartiles of F, then $Q_3 - Q_2 = Q_2 - Q_1$. Since $Q_3 - Q_1 = (Q_3 - Q_2) + (Q_2 - Q_1)$ then $Q_3 - Q_1 = 2(Q_3 - Q_2)$ and thus

$$F_U\,(common) = Q_3 + 1.5(Q_3 - Q_1) = Q_3 + 1.5(2(Q_3 - Q_2) = Q_3 + 3(Q_3 - Q_2) = F_U\,(skewed)$$

Noting that $F_U\,(skewed)$ can be written as $Q_2 + 4(Q_3 - Q_2)$ and that

$$F_U\,(signed) = Q_2 + 4((.845)T_{1,+}\,(F))$$

where $T_{1,+}\,(F)$ is the limiting value of Abs^+ defined in Section 1.2.1 on page 5 with $\mu\,(F) = F^{-1}\,(.5)$, then $F_U\,(skewed) \leq F_U\,(signed)$ if $(.845)T_{1,+}\,(F) \geq Q_3 - Q_2$. Using the limiting values of Abs^+ given in Table 2.8 and the limiting values of Q_2 and Q_3 given in Table 3.1 for the four symmetric distributions considered here, we obtain $F_U\,(skewed) \leq F_U\,(signed)$. A similar analysis shows that the relationships between the lower fences also holds.

All of these relationships suggest that the signed boxplot accentuates skewness when the underlying distribution is skewed, and accentuates heavy-tailedness when the data come from a symmetric heavy-tailed distribution. The visual differences between the three boxplot rules is next examined using three real datasets.

Table 3.1: Limiting Values For The Common and Skew Adjusted Boxplots

	Common					Skew				
Distribution	Q_1	Q_2	Q_3	F_L	F_U	Q_1	Q_2	Q_3	F_L	F_U
Uniform(0,1)	0.25	0.50	0.75	-0.50	1.50	0.25	0.50	0.75	-0.50	1.50
Normal(0,1)	-0.67	0.00	0.67	-2.70	2.70	-0.67	0.00	0.67	-2.70	2.70
Logistic(0,1)	-1.10	0.00	1.10	-4.39	4.39	-1.10	0.00	1.10	-4.39	4.39
Laplace(0,1)	-0.69	0.00	0.69	-2.77	2.77	-0.69	0.00	0.69	-2.77	2.77
t_5	-0.73	0.00	0.73	-2.91	2.91	-0.73	0.00	0.73	-2.91	2.91
Ex.Val.(0,1)	-0.33	0.37	1.25	-2.69	3.60	-0.33	0.37	1.25	-2.41	3.88
χ^2_4	1.92	3.36	5.39	-3.27	10.58	1.92	3.36	5.39	-2.38	11.47
Exp.(1)	0.29	0.69	1.39	-1.36	3.03	0.29	0.69	1.39	-0.93	3.47
Lognorm.(0,1)	0.51	1.00	1.96	-1.67	4.14	0.51	1.00	1.96	-0.96	4.85

All values rounded to two decimal places. F_L and F_U are calculated from the unrounded values for Q_1, Q_2, and Q_3.

Table 3.2: Limiting Values For the Signed Boxplot

	Signed				
Distribution	Q_1	Q_2	Q_3	F_L	F_U
Uniform(0,1)	0.29	0.50	0.71	-0.35	1.35
Normal(0,1)	-0.67	0.00	0.67	-2.70	2.70
Logistic(0,1)	-1.17	0.00	1.17	-4.69	4.69
Laplace(0,1)	-0.85	0.00	0.85	-3.38	3.38
t_5	-0.80	0.00	0.80	-3.21	3.21
Extreme Value(0,1)	-0.27	0.37	1.36	-2.19	4.35
χ^2_4	2.12	3.36	5.68	-1.58	12.64
Exponential(1)	0.37	0.69	1.54	-0.61	4.07
Lognormal(0,1)	0.60	1.00	2.50	-0.61	7.00

All values rounded to two decimal places. F_L and F_U are calculated from the unrounded values for Q_1, Q_2, and Q_3.

3.2.1 Example 1

Example 1 considers data (Barnett 1994, p.14) on annual incomes, in pounds, of 69 scientific and literary societies in England during 1840 (organizations with incomes less than 75 pounds are excluded, and incomes are rounded to the nearest pound). These data are listed in Table 3.3 and presented in Figure 3.4, in the form of boxplots, using each boxplot method.

Table 3.3: Annual Incomes of 69 Scientific and Literary Societies in England in 1840

77	90	112	135	169	200	249	404	700	1400
77	90	115	136	170	201	290	431	800	1878
79	90	120	138	170	206	300	445	844	2000
80	92	120	140	190	208	309	456	900	2363
80	100	120	147	200	230	335	500	900	3000
84	102	125	150	200	230	350	650	1050	7000
87	110	130	150	200	237	400	650	1300	

No incomes below the median income are presented as being "outside." All outside values are above the upper fence. For the common boxplot, eight incomes above 900 pounds are labeled as outsiders. For the skew adjusted boxplot, seven incomes above 1050 pounds are labeled as outsiders. For the signed boxplot, only the 3000 and 7000 pound incomes are labeled as outsiders. The Common and Skew Adjusted boxplots visually appear to provide almost identical information about skewness, while the signed boxplot suggests a more highly skewed distribution.

3.2.2 Example 2

Example 2 consists of data on the capacities, in cubic-centimeters, of seventeen male Moriori skulls (Barnett 1994, p.40). Table 3.4 lists these data, while the data are presented in Figure 3.5 using the three boxplot methods.

Table 3.4: Cranial Capacity of 17 Male Moriori Skulls

1230	1318	1380	1420	1630	1378
1348	1380	1470	1445	1360	1410
1540	1260	1364	1410	1545	

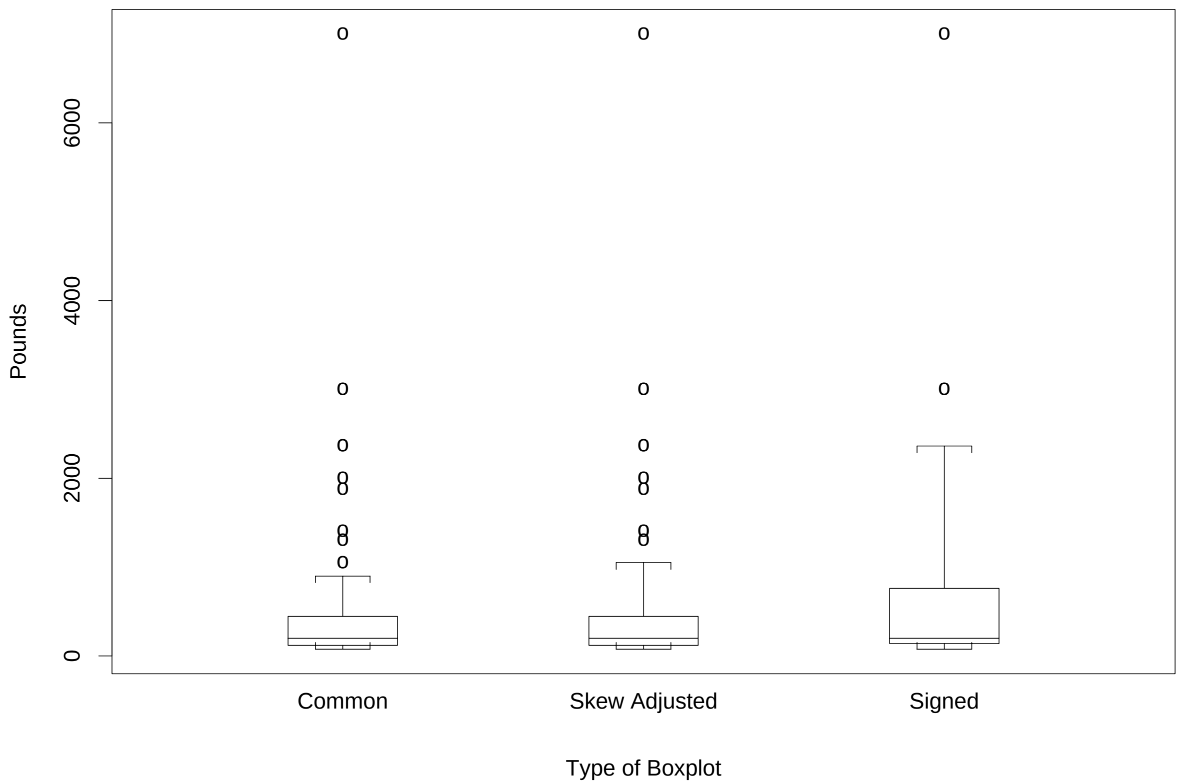

Figure 3.4: Incomes of the Top 69 Scientific and Literary Societies of England in 1840.

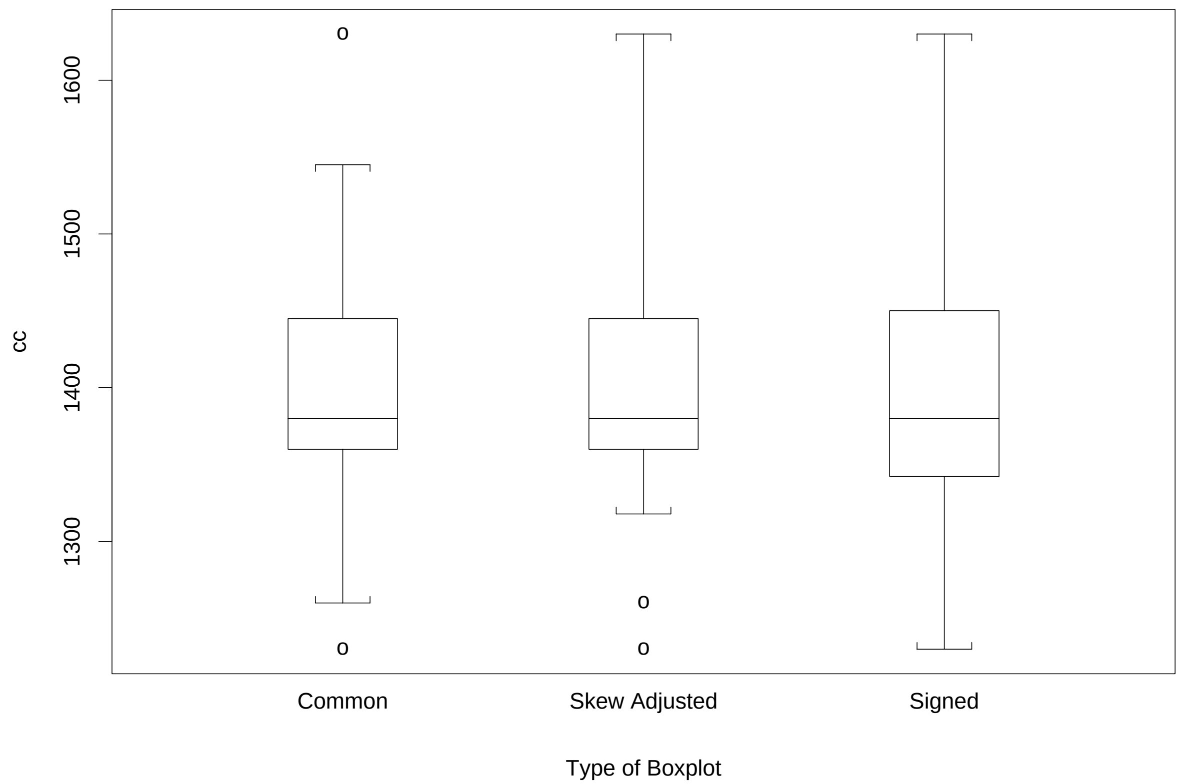

Figure 3.5: Cranial Capacity of 17 Male Moriori Skulls.

The common boxplot flags the lowest and highest values, 1230 and 1630, respectively, as outsiders. The skew adjusted boxplot does not label any observations above the upper fence as outsiders. However, the skew adjusted boxplot does flag the two smallest observations, 1230 and 1260, as outsiders. Lastly, the signed boxplot does not label any points as outsiders. While Barnett et al. (1994) report that the 1630 observation was suspected as being "too large," the only boxplot rule that flags this observation is the common boxplot. The skew adjusted boxplot even points out the two smallest observations. The shape of the box for the common boxplot suggests a skewed distribution, yet the almost identical length of the whiskers suggests a symmetric distribution. The skew adjusted boxplot suggests skewness both through the shape of its box and through the difference in lengths of its whiskers. The signed boxplot suggests a slightly skewed distribution via the shape of the box and the relative length of the whiskers.

3.2.3 Example 3

Example 3 consists of the residuals, about a simple model, of 15 observations made by Lt. Herndom in 1846 of the vertical semi-diameter of the planet Venus (Barnett 1994, p.38). The data are listed in Table 3.5 and presented in Figure 3.6 using each of the three boxplot methods.

Table 3.5: 15 Observations of the Vertical Semi-diameter of the Planet Venus

-0.30	0.48	0.63	-0.22	0.18
-0.44	-0.24	-0.13	-0.05	0.39
1.01	0.06	-1.40	0.20	0.10

All three boxplots label the observation -1.4 as an outsider. Only the skew adjusted boxplot labels any point above the upper fence as an outsider. Barnett et al. (1994) found -1.4 to not quite be an "outlier." The common and skew adjusted boxplots have identical boxes while the signed boxplot has a box which is more elongated along the "Seconds" axis. Additionally, the common and signed boxplots have identical whiskers while the skew adjusted boxplot has a much shorter upper whisker than either of them. All three boxplots suggests symmetry of the data near the median, though the signed boxplot suggests more variability about the median than the others, but while the common and signed boxplots indicate a skewed right distribution, as indicated by the length of their upper whiskers, the skew adjusted boxplot indicate symmetry with its whiskers.

As seen by these three examples, the information on skewness and outside observations provided by the three type of boxplots can vary from sample to sample. Additionally, each boxplot rule can provide a different picture of a set of data. It is difficult to say which boxplot is the most informative

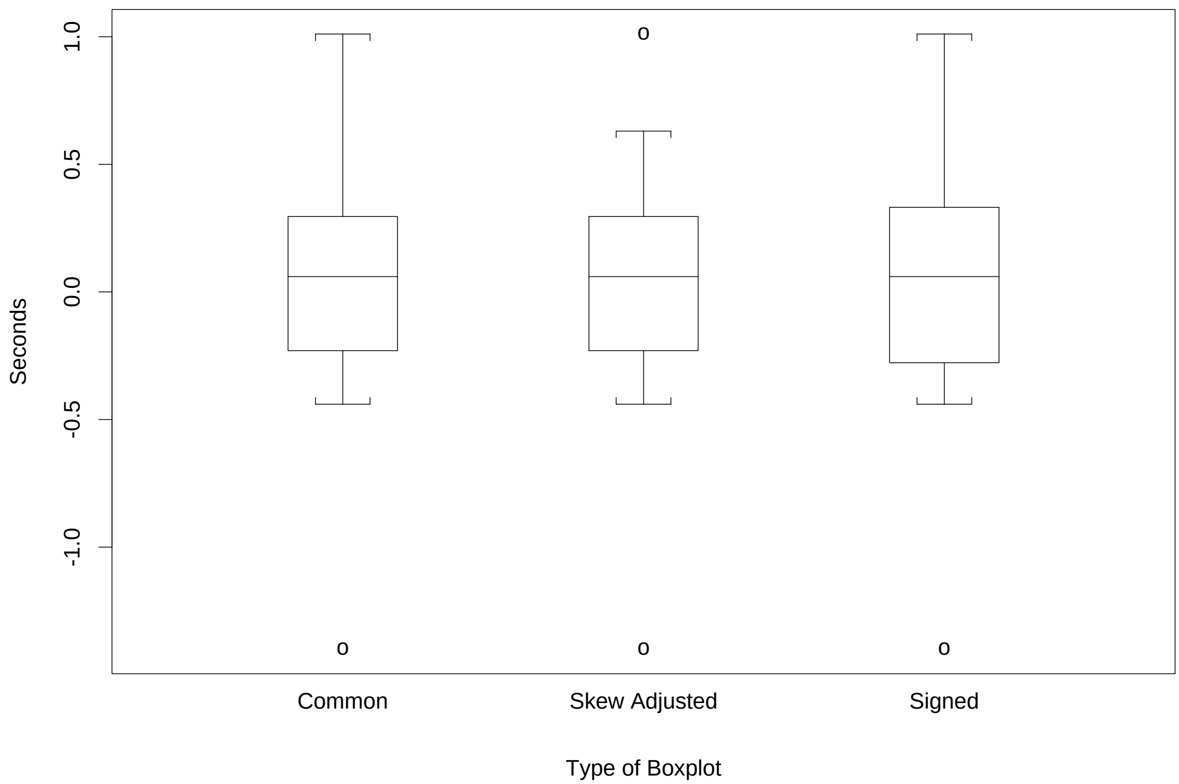

Figure 3.6: Vertical Semi-Diameter Measurement Residuals of the Planet Venus.

or most reliable boxplot to use based on just these three examples. The following section addresses this by examining the behavior of the three boxplots under controlled conditions.

3.3 Detecting Interesting Observations In Small Samples

Aside from providing information about location, scale, and skewness, boxplots also provide information about strange or unusual data. According to Tukey (1977, p.44), "values between an inner fence and its neighboring outer fence are outside. Values beyond outer fences are 'far out'." Tukey's "outer fence" refers to fences constructed by using $k = 3.0$ and his "inner fence" refers to fences constructed by using $k = 1.5$. Since the boxplots considered here use $k = 1.5$ in construction of fences and do not use the notion of outer fences, values beyond their fences will correspond to either Tukey's 'outside' values or 'far out' values. Values outside the fences of the three boxplots considered will be referred to as 'outside values.'

In order to describe the performance of the three boxplot rules in detecting outsiders, the 'all-inside rate per sample' and the 'outside rate per observation,' both introduced by Hoaglin et al. (1986), are examined for different sample sizes and distributions. The all-inside rate per sample, denoted $B(n)$, is the probability that a sample of size n from a given distribution contains no observations beyond the fences of the boxplot. The outside rate per observation, denoted $A(n)$, is the probability that an observation from a sample of size n would be labeled as an outsider simply by chance.

Since the all-inside rate per sample and the outside rate per observation are not amenable to analytic calculation in small samples, a Monte Carlo simulation was performed. Splus version 4.0 for Windows95 on an AMD based PC was used to generate samples from Normal(0,1), Uniform(0,1), t_5, χ^2_4, Exponential(1), Logistic(0,1), Lognormal(0,1), Laplace(0,1), and Extreme Value(0,1) distributions. For $n = 9, 10, \ldots, 30$, one thousand samples were simulated for each distribution and the all-inside rate per sample and the outside rate per observation were estimated.

3.3.1 All-Inside Rate Per Sample

Given a random sample $X_1, \ldots, X_n$ of size n from distribution F and any type of boxplot, the all-inside rate per sample, $B(n)$, is the probability that all observations are within the fences of the boxplot. For the common boxplot described in Section 3.1, $B(n)$ can be written as

$$\begin{aligned} B(n) &= P\{All\ observations\ within\ fences\} \\ &= P\left\{2.5Q_1 - 1.5Q_3 \leq \min_i X_i \quad \cap \max_i X_i \leq 2.5Q_3 - 1.5Q_1\right\} \end{aligned}$$

In order to calculate this probability exactly, the joint distribution of the first and third sample quartiles (Q_1 and Q_3), the min X_i, and the max X_i would have to be calculated. Instead of doing this, Monte Carlo simulation was used to estimate $B\left(n\right)$ for various n, the three boxplots, and nine distributions of interest. For a given sample size n and specific distribution, one thousand samples of size n were generated from that distribution. For each generated sample, three indicators were created, one for each boxplot type, where an indicator for a particular boxplot was set to 1 if all observations were within its fences, and 0 otherwise. Each indicator was then averaged over the one thousand generated samples in order to estimate the all-inside rate per sample for a particular boxplot.

For continuous distributions on $\Re$, $B(n)$ goes to 0 as n goes to infinity. In the case of the Uniform(0,1) distribution, $B\left(n\right)$ converges to 1 since the upper and lower fences converge to values outside the range of the distribution. For a given distribution and fixed sample size n, if $B_1(n) < B_2(n)$, where $B_1\left(n\right)$ and $B_2\left(n\right)$ are derived from two types of boxplot methods, the type 1 boxplot covers a smaller probability region than the type 2 boxplot.

Figures 3.7 and 3.8 show the plots of $B(n)$ (multiplied by 100 to create a percentage scale) versus n for the three boxplot rules and for the five symmetric sampling distributions. The flat solid line at $B\left(n\right) = 100$ represents the asymptotic limit of $B\left(n\right)$ under the Uniform(0,1) distribution for all boxplot rules, while the flat solid line at $B(n) = 0$ represents the asymptotic limit of $B(n)$ under the other symmetric distributions for all boxplot rules. Estimates of 95% confidence intervals derived from a normal approximation are provided for each estimate of $B(n)$. Specifically, the confidence limits are set as

$$(100)B\left(n\right) \pm (100)(1.96)\sqrt{\frac{B(n)[1 - B(n)]}{1000}}.$$

When the 95% confidence limits extend beyond the range of (0,1), the confidence limits are adjusted so that they are constrained to be between 0 and 1. In all cases, the signed boxplot has a higher all-inside rate per sample point estimate than the other boxplot types, while the common boxplot has a higher all-inside rate per sample point estimate than the skew adjusted boxplot. When the confidence limits are taken into consideration, the differences between the signed and common boxplot estimates of $B(n)$ disappear for $n \geq 18$ under the Uniform(0,1) distribution while the signed boxplot retains the highest all-inside rate per sample under the other symmetric distributions. The difference in $B\left(n\right)$ between the common and the skew adjusted boxplots disappears for $n \geq 23$ under the t_5 distribution and for $n \geq 20$ under the Laplace(0,1) distribution. It also seems that $B(n)$ converges more quickly for the "less normal" or more heavy tailed distributions.

The behavior of $B\left(n\right)$ under the Uniform(0,1) distribution is different than under the other four symmetric distributions. Under the Uniform(0,1) distribution, the all inside rate per sample

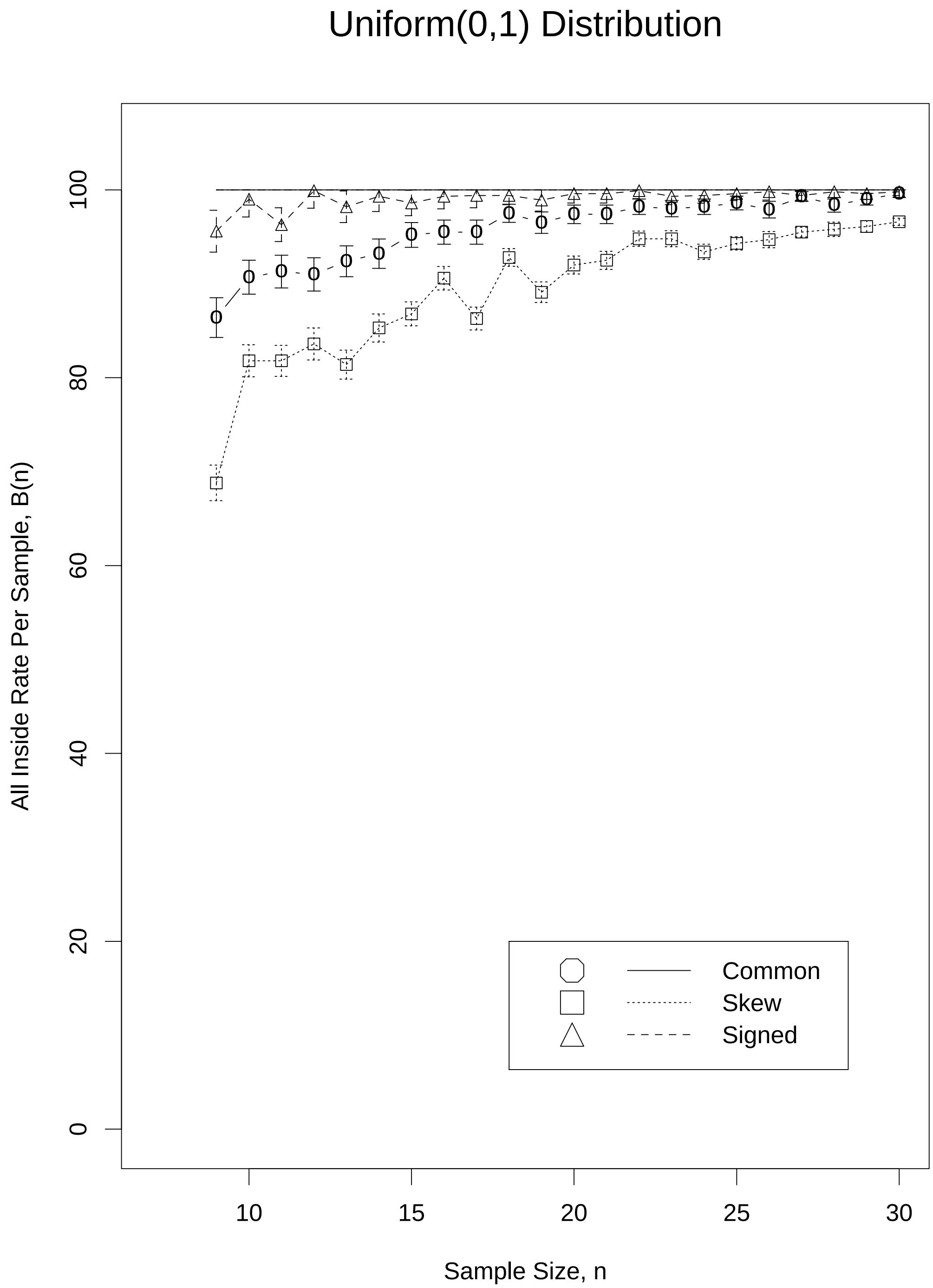

Figure 3.7: Plots of the All Inside Rate Per Sample Versus Sample Size for the Uniform Distribution. The bars represent estimates of the 95 percent confidence interval for B(n). The horizontal line at B(n)=100 indicates the asymptotic limit of B(n).

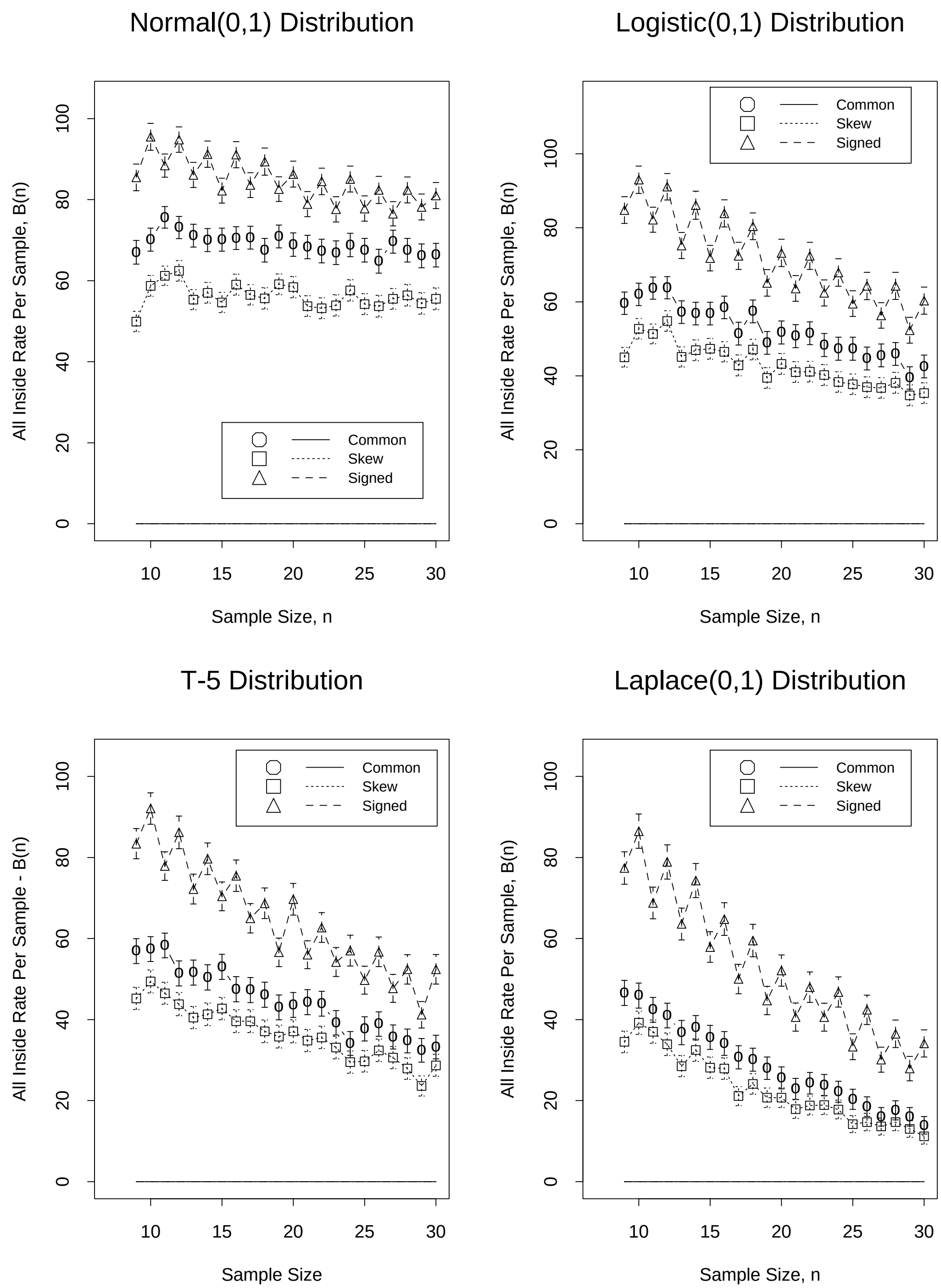

Figure 3.8: Plots of the All Inside Rate Per Sample Versus Sample Size for Symmetric Distributions. The bars represent estimates of the 95 percent confidence interval for B(n). The horizontal line at B(n)=0 indicates the asymptotic limit of B(n).

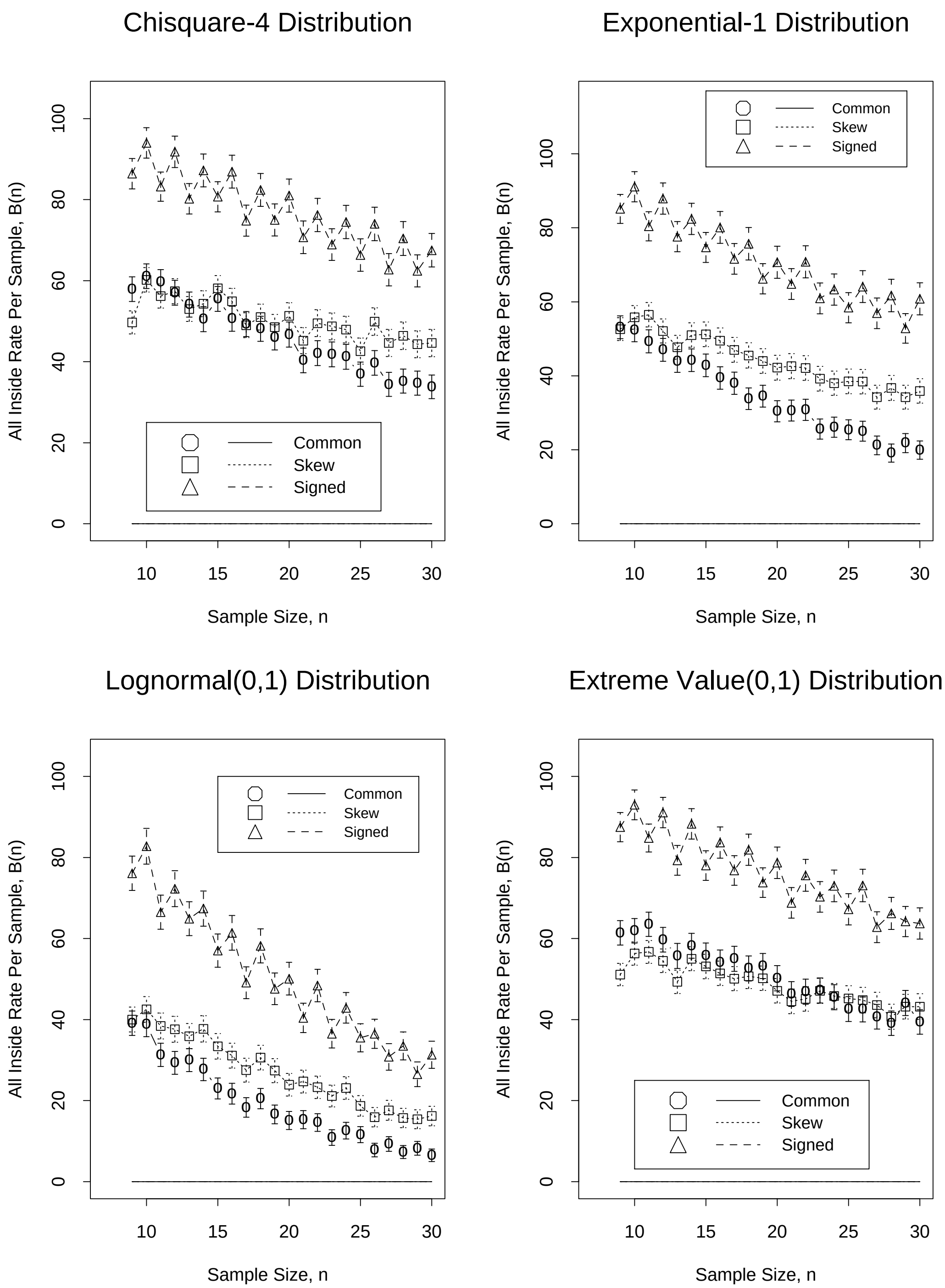

Figure 3.9: Plots of the All Inside Rate Per Sample Versus Sample Size for Skewed Distributions. The bars represent estimates of the 95 percent confidence interval for B(n). The horizontal line at B(n)=0 indicates the asymptotic limit of B(n).

increases towards 100 as the sample size increases while for the other symmetric distributions the all inside rate per sample decreases towards 0 as the sample size increases. For a given sample size, the (simulated) fences for the three boxplots quite often lie outside the range $(0, 1)$ of the Uniform(0,1) distribution. This means that for many of the Monte Carlo samples, the entire sample is contained within the estimated fences, resulting in a large estimate of $B(n)$. Since the asymptotic limits of the fences for the three boxplots lay outside the range (0,1), then asymptotic limit of the all inside rate per sample for the three boxplots is 100.

Figure 3.9 shows plots of $B(n)$ versus n for the three boxplot rules and for the four skewed distributions. Based on the point estimates and confidence limits, the signed boxplot yields larger estimates of $B(n)$, under the skewed distributions, than either the skew adjusted or common boxplot. In general, the common boxplot yields a smaller point estimate of $B(n)$ than the skew adjusted boxplot. Considering the confidence limits under the Extreme Value(0,1) distribution, the common and skew adjusted estimates of $B(n)$ are not different except when $n = 9$ and $n = 11$. For the remaining skewed distributions, the estimates of $B(n)$ under the common and skew adjusted boxplots become significantly different as n increases. The common and skew adjusted estimates of $B(n)$ differ for $n \geq 26$ under the χ^2_4 distribution, for $n \geq 15$ under the Exponential(1) distribution, and for $n \geq 13$ under the Lognormal(0,1) distribution.

Under every distribution, except the Uniform(0,1), the signed boxplot gives a higher probability that a sample of size n will be captured between its fences than either the skew adjusted or common boxplots. This holds under the Uniform(0,1) distribution except when $n \geq 18$, where the common and signed boxplots yield the same probability. Under every symmetric distribution except for the Laplace(0,1), the common boxplot has a higher probability that a sample of size n will be captured between its fences than the skew adjusted boxplot. Under the Laplace(0,1) distribution, the skew adjusted and common boxplots have indistinguishable all inside rates. Under every skewed distribution except for the Extreme Value(0,1) and χ^2_4 distributions, the skew adjusted boxplot has a higher all-inside rate than the common boxplot for almost every n. Under the χ^2_4, the skew adjusted and common boxplot have indistinguishable all-inside rates for $n \leq 25$. Under the Extreme Value(0,1) distribution, the common and skew adjusted boxplots have similar all-inside rates except for $n = 9$. A general rule of thumb is that the signed boxplot has the highest all-inside rate under every distribution, followed by the common boxplot and then the skew adjusted boxplot under the symmetric distributions, and followed by the skew adjusted and then the common boxplot under the skewed distributions.

It should be noted that due to the nature of the sampling methodology, there are no interesting observations to be detected. The fact that the signed boxplot has a higher all-inside rate means that the signed boxplot does a better job of not inadvertently identifying observations as interesting.

3.3.2 Outside Rate Per Observation

Given a random sample $X_1, \ldots, X_n$ of size n from distribution F and any type of boxplot, the outside rate per observation, $A(n)$, is the probability that an observation lies outside the fences of the boxplot. For the common boxplot described in Section 3.1, $A(n)$ can be written as

$$\begin{aligned} A(n) &= P\{an\ observation\ lies\ outside\ the\ fences\} \\ &= P\{[X_1 \leq 2.5Q_1 - 1.5Q_3] \quad \cup \quad [X_1 \geq\ 2.5Q_3 - 1.5Q_1]\} \end{aligned}$$

In order to calculate this probability exactly, the joint distribution of the first and third sample quartiles (Q_3 and Q_1) and X_1 would have to be calculated. Instead of doing this, Monte Carlo simulation was used to estimate $A(n)$ for various n, the three boxplots, and nine distributions of interest. For a given sample size n and specific distribution, one thousand samples of size n were generated from that distribution. For each generated sample, three indicators were created, one for each boxplot type, where an indicator for a particular boxplot was set to the percentage of observations outside its fences. Each indicator was then averaged over the one thousand generated samples in order to estimate the outside rate per observation for a particular boxplot.

For the all-inside rate per sample, the limiting value of $B(n)$ was either 0 or 1, depending upon the distribution. The limiting value of $A(n)$ differs across distribution and boxplot type. The limiting value of $A(n)$ can be explicitly calculated, and in the case of the common boxplot, the limiting value of $A(n)$, for continuous distributions F defined on $\Re$, is

$$\int_{-\infty}^{2.5Q_1-1.5Q_3} dF + \int_{2.5Q_3-1.5Q_1}^{\infty} dF$$

where Q_1 and Q_3 represent the limiting values of the estimators used in the common boxplot. In this case, Q_1 is the first quartile of F and Q_3 is the third quartile of F.

Table 3.6 lists the asymptotic values of $A(n)$ for the nine sampling distributions and three boxplot types considered in this paper; details are given in Appendix A.3.

Under the Uniform(0,1) and Normal(0,1) distributions, all three boxplots are equally likely to label an observation as an outsider by chance, while under the other symmetric distributions, the common and skew adjusted boxplots are just as likely to label an observation as an outsider and more likely than the signed boxplot. For the skewed distributions, the common boxplot is the most likely to label an observation as an outsider by chance, followed by the skew adjusted boxplot, and then by the signed boxplot.

Table 3.6: Population Outside Rate Per Observation, as a Percentage

Distribution	Common	Skew Adjusted	Signed
Uniform(0,1)	0.0	0.0	0.0
Normal(0,1)	0.70	0.70	0.70
Logistic(0,1)	2.44	2.44	1.83
Laplace(0,1)	6.25	6.25	3.43
t_5	3.35	3.35	2.39
Extreme Value(0,1)	2.68	2.04	1.30
χ^2_4	3.17	2.18	1.32
Exponential(1)	4.81	3.12	1.72
LogNormal(0,1)	7.76	5.71	2.61

Figures 3.10 and 3.11 plot the estimates of $A(n)$ (multiplied by 100 in order to be on a percentage scale) versus n for the five symmetric distributions and for the three boxplot rules.

The horizontal solid line represents the asymptotic values of $A(n)$ for the common and skew adjusted boxplots while the horizontal dashed line represents the asymptotic value for the signed boxplot. These two lines overlap for the Uniform(0,1) and Normal(0,1) distributions. Estimates of 95% confidence intervals derived from a normal approximation are provided for each estimate of $A(n)$. When the 95% confidence limits extend beyond the range of (0,1), the confidence limits are adjusted so that they are constrained to be between 0 and 1. In terms of point estimates, the skew adjusted boxplot has the highest values of $A(n)$, followed by the common boxplot, and then by the signed boxplot. When the confidence limits are factored into the comparisons, the three boxplots provide different estimates of $A\left(n\right)$ for every distribution except the Uniform(0,1) distribution, though the common and skew adjusted boxplots yield very close estimates under the Laplace(0,1) distribution. Under the Uniform(0,1) distribution, the skew adjusted boxplot yields higher estimates of $A\left(n\right)$ than the other two boxplots but the common and signed estimates become indistinguishable for $n \geq 22$.

Figure 3.12 plots estimates of $A(n)$ (multiplied by 100 in order to be on a percentage scale) versus n for the four skewed distributions and for the three boxplot types.

The asymptotic limits for $A(n)$ for the common, skew adjusted, and signed boxplots are represented by the solid, dotted, and dashed horizontal lines, respectively. The signed boxplot has the smallest estimate of $A\left(n\right)$ under each skewed distribution. The common and skew adjusted boxplots have similar estimates of $A\left(n\right)$ under the χ^2_4 and Extreme Value(0,1) distributions. The common and

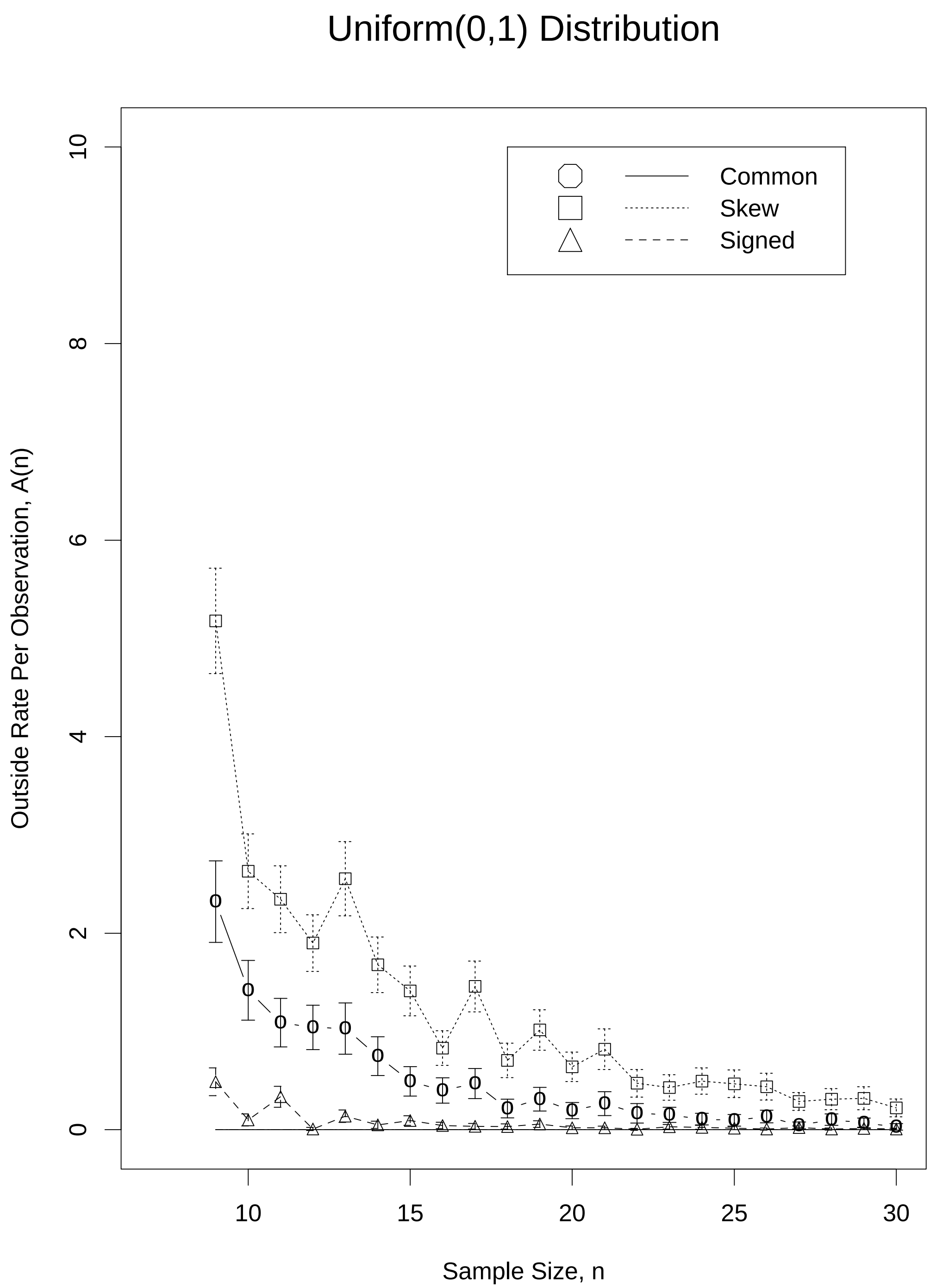

Figure 3.10: Outside Rate Per Observation versus Sample Size for the Uniform Distribution. The bars represent estimates of the 95 percent confidence interval for $A(n)$. The horizontal line at $A(n) = 0$ indicates the asymptotic value of $A(n)$.

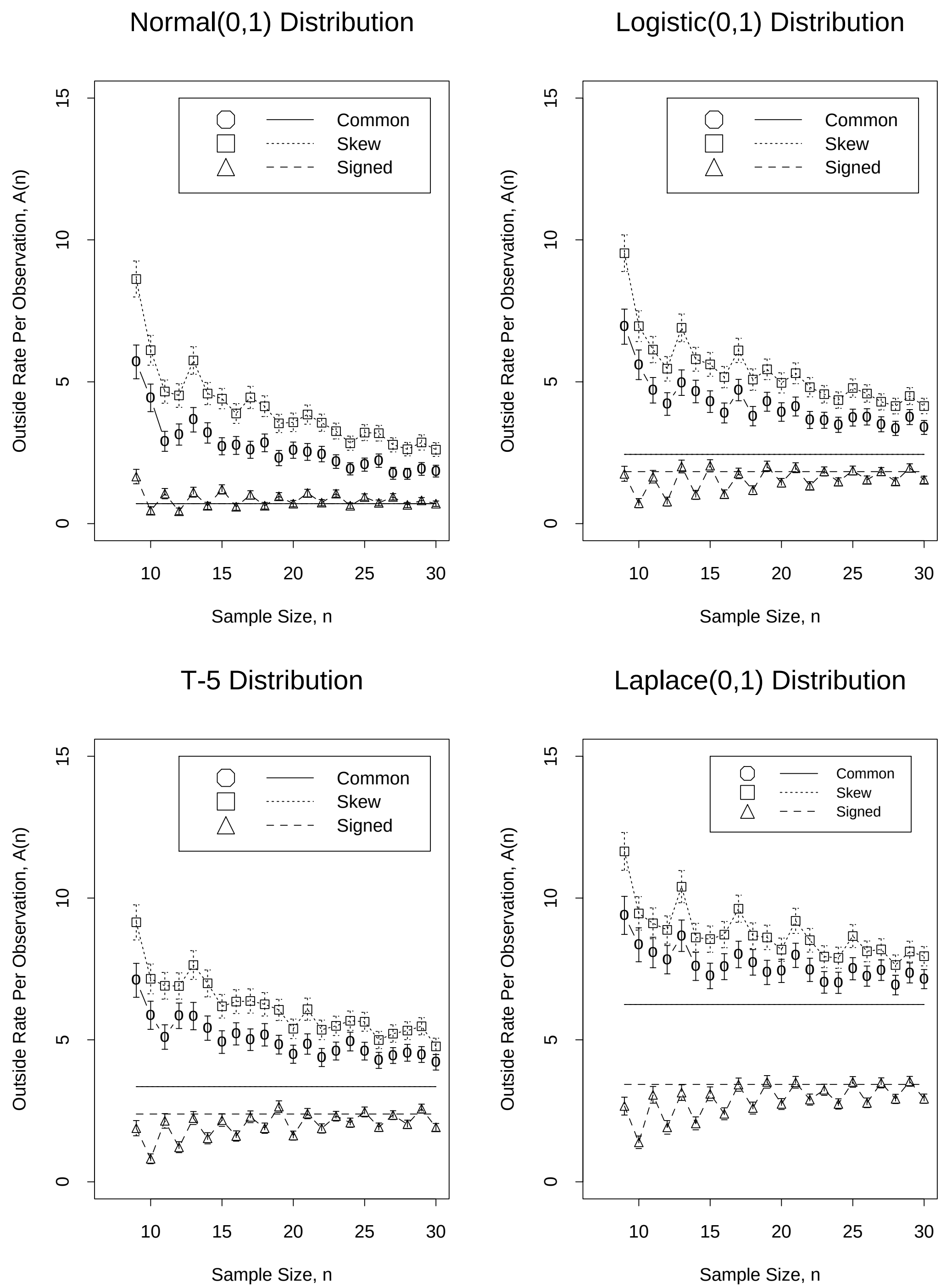

Figure 3.11: Outside Rate Per Observation versus Sample Size for the Symmetric Distributions. The bars represent estimates of the 95 percent confidence interval for $A(n)$. The horizontal solid line indicates the asymptotic value of $A(n)$ for the common and skew adjusted boxplots. The horizontal dashed line indicates the asymptotic value of $A(n)$ for the signed boxplot.

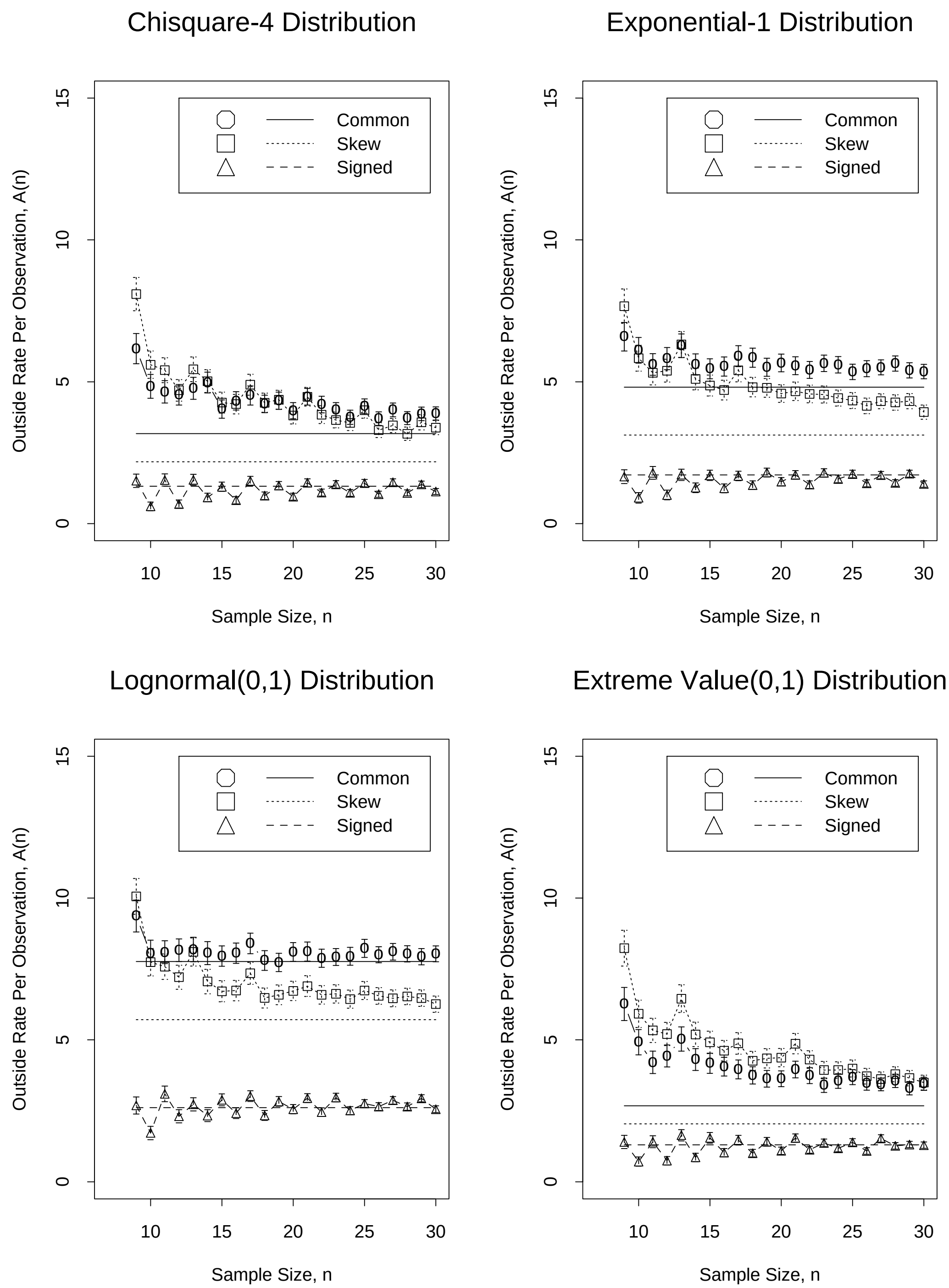

Figure 3.12: Outside Rate Per Observation versus Sample Size for the Skewed Distributions. The bars represent estimates of the 95 percent confidence interval for $A(n)$. The asymptotic limits of $A(n)$ for the common, skew adjusted, and signed boxplots are represented by the solid, dotted, and dashed horizontal lines, respectively.

skew adjusted boxplots estimates of $A(n)$ differ for $n \geq 18$ under the Exponential(1) distribution and for $n \geq 14$ under the Lognormal(0,1) distribution.

The signed boxplot has the lowest probability of labelling an observation an outsider, by chance, under any of the nine distributions considered. In general, the common boxplot has the highest probability of labelling an observation an outsider by chance, particularly under the extremely heavy tailed distributions, namely the Lognormal(0,1) and Exponential(1).

It should be noted that due to the nature of the sampling methodology, there are no interesting observations to be detected. The fact that the signed boxplot has the lowest outside rate per observation means that the signed boxplot has the lowest probability of labelling a non-outlier value from the underlying sampling distribution as interesting.

3.3.3 Outlier Detection

The previous two sections examined the rates at which the three boxplot types label non-outlier observations as interesting. This section examines the ability of the three boxplots to accurately label an observation that is a true outlier. Samples of size n =10, 30, and 100 were generated from a Normal(0,1) distribution. For a given sample $S_{n,i}$, $i = 1, 2, 3, \ldots$, the value Delta=$(i - 1)$ was added to the maximum value of the sample. This new value will be referred to as the outlier. The three boxplots were plotted for each sample $S_{n,i}$ and examined to see if they correctly identified the outlier by locating the outlier beyond the fences. For a given sample size, the values of i for which each boxplot identified the outlier were noted. The plots associated with these values of i are given below.

Figure 3.13 shows the three boxplots plotted for the original sample of size 10. No observations are labelled as interesting.

Figure 3.14 shows the three boxplots plotted for the sample of size 10 with Delta=2 added to the maximum value. This plot indicates that the first boxplot to identify the outlier is the common boxplot. The skew adjusted and signed boxplots do not yet detect the outlier.

Figure 3.15 shows the three boxplots plotted for the sample of size 10 with Delta=3 added to the maximum value. This plot indicates that the second boxplot to identify the outlier is the skew adjusted boxplot. The signed boxplot does not yet detect the outlier.

Figure 3.16 shows the three boxplots plotted for the sample of size 10 with Delta=6 added to the maximum value. This plot indicates that the signed boxplot identified the outlier only when six standard deviations was added to the maximum value of the sample. This plot also reveals a difference between the signed boxplot and the other boxplots. Unlike the common and skew adjusted boxplots, there are no observations that occur between the top of the signed boxplot and the outlier.

This means that there is no upper whisker for the signed boxplot. The large magnitude of the outlier inflates the value of Abs^+ so that the top half of the box is inflated beyond any observations other than the outlier. For the purposes of illustration, the upper whisker was drawn at the top of the box.

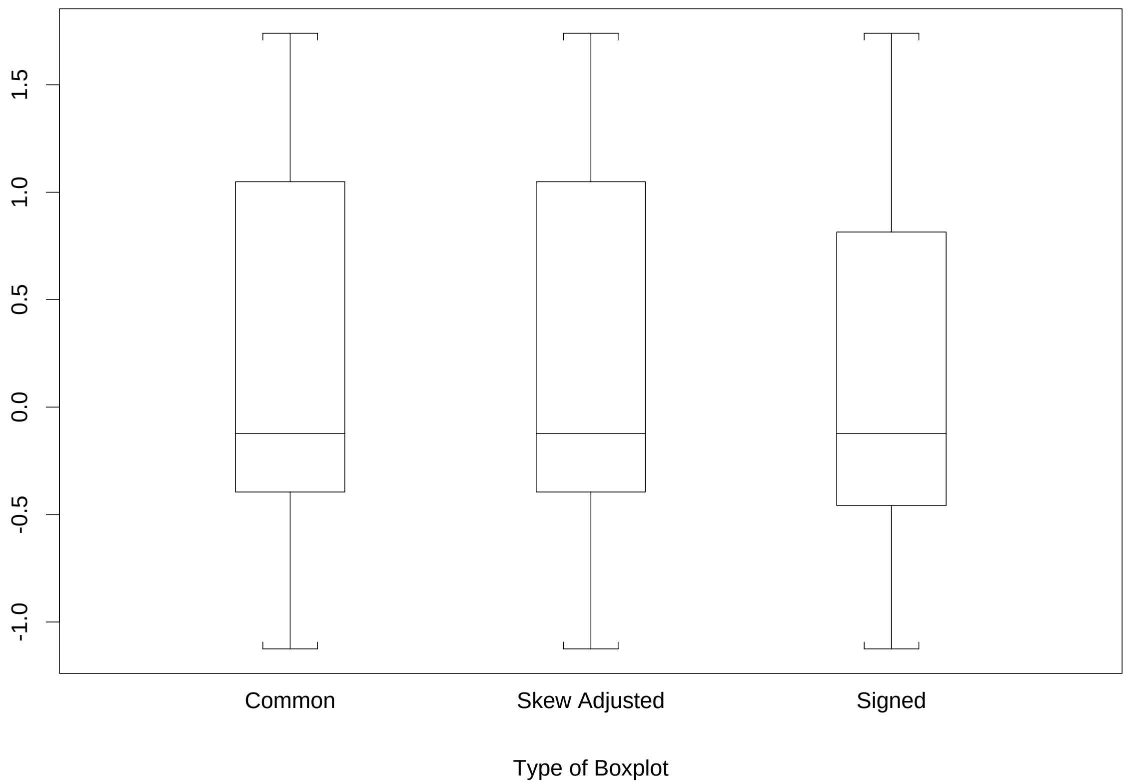

Figure 3.13: Outlier Detection Comparisons – N=10, Delta=0.

Figure 3.17 shows the three boxplots plotted for the original sample of size 30. This plot indicates that the common and skew adjusted boxplots identify the maximum value of the sample as being interesting though this value is not a true outlier. The signed boxplot does label any observation as interesting.

Figure 3.18 shows the three boxplots plotted for the sample of size 30 with Delta=1 added to the maximum value. This plot indicates that the signed boxplot identifies the outlier after one standard deviation is added to the maximum value.

Figure 3.19 shows the three boxplots plotted for the original sample of size 100. This plot

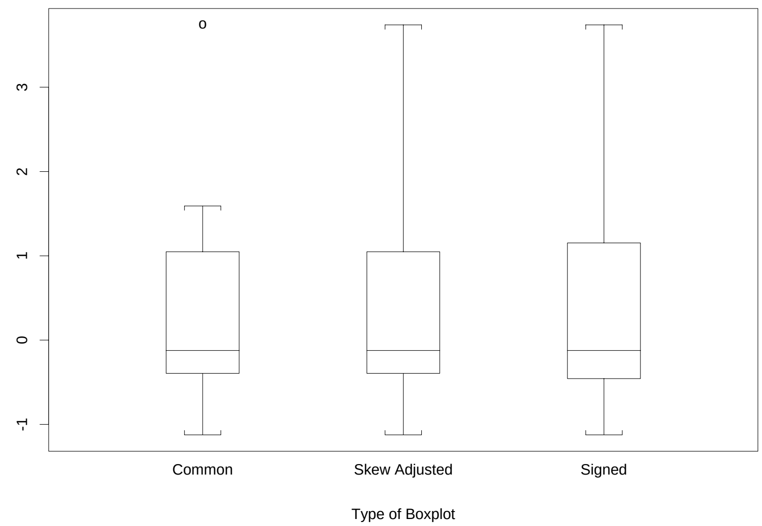

Figure 3.14: Outlier Detection Comparisons – N=10, Delta=2.

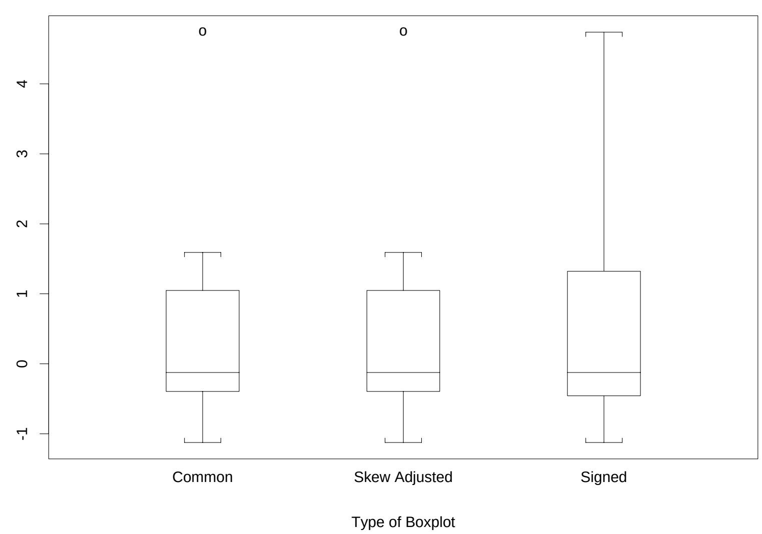

Figure 3.15: Outlier Detection Comparisons – N=10, Delta=3.

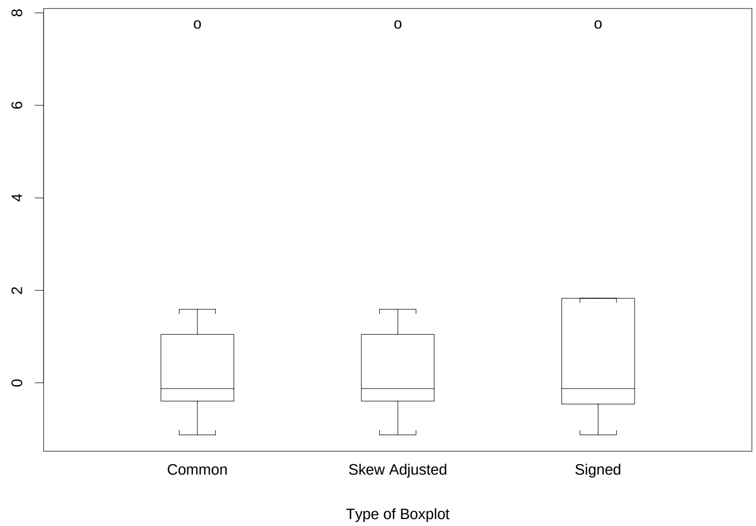

Figure 3.16: Outlier Detection Comparisons – N=10, Delta=6.

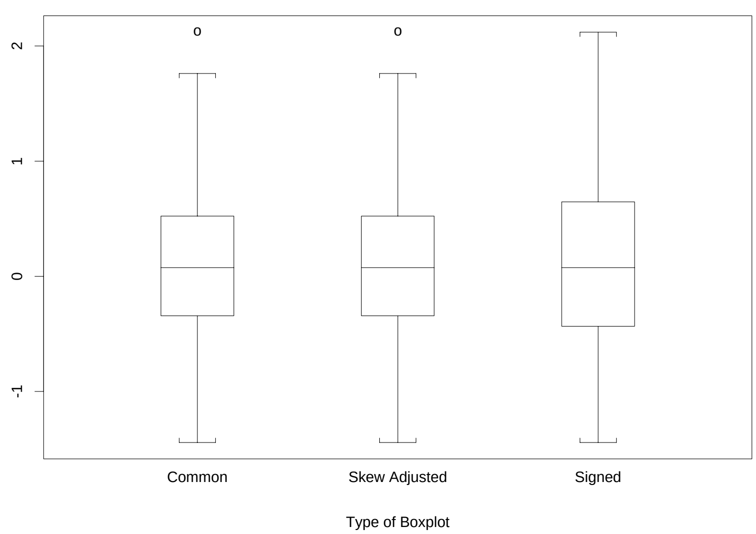

Figure 3.17: Outlier Detection Comparisons – N=30, Delta=0.

indicates that the common boxplot and skew adjusted boxplots identify the maximum value of the sample as being interesting though this value is not a true outlier. The signed boxplot does not label any observation as interesting. All three boxplots do label other points as interesting though not true outliers.

Figure 3.20 shows the three boxplots plotted for the sample of size 100 with Delta=1 added to the maximum value. This plot indicates that the signed boxplot identifies the outlier after one standard deviation is added to the maximum value.

For the original samples of size 30 and 100, the common and skew adjusted boxplots identify the maximum value of the samples as being interesting though these values are not true outliers. The signed boxplot identifies the outlier after one standard deviation is added to the maximum values of the samples. It should be noted that, for the original samples of size 100, all three boxplots label observations other than the maximum value as interesting though these observations are not true outliers.

For the original sample of size 10, none of the boxplots label any observations as interesting. The outlier is detected by the common boxplot when Delta=1, by the skewed boxplot when Delta=2, and by the signed boxplot when Delta=6. The upper half of the box of the signed boxplot is affected by the large magnitude of the outlier more than the upper halves of the other boxplots. Additionally, the large magnitude of the outlier causes the upper half of the box of the signed boxplot to be located between the outlier and all other observations. This makes an upper whisker unable to be plotted and removes the information about scale provided by the upper whisker. This effect on the upper half of the box does provide more information regarding the existence of an outlier though at the expense of information about scale.

3.3.4 Practical Consequences of the Signed Boxplot

The results of the previous sections show that the signed boxplot is less likely to label non-outlier observations as interesting. The common and skew adjusted boxplots are more likely to label non-outlier observations as being interesting. As seen in the previous section, the signed boxplot has the hardest time identifying a true outlier for small samples. For large samples, all three boxplots identify the outlier when Delta=1.

For moderate to large sample sizes all three boxplots correctly identify an outlier for the value Delta=1. The common and skew adjusted boxplots label the maximum value of the original sample as interesting while the signed boxplot does not.

For small sample sizes, the common boxplots detects an outlier first (Delta=2), followed by the skew adjusted boxplot (Delta=3) and the signed boxplot (Delta=6).

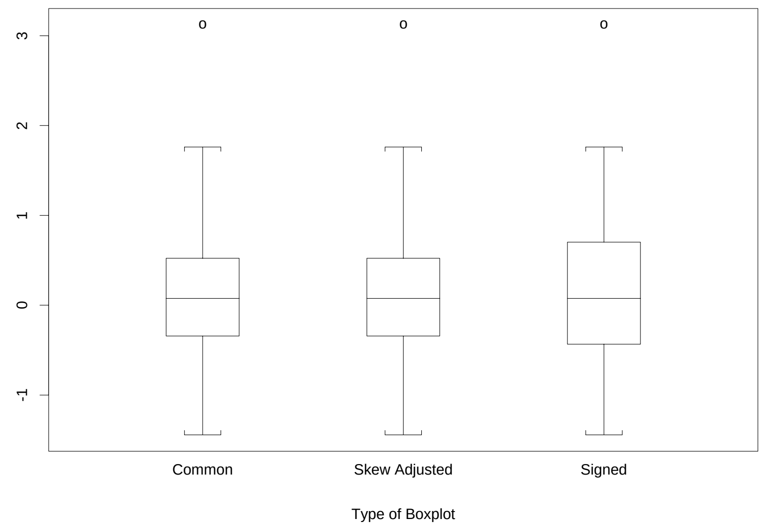

Figure 3.18: Outlier Detection Comparisons – N=30, Delta=1.

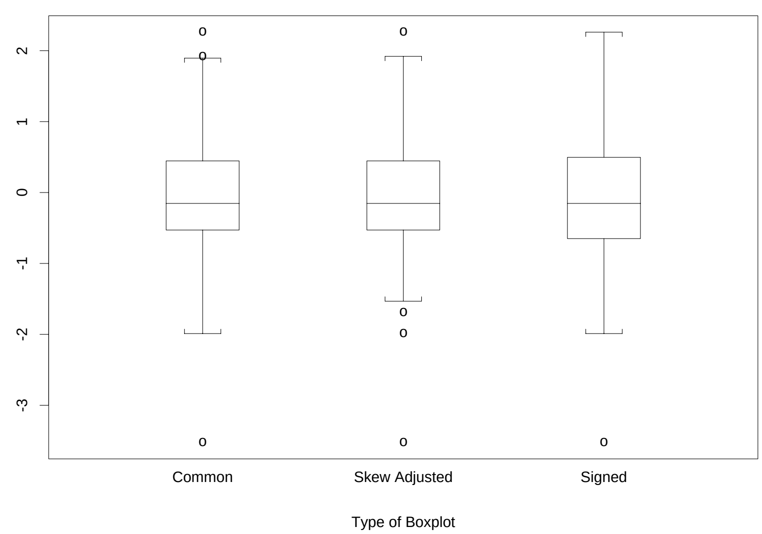

Figure 3.19: Outlier Detection Comparisons – N=100, Delta=0.

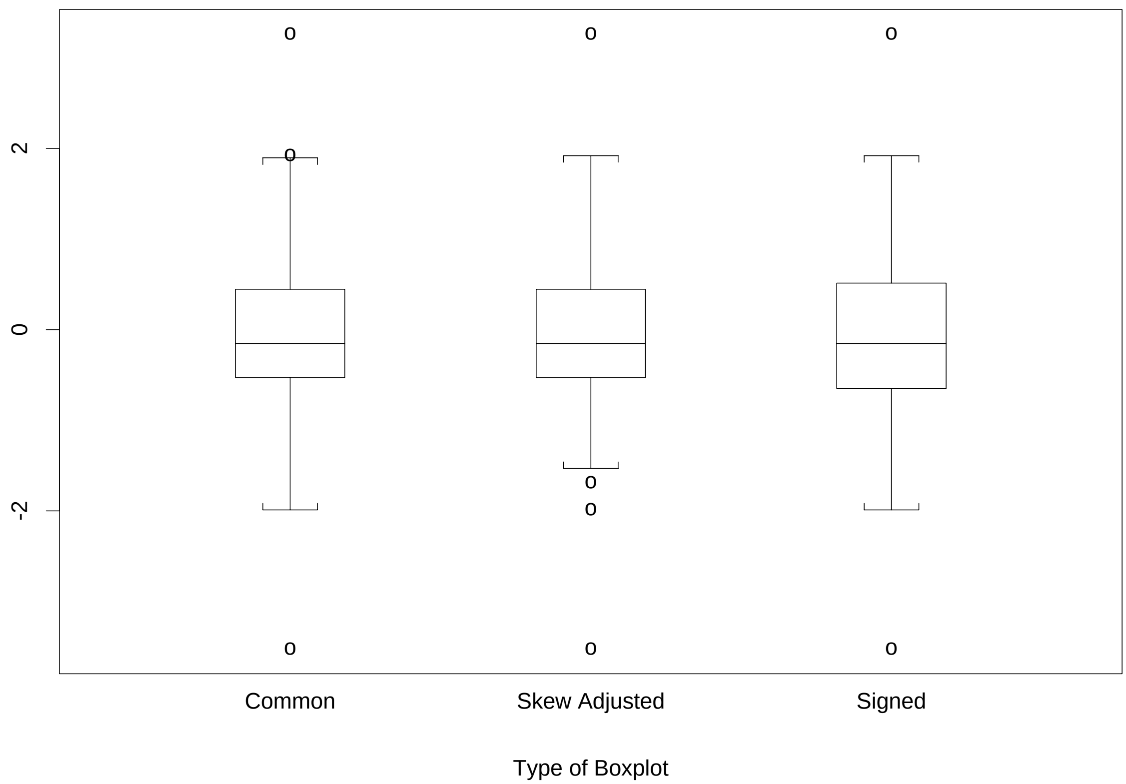

Figure 3.20: Outlier Detection Comparisons – N=100, Delta=1.

The effect on the shape of the signed boxplot for small samples in the presence of an outlier suggests that the signed boxplot may not be the best boxplot to use if the goal is to detect skewness or scale information in the presence of outliers. The signed boxplot is not able to detect an outlier in small samples until the magnitude of the outlier is quite large. The magnitude of the outlier affects the box of the boxplot and can lead to a suggestion of skewness when no skewness exists.

3.4 Skewness Information

Interest in the notion and measurement of skewness has existed since at least the late nineteenth century. Since that time, various "skewness measuring" coefficients, such as the standardized third central moment $\gamma_1 = \frac{\upsilon_3}{\sigma^3} = \frac{E(X-\mu)^3}{\left[E(X-\mu)^2\right]^{3/2}}$, the Bowley coefficient of skewness (Bowley 1920) given by

$$b_1 = \frac{F^{-1}(.75) + F^{-1}(.25) - 2F^{-1}(.5)}{F^{-1}(.75) - F^{-1}(.25)},$$

and the "Pearson" measure of skewness $q = \frac{\mu - F^{-1}(.5)}{\sigma}$, have been introduced without having a formal concept of skewness (Van Zwet 1964). Van Zwet (1964) used convex transformations to formalize the concepts of skewness, suggesting that such concepts require meaningful orderings of distributions, which measures of skewness must then preserve. Van Zwet (1964) proposed a method of ordering two distributions with respect to skewness based on the following proposition.

Proposition 16 *Assume that the random variables X and Y have continuous distribution functions $F(x)$ and $G(y)$, respectively, with differentiable density functions $f(x) > 0$ and $g(y) > 0$ on an interval $(a,b) \subseteq \Re$. Then, $G(t)$ is at least as skew to the right as $F(t)$ if $G^{-1}[F(t)] = R(t)$, for $t \in (a,b)$, is convex on (a,b). If this is true, then write $F <_c G$, and say "F c-precedes G".*

Building upon the work of Van Zwet (1964), Groeneveld and Meeden (1984) suggested four properties that any skewness coefficient γ should satisfy. These properties are:

1. A location or positive scale change for a random variable does not alter γ. That is, if $Y = cX+d$ for $c > 0$ and $d \in \Re$ then $\gamma(X) = \gamma(Y)$.

2. $\gamma = 0$ for a symmetric distribution.

3. If $Y = -X$ then $\gamma(Y) = -\gamma(X)$.

4. If F and G are c.d.f.s for X and Y as above and $F <_c G$ then, $\gamma(X) \leq \gamma(Y)$.

The common and skew adjusted boxplots provide information about skewness via the statistic

$$Q_C = \frac{IQR^+ - IQR^-}{IQR^+ + IQR^-} \ ,$$

which is an estimator of the skewness coefficient

$$\gamma_1 = \frac{\left[F^{-1}(.75) - F^{-1}(.5)\right] - \left[F^{-1}(.5) - F^{-1}(.25)\right]}{\left[F^{-1}(.75) - F^{-1}(.5)\right] + \left[F^{-1}(.5) - F^{-1}(.25)\right]} = \frac{F^{-1}(.75) + F^{-1}(.25) - 2F^{-1}(.5)}{F^{-1}(.75) - F^{-1}(.25)} \ ,$$

or Bowley's coefficient. The signed boxplot provides information about skewness via the statistic

$$Q_{Abs} = \frac{.845Abs^+ - .845Abs^-}{.845Abs^+ + .845Abs^-} = \frac{Abs^+ - Abs^-}{Abs^+ + Abs^-},$$

which is an estimator of the skewness coefficient

$$\gamma_2 = \frac{T_{1,+} - T_{1,-}}{T_{1,+} + T_{1,-}}.$$

Verification of the four properties for γ_1 was provided by Groeneveld and Meeden (1984). Verification of the four properties for γ_2 proceeds as follows. Consider property 1. Let Y and X be as given and without loss of generality assume X has median 0. Then,

$$\gamma_2(Y) = \frac{\int_d^\infty |y-d|\, dF_Y - \int_{-\infty}^d |y-d|\, dF_Y}{\int_{-\infty}^\infty |y-d|\, dF_Y}.$$

Using the change of variable $x = (y-d)/c$,

$$\gamma_2(Y) = \frac{\int_d^\infty |y-d|\, dF_Y - \int_{-\infty}^d |y-d|\, dF_Y}{\int_{-\infty}^\infty |y-d|\, dF_Y} = \frac{c\int_0^\infty |cx|\, dF_X - c\int_{-\infty}^0 |cx|\, dF_X}{c\int_{-\infty}^\infty |cx|\, dF_X} =$$

$$\frac{\int_0^\infty |x|\, dF_X - \int_{-\infty}^0 |x|\, dF_X}{\int_{-\infty}^\infty |x|\, dF_X} = \gamma_2(X).$$

Next, consider property 2. Suppose the random variable X follows a symmetric distribution, and without loss of generality, about 0. Then,

$$\gamma_2(X) = \frac{\int_0^\infty |x|\, dF_X - \int_{-\infty}^0 |x|\, dF_X}{\int_{-\infty}^\infty |x|\, dF_X} = 0$$

since $\int_0^\infty |x|\, dF_X = \int_{-\infty}^0 |x|\, dF_X$. For property 3, let X be a random variable with median 0. Then note that $Y = -X$ has median 0 and

$$\gamma_2\left(Y\right) = \frac{\int_0^\infty \left(|y|\right) dF_Y - \int_{-\infty}^0 \left(|y|\right) dF_Y}{\int_{-\infty}^\infty |y|\, dF_Y} = \frac{\int_{-\infty}^0 |x|\, dF_X - \int_0^\infty |x|\, dF_X}{\int_{-\infty}^\infty |x|\, dF_X} = -\gamma_2\left(X\right).$$

Lastly, consider property 4 and assume, without loss of generality, that X and Y both have median 0. It must be shown that

$$\gamma_2\left(X\right) = \frac{\int_0^\infty |x|\, dF - \int_{-\infty}^0 |x|\, dF}{\int_{-\infty}^\infty |x|\, dF} \leq \gamma_2\left(Y\right) = \frac{\int_0^\infty |y|\, dG - \int_{-\infty}^0 |y|\, dG}{\int_{-\infty}^\infty |y|\, dG}.$$

Using the substitutions $t = F\left(x\right)$ and $s = G\left(y\right)$, the above inequality can be written as

$$\frac{\int_{.5}^1 \left|F^{-1}\left(t\right)\right| dt - \int_0^{.5} \left|F^{-1}\left(t\right)\right| dt}{\int_0^{.5} \left|F^{-1}\left(t\right)\right| dt + \int_{.5}^1 \left|F^{-1}\left(t\right)\right| dt} \leq \frac{\int_{.5}^1 \left|G^{-1}\left(s\right)\right| ds - \int_0^{.5} \left|G^{-1}\left(s\right)\right| ds}{\int_0^{.5} \left|G^{-1}\left(s\right)\right| ds + \int_{.5}^1 \left|G^{-1}\left(s\right)\right| ds}.$$

This inequality is of the form

$$\frac{A - B}{A + B} \leq \frac{C - D}{C + D},$$

where $A = \int_{.5}^1 \left|F^{-1}\left(t\right)\right| dt$, $B = \int_0^{.5} \left|F^{-1}\left(t\right)\right| dt$, $C = \int_{.5}^1 \left|G^{-1}\left(s\right)\right| ds$, and $D = \int_0^{.5} \left|G^{-1}\left(s\right)\right| ds$. Cross multiplying the denominators of an inequality of this form yields

$$(A - B)(C + D) \leq (A + B)(C - D).$$

Carrying out the multiplication on each side of the inequality and then combining all terms on the left side yields

$$2DA - 2BC \leq 0$$

or, alternatively,

$$\frac{A}{C} \leq \frac{B}{D}.$$

Substituting values in for A, B, C, and D yields

$$\frac{\int_{.5}^1 \left|F^{-1}\left(t\right)\right| dt}{\int_{.5}^1 \left|G^{-1}\left(s\right)\right| ds} \leq \frac{\int_0^{.5} \left|F^{-1}\left(t\right)\right| dt}{\int_0^{.5} \left|G^{-1}\left(s\right)\right| ds}.$$

$F^{-1}(t)$ and $G^{-1}(s)$ are positive (or 0) for $t \geq .5$ and $s \geq .5$, respectively, since the medians of F and G are 0. This positivity allows for the removal of the absolute value bars on the left side of the previous inequality and yields

$$\frac{\int_{.5}^{1} F^{-1}(t)\,dt}{\int_{.5}^{1} G^{-1}(s)\,ds} \le \frac{\int_{0}^{.5} \left|F^{-1}(t)\right| dt}{\int_{0}^{.5} \left|G^{-1}(s)\right| ds}.$$

By the generalized mean value theorem, there exist α_1 and α_2, $0 < \alpha_1 < .5 < \alpha_2 < 1$ such that the inequality given above may be replaced by

$$\frac{F^{-1}(\alpha_2)}{G^{-1}(\alpha_2)} \le \frac{\left|F^{-1}(\alpha_1)\right|}{\left|G^{-1}(\alpha_1)\right|}.$$

Groeneveld and Meeden (1984) showed that when $F <_c G$, the function $\frac{F^{-1}(\alpha)}{G^{-1}(\alpha)}$ is non-increasing in α. As such,

$$\frac{F^{-1}(\alpha_2)}{G^{-1}(\alpha_2)} \le \frac{F^{-1}(\alpha_1)}{G^{-1}(\alpha_1)} \le \left|\frac{F^{-1}(\alpha_1)}{G^{-1}(\alpha_1)}\right| = \frac{\left|F^{-1}(\alpha_1)\right|}{\left|G^{-1}(\alpha_1)\right|},$$

showing that property 4 holds.

Even though both Q_C and Q_{Abs} are skewness measures, they are not necessarily equal, neither for small samples sizes nor asymptotically. In order to explore the relationship between Q_C and Q_{Abs}, Table 3.7 lists the asymptotic (and small sample) means for the nine distributions of interest.

Since skewness measures are 0 for symmetric distributions, both Q_C and Q_{Abs} are 0 for the Uniform, Laplace, Logistic, Normal, and T distributions. For every skewed distribution, the common boxplot measure of skewness is smaller in magnitude than the signed boxplot measure. Q_{Abs} is anywhere from 1.6 to 1.9 times as large as Q_C.

In order to examine the small sample properties of the two skewness measures, 5,000 Monte Carlo samples of sizes n=10, 20, and 30 were simulated for each distribution, and the means and relative efficiencies were estimated. Table 3.7 lists the means for each sample size, each skewed distribution, and both skewness measures. Since skewness measures for symmetric distributions are zero, comparisons of the means and relative efficiencies of Q_C and Q_{Abs} are non-trivial only for the skewed distributions. For every sample size and skewed distribution, Q_{Abs} is roughly twice Q_C. The last row of Table 3.7 shows the range of standard error estimates across the skewed distributions for each skewness measure and each sample size. The standard error estimates for Q_C for samples of size 20 are the same for every skewed distribution. The standard error estimates for Q_C for samples of size 30 are the same for every skewed distribution. The standard error estimates for Q_{ABS} for samples of size 10 are the same for every skewed distribution. For the remaining sample sizes, the standard error estimates (for a given sample size) of a particular skewness measure are within 0.01 of each other. Q_{Abs} is less variable than Q_C under every sample size.

Since the standardized variances are not calculable for the symmetric distributions, the relative efficiencies of the two skewness measures were only examined for the four skewed distributions. The relative efficiencies under these four distributions, as determined by the ratio of the standardized variance of Q_C to the standardized variance of Q_{Abs}, denoted

$$\frac{SV\left(Q_C\right)}{SV\left(Q_{Abs}\right)}$$

is provided in Table 3.8.

Q_{Abs} is anywhere from five to seven times more efficient than Q_C. Q_{Abs} is five times more efficient than Q_C even for the samples of size 10.

The efficiency of Q_{Abs} over Q_C means that inference about skewness derived from the shape of the box part of a boxplot is more reliable for the signed boxplot than it is for either the common or skew adjusted boxplots.

Table 3.7: Asymptotic and Estimated Means of Q_C and Q_{Abs}

	Q_C				Q_{Abs}			
	10	20	30	∞	10	20	30	∞
Uniform(0,1)	0.007	0.001	-0.002	0	0.008	-0.001	0.002	0
Normal(0,1)	0.008	0.007	0.008	0	0.009	0.001	0.002	0
Logistic(0,1)	-0.01	-0.01	-0.01	0	-0.001	-0.005	-0.005	0
Laplace(0,1)	-0.0008	0.01	0.008	0	0.007	0.003	0.006	0
t_5	-0.007	0.003	0.007	0	-0.006	0.0007	0.008	0
Ex. Val.(0,1)	0.09	0.10	0.10	0.12	0.18	0.20	0.21	0.22
χ^2_4	0.13	0.15	0.15	0.17	0.26	0.29	0.30	0.31
Exp.(1)	0.19	0.22	0.24	0.26	0.36	0.41	0.42	0.44
Lognormal(0,1)	0.24	0.28	0.30	0.33	0.46	0.52	0.54	0.58
S.E. Range	0.35-0.36	0.27	0.23		0.26	0.19-0.20	0.15-0.16	

Table 3.8: Estimated Relative Efficiencies for the Skewness Measures Q_C and Q_{Abs}, as Measured by $\frac{SV(Q_C)}{SV(Q_{Abs})}$

N	10	20	30
Ex. Val.(0,1)	6.02	6.62	6.87
χ^2_4	5.60	6.16	6.34
Exp.(1)	5.15	5.66	5.92
Lognormal(0,1)	5.72	6.74	7.02

Chapter 4

SUMMARY

The motivation behind this dissertation was the common boxplot. The boxplot attempts to provide information about location, scale, skewness, and unusual observations. Construction of the common boxplot is based upon the interquartile range, or for the skew adjusted boxplot, based up IQR^+ and IQR^-. However, the interquartile range, as well as IQR^+ and IQR^-, are inefficient measures. As such, other possible measures were considered that would be more efficient and, when used in construction of a boxplot, would be more reliable in providing information about scale, skewness, and unusual observations.

The purpose of this dissertation was to introduce the notion of signed scale parameters and estimators, and to offer an example where the use of signed scale estimators provided more reliable inference. Signed scale parameters measure the amount of variability of a distribution above and below a certain point. The positive signed scale parameter measures variability above a point while the negative signed scale parameter measures variability below a point. Taken together, these parameters also provide information about skewness when their relative magnitudes are considered. If the positive signed scale parameter is much larger than the negative signed scale parameter, then the underlying distribution is skewed right. On the other hand, if the negative signed scale parameter is much larger than the positive signed scale parameter, then the underlying distribution is skewed left. If the positive and negative signed scale parameters are equal, then the distribution is symmetric.

Chapter 1 introduced the notion of signed scale parameters as related to regular scale parameters and defined several classes of signed scale parameters. Classes of signed scale estimators derived from the classes of signed scale parameters were also introduced. Three properties that signed estimators must satisfy in order to be signed scale estimators were listed and the introduced classes of signed estimators were shown to satisfy these properties.

Chapter 2 selected seven specific pairs of signed scale estimators from the classes of signed scale estimators introduced in Chapter 1. Asymptotic consistency was shown to hold for these seven pairs, and these pairs were shown to be asymptotically normally distributed. The asymptotic efficiencies of the seven positive signed scale estimators were calculated relative to IQR^+, while the asymptotic efficiencies of the seven negative signed scale estimators were calculated relative to IQR^-. For every distribution considered, all other positive signed scale estimators were more asymptotically efficient than IQR^+, and all other negative signed scale estimators were more asymptotically efficient than IQR^-.

Chapter 2 also included small sample comparisons of the efficiency of the seven pairs of signed scale estimators. The $\text{Sd}^{+/-}$ and $\text{Sdtrim}^{+/-}$ estimators appeared to perform equally well and were the best under symmetric distributions with low kurtosis. The $S_{Bi,20,+/-}$ pair of estimators performed almost as well under the symmetric distributions with low kurtosis and had the best performance under the symmetric distributions with large kurtosis. Sd^- and $S_{Bi,20,-}$ performed the best of the other negative signed estimators under the skewed distributions. Abs^+ generally performed the best of the other positive signed estimators under the skewed distributions. The performance of Abs^+ under the skewed distributions led to its use in constructing a new type of boxplot introduced in Chapter 3.

Chapter 3 reviewed the notion of the common boxplot and the method of its construction. Problems with using the common boxplot to get information about skewness, scale, and unusual observations were noted and an alternative type of boxplot was reviewed. This boxplot, or skew adjusted boxplot, does provide somewhat better information about skewness, yet it also has flaws. The flaws inherent in both the common and skew adjusted boxplots revolve around the use of IQR^+ and IQR^- in their construction. An alternative boxplot construction method, the signed boxplot, was introduced. The signed boxplot uses $\text{Abs}^{+/-}$ in its construction in place of $\text{IQR}^{+/-}$. $\text{Abs}^{+/-}$ were shown to be more efficient than $\text{IQR}^+/\text{IQR}^-$ in Chapter 2.

Three examples were given in Chapter 3 where three sets of data were compared using the common, skew adjusted, and signed boxplots. In all cases, the signed boxplot was less likely to label an observation as unusual, and appeared to provide more reliable information about skewness.

The rate at which the three boxplots labelled non-outlier observations as interesting was also examined. Monte Carlo simulations were run to estimate the all-inside rate per sample and the outside rate per observation for each boxplot across a variety of sample sizes and distributions. The signed boxplot was shown to be less likely to label a non-outlier observation as unusual than either the common or skew adjusted boxplot.

The ability of the boxplots to detect a true outlier was investigated. For small samples, the signed boxplot was not as able to detect a true outlier as the common and skew adjusted boxplots. The shape of the signed boxplot was also affected, by an outlier, more than the shapes of the common and skew adjusted boxplots. This observation suggests that the signed boxplot may not be the best choice if the goal is to provide information about scale and skewness in the presence of outliers in small samples. For large samples, all three boxplots were able to detect an outlier to the same degree. The common and skew adjusted boxplots labeled more non-outlier observations as interesting than did the signed boxplot.

The skewness information provided by the common and signed boxplots were also examined in Chapter 3. Measures of skewness were defined for each boxplot and shown to satisfy the historical definition of a skewness measure. A Monte Carlo simulation was performed to examine the small sample performance of each skewness measure. The skewness information derived from the signed boxplot was shown to be approximately 30 percent more efficient than the common boxplot.

There is a tradeoff when using the signed boxplot. For moderate to large sample sizes, the signed boxplot appears to provide the best information about skewness and scale and provides the same ability to detect outliers as the common and skew adjusted boxplots. For small samples, the signed boxplot does provide the best information about scale and skewness when there are no outliers present. However, the signed boxplot is more affected by outliers than the other boxplots. The signed boxplot cannot detect outliers in small samples before the other boxplots and large outliers can have a significant effect on the shape of the boxplot. The effect of a large outlier on the shape of the signed boxplot causes information on scale and skewness to be distorted or misleading.

REFERENCES

Barnett, V., and Lewis, T. (1994), *Outliers in Statistical Data* (3rd ed.), New York: John Wiley and Sons, Inc.

Bickel, P.J., and Lehmann, E.L. (1976), Descriptive Statistics For Nonparametric Models. III. Dispersion., *The Annals of Statistics*, Vol 4, No. 6, 1139-1158.

Bowley, A.L. (1920), *Elements of Statistics* (4th ed.), New York: Charles Scribner's Sons.

Ferguson, T.S. (1967), *Mathematical Statistics: A Decision Theoretic Approach*, San Diego: Academic Press, Inc.

Frigge, M., Hoaglin, D.C., and Iglewicz, B. (1989), Some Implementations of the Boxplot, *The American Statistician*, vol. 43, no. 1, 50-54.

Groeneveld, R.A., and Meeden G. (1984), Measuring Skewness and Kurtosis, *The Statistician*, 33, 391-399.

Hoaglin, D.C., Iglewicz, B., and Tukey, J.W. (1986), Performance of Some Resistant Rules For Outlier Labeling, *Journal of the American Statistical Association*, vol. 81, 991-999.

Huber, P.J. (1964), Robust Estimation of A Location Parameter, *The Annals of Mathematical Statistics*, Vol. 35, No. 1, 73-101.

Huber, P.J. (1967), The Behavior of Maximum Likelihood Estimates Under Non-standard Conditions, *Proceedings of the 5th Berkeley Symposium*,1,221-233.

Iverson, H.K., and Randles, R.H. (1989) The Effects on Convergence of Substituting Parameter Estimates into U-Statistics and Other Families of Statistics, *Probability Theory and Related Fields*, vol. 81, 453-471.

Kimber, A.C. (1990) Exploratory data analysis for possibly censored data from skewed distributions, *Applied Statistics*, Vol. 39, No. 1, 21-30.

Lax, D.A. (1985) Robust Estimators of Scale: Finite-Sample Performance in Long-Tailed Symmetric Distributions, *Journal of the American Statistical Association*, Vol. 80, No. 391, 736-741.

Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: John Wiley & Sons, Inc.

Stefanski, L. (1997), Convergence of Averages of Random Variables That Are Functions of Estimators of Population Parameters, unpublished theorem.

Tukey, J.W. (1977), *Exploratory Data Analysis*, Reading: Addison-Wesley Publishing Company.

Van Zwet, W.R. (1964), Convex Transformations of Random Variables, *Mathematics Centre Tract 7*, Mathematisch Centrum, Amsterdam.

APPENDICES

A.1 SPLUS Code For Signed Scale Estimators

```
#Abs+ Function #
sigplus<-function(x){
(1/(length(x[x>=median(x)])))*sum(abs((x-median(x))[x>=median(x)]))
}

#Abs- Function#
sigminus<-function(x){
(1/(length(x[x<=median(x)])))*sum(abs((x-median(x))[x<=median(x)]))
}

#IQR+ Function#
iqrplus <- function(x) {
diff(quantile(x,c(.5,.75)))
}

#IQR- Function#
iqrminus<-function(x) {
diff(quantile(x,c(.25,.5)))
}

#SD+ Function#
sdplus<-function(x){
sqrt((1/(length(x[x>=mean(x)])))*sum(((x-mean(x))^2)[x>=mean(x)]))
}

#SD- Function#
sdminus<-function(x){
sqrt((1/(length(x[x<=mean(x)])))*sum(((x-mean(x))^2)[x<=mean(x)]))
}

# This Function calculates the MAD statistic #
madd <-function(x){
m<-median(x) median(abs(x-m))
}
```

```
# Bisquare+ Function (Positive A-estimator) #
bisquareplus<-function(x,c) {
n<-length(x)
t<-median(x)
s<-madd(x)
u<-((x-t)/(c*s))
top<-sqrt(.5*sum( (((c*s*u)^2)*((1-u^2)^4))[u>=0& u<=1]))
bottom<-abs(sum(((1-u^2)*(1-5*u^2))[u>=0 & u<=1]))
answer<-sqrt(n)*top/bottom
if(is.na(answer)) 0 else answer
}

# Bisquare- Function (Negative A-estimator) #
bisquareminus<-function(x,c) {
n<-length(x)
t<-median(x)
s<-madd(x)
u<-((x-t)/(c*s))
top<-sqrt(.5*sum((((c*s*u)^2)*((1-u^2)^4)) [u<=0 & u>=-1]))
bottom<-abs(sum(((1-u^2)*(1-5*u^2))[u<=0 & u>=-1]))
answer<-sqrt(n)*top/bottom
if(is.na(answer)) 0 else answer
}

# Trimmed Abs+ Function #
trimmedsigplus<-function(x) {
median<-median(x)
x<-sort(x)
xnew<-x[quantile(x,.1)<=x & x<=quantile(x,.9)]
nstar<-length(x[median<=x & x<=quantile(x,.9)])
answer<-(1/(nstar))*sum(abs((xnew-median)[xnew>=median]))
if(is.na(answer)) 0 else answer
}

# Trimmed Abs- Function #
trimmedsigminus<-function(x) {
median<-median(x)
x<-sort(x)
```

```
xnew<-x[quantile(x,.1)<=x & x<=quantile(x,.9)]
nstar<-length(x[quantile(x,.1)<=x & x<=median])
answer<-(1/(nstar))*sum(abs((xnew-median)[xnew<=median]))
if(is.na(answer)) 0 else answer
}

#Trimmed Sd+ Function #
trimmedsdplus<-function(x) {
x<-sort(x)
if(mean(x)>quantile(x,.9)) middle<-median(x)
   else middle<-mean(x)
xnew<-x[quantile(x,.1)<=x & x<=quantile(x,.9)]
nstar<-length(x[middle<=x & x<=quantile(x,.9)])
answer<-sqrt((1/(nstar))*sum(((xnew-middle)^2)[xnew>=middle]))
if(is.na(answer)) 0 else answer
}

#Trimmed Sd- Function #
trimmedsdminus<-function(x) {
x<-sort(x)
if(mean(x)>quantile(x,.9)) middle<-median(x)
   else middle<-mean(x)
xnew<-x[quantile(x,.1)<=x & x<=quantile(x,.9)]
nstar<-length(x[quantile(x,.1)<=x & x<=middle])
answer<-sqrt((1/(nstar))*sum(((xnew-middle)^2)[xnew<=middle]))
if(is.na(answer)) 0 else answer
}
```

A.2 Deriving the Limiting Values of the Boxplot Fences

The limiting values of the upper (F_U) and lower (F_L) fences for the three types of boxplots can be derived from five quantities. The upper fence for the common boxplot is defined as

$$F_U = Q_3 + 1.5(Q_3 - Q_1)$$

and the lower fence is defined as

$$F_L = Q_1 - 1.5(Q_3 - Q_1).$$

The upper and lower fences for the skew adjusted boxplot are, respectively,

$$F_U = Q_3 + 3(Q_3 - Q_2)$$

and

$$F_L = Q_1 - 3(Q_2 - Q_1).$$

The upper and lower fences for the signed boxplot are, respectively,

$$F_U = Q_2 + 4(.845 Abs^+)$$

and

$$F_L = Q_2 - 4(.845 Abs^-).$$

Q_1, Q_2, and Q_3 are the 25^{th}, 50^{th}, and 75^{th} sample quantiles, respectively. These three quantities allow for the calculation of the upper and lower fences for the common and skew adjusted boxplots. The signed boxplot requires two more quantities, Abs^+ and Abs^-, in order to calculate the upper and lower fences.

Since Theorems 2, 3, and 4 in Chapter 2 show that these five quantities (Q_1, Q_2, Q_3, Abs^+, and Abs^-) converge in probability, the limiting values of the upper and lower fences are calculated by substituting in the population (or asymptotic) values for these statistics.

The limiting values of Q_1, Q_2, and Q_3 given in Tables 3.1 and 3.2 were derived from calculating $F^{-1}(.25)$, $F^{-1}(.50)$, and $F^{-1}(.75)$, respectively, for each of the nine distributions under consideration. These quantities were calculated as the solution, c, to equations of the form $\int_{-\infty}^{c} dF(x) - p = 0$, where F represents one of the nine distributions, $c \in \Re$, and p is .25, .50, or .75. The limiting values of Abs^+ and Abs^- are given in Tables 2.7 and 2.8 as derived from formulas following the methodology derived in Chapter 2. Specifically, the limiting value of Abs^{+} is $2\int_{\mu}^{\infty} |x-\mu|\, dF(x)$, for a random variable X with cdf F and $\mu = F^{-1}(.5)$, as given in Section 2.2.2 on page 28, and the limiting value for Abs^{-} is $2\int_{-\infty}^{\mu} |x-\mu|\, dF(x)$.

A.3 Deriving the Population Outside Rate Per Observation

For any of the three boxplots considered in this dissertation, the quantity $A(n)$, the outside rate per observation, is the probability that an observation (from a sample $X_1, \ldots, X_n$) lies outside the boxplot fences.

The limiting values of $A(n)$ depend upon the boxplot type and the underlying distribution. Table 3.6 lists the population outside rate per observation (as a percentage) for each of the boxplots and for nine distributions.

The formula for calculating the population outside rate per observation is

$$1 - \int_{F_L}^{F_U} dF$$

where F specifies the underlying distribution, F_L is the limiting value of the lower fence for a particular boxplot, and F_U is the limiting value of the upper fence for a particular boxplot.

Tables 3.1 and 3.2 provide the limiting values (rounded to two decimal places) for the lower and upper fences of all three boxplots for nine distributions. Using the above formula, these limiting values, and the formulas for any of the nine distributions, the population outside rate can be calculated.

For example, if the underlying distribution is Exponential(1), then the population outside rate per observation can be written

$$1 - \int_{F_L}^{F_U} e^{-x} dx$$

Using the appropriate values in Tables 3.1 and 3.2, the population outside rate per observation for the common boxplot is

$$1 - \int_{0}^{3.03} e^{-x} dx = e^{-3.03} = 0.0483.$$

The lower limit of the integral is replaced by 0 since the limiting value of the lower fence is negative and the exponential distribution gives zero probability to negative values.

Since the values in Tables 3.1 and 3.2 are rounded to two decimal places, using these values to reproduce the values in Table 3.6 will not result in perfect agreement. For example, the limiting value for the upper fence when the underlying distribution is exponential(1) is, to five decimal places, 3.03421. Using this value in the formula for the population outside rate per observation yields

$$e^{-3.03421} = 0.0481$$

which is the value reported in Table 3.6.