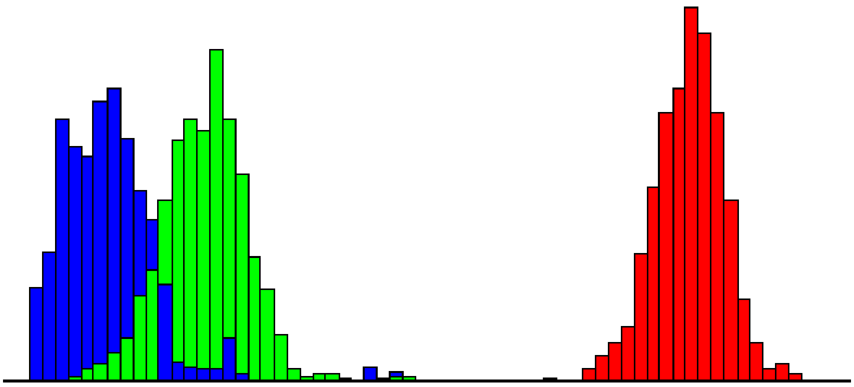
Charles University, Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics

Robust Sequential Methods



Zdeněk Hlávka

Advisor: Prof. RNDr. Marie Hušková, DrSc.

Univerzita Karlova, Matematicko-fyzikální fakulta
Katedra pravděpodobnosti a matematické statistiky

Robust Sequential Methods

Disertační práce

Zdeněk Hlávka

Obor: M9 – Matematická statistika
Školitel: Prof. RNDr. Marie Hušková, DrSc.

Acknowledgements

First of all, I would like to express my thanks to my advisor, Marie Hušková. Without her constant support, I would never start and I would never be able to finish this work.

I want to thank also to all my teachers from Charles University in Prague in Czech Republic and from Limburgs Universitair Centrum in Diepenbeek in Belgium for introducing me into the world of statistics. In particular, I want to thank Noël Veraverbeke for giving me the possibility to visit Diepenbeek in February this year, Jaromír Antoch for his advice concerning random number generators, and Marc Aerts for the copy of his PhD. thesis.

Last, but certainly not the least, I want to thank also to Wolfgang Härdle and all my colleagues from the Faculty of Economics at Humboldt University for enabling me to finish this work during my stay in Berlin.

This work was sponsored by a gift from the foundation "Nadání Josefa, Marie a Zdeňky Hlávkových" in Prague. I gratefully acknowledge their support.

Berlin, March 2000

Contents

Chapter 1

Preliminary Notions

In this introductory chapter, we intend to provide an overview of the methods for construction of fixed-width confidence intervals. The sequential procedures based on sample mean will be systematically presented in Section 1.1 whereas Section 1.2 concerns M-estimators and existing robust sequential methods. These methods will be further investigated and developed in Chapters 4–6.

1.1 Sequential Methods

The problem of constructing fixed-width confidence intervals is often studied in the literature. Suppose that we have i.i.d. observations $X_1, X_2, \ldots, X_n$ with common distribution function $F(., \theta)$. The estimate $\hat{\theta}_n$ of the unknown parameter θ has typically asymptotically normal distribution and its variance depends on the sample size n. The calculation of the confidence intervals (1.1) based on the asymptotic normal distribution is very easy.

In order to obtain a $1-\alpha$ confidence interval shorter than some prescribed length $2d$, we basically need enough observations in order to get sufficiently small variance of $\hat{\theta}_n$. Thus the problem of constructing fixed width confidence intervals reduces usually to the problem of determining sufficient sample size.

The most popular sequential methods for construction of fixed-width confidence sets for the parameter of location will be described in this section.

1.1.1 Chow-Robbins Procedure

Consider a random sample $X_1, \ldots, X_n$ from Normal distribution $N(\theta, \sigma^2)$. Suppose that the parameter σ^2 is known. The $1-\alpha$ confidence interval for θ is given by

$$\left(\bar{X}_n - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right), \tag{1.1}$$

where $\bar{X}_n$ is the sample mean of $X_1, \ldots, X_n$ and where $u_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard Normal distribution. The length of the confidence interval (1.1) is

$$2u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}. \tag{1.2}$$

It is immediate that in order to obtain $1-\alpha$ confidence interval shorter than $2d$, we need the number of observations to exceed

$$c(d) = \left(\frac{u_{1-\alpha/2}\sigma}{d}\right)^2. \tag{1.3}$$

The Chow-Robbins procedure concerns the situation with the unknown variance. The number of observations (stopping rule) N is given as the smallest integer exceeding $c(d)$, i.e.,

$$N = \inf\left\{n \geq m : n \geq \left(\frac{u_{1-\alpha/2}S_n}{d}\right)^2\right\}, \tag{1.4}$$

where $m \geq 2$ is the initial sample size and $S_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2/(n-1)$ is the sample variance of X_i's.

Basic asymptotic properties are given in the following Theorem 1.1.1.

THEOREM **1.1.1** *For the Chow-Robbins procedure (1.4), if $0 < \sigma^2 < \infty$, then*

$$\begin{aligned}
&(i) && P_{\mu,\sigma^2}(N < \infty) = 1; && (1.5)\\
&(ii) && E_{\mu,\sigma^2}(N) = n_0 + 1 + c; && (1.6)\\
&(iii) && N \equiv N(d) \downarrow a.s.\ in\ d; N \to \infty\ a.s.\ as\ d \to 0+; && \\
& && E_{\mu,\sigma^2}(N) \to \infty\ \ as\ d \to 0+; && (1.7)\\
&(iv) && N/c \to 1\ a.s.\ as\ d \to 0+; && (1.8)\\
&(v) && \lim_{d\to 0+} E_{\mu,\sigma^2}(N/c) = 1; && (1.9)\\
&(vi) && \lim_{d\to 0+} P_{\mu,\sigma^2}(\bar{X}_N - d \leq \mu \leq \bar{X}_N + d) = 1-\alpha. && (1.10)
\end{aligned}$$

PROOF: The proof is given in Ghosh, Mukhopadhyay, and Sen (1997), among others.

□

In Theorem 1.1.1, the comparison between Chow-Robbins stopping time $N(d)$ and the optimal fixed-sample size $c(d)$ given by the formula (1.3) is carried out. The last part of Theorem 1.1.1 asserts that the asymptotic coverage probability of the resulting confidence interval is really $1-\alpha$.

Notation: The symbols P_{μ,σ^2} and E_{μ,σ^2} denote respectively the probability and the expected value calculated for any fixed μ and σ^2. For simplicity of notation, the subscript μ, σ^2 will be omitted in the following text. We will also write N instead of $N(d)$ and c instead of $c(d)$. The full notation $N(d)$ and $c(d)$ will be used, rather arbitrarily, whenever we decide to stress the dependence of the sample size on the desired length of the confidence interval.

1.1.2 Two-stage Procedures

In some situations, Chow-Robbins procedure can not be recommended and a different procedure has to be considered. Sometimes, the data are obtained in batches (and very often the price is proportional to the number of batches instead to the number of observations) and it might be more appropriate to use methods which do not require to include only one new observation at time. Stein's method which will be shortly described in this section requires only two batches of data.

Suppose the same setup as before with the optimal fixed-sample size $c(d)$ defined in (1.3).

A procedure consisting of two steps has been suggested by Stein (1945,1949). In the first step, observations $X_1, \ldots, X_m$ from $\mathrm{N}(\mu, \sigma^2)$ are drawn, giving the sample mean $\bar{X}_m$ and the sample variance S_m^2.

In the second step, $N - m$ additional observations are drawn,

$$N = N(d) = \max\left\{m, \left[\left(\frac{t_{m-1}(\alpha)S_m}{d}\right)^2\right]^\circ + 1\right\}, \tag{1.11}$$

where $[x]^\circ$ denotes the integer part of x and $t_{m-1}(\alpha)$ is the $1-\alpha/2$ quantile of the Student's t-distribution with $m-1$ degrees of freedom.

THEOREM 1.1.2 *For Stein's two-stage procedure (1.11), we have*

$$(i) \qquad \frac{t_{m-1}(\alpha)\sigma^2}{d^2} \leq E(N) \leq m + \frac{t_{m-1}(\alpha)\sigma^2}{d^2}; \tag{1.12}$$

$$(ii) \qquad \lim_{d\to 0+} E\left(\frac{N}{c}\right) = \frac{t^2_{m-1}(\alpha)}{u^2_{1-\alpha/2}}; \tag{1.13}$$

$$(iii) \qquad P(\bar{X}_N - d \leq \mu \leq \bar{X}_N + d) \geq 1 - \alpha \text{ for all } \mu \text{ and } \sigma^2; \tag{1.14}$$

$$(iv) \qquad \lim_{d\to 0+} P(\bar{X}_N - d \leq \mu \leq \bar{X}_N + d) = 1 - \alpha \text{ for all } \mu \text{ and } \sigma^2. \tag{1.15}$$

PROOF: The proof can be found in Ghosh, Mukhopadhyay, and Sen (1997).

□

Theorem 1.1.2 states that Stein's two-stage procedure tends to overestimate the optimal sample size, even asymptotically. This negative feature is due to the fixed starting-sample size m, because for small d can a small change in the estimate of σ^2 cause big differences in the final sample size N. A modified two-stage procedure which improves this feature was suggested by Mukhopadhyay (1980).

The modified two-stage procedure varies the starting-sample size m according to the desired width of the confidence interval $2d$.

We choose a real number $\gamma > 0$, and define

$$m = m(d) = \max\left\{2, \left[\left(\frac{u_{1-\alpha/2}}{d}\right)^{2/(1+\gamma)}\right]^\circ + 1\right\}. \tag{1.16}$$

We start the experiment with this number of observations and continue as in the Stein's procedure (1.11) to obtain N and confidence interval $(\bar{X}_N - d, \bar{X}_N + d)$. The asymptotic properties of this procedure are stated in Theorem 1.1.3. Recall that $c = c(d)$, defined by (1.3), denotes the asymptotically optimal number of observations.

THEOREM 1.1.3 *The following statements hold for the modified two-stage procedure.*

$$(i) \qquad \lim_{d\to 0}\left(\frac{N}{c}\right) = 1 \ a.s. \tag{1.17}$$

$$(ii) \qquad \lim_{d\to 0} E\left(\frac{N}{c}\right) = 1 \tag{1.18}$$

$$(iii) \qquad P\{(\bar{X}_N - d, \bar{X}_N + d) \ni \mu\} \geq 1 - \alpha, \text{ for all } \mu \text{ and } \sigma^2, \tag{1.19}$$

$$(iv) \qquad \lim_{d\to 0} P\{(\bar{X}_N - d, \bar{X}_N + d) \ni \mu\} = 1 - \alpha, \text{ for all } \mu,\ \sigma^2. \tag{1.20}$$

PROOF: See Ghosh, Mukhopadhyay, and Sen (1997).

□

Theorem 1.1.3 shows the good properties of the modified two-stage procedure, namely the first order asymptotic efficiency in (1.18). An unpleasant property is given in Theorem 1.1.4.

THEOREM **1.1.4** *For the modified two-stage procedure, we have*

$$\liminf_{d\to 0+} E(N-c) = \infty. \tag{1.21}$$

PROOF: The proof is given in Ghosh, Mukhopadhyay, and Sen (1997).

□

This undesirable property led Ghosh and Mukhopadhyay (1981) to define the second order asymptotic efficiency. The procedure is called asymptotically second order efficient if $\lim_{d\to 0+} E(N-c)$ is a finite constant.

1.1.3 Three-stage Procedure

The asymptotic behaviour of Stein's procedure can be improved by adding one additional sampling stage to the algorithm. The three-stage method goes as follows.

In the first stage we draw

$$m = m(d) = \max\left\{2, \left[\left(\frac{u_{1-\alpha/2}}{d}\right)^{2/(1+\gamma)}\right]^{\circ} + 1\right\} \tag{1.22}$$

observations. The number of observations $m(d)$ is controlled by the tuning parameter $\gamma > 0$. The exponent $2/(1+\gamma)$ has to lie in the interval $(0,2)$ in order to guarantee the desired asymptotic properties. We obtain the sample variance S_m^2 which is used to determine the intermediate sample size N_1.

$$N_1(d) = \max\left\{m, \left[\left(k\frac{u_{1-\alpha/2}S_m}{d}\right)^2\right]^{\circ} + 1\right\}, \tag{1.23}$$

where $0 < k < 1$ is a parameter controlling the sample size in this stage.

Additionally, we draw $N_2 - N_1$ observations, where

$$N_2(d) = \max\left\{N_1, \left[\left(\frac{u_{1-\alpha/2}S_{N_1}}{d}\right)^2\right]^{\circ} + 1\right\} \tag{1.24}$$

and obtain the $1-\alpha$ confidence interval $(\bar{X}_{N_2} - d, \bar{X}_{N_2} + d)$.

The three-stage procedure has better asymptotic properties than the two-stage procedure. Namely, it is asymptotically second order efficient.

THEOREM **1.1.5** *For the three-stage procedure defined by (1.22)— (1.24), we have that*

$$\lim_{d\to 0+} E(N_2 - c) = \frac{1}{2} - 2k^{-1}. \tag{1.25}$$

PROOF: See Section 6.3 in Ghosh, Mukhopadhyay, and Sen (1997).

□

1.1.4 Three-stage Procedure Based on Bootstrap

All methods presented in the previous sections heavily depend on the accuracy of the normal approximation. It is useful to consider approach based on bootstrap which was introduced by Swanepoel, van Wijk, and Venter (1984). They generalized the Chow-Robbins procedure in the following way.

Given $X_1, \dots, X_n$ we have the empirical distribution function $F_n(x)$. We draw the bootstrap sample $X_1^*, \dots, X_n^*$ from $F_n(x)$ and calculate the sample mean $\bar{X}_n^*$. Denote by P_n^* the conditional probability measure corresponding to the empirical distribution function $F_n(x)$, i.e.,

$$P_n^*(A^*) = P(A^*|X_1, X_2, \dots, X_n). \tag{1.26}$$

Then we can calculate the conditional coverage probability

$$P_n^*(d) = P_n^*(\bar{X}_n^* - d < \bar{X}_n < \bar{X}_n^* + d). \tag{1.27}$$

which leads to the stopping time

$$N(d) = \inf(n \geq m; P_n^*(d) \geq 1 - \alpha). \tag{1.28}$$

Three-stage procedure based on bootstrap method was considered by Aerts and Gijbels (1993). They use standardized and studentized bootstrap method.

- In the standardized bootstrap, the normal approximation $u_{1-\alpha/2}$ of the upper $1-\alpha/2$ quantile of the standardized distribution of the parameter of interest is substituted by the $1-\alpha$ quantile of the (centered and standardized) conditional distribution of

$$\sqrt{n}\frac{|\bar{X}_n^* - \bar{X}_n|}{S_n}. \tag{1.29}$$

- In the studentized version of the bootstrap method, the normal quantile $u_{1-\alpha/2}$ is replaced by the $1-\alpha$ quantile of the (centered and studentized) conditional distribution of

$$\sqrt{n}\frac{|\bar{X}_n^* - \bar{X}_n|}{S_n^*}. \tag{1.30}$$

The intermediate sample size N_1 and the final sample size N_2 are calculated similarly as in the formulas (1.23) and (1.24), with the bootstrap critical points replacing the quantile $u_{1-\alpha/2}$. The asymptotic properties of this three-stage procedure are established in Theorem 1.1.6. Recall that $c(d)$ denotes the asymptotically optimal stopping time, see formula (1.3).

Theorem **1.1.6** *Suppose that $E(X^2) < \infty$ and $\sigma^2 > 0$. Then the three-stage procedure based on bootstrap has the following properties:*

$$(i) \qquad \lim_{d\to 0+} N_2(d) = \infty \qquad [P] \text{ a.s.}, \tag{1.31}$$

$$(ii) \qquad \lim_{d\to 0+} \frac{N_2(d)}{c(d)} = 1 \qquad [P] \text{ a.s.}, \tag{1.32}$$

$$(iii) \qquad \lim_{d\to 0+} P(\bar{X}_{N_2(d)} - d < \mu < \bar{X}_{N_2(d)} + d) = 1 - \alpha, \tag{1.33}$$

$$(iv) \qquad \text{if moreover } E|X|^{2+\delta} < \infty \text{ for some } \delta > 0 \text{ then also} \tag{1.34}$$

$$\lim_{d\to 0+} E\left(\frac{N_2(d)}{c(d)}\right) = 1. \tag{1.35}$$

Proof: See Aerts and Gijbels (1993).

□

1.2 M-estimators

Our intention is to robustify the three-stage methods. One possiblity is to consider the M-estimators which are very well described in the literature. In this section we will give the definition and state basic properties of M-estimators.

1.2.1 Definition and Basic Properties

Let $X_1, X_2, \ldots, X_n$ be independent, identically distributed (iid) random variables with common unknown distribution function $F(x) = F_\theta(x) = F_0(x-\theta)$ and let us assume that

$$\theta = M(F) = \arg\min_{t\in R} \int \rho(x-t)dF(x). \tag{1.36}$$

Assume that the distribution function $F_0(x)$ is symmetric about 0 and that the function $\rho(x)$ is convex and symmetric. These conditions imply that the parameter θ corresponds to the centre of symmetry of $F(.)$. The natural estimator for θ based on the observations $X_1, \ldots, X_n$ is then

$$M_n = M(F_n) = \arg\min_{t\in R} \int \rho(x-t)dF_n(x) = \arg\min_{t\in R} \sum_{i=1}^{n} \rho(X_i - t), \tag{1.37}$$

where $F_n(.)$ is the empirical distribution function of $X_1, \ldots, X_n$, $F_n(x) = (1/n)\sum_{i=1}^n I(X_i \le x)$. The random variable $M(F_n)$ is called the M-estimator of the parameter θ.

If the function $\rho(.)$ admits a derivative $\rho'(.) = \psi(.)$, then the M-estimator defined in (1.37) solves the equation

$$\sum_{i=1}^{n} \psi(X_i - t) = 0. \tag{1.38}$$

However, it can happen that the equation (1.38) has more than one solution. with only one solution corresponding to the desired global minimum. If this is the case and if the function $\psi(.)$ is nondecreasing, we may define M_n in a unique way as

$$M_n = \frac{1}{2}(M_n^+ + M_n^-), \tag{1.39}$$

where

$$M_n^- = \sup\left\{\sum_{i=1}^{n} \psi(X_i - t) > 0\right\}, \tag{1.40}$$

$$M_n^+ = \inf\left\{\sum_{i=1}^{n} \psi(X_i - t) < 0\right\}. \tag{1.41}$$

The assumed convexity of function ρ implies that the function ψ is non-decreasing.

There is a huge amount of work on M-estimators. Huber (1981) is the classical reference. Jurečková and Sen (1995), Hampel, Ronchetti, Rousseeuw, and Stahel (1990) and Rieder (1994) are more recent monographies connected with this subject. An introduction to the theory of M-estimators is given e.g. in Serfling (1980).

Basic asymptotic properties of M-estimators are stated in Theorems 1.2.1 and 1.2.2 below. Theorem 1.2.2 states sufficient conditions for asymptotic normality of the M-estimator generated by non-decreasing score function $\psi(.)$. Theorem 1.2.1 concerns strong consistency of the M-estimator. We use the notation

$$\lambda_F(t) = \lambda_{F_\theta}(t) = \int_{-\infty}^{\infty} \psi(x-t)dF(x), \qquad \text{for } -\infty < t < \infty. \tag{1.42}$$

We use the subscript θ and write $F_\theta(.)$ instead of $F(.)$ whenever we want to stress (or fix) the value of the parameter.

THEOREM **1.2.1** *Let θ be an isolated root of $\lambda_F(t) = 0$. Let $\psi(x-t)$ be monotone in t. Then θ is unique and any solution sequence $\{M_n\}$ of the empirical equation $\lambda_{F_n}(t) = 0$ converges to θ almost surely. If, further, $\psi(t)$ is continuous in t in a neighborhood of θ, then there exists such a solution sequence.*

PROOF: See Serfling (1980).

□

THEOREM **1.2.2** *Let θ be an isolated root of $\lambda_F(t) = 0$. Let $\psi(x-t)$ be monotone in t. Suppose that $\lambda_F(t)$ is differentiable at $t = \theta$, with $\lambda'_F(\theta) \neq 0$. Suppose that $\int \psi_F^2(x-t)dF(x)$ is finite for t in a neighborhood of θ and is continuous at $t = \theta$. Then any solution sequence $\{M_n\}$ of the empirical equation $\lambda_{F_n}(t) = 0$ is asymptotically normal, $AN(\theta, \sigma^2(\psi, F))$, where*

$$\sigma^2(\psi, F) = \frac{\int \psi^2(x-\theta)dF(x)}{[\lambda'_F(\theta)]^2}. \tag{1.43}$$

PROOF: See Serfling (1980).

□

1.2.2 Robust Sequential Procedure

Robust sequential procedure has been suggested by Jurečková and Sen (1978).

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables with a common continuous distribution function $F_\theta(x)$. Fix some function $\psi(.)$ which defines the M-estimator and assume that the conditions which will be given in Section 4.1 are fulfilled.

Define

$$\begin{aligned}
\hat{\theta}_n^- &= \sup\left\{t : \hat{r}_n^{-1} n^{-1/2} \sum_{i=1}^n (X_i - t) > u_{1-\alpha/2}\right\} \\
\hat{\theta}_n^+ &= \inf\left\{t : \hat{r}_n^{-1} n^{-1/2} \sum_{i=1}^n (X_i - t) < u_{1-\alpha/2}\right\},
\end{aligned}$$

where

$$\hat{r}_n^2 = \frac{1}{n} \sum_{i=1}^n \psi^2(X_i - \bar{X}_n) - \left(\sum_{i=1}^n \psi(X_i - \bar{X}_n)\right)^2.$$

The sequential procedure is defined as follows: for fixed $d > 0$, let $N(d)$ be the first integer $n \geq n_0$ (initial sample size) for which

$$L_n = \hat{\theta}_n^+ - \hat{\theta}_n^- \leq 2d.$$

The corresponding confidence interval is $\left(\hat{\theta}^{-}_{N(d)}, \hat{\theta}^{+}_{N(d)}\right)$ based on $X_1, \ldots, X_{N(d)}$.

Similarly, as for the sample mean, we define the asymptotically optimal sample size $c_M(d)$ for the M-estimator:

$$c_M(d) = \left(\frac{u_{1-\alpha/2}\sigma(\psi, F)}{d}\right)^2, \tag{1.44}$$

where $\sigma^2(\psi, F)$ is the variance of the asymptotic normal distribution of the M-estimator given by (1.43). Jurečková and Sen (1978) proved the following properties of their fully sequential procedure.

Theorem **1.2.3** *We have the following for the sequential procedure based on M-estimators.*

$$(i) \qquad \lim_{d \to 0} \left(\frac{N(d)}{c_M(d)}\right) = 1 \ a.s. \tag{1.45}$$

$$(ii) \qquad \lim_{d \to 0} P\{(\bar{X}_{N(d)} - d, \bar{X}_{N(d)} + d) \ni \mu\} = 1 - \alpha, \ \textit{for all } \mu, \ \sigma^2. \tag{1.46}$$

$$(iii) \qquad \sqrt{n}\left(\hat{\theta}^{-}_n - \theta\right) \sim N(-\sigma(\psi, F)u_{1-\alpha/2}, \sigma^2(\psi, F)), \tag{1.47}$$

$$\sqrt{n}\left(\hat{\theta}^{+}_n - \theta\right) \sim N(\sigma(\psi, F)u_{1-\alpha/2}, \sigma^2(\psi, F)). \tag{1.48}$$

Proof: See Theorem 3.1 and Lemma 3.1 in Jurečková and Sen (1978).

□

Chapter 2

Aim of Thesis

The sequential methods based on the sample mean which are presented in Section 1.1 are non-robust and can produce misleading results if the real distribution does not fulfil assumptions of normality. One possibility is to consider M-estimators which may work better in some situations.

The fully sequential methods based on M-estimators are computationally very intensive. On the other hand, the two-stage procedure has some undesirable asymptotic properties. Therefore, we focus on the three-stage procedure. We will generalize the sequential procedure based on bootstrapping sample mean suggested by Aerts and Gijbels (1993). We suggest to base the sequential method on the more general and more robust M-estimators. We suggest (and compare) three types of approximations of the distribution of the M-estimator: approach based on asymptotic normality, approach based on standardized bootstrap, and approach based on studentized bootstrap.

Asymptotic properties will be established in order to provide a theoretical background for the suggested methods. This includes the description of the behaviour of the bootstrap for M-estimators and the description of the behaviour of the stopping time of the sequential procedure based on the bootstrap critical points.

In Chapter 4, the three-stage procedure based on bootstrap critical points for fixed-width confidence intervals will be defined for M-estimators of the location parameter. Theorem 4.3.3 states that the asymptotic distribution of the bootstrap M-estimator is the same as the asymptotic distribution of the original M-estimator. The robust three-stage procedure is defined in Section 4.4 and its basic asymptotic properties are stated in Theorem 4.4.1.

Under additional assumptions, we are able to investigate the asymptotics for the three-stage procedure in more detail. The asymptotic distribution of the final sample size will be derived for the procedure based on sample mean in Section 4.5 and for the robust procedure based on bootstrap for M-estimators in Section 4.6.

In order to investigate the behaviour of our method for smaller sample size, we will carry out also the simulations. In Chapter 5, we will simulate random samples from standard Normal, Cauchy, Double Exponential, and from a mixture of Normal distributions. On the simulations from Normal distribution we want to see how well do the confidence intervals based on the normal critical points compare to the confidence intervals based on bootstrap critical points — we expect than in this case they should behave very similarly. On the other hand, for the heavy-tailed Cauchy and Double Exponential distributions, we expect to see some difference between the methods based on the asymptotic normality and the methods based on the bootstrap approximation. We expect that the coverage probability of the $1 - \alpha$ confidence intervals based on bootstrap will be closer to the desired value of $1 - \alpha$, especially for small sample sizes.

We do not intend to restrict ourselves only to the estimation of the parameter of location. In Chapter 6, we will generalize our method and define the fixed-width confidence intervals based on the bootstrap for the least squares regression parameters and we will discuss also other possible future generalizations, namely robust regression and generalized M-estimators.

In Appendix B, we present some useful tables. In Tables B.2 and B.3, we tabulate the values of the starting sample size $m(d)$ defined by (1.22) for different values of d and γ for $\alpha = 0.05$ (Table B.2) and $\alpha = 0.01$ (Table B.3). In Tables B.4–B.9, we list the asymptotically optimal sample sizes for the fixed-width confidence intervals based on M-estimators for $\alpha = 0.05$ and $\alpha = 0.01$ and for different score (Huber's type) score functions.

Chapter 3

Methods and Tools

Our main aim is to construct robust fixed-width confidence intervals. It can be easily shown that, if the scale of the observations is unknown, there is no method for constructing fixed-width confidence interval with fixed number of observations. Therefore, sequential methods are a very important area of research. Systematic overview of sequential analysis is given for instance in Ghosh, Mukhopadhyay, and Sen (1997). We present shortly some of the sequential methods in Chapter 1. These methods will be robustified by considering both M-estimators and bootstrap critical points.

M-estimators are a generalization of the least squares and the maximum likelihood estimators. They have some very appealing properties. M-estimators are usually defined as a solution of an implicit equation involving the score function $\psi(.)$, see Section 1.2. The properties of M-estimators can be tuned by choosing appropriate score function. By choosing $\psi(x) = x$, we obtain the usual least squares estimator.

It is well known (see e.g. Huber (1981) or Jurečková and Sen (1995)) that M-estimators are (under some conditions) asymptotically normal. It is possible to base the construction of the confidence intervals on the asymptotic normality. However, the approach based on the bootstrap critical points seems to be in this situation more appropriate, especially for small sample sizes.

The bootstrap method was invented by Efron (1979) and it has become standard statistical tool because of its usefulness and the ease of its implementation. A nice overview of the theory of bootstrap can be found in Shao and Tu (1995). There are many types of bootstrap critical points, but we will concentrate only on the standardized and studentized version. In the future, it is possible to refine our methods by considering some bias-corrections, smoothed bootstrap or some other refinement of the naive Efron's bootstrap.

The main results are Theorem 4.3.3 which says that bootstrap works under quite general assumptions, Theorem 4.4.1 which states the basic asymptotic properties of our robust three-stage procedure based on bootstrap, and Theorems 4.5.6 and 4.6.4 which describe the asymptotic distribution of $\sqrt{N_2}$ for the three-stage procedure based on the sample mean by Aerts and Gijbels (1993) and for the robust three-stage procedure proposed in Chapter 4, respectively. The main tools which were used in the proofs of these theorems are collected in Appendix A. These are namely the basic probability inequalities (Markov, Hoeffding), inequalities connected with Central Limit Theorem (Edgeworth expansions), and random Central Limit Theorem by Anscombe (1952). The asymptotic linearity for M-estimators as well as other results by Jurečková and Sen (1982) are restated and generalized in Section 4.2 and used in the proof of Theorem 4.4.1. Results of Hall (1992) and Lahiri (1992) concerning the Edgeworth expansion of bootstrap critical points are used in the proofs of Theorems 4.5.6 and 4.6.4.

Chapter 4

Location Parameter

In this chapter, we will introduce the robust three-stage procedure for fixed-width confidence intervals. The assumptions are stated in Section 4.1. Some interesting and useful properties of M-estimators based on results of Jurečková and Sen (1982) can be found in Section 4.2. In Section 4.3, we will describe the "naive" bootstrap for M-estimators, we show that it "works", and we prove some lemmas which will be used in the proof of Theorem 4.4.1. In Section 4.5, we will turn our attention back to the (non-robust) procedure based on bootstrap for sample mean by Aerts and Gijbels (1993) and we will investigate the asymptotic distribution of its stopping time. Section 4.6 contains results on asymptotic distribution of the stopping time for the robust three-stage procedure based on bootstrap critical points.

4.1 Assumptions

Let us now formulate the regularity conditions on the score function $\psi(.)$ and on the distribution function $F(.) = F_\theta(.) = F_0(. - \theta)$ of the observations X_i.

Assumptions on F: $F_0(.) = F(. + \theta)$ has an absolutely continuous density $f_0(.)$ such that

$$f_0(x) = f_0(-x),\ \forall x \in \Re \tag{4.1}$$

and

$$f_0(x) \text{ is decreasing in } x \text{ for } x \geq 0. \tag{4.2}$$

Moreover, $F(.)$ has the finite Fisher information, i.e.,

$$0 < I(F) = \int_{-\infty}^{\infty} \{f'(x)/f(x)\}^2 dF(x) < \infty \tag{4.3}$$

and there exists $l > 0$ such that

$$E|X_1|^l = \int_{-\infty}^{\infty} |x|^l dF(x) < \infty. \tag{4.4}$$

Put

$$c_l(x) = |x|^l F(x)(1 - F(x)), \qquad x \in \Re, \tag{4.5}$$

$$c_l^* = \sup_{x\in\Re} c_l(x), \tag{4.6}$$

where l is given by (4.4). Then

$$c_l^* < \infty, \ \lim_{x\pm\infty} c_l(x) = 0, \text{ and } \int_{-\infty}^{\infty} \{F(x)(1-F(x))\}^b \, dx < \infty \qquad \forall b > \frac{1}{l} > 0. \tag{4.7}$$

Assumptions on ψ: The score function $\psi(.)$ is nondecreasing and skew–symmetric, i.e.

$$\psi(x) = -\psi(-x) \text{ is } \nearrow \text{ in } x \in [0,\infty), \tag{4.8}$$

and that there exists a positive number h such that

$$\psi(x) = \psi(h)\,\mathrm{sign}(x) \qquad \text{for } |x| \geq h. \tag{4.9}$$

$\psi(.)$ can be decomposed into the absolutely continuous and the step components,

$$\psi(x) = \psi_1(x) + \psi_2(x) \qquad \forall x \in \Re, \tag{4.10}$$

where $\psi_1(x)$ and $\psi_2(x)$ are respectively the absolutely continuous and the step component. Assume that the step component ψ_2 can be written as

$$\psi_2(x) = \beta_j \quad \text{for } a_{j-1} < x < a_j, \tag{4.11}$$

where $j = 1,\ldots,m+1$, $a_0 = -h$, $a_{m+1} = h$.

Assumptions on $\lambda_{F_\theta}(t)$: The function $\lambda_{F_\theta}(t)$ is differentiable at $t = \theta$,

$$\lambda_{F_\theta}(\theta) = 0, \tag{4.12}$$

$$\gamma(\psi, F_\theta) \equiv \lambda'_{F_\theta}(\theta) \neq 0, \tag{4.13}$$

and its derivative $\lambda'_{F_\theta}(t)$ is a continuous function of t in some neighbourhood of θ. Notice that, under our assumptions on the function $\psi(.)$, we have

$$\lambda'_{F_\theta}(\theta) = \int_{-\infty}^{\infty} \psi'_1(x)dF_\theta(x) + \sum_{j=1}^{m+1} (\beta_j - \beta_{j-1}) f(a_j). \tag{4.14}$$

4.2 Useful Properties of M-estimators

Jurečková and Sen (1982) investigated various asymptotic properties of M-estimators. Theorem 4.2.1 below deals with the asymptotic linearity of the M-estimator. For simplicity of notation we define, similarly as in (1.42),

$$\lambda_{F_n}(t) = \int_{-\infty}^{\infty} \psi(x-t)dF_n(x) = \sum_{i=1}^{n} \psi(X_i - t), \qquad \text{for } -\infty < t < \infty. \tag{4.15}$$

The asymptotic linearity for M-estimator is very useful because it allows us to investigate the behaviour of the M-estimators with the same methods which are usually used for the sample mean.

THEOREM **4.2.1** *(Asymptotic Linearity for M-estimator) Let us assume that the conditions of Section 4.1 hold. Then, for every fixed $K < \infty$, for every $\varepsilon > 0$ and $\delta > 0$, there exists n_0 such that*

$$P\left\{\sup_{|t|<K} |n^{-1/2}\left[\sum_{i=1}^{n} \psi(X_i - \theta - tn^{-1/2}) - \sum_{i=1}^{n} \psi(X_i - \theta)\right] - t\lambda_F'(\theta)| > \varepsilon\right\} \leq cn^{-1-\delta}, \ \forall n > n_0. \tag{4.16}$$

PROOF: See Jurečková and Sen (1982).

□

Theorem 4.2.1 immediately implies the almost sure convergence of the supremum inside the probability in (4.16) to zero. This result is stated in the following Corollary 4.2.1.

COROLLARY **4.2.1** *Under the assumptions of Section 4.1, we have for every fixed $0 < K < \infty$*

$$\sup_{|t|<K} |n^{1/2}\left[\lambda_{F_n}(\theta + tn^{-1/2}) - \lambda_{F_n}(\theta)\right] - t\lambda_F'(\theta)| \to 0 \qquad [P] \ a.s. \tag{4.17}$$

PROOF: By Theorem 4.2.1, for every fixed $K < \infty$, for any fixed $\delta > 0$ and $\varepsilon > 0$ there exists n_0 such that

$$\sum_{n=n_0+1}^{\infty} P\left\{\sup_{|t|<K} |n^{1/2}\left[\lambda_{F_n}(\theta + tn^{-1/2}) - \lambda_{F_n}(\theta)\right] - t\lambda_F'(\theta)| > \varepsilon\right\} \leq \sum_{n=n_0+1}^{\infty} cn^{-1-\delta}.$$

This yields that the series on the left hand side is convergent and this implies (by Theorem 1.3.4 in Serfling (1980)) the almost sure convergence of the sequence

$$\sup_{|t|<K} |n^{1/2}\left[\lambda_{F_n}(\theta + tn^{-1/2}) - \lambda_{F_n}(\theta)\right] - t\lambda_F'(\theta)|$$

in (4.17).

□

For proving the asymptotic properties of the bootstrap procedure under our conditions, we will need stronger version of Theorem 4.2.1.

LEMMA **4.2.1** *Under the assumptions of Section 4.1, for every fixed $0 < K < \infty$, $\varepsilon > 0$, $\delta > 1/4$, and $\eta > 0$, there exists n_0 such that*

$$P\left\{\sup_{|t|<K\log n} |n^{1-\delta}\left[\lambda_{F_n}(\theta + tn^{-1/2}) - \lambda_{F_n}(\theta)\right] - n^{1-\delta}\left[\lambda_F(\theta + tn^{-1/2}) - \lambda_F(\theta)\right]| \geq \varepsilon\right\}$$

$$\leq cn^{-1-\eta}, \ \forall n > n_0. \tag{4.18}$$

PROOF: We assume, without loss of generality, that $\theta = 0$ and we define equidistant grid points $-K\log n = t_1 < t_2 < \cdots < t_{L_n} = K\log n$. We denote the number of the grid points by L_n. For all $t \in \langle t_1, t_2\rangle$ we have that

$$n^{1-\delta}\left[\lambda_{F_n}(tn^{-1/2}) - \lambda_F(tn^{-1/2}) - \lambda_{F_n}(0)\right] \leq n^{1-\delta}\left[\lambda_{F_n}(t_1 n^{-1/2}) - \lambda_F(t_2 n^{-1/2}) - \lambda_{F_n}(0)\right]$$

$$\leq n^{1-\delta}\left[\lambda_{F_n}(t_1n^{-1/2})-\lambda_{F_n}(0)-\left\{\lambda_F(t_1n^{-1/2})-\lambda_F(0)\right\}+\lambda_F(t_1n^{-1/2})-\lambda_F(t_2n^{-1/2})\right]$$

$$=n^{1-\delta}\left[\lambda_{F_n}(t_1n^{-1/2})-\lambda_{F_n}(0)-\left\{\lambda_F(t_1n^{-1/2})-\lambda_F(0)\right\}\right]+n^{1-\delta}\left[\lambda_F(t_1n^{-1/2})-\lambda_F(t_2n^{-1/2})\right]$$

$$\leq n^{1-\delta}\left[\lambda_{F_n}(t_1n^{-1/2})-\lambda_{F_n}(0)-\left\{\lambda_F(t_1n^{-1/2})-\lambda_F(0)\right\}\right]+\varepsilon/2, \tag{4.19}$$

for all n greater than some n_0 if we choose t_1 and t_2 such that

$$(t_2-t_1)n^{-1/2+1-\delta}=2Kn^{1/2-\delta}\log n/L_n\to 0$$

as $n\to\infty$. This can be achieved by choosing e.g. $L_n=n^{1/2-\delta/2}$. It follows from (4.19) and from the Markov inequality that

$$P\left\{\sup_{|t|<K\log n}|n^{1/2}\left[\lambda_{F_n}(tn^{-1/2})-\lambda_{F_n}(0)\right]-n^{1/2}\left[\lambda_F(tn^{-1/2})-\lambda_F(0)\right]|\geq\varepsilon n^{-1/2+\delta}\right\}$$

$$\leq P\left\{\max_{t_i,i\in 1,\dots,L_n-1}|n^{1/2}\left[\lambda_{F_n}(t_in^{-1/2})-\lambda_{F_n}(0)\right]-n^{1/2}\left[\lambda_F(t_in^{-1/2})-\lambda_F(0)\right]|\geq\varepsilon n^{-1/2+\delta}/2\right\}$$

$$\leq\sum_{i=1}^{L_n-1}P\left\{|n^{1/2}\left[\lambda_{F_n}(t_in^{-1/2})-\lambda_{F_n}(0)\right]-n^{1/2}\left[\lambda_F(t_in^{-1/2})-\lambda_F(0)\right]|\geq\varepsilon n^{-1/2+\delta}/2\right\}$$

$$\leq\sum_{i=1}^{L_n-1}\frac{n^{q-2q\delta}}{(\varepsilon/2)^{2q}}E\left(\sqrt{n}\left[\lambda_{F_n}(t_in^{-1/2})-\lambda_{F_n}(0)-\lambda_F(t_in^{-1/2})\right]\right)^{2q}$$

$$=\frac{n^{q-2q\delta}}{(\varepsilon/2)^{2q}}n^{-q}\sum_{i=1}^{L_n-1}E\left[\sum_{j=1}^{n}Z_j(t_i,n)\right]^{2q}\qquad\text{for } q=1,2,\dots, \tag{4.20}$$

where $Z_j(t_i,n)$ denotes $\psi(X_j-t_in^{-1/2})-\psi(X_j)-E\psi(X_j-t_in^{-1/2})$. Clearly, we have that $EZ_j(t_i,n)=0$ and $|Z_j(t_i,n)|\leq 2K$ for all i and j.

$$E\left[\sum_{j=1}^{n}Z_j(t_i,n)\right]^{2q}=\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\cdots\sum_{j_{2q}=1}^{n}EZ_{j_1}(t_i,n)Z_{j_2}(t_i,n)\dots Z_{j_{2q}}(t_i,n)$$

Notice that $Z_j(t_i,n)$, $j\in 1,\dots,n$ are independent and bounded random variables. It follows that

$$EZ_{j_1}(t_i,n)Z_{j_2}(t_i,n)\dots Z_{j_{2q}}(t_i,n)=0$$

if there exists index j_k different from all other indeces and

$$EZ_{j_1}(t_i,n)Z_{j_2}(t_i,n)\dots Z_{j_{2q}}(t_i,n)\leq[EZ_1^2(t_i,n)]^jK^{2q-2j}$$

if there are j different indeces among $j_1,\dots,j_{2q}$, each of them appearing at least twice. This gives that

$$E\left[\sum_{j=1}^{n}Z_j(t_i,n)\right]^{2q}\leq\sum_{j=1}^{q}n^j\left[EZ_1^2(t_i,n)\right]^jK^{2q-2j}=n^q\left[EZ_1^2(t_i,n)\right]^q+O(n^{q-1}) \tag{4.21}$$

Let's now investigate establish the asymptotic properties of $EZ_1^2(t_i, n)$. We have

$$EZ_1^2(t_i, n) = E\left[\psi(X_1 - t_i n^{-1/2}) - \psi(X_1) - E\psi(X_1 - t_i n^{-1/2})\right]^2$$

$$= E\left[\psi(X_1 - t_i n^{-1/2}) - \psi(X_1)\right]^2 - \left[E\psi(X_1 - t_i n^{-1/2})\right]^2$$

$$\leq 2\psi(h)E\left[\psi(X_1 - t_i n^{-1/2}) - \psi(X_1)\right] - \left[\lambda'_F(\theta)t_i n^{-1/2} + o(t_i n^{-1/2})\right]^2$$

$$= 2\psi(h)\lambda'_F(\theta)t_i n^{-1/2} + o(t_i n^{-1/2}) = O(t_i n^{-1/2}). \tag{4.22}$$

Combining (4.20) with (4.21) and (4.22) gives that there exists n_1 such that for all $n > n_1$ we have that

$$P\left\{\sup_{|t|<K\log n} |n^{1-\delta}\left[\lambda_{F_n}(tn^{-1/2}) - \lambda_{F_n}(0)\right] - n^{1-\delta}\left[\lambda_F(tn^{-1/2}) - \lambda_F(0)\right]| \geq \varepsilon\right\}$$

$$\leq \frac{n^{q-2q\delta}}{(\varepsilon/2)^{2q}} n^{-q} \sum_{i=1}^{L_n-1} n^q \left[EZ_1^2(t_i, n)\right]^q = \frac{n^{q-2q\delta}}{(\varepsilon/2)^{2q}} \sum_{i=1}^{L_n-1} \left[EZ_1^2(t_i, n)\right]^q$$

$$= L_n O(n^{q/2-2q\delta}\log n) = O(n^{q/2-2q\delta+1-\delta/2}\log n), \tag{4.23}$$

where the number of grid points was set to $L_n = n^{1/2-\delta/2}$. Notice that it is necessary to have $\delta > 1/4$ if we want that the term on the right hand side of (4.23) converges to 0.

To finish the proof we choose

$$q > \frac{3/2 + \eta + \delta/2}{2\delta - 1/2}$$

which, together with $\delta > 1/4$, guarantees that

$$q/2 - 2q\delta + 1 - \delta/2 < -1 - \eta$$

which gives that the term on the right hand side of (4.23) converges to zero faster than $n^{-1-\eta}$.

Notice that the supremum in the lemma is taken over the set $\{t : |t| < K\log n\}$, because we need $tn^{-1/2} \to 0$ in order to have the property (4.22).

□

The following Lemma 4.2.2 gives us exponential upper bound for the probability that the M-estimator is far from the true parameter θ.

Lemma 4.2.2 *Let us assume that the conditions of Section 4.1 hold. Then, for every $c_1 > 0$ and $0 < t < \sqrt{n}c_1$, we have that*

$$P_\theta\{\sqrt{n}|M_n - \theta| > t\} \leq 2e^{-c_2 t^2}, \tag{4.24}$$

where

$$c_2 \geq 2[f(K + c_1)]^2. \tag{4.25}$$

PROOF: See Lemma 3.1 in Jurečková and Sen (1982).

□

Combining Lemmas 4.2.1 and 4.2.2, we get the following Corollary 4.2.2 which we will use in the proof of Theorem 4.3.3.

COROLLARY **4.2.2** *Under the assumptions of Section 4.1, we get for every $t \in \Re$ fixed and for every sequence of random variables $\{S_n\}_{n=1}^n$, $S_n \to S$, $[P]$ a.s. that*

$$\lim_{n\to\infty} n^{1/2}\lambda_{F_n}(M_n + S_n t n^{-1/2}) = St\lambda'_F(\theta) \qquad [P] \ a.s. \tag{4.26}$$

PROOF: Notice that

$$n^{1/2}\lambda_{F_n}(M_n + S_n t n^{-1/2})$$
$$= n^{1/2}\left[\lambda_{F_n}\left(\theta + \left\{\sqrt{n}(M_n - \theta) + (S_n - S)t + Stn^{-1/2}\right\}\right) - \lambda_{F_n}(\theta + (M_n - \theta))\right].$$

We can use Lemma 4.2.2 with $c_1 = 1$, because there always exists n_0 such that for all $n > n_0$ we have

$$0 < \sqrt{\frac{2}{c_2}\log n} < \sqrt{n}.$$

In this situation, Lemma 4.2.2 leads that

$$P_\theta\left\{\sqrt{n}(M_n - \theta) > \sqrt{\frac{2}{c_2}\log n}\right\} < 2\exp\{-2\log n\} = 2n^{-2}$$

which further implies that there exists n_0 such that for all $n > n_0$ we have, $[P]$ *a.s.*,

$$\sqrt{n}(M_n - \theta) < \sqrt{\frac{2}{c_2}\log n} < \log n.$$

Using the assumptions, we can establish similar result for $tn^{-1/2}(S_n - S)$. This means that we can use Lemma 4.2.1. We get that there exists n_0 such that for all $n > n_0$

$$P\left\{\sup|n^{1/2}\left[\lambda_{F_n}(M_n + S_n t n^{-1/2}) - \lambda_{F_n}(M_n)\right] - n^{1/2}\left[\lambda_F(M_n + S_n t n^{-1/2}) - \lambda_F(M_n)\right]| \geq \varepsilon\right\}$$
$$< n^{1-\delta}$$

which implies that, $[P]$ a.s.,

$$|n^{1/2}\left[\lambda_{F_n}(M_n + S_n t n^{-1/2}) - \lambda_{F_n}(M_n)\right] - n^{1/2}\left[\lambda_F(M_n + S_n t n^{-1/2}) - \lambda_F(M_n)\right]| \to 0 \tag{4.27}$$

Using the differentiability of $\lambda_F(.)$ in some neighbourhood of θ, the almost sure convergence of M_n and S_n, and the Taylor expansion for the second difference in (4.27), and letting n tend to infinity concludes the proof.

□

In order to investigate the asymptotic properties of the M-estimators, we have to establish also some basic properties of estimators of the asymptotic variance of the M-estimator. Some of these properties (under our assumptions) were investigated in Jurečková and Sen (1982).

LEMMA **4.2.3** *Let us assume that the conditions of Section 4.1 hold. For any $\varepsilon > 0$ and $\delta > 0$, there exist positive constants c and n_0 such that*

$$P\left\{ \left|\frac{1}{n}\sum_{i=1}^{n}\psi^2(X_i - M_n) - \int_{-\infty}^{\infty}\psi^2(x-\theta)dF(x)\right| > \varepsilon \right\} \leq cn^{-1-\delta} \qquad \forall n > n_0 \tag{4.28}$$

PROOF: See Lemma 3.6 in Jurečková and Sen (1982).

□

For being able to prove that bootstrap for M-estimators works, we need the following modification of Lemma 4.2.3.

LEMMA **4.2.4** *Let us assume that the conditions of Section 4.1 hold. For any $\varepsilon > 0$ and $\delta > 0$ and for any sequence of random variables $\{S_n\}$ such that $S_n \to S$, $[P]$ a.s., there exist positive constants c and n_0 such that*

$$P\left\{ \left|\frac{1}{n}\sum_{i=1}^{n}\psi^2(X_i - M_n - zS_n n^{-1/2}) - \int_{-\infty}^{\infty}\psi^2(x-\theta)dF(x)\right| > \varepsilon \right\} \leq cn^{-1-\delta} \qquad \forall n > n_0. \tag{4.29}$$

PROOF: This can be seen immediately after a straightforward minor modification of the proof of Lemma 3.6 in Jurečková and Sen (1982).

□

Let's define

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\psi^2(X_i - M_n), \tag{4.30}$$

the following corollary gives us the almost sure convergence of this estimator to $\int \psi^2(x-\theta)dF(x)$.

COROLLARY **4.2.3** *Let us assume that the assumptions of Section 4.1 are fulfilled and that $\{S_n\}$ is sequence of random variables such that $S_n \to S$, $[P]$ a.s.. Then we have*

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\psi^2(X_i - M_n) = \int_{-\infty}^{\infty}\psi^2(x-\theta)dF(x) \quad [P]\ a.s. \tag{4.31}$$

and

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\psi^2(X_i - M_n - zS_n n^{-1/2}) = \int_{-\infty}^{\infty}\psi^2(x-\theta)dF(x) \quad [P]\ a.s. \tag{4.32}$$

PROOF: It follows from Lemmas 4.2.3 and 4.2.4. The proof is exactly the same as the proof of Corollary 4.2.1.

□

4.3 Bootstrap for M-estimators

The bootstrap for M-estimators goes along the same lines as the bootstrap for the sample mean introduced in section 1.1.4. As before, the unknown exact distribution of the M-estimator M_n is approximated by the distribution of M_n^* which is the M-estimator calculated from the bootstrap sample $X_1^*, \ldots, X_n^*$.

The M-estimator M_n is calculated from the measurements $X_1, X_2, \ldots, X_n$. Denote by F_n the empirical distribution function of X_i's, $F_n(x) = (1/n) \sum_{i=1}^n I(X_i \leq x)$ and let us denote by P^* the corresponding probability measure (conditional on $X_1, \ldots, X_n$).

Now random sample $X_1^*, \ldots, X_n^*$ from the distribution with d.f. F_n is drawn. Let us denote by M_n^* the M-estimator calculated from $X_1^*, \ldots, X_n^*$. In a number of situations, the distribution of $\sqrt{n}(M_n^* - M_n)$ approximates the distribution of $\sqrt{n}(M_n - \theta)$ with great accuracy. Sometimes, explicit formula for the distribution of $\sqrt{n}(M_n^* - M_n)$ exists, but more often approximations or simulation approaches are needed.

The critical points of the bootstrap distribution of M_n^* can be used as approximation of the unknown critical points of the distribution of the M-estimator M_n. We will use the bootstrap critical points which will be defined in Section 4.4 for the construction of the confidence regions which should be more accurate than the confidence regions based on the asymptotic normal distribution of M_n.

The properties of bootstrap for M-estimators were already investigated e.g. by Lahiri (1992) or Shorack (1982). Some results can be found also in Shao and Tu (1995). Unfortunatelly, the usual assumptions used in the literature do not cover M-estimators with discontinuous score functions which will be investigated here.

Theorem 4.3.1 below states the almost sure consistency of the conditional distribution of the bootstrap M-estimator M_n^*.

THEOREM **4.3.1** *Let the assumptions of Section 4.1 be satisfied. Let M_n^* be the M-estimator based on the bootstrap sample from the empirical distribution $F_n(x)$ and let M_n be the M-estimator based on the random sample from the distribution $F(x - \theta)$. Then*

$$\lim_{n\to\infty} P^* \{|M_n^* - M_n| > \varepsilon\} = 0 \qquad [P] \ a.s. \tag{4.33}$$

PROOF: Let us investigate

$$\limsup_{n\to\infty} P^*\{M_n^* \leq M_n - \varepsilon\} \tag{4.34}$$

By monotonicity of $\psi(x)$ and the definition of M_n^*, we can write that

$$\limsup_{n\to\infty} P^*\{M_n^* \leq M_n - \varepsilon\}$$

$$= \limsup_{n\to\infty} P^* \left\{ \frac{1}{n} \sum_{i=1}^n \psi(X_i^* - M_n + \varepsilon) \leq \frac{1}{n} \sum_{i=1}^n \psi(X_i^* - M_n^*) \right\}$$

$$= \limsup_{n\to\infty} P^* \left\{ \frac{1}{n} \sum_{i=1}^n \psi(X_i^* - M_n + \varepsilon) \leq 0 \right\}. \tag{4.35}$$

Tchebyschev's inequality and standard tools give

$$P^*\left\{\left|\frac{1}{n}\sum_{i=1}^n[\psi(X_i^*-M_n+\varepsilon)-E^*\psi(X_i^*-M_n+\varepsilon)]\right|>\eta\right\}$$

$$\le\frac{1}{\eta^2}Var^*\left\{\frac{1}{n}\sum_{i=1}^n\psi(X_1^*-M_n+\varepsilon)\right\}=\frac{1}{\eta^2 n}Var^*\left\{\psi(X_1^*-M_n+\varepsilon)\right\}$$

$$=\frac{1}{\eta^2 n}\left\{E^*[\psi(X_1^*-M_n+\varepsilon)]^2-[E^*\psi(X_1^*-M_n+\varepsilon)]^2\right\}$$

$$=\frac{1}{\eta^2 n}\left\{\frac{1}{n}\sum_{i=1}^n[\psi(X_i-M_n+\varepsilon)]^2-\left[\frac{1}{n}\sum_{i=1}^n\psi(X_i-M_n+\varepsilon)\right]^2\right\}$$

$$\le\frac{1}{\eta^2 n}\left\{\frac{1}{n}\sum_{i=1}^n[\psi(X_i-M_n+\varepsilon)]^2\right\}\le\frac{1}{\eta^2 n}K^2\to 0,\ \text{as } n\to\infty. \tag{4.36}$$

Notice that for $\varepsilon>0$ small enough, as $n\to\infty$,

$$E^*\psi(X_i^*-M_n+\varepsilon)=\frac{1}{n}\sum_{i=1}^n\psi(X_i-M_n+\varepsilon)\to E\psi(X_1-\theta+\varepsilon)>0\ \ [P]\ a.s. \tag{4.37}$$

Combining (4.36) and (4.37) gives that the limit considered in (4.35) is 0, $[P]$ almost surely. Similarly, it can be shown that also $P^*\{M_n^*>M_n+\varepsilon\}\to 0$, $[P]$ almost surely. Combining this and (4.34) yields the desired result in (4.33).

□

4.3.1 Asymptotic Properties of Bootstrap for M-estimators

In Theorem 4.3.2 we will state the almost sure convergence of the estimate of the asymptotic variance of the M-estimator. Define

$$\hat\sigma_n^2=\frac{(1/n)\sum_{i=1}^n\psi^2(X_i-M_n)}{\left(\sqrt{n}[\lambda_{F_n}(M_n-tn^{-1/2})-\lambda_{F_n}(M_n+tn^{-1/2})]/2t\right)^2}. \tag{4.38}$$

Recall that the numerator of (4.38) has been already denoted by S_n^2, see (4.30).

Theorem **4.3.2** *Let the assumptions of Section 4.1 be satisfied. Then we have, for any fixed* $0<t<\infty$,

$$\sqrt{n}[\lambda_{F_n}(M_n+tn^{-1/2})-\lambda_{F_n}(M_n-tn^{-1/2})]\to 2t\lambda_F'(\theta)\qquad [P]\ a.s., \tag{4.39}$$

$$S_n^2=\frac{1}{n}\sum_{i=1}^n\psi^2(X_i-M_n)\to\sigma^2(\psi,F)\lambda_F'(\theta)\qquad [P]\ a.s., \tag{4.40}$$

and

$$\hat\sigma_n^2\to\sigma^2(\psi,F)\qquad [P]\ a.s. \tag{4.41}$$

PROOF: The first part (4.39) of the theorem is a direct consequence of the Corollary 4.2.2. The almost sure convergence in (4.40) and (4.41) follows immediately from Corollary 4.2.3 and (4.39). □

Remark 1 *Assuming that the conditions of Section 4.1 hold with the jump component* $\psi_2(x) \equiv 0$. *Assume that the continuous component* $\psi_1(x)$ *has bounded derivative inside the interval* $(-h, h)$, *where* h *is given by (4.9). Then we have*

$$\frac{1}{n}\sum_{i=1}^{n} \psi'(X_i - M_n) \to \lambda'_F(\theta) \qquad [P]\ a.s. \tag{4.42}$$

This result is valid because of the assumed symmetry. Notice that this remark gives an estimator of $\lambda'_F(\theta)$ *for the Huber's score function.*

The main assertion in this section says that the bootstrap "works" in our situation.

THEOREM **4.3.3** *Let the assumptions of Section 4.1 be satisfied. Let* M_n^* *be the M-estimator based on the bootstrap sample from the empirical distribution* $F_n(x)$ *and let* M_n *be the M-estimator based on the random sample from the distribution* $F(x)$. *Then*

$$\lim_{n\to\infty} \sup_{x\in\Re} \left|P^*\left\{\sqrt{n}(M_n^* - M_n) < x\right\} - \Phi[x/\hat{\sigma}_n]\right| = 0 \qquad [P]\ a.s., \tag{4.43}$$

$$\lim_{n\to\infty} \sup_{x\in\Re} \left|P^*\left\{\sqrt{n}(M_n^* - M_n) < x\right\} - \Phi[x\lambda'_{F_\theta}(\theta)/S_n]\right| = 0 \qquad [P]\ a.s., \tag{4.44}$$

and

$$\lim_{n\to\infty} \sup_{x\in\Re} \left|P^*\left\{\sqrt{n}(M_n^* - M_n) < x\right\} - P\left\{\sqrt{n}(M_n - \theta) < x\right\}\right| = 0 \qquad [P]\ a.s., \tag{4.45}$$

where $\Phi(x)$ *denotes the distribution function of the standard normal distribution and where* $\hat{\sigma}_n^2$ *is defined by (4.38).*

PROOF: Let us investigate the behaviour of $P^*(\sqrt{n}(M_n^* - M_n) < x)$. First recall the definition (1.42) of $\lambda_{F_\theta}(.)$. It immediately follows from the definition of the $M-$estimators and from the assumed monotonicity of $\psi(x-t)$ that

$$\{\lambda_{F_n^*}(t) < 0\} \subseteq \{M_n^* \le t\} \subseteq \{\lambda_{F_n^*}(t) \le 0\}, \tag{4.46}$$

for all $t \in \Re$, where F_n^* denotes the empirical distribution function of $X_1^*, X_2^*, \ldots, X_n^*$, which implies that

$$\begin{aligned} &\left\{\lambda_{F_n^*}(M_n + z\hat{\sigma}_n n^{-1/2}) < 0\right\} \\ \subseteq &\left\{n^{1/2}(M_n^* - M_n) \le z\hat{\sigma}_n\right\} \\ \subseteq &\left\{\lambda_{F_n^*}(M_n + z\hat{\sigma}_n n^{-1/2}) \le 0\right\}. \end{aligned}$$

In order to simplify notation, we denote

$$T_{z,n} = M_n + z\hat{\sigma}_n n^{-1/2}. \tag{4.47}$$

We investigate the limit of

$$P^* \left\{\lambda_{F_n^*}(T_{z,n}) < 0\right\} = P^* \left\{\sqrt{n}\lambda_{F_n^*}(T_{z,n}) < 0\right\}$$

$$= P^* \left\{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i^* - T_{z,n}) < 0\right\}$$

$$= P^* \left\{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\psi(X_i^* - T_{z,n}) - \lambda_{F_n}(T_{z,n})}{\sqrt{Var^*\psi(X_1^* - T_{z,n})}} < -\frac{\sqrt{n}\lambda_{F_n}(T_{z,n})}{\sqrt{Var^*\psi(X_1^* - T_{z,n})}}\right\} \tag{4.48}$$

We start with separate treatment of the term

$$-\frac{\sqrt{n}\lambda_{F_n}(T_{z,n})}{\sqrt{Var^*\psi(X_1^* - T_{z,n})}}. \tag{4.49}$$

Concerning the numerator of (4.49), we have

$$\sqrt{n}\left(\lambda_{F_n}(T_{z,n}) - \lambda_{F_n}(M_n)\right) = \sqrt{n}\left(\lambda_{F_n}(M_n + z\hat{\sigma}_n n^{-1/2}) - \lambda_{F_n}(M_n)\right). \tag{4.50}$$

Now we use the Corollary 4.2.2 and the a.s. convergence of $\hat{\sigma}_n^2$ guaranteed by Theorem 4.3.2 to claim that the limit of (4.50) is equal to

$$z\sigma(\psi, F)\lambda_F'(\theta) \qquad [P]\ a.s. \tag{4.51}$$

It remains to find the limit of the denominator of (4.49)

$$Var^*\psi(X_1^* - T_{z,n})$$

$$= \frac{1}{n}\sum_{i=1}^{n} \psi^2(X_i - T_{z,n}) - \left[\frac{1}{n}\sum_{i=1}^{n} \psi(X_i - T_{z,n})\right]^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} \psi^2(X_i - T_{z,n}) - \left[\lambda_{F_n}(T_{z,n})\right]^2. \tag{4.52}$$

The first term on the right hand side of (4.52) has been already investigated in the Corollary 4.2.3 and it has been shown that it converges to $E\psi^2(X_1 - \theta)$, $[P]$ a.s. The second part on the right hand side of (4.52) converges to 0 (see formulas (4.50) and (4.51)). Therefore, we can write

$$\lim_{n\to\infty} Var^*\psi(X_1^* - T_{z,n}) = E\psi^2(X_1 - \theta) = \int_{\Re} \psi^2(x-\theta)dF(x-\theta) = [\lambda_F'(\theta)]^2\sigma^2(\psi, F) \quad [P]\ a.s. \tag{4.53}$$

Combining (4.51) and (4.53) (notice that $\lambda'_F(\theta)$ is negative) yields that

$$\lim_{n\to\infty} -\frac{\sqrt{n}\lambda_{F_n}(T_{z,n})}{\sqrt{Var^*\psi(X_1^*-T_{z,n})}} = z \qquad [P]\ a.s. \tag{4.54}$$

This gives that

$$\begin{aligned}
&|P^*\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\psi(X_i^*-T_{z,n})-\lambda_{F_n}(T_{z,n})}{\sqrt{Var^*\psi(X_1^*-T_{z,n})}} < -\frac{\sqrt{n}\lambda_{F_n}(T_{z,n})}{\sqrt{Var^*\psi(X_1^*-T_{z,n})}}\right\} \\
&\qquad -P^*\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\psi(X_i^*-T_{z,n})-\lambda_{F_n}(T_{z,n})}{\sqrt{Var^*\psi(X_1^*-T_{z,n})}} < z\right\}| \\
&\le P^*\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\psi(X_i^*-T_{z,n})-\lambda_{F_n}(T_{z,n})}{\sqrt{Var^*\psi(X_1^*-T_{z,n})}} \in \left(z-|z+\frac{\sqrt{n}\lambda_{F_n}(T_{z,n})}{\sqrt{Var^*\psi(X_1^*-T_{z,n})}}|,\right.\right. \\
&\qquad \left.\left. z+|z+\frac{\sqrt{n}\lambda_{F_n}(T_{z,n})}{\sqrt{Var^*\psi(X_1^*-T_{z,n})}}|\right)\right\}
\end{aligned} \tag{4.55}$$

This means that in order to establish the asymptotic properties of (4.48), it is sufficient to investigate the properties of

$$P^*\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\psi(X_i^*-T_{z,n})-\lambda_{F_N}(T_{z,n})}{\sqrt{Var^*\psi(X_1^*-T_{z,n})}} < z\right\}. \tag{4.56}$$

The summands in (4.56) are i.i.d. random variables with mean 0 and variance 1. We may use the Lindeberg–Feller Triangular Array Central Limit Theorem. Let us verify the Lindeberg's condition

$$\lim_{n\to\infty} \int\limits_{|\frac{\psi(x-T_{z,n})-\lambda_{F_n}(T_{z,n})}{\sqrt{Var^*\psi(X_1^*-T_{z,n})}}|>\varepsilon\sqrt{n}} \left[\frac{\psi(x-T_{z,n})-\lambda_{F_n}(T_{z,n})}{\sqrt{Var^*\psi(X_1^*-T_{z,n})}}\right]^2 dF_n(x) = 0 \tag{4.57}$$

or equivalently using (4.53)

$$\lim_{n\to\infty} \int\limits_{|\psi(x-T_{z,n})-\lambda_{F_n}(T_{z,n})|>\varepsilon\sqrt{n}} [\psi(x-T_{z,n})-\lambda_{F_n}(T_{z,n})]^2 dF_n(x) = 0. \tag{4.58}$$

It is easy to see that this condition is fulfilled, because the function ψ is bounded by the assumptions. Thus the Lindeberg's condition is fulfilled and

$$P^*\{\sqrt{n}(M_n^*-M_n) \le z\hat{\sigma}_n\} \to \Phi(z) \qquad [P]\ a.s.$$

for every z fixed. By the monotonicity of

$$P^*\{\sqrt{n}(M_n^*-M_n) \le z\hat{\sigma}_n\}$$

as a function of z together with the continuity, monotonicity and existence and boundedness of the derivative of its limit $\Phi(z)$ we get that

$$\lim_{n\to\infty} \sup_{x\in\Re} |P^*\{\sqrt{n}(M_n^* - M_n) \leq z\hat{\sigma}_n\} - \Phi(z)| = 0 \qquad [P]\ a.s. \tag{4.59}$$

and we obtain formula (4.43) if we put $x = z\hat{\sigma}_n$. □

Remark 2 *The estimate $\hat{\sigma}_n^2$ can be replaced by any estimate of the form*

$$S^2(\psi, F_n) = \frac{\int \psi^2(x-\theta)dF_n(x)}{\left[\hat{\lambda}_n'\right]^2}, \tag{4.60}$$

where $\hat{\lambda}_n'$ is some strongly consistent estimator of $\lambda_F'(\theta)$, see formula (4.39) above.

Remark 3 *Theorem 4.3.3 concerns the standardized bootstrap. A similar theorem for the studentized bootstrap can be proved only with minor modification of the proof of Theorem 4.3.3. We get that*

$$\lim_{n\to\infty} \sup_{x\in\Re} |P^*\left(\sqrt{n}\frac{M_n^* - M_n}{S(\psi, F_n^*)} < x\right) - \Phi(x)| = 0 \qquad [P]\ a.s., \tag{4.61}$$

where

$$S(\psi, F_n^*) = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^n [\psi(X_i^* - M_n^*)]^2}{\left[\hat{\lambda}_{F_n^*}'(M_n^*)\right]^2}}. \tag{4.62}$$

PROOF: Theorem 4.3.1 implies that

$$\lim_{n\to\infty} P^*\left\{|S(\psi, F_n^*) - \hat{\sigma}_n| > \varepsilon\right\} = 0 \qquad [P]\ a.s. \tag{4.63}$$

Thus we can write that

$$\lim_{n\to\infty} P^*\left(\sqrt{n}\frac{M_n^* - M_n}{S(\psi, F_n^*)} < x\right) = \lim_{n\to\infty} P^*\left(\sqrt{n}\frac{M_n^* - M_n}{\hat{\sigma}_n} < x\right), \tag{4.64}$$

notice that the properties of right hand side of (4.64) have been already established in Theorem 4.3.3.

□

Remark 4 *It is also easy to prove similar theorem for the studentized M-estimator. The studentization ensures that the M-estimator is not scale dependent. This is achieved by calculating the scale estimate, e.g. the Mean Absolute Deviation (MAD),*

$$S_{n,MAD} = MAD_n/\Phi^{-1}(0.75) = \mathrm{med}_i\left\{|X_i - \mathrm{med}_j(X_j)|\right\}/\Phi^{-1}(0.75) \tag{4.65}$$

and by defining the M-estimator as the solution of the studentized equation

$$\sum_{i=1}^n \psi\left(\frac{X_i - t}{S_{n,MAD}}\right) = 0. \tag{4.66}$$

The following two lemmas will be useful in the next section in the proof of Theorem 4.4.1 which will give the basic asymptotic properties of the robust three-stage procedure.

LEMMA **4.3.1** *Under the assumptions of Section 4.1, for every* $0 < c_1 < \infty$*,* $0 < \alpha < 1$*, and* $c_3 < t < \sqrt{m}c_4$*, we have*

$$P[P^*(\sqrt{m}|M_m^* - M_m| > t) > \alpha] \le 2\exp\left\{-\left[-\frac{\sqrt{-\log\alpha}}{\sqrt{2}\psi(h)} - \frac{\sqrt{c_2}}{2}t\right]^2\right\} + 4\exp\{-c_2t^2/4\},$$

where $c_2 = 2[f(h+c_1)]^2 > 0$*, and where the constant* $\psi(h)$ *is given by (4.9),*

$$c_3 = \frac{\sqrt{-\log\alpha}}{\sqrt{c_2/2}\psi(h)}, \tag{4.67}$$

and

$$c_4 = \min\left\{2c_1, \frac{\sqrt{2}c_1}{\sqrt{c_2}\psi(h)}\right\}. \tag{4.68}$$

PROOF:

$$P^*(\sqrt{m}|M_m^* - M_m| > t) = P^*(M_m^* > M_m + tm^{-1/2}) + P^*(M_m^* < M_m - tm^{-1/2}) \tag{4.69}$$

We investigate only the first term on the right hand side of (4.69), the treatment of the other term is similar and it gives the same result. Using the definition of the M-estimator and Theorem 2 of Hoeffding (1963) (see Lemma A.2.5 in Appendix) we have that

$$\begin{aligned}
&P^*(M_m^* > M_m + tm^{-1/2})\\
&\le P_m^*\left\{\frac{1}{m}\sum_{i=1}^m \psi(X_i^* - M_m^*) \le \frac{1}{m}\sum_{i=1}^m \psi(X_i^* - M_m - tm^{-1/2})\right\}\\
&= P_m^*\left\{\frac{1}{m}\sum_{i=1}^m \psi(X_i^* - M_m - tm^{-1/2}) \ge 0\right\}\\
&= P_m^*\left\{\frac{1}{m}\sum_{i=1}^m \left[\psi(X_i^* - M_m - tm^{-1/2}) - E_m^*\psi((X_i^* - M_m - tm^{-1/2})\right]\right.\\
&\qquad \left.\ge \left[-\frac{1}{\sqrt{m}}\sum_{i=1}^m \psi(X_i - M_m - tm^{-1/2})\right]/\sqrt{m}\right\}\\
&\le \exp\left\{-\left[\frac{1}{\sqrt{m}}\sum_{i=1}^m \psi(X_i - M_m - tm^{-1/2})\right]^2\right\},
\end{aligned}$$

for all $0 \le t \le c_1\sqrt{m}$. This means that

$$P[P^*(M_m^* > M_m + t) > \alpha]$$

$$
\begin{aligned}
&\le P\left[\exp\left\{-\left[\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\psi(X_i - M_m - tm^{-1/2})\right]^2\right\} > \alpha\right] \\
&= P\left[-\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\psi(X_i - M_m - tm^{-1/2}) < \sqrt{-\log\alpha}\right] \\
&\le P\left[\left\{-\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\psi(X_i - \theta - tm^{-1/2}/2) < \sqrt{-\log\alpha}\right\}\bigcap\left\{\sqrt{m}|M_m-\theta| < t/2\right\}\right] \\
&+P\left[\left\{-\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\psi(X_i - M_m - tm^{-1/2}) < \sqrt{-\log\alpha}\right\}\bigcap\left\{\sqrt{m}|M_m-\theta| \ge t/2\right\}\right] \\
&\le P\left[-\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\psi(X_i - \theta - tm^{-1/2}/2) < \sqrt{-\log\alpha}\right] + P\left[\sqrt{m}|M_m-\theta| \ge t/2\right] \qquad (4.70)
\end{aligned}
$$

Using Lemma 4.2.2, the second probability in (4.70) is bounded by

$$P\left[\sqrt{m}|M_m-\theta| \ge t/2\right] \le 2\exp\left\{-c_2t^2/4\right\}. \qquad (4.71)$$

for $0 < t < 2c_1\sqrt{m}$. It remains to investigate the behaviour of the first term in (4.70). The following inequality was established in the proof of Lemma 3.1 in Jurečková and Sen (1982):

$$-E\psi(X_1 - \theta - \frac{t}{2}m^{-1/2}) \ge \sqrt{2c_2}\psi(h)\frac{t}{2}m^{-1/2}. \qquad (4.72)$$

Using (4.72) and the Hoeffding inequality, we have that

$$
\begin{aligned}
&P\left[-\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\psi(X_i - \theta - \frac{t}{2}m^{-1/2}) < \sqrt{-\log\alpha}\right] \\
&= P\left[\frac{1}{m}\sum_{i=1}^{m}\psi(X_i - \theta - \frac{t}{2}m^{-1/2}) - E\psi(X_i - \theta - \frac{t}{2}m^{-1/2})\right. \\
&\qquad \left. > -\sqrt{-\frac{\log\alpha}{m}} - E\psi(X_i - \theta - \frac{t}{2}m^{-1/2})\right] \\
&\le P\left[\frac{1}{m}\sum_{i=1}^{m}\psi(X_i - \theta - \frac{t}{2}m^{-1/2}) - E\psi(X_i - \theta - \frac{t}{2}m^{-1/2}) > -\sqrt{-\frac{\log\alpha}{m}} + \sqrt{2c_2}\psi(h)\frac{t}{2}m^{-1/2}\right] \\
&\le \exp\left\{\frac{-2\left[-\sqrt{-\log\alpha} + \sqrt{2c_2}\psi(h)t/2\right]^2}{[2\psi(h)]^2}\right\} = \exp\left\{-\left[-\frac{\sqrt{-\log\alpha}}{\sqrt{2}\psi(h)} + \frac{\sqrt{c_2}}{2}t\right]^2\right\} \qquad (4.73)
\end{aligned}
$$

for all t such that

$$0 < t < 2c_1\sqrt{m}$$

and

$$0 < -\sqrt{-\log\alpha} + \sqrt{2c_2}\psi(h)t/2 < c_1\sqrt{m},$$

where h (and $\psi(h)$) are given by (4.9). Combining (4.69), (4.70), (4.71), and (4.73) leads the desired result.

□

LEMMA **4.3.2** *Under assumptions of Section 4.1, there exists m_0 such that for all $m > m_0$ and $t > 2hm^{1/2}$*

$$P\left(P^*\left\{\sqrt{m}|M_m^* - M_m| > t\right\} > \alpha\right) \le 3\left[4F(\theta - tm^{-1/2} - h)(1 - F(\theta - tm^{-1/2}/2 - h))\right]^m,$$

where h is given by (4.9).

PROOF: Using the technique of proof of Lemmas 3.2 and 3.3. in Jurečková and Sen (1982), we can write

$$\begin{aligned} P^*\left\{\sqrt{m}|M_m^* - M_m| > t\right\} &\le 2P^*\left\{X^*_{m,m/2+1} \ge M_m + tm^{-1/2} - h\right\} \\ &\le 2\left[4F_m(M_m + tm^{-1/2} - h)\{1 - F_m(M_m + tm^{-1/2} - h)\}\right]^m \end{aligned}$$

It follows that

$$\begin{aligned} &P\left(P^*\left\{\sqrt{m}|M_m^* - M_m| > t\right\} > \alpha\right) \\ &\le P\left(\left[4F_m(M_m + tm^{-1/2} - h)\{1 - F_m(M_m + tm^{-1/2} - h)\}\right]^m > \alpha/2\right) \\ &\le P\left[1 - F_m(M_m + tm^{-1/2} - h) > \frac{1}{4}\left(\frac{\alpha}{2}\right)^{1/m}\right] \\ &\le P\left[1 - F_m(\theta + tm^{-1/2}/2 - h) > \frac{1}{4}\left(\frac{\alpha}{2}\right)^{1/m}\right] + P\left[\sqrt{n}|M_m - \theta| \ge t/2\right] \\ &= P\left[F_m(\theta + tm^{-1/2}/2 - h) < 1 - \frac{1}{4}\left(\frac{\alpha}{2}\right)^{1/m}\right] + P\left[\sqrt{n}|M_m - \theta| \ge t/2\right] \end{aligned}$$

$$\le P\left[X_{(m:[m(1-\varepsilon)]^\circ)} > \theta + tm^{-1/2}/2 - h\right] + P\left[\sqrt{n}|M_m - \theta| \ge t/2\right] \tag{4.74}$$

for all $m > m_0$, where $0 < \varepsilon < 1/4$ and m_0 are such that

$$1 - \frac{1}{4}\left(\frac{\alpha}{2}\right)^{1/m_0} < 1 - \varepsilon.$$

Jurečková and Sen (1982) investigated the behaviour of the second probability in (4.74). They showed that, for $t > 2hm^{1/2}$, it is bounded by

$$2\left[4F(\theta - tm^{-1/2}/2 - h)(1 - F(\theta - tm^{-1/2}/2 - h))\right]^m. \tag{4.75}$$

Using their technique, we investigate also the first probability in (4.74). For the simplicity of notation, we denote

$$a = F(\theta + tm^{-1/2}/2 - h).$$

Notice that $a > 1/2$ for $t > 2hm^{1/2}$. Using Theorem 1 of Hoeffding (1963) with the bounds for $a > 1/2$, we have

$$P\left[X_{(m:[m(1-\varepsilon)]^\circ)} > a\right] \le P\left[\frac{1}{m}Bi(m,a) - a > 1 - \varepsilon\right]$$

$$\le \exp\left\{-\frac{m(1-\varepsilon)^2}{2a(1-a)}\right\} \le \left[\exp\{-1/4a(1-a)\}\right]^m \le [4a(1-a)]^m \tag{4.76}$$

Combining (4.74), (4.75), and (4.76) yields the lemma.

□

4.4 Three-stage Procedure Based on Bootstrap

For the sequential procedure, we need to estimate quantiles of the distribution of the parameter of interest. The easiest approach is to use the normal approximation. The asymptotic normal critical point $\xi^N(\alpha)$ is defined as the $1-\alpha/2$ quantile of the standard Normal distribution (i.e. $u_{1-\alpha/2}$ in the above used notation).

Using bootstrap approximations, we can estimate the quantiles of the distribution of the standardized (or studentized) M-estimator more closely. The standardized bootstrap critical point $\xi_m^A(\alpha)$ is defined as $1-\alpha$ quantile of the (centered and standardized) conditional distribution of

$$\sqrt{m}\frac{|M_m^* - M_m|}{\hat{\sigma}_m}, \tag{4.77}$$

where $\hat{\sigma}_m^2$ is the estimate of the asymptotic variance of the M-estimator which has been already defined in the formula (4.38).

The studentized bootstrap critical point $\xi_m^U(\alpha)$ is defined as $1-\alpha$ quantile of the (centered and studentized) conditional distribution of

$$\sqrt{m}\frac{|M_m^* - M_m|}{\hat{\sigma}_m^*}. \tag{4.78}$$

These critical points can be used to define the following three-stage procedure. In the first stage we fix the parameter which controls the sample size in the first stage $\gamma > 0$, and we draw

$$m = m(d) = \max\left\{2, \left[\left(\frac{u_{1-\alpha/2}}{d}\right)^{2/(1+\gamma)}\right]^\circ + 1\right\} \tag{4.79}$$

observations. These observations are used to determine $\hat{\sigma}_m^2$ and the standardized bootstrap critical points $\xi_m^A(\alpha)$ given by (4.77).

The intermediate sample size $N_1(d)$ based on the standardized bootstrap critical points is then given as

$$N_1(d) = \max\left\{m, \left[k\left(\frac{\xi_m^A(\alpha)\hat{\sigma}_m}{d}\right)^2\right]^\circ + 1\right\}, \tag{4.80}$$

where $0 < k < 1$ is the parameter controlling the sample size in this stage.

Finally, we draw next $N_2(d) - N_1(d)$ observations, where

$$N_2(d) = \max\left\{N_1(d), \left[\left(\frac{\xi_{N_1}^A(\alpha)\hat{\sigma}_{N_1}}{d}\right)^2\right]^\circ + 1\right\} \tag{4.81}$$

and obtain the $1-\alpha$ confidence interval $(M_{N_2} - d, M_{N_2} + d)$ based on the standardized bootstrap critical points.

The construction of fixed-width confidence intervals based on studentized bootstrap critical points (4.78) is analogous.

4.4.1 Basic Properties

Theorem 4.4.1 states basic asymptotic properties of the robust three-stage procedure (4.79)–(4.81) based on the standardized bootstrap critical points.

THEOREM **4.4.1** *Suppose that the conditions of Section 4.1 hold. Then the three-stage procedure based on bootstrapping M-estimators has following asymptotic properties:*

$$(i) \qquad \lim_{d\to 0+} N_2(d) = \infty \qquad [P] \text{ a.s.}, \tag{4.82}$$

$$(ii) \qquad \lim_{d\to 0+} \frac{N_2(d)}{c_M(d)} = 1 \qquad [P] \text{ a.s.}, \tag{4.83}$$

$$(iii) \qquad \lim_{d\to 0+} P(M_{N_2(d)} - d < \theta < M_{N_2(d)} + d) = 1 - \alpha, \tag{4.84}$$

$$(iv) \qquad \lim_{d\to 0+} E\left(\frac{N_2(d)}{c_M(d)}\right) = 1, \tag{4.85}$$

where $c_M(d)$ is the asymptotically optimal number of observations for fixed-width confidence intervals based on M-estimators given by (1.44).

PROOF: Denote by

$$\mathcal{L}_n^A(x) = P_n^*\left\{\sqrt{n}\frac{|M_n^* - M_n|}{S(\psi, F_n)} \le x\right\} \tag{4.86}$$

the distribution function of the absolute value of the standardized bootstrap statistics. By Theorem 4.3.3 we have that

$$\lim_{n\to\infty} \mathcal{L}_n^A(x) = \begin{cases} 2\Phi(x) - 1, & \text{if } x \ge 0 \\ 0 & \text{otherwise.} \end{cases} \qquad [P] \text{ a.s.} \tag{4.87}$$

Recall that $\xi_n^A(\alpha)$ denotes the $1-\alpha$ quantile of $\mathcal{L}_n^A$. By the strict monotonicity and continuity of $\Phi(x)$ and by e.g. Lemma 1.5.6 in Serfling (1980) we have for all $t \in (0,1)$

$$\lim_{n\to\infty} \xi_n^A(t) = u_{1-t/2} \qquad [P] \text{ a.s.} \tag{4.88}$$

where the symbol $u_{1-t/2}$ denotes the $1-t/2$ quantile of the standard normal distribution. Notice that $\lim_{d\to 0+} m(d) = \infty$ and therefore

$$\lim_{d\to 0+} \xi_{m(d)}^A(\alpha) = u_{1-\alpha/2} \qquad [P] \text{ a.s.} \tag{4.89}$$

This, together with consistence of $\hat\sigma^2_{m(d)}$ and the definition of $N_1(d)$ implies that

$$\lim_{d\to 0+} N_1(d) = \infty \qquad [P] \text{ a.s.} \tag{4.90}$$

which in turn implies that also

$$\lim_{d\to 0+} N_2(d) = \infty \qquad [P] \text{ a.s.} \tag{4.91}$$

The definition (4.81) of the stopping time $N_2(d)$ implies following inequalities.

$$\left[\frac{\xi_{N_1}^A(\alpha)\hat\sigma_{N_1}}{d}\right]^2 < N_2(d) \le \left[\frac{\xi_{N_1}^A(\alpha)\hat\sigma_{N_1}}{d}\right]^2 + 1$$

$$+N_1(d)I\left[(\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1})^2 \le k(\xi^A_m(\alpha)\hat{\sigma}_m)^2 + md^2 + d^2\right] \tag{4.92}$$

Next, by the strong consistency of $\xi^A_n\hat{\sigma}_n$ as an estimator of $u_{1-\alpha/2}\sigma(\psi, F)$ we have for every $\varepsilon > 0$

$$\lim_{m\to\infty} P(|\xi^A_n\hat{\sigma}_n - u_{1-\alpha/2}\sigma| < \varepsilon, \forall n > m) = 1, \tag{4.93}$$

where we abbreviate $\sigma(\psi, F)$ by σ. Let us fix $\delta > 0$. We can choose $d_0 > 0$ such that

$$P\left\{|(\xi^A_n(\alpha)S(\psi, F_n))^2 - (u_{1-\alpha/2}\sigma)^2| < \frac{(1-k)(u_{1-\alpha/2}\sigma)^2}{3}, \forall n > m(d_0)\right\} \ge 1-\delta \tag{4.94}$$

and

$$m(d_0)d_0^2 + d_0^2 \le \frac{(1-k)(u_{1-\alpha/2}\sigma)^2}{3}. \tag{4.95}$$

Now, using the facts that $N_1 \ge m(d)$ and $k \in (0,1)$ we have

$$\begin{aligned}
&P\left\{I\left[(\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1})^2 \le k(\xi^A_m(\alpha)\hat{\sigma}_m)^2 + md^2 + d^2\right] = 0, \forall d > d_0\right\}\\
&= P\left\{(\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1})^2 > k(\xi^A_m(\alpha)\hat{\sigma}_m)^2 + md^2 + d^2, \forall d > d_0\right\}\\
&\ge P\left\{(u_{1-\alpha/2}\sigma)^2 - \frac{(1-k)(u_{1-\alpha/2}\sigma)^2}{3} \ge k\left[(u_{1-\alpha/2}\sigma)^2 + \frac{(1-k)(u_{1-\alpha/2}\sigma)^2}{3}\right]\right.\\
&\left. + \frac{(1-k)(u_{1-\alpha/2}\sigma)^2}{3}\right\} \times P\left\{|(\xi^A_n S(\psi, F_n))^2 - (u_{1-\alpha/2}\sigma)^2| < \frac{(1-k)(u_{1-\alpha/2}\sigma)^2}{3}, \forall n > m(d_0)\right\}\\
&\ge 1-\delta
\end{aligned} \tag{4.96}$$

This says that

$$P\left\{I\left[(\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1})^2 \le k(\xi^A_m(\alpha)\hat{\sigma}_m)^2 + md^2 + d^2\right] = 0, \forall d < d_0\right\} \to 1$$

as $d_0 \to 0+$ and that is equivalent to

$$\lim_{d\to 0} I\left[(\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1})^2 \le k(\xi^A_m(\alpha)\hat{\sigma}_m)^2 + md^2 + d^2\right] = 0 \qquad [P]\ a.s.$$

which in turn implies that there exists d_0 such that for all $d > d_0$

$$I\left[(\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1})^2 \le k(\xi^A_m(\alpha)\hat{\sigma}_m)^2 + md^2 + d^2\right] = 0 \qquad [P]\ a.s. \tag{4.97}$$

Combining (4.92) and (4.97) leads part (ii) of the theorem.

Part (iii) of the theorem follows from the Slutzky Theorem, Anscombe Theorem and part (ii).

To verify part (iv) of the theorem it is sufficient to prove uniform integrability of the set $\{N_2(d)d^2\}_{d>0}$. It suffices to show that there exists $d_0 > 0$ such that

$$\sum_{l=1}^{\infty} \sup_{0<d<d_0} P\{N_2(d)d^2 > l\} < \infty. \tag{4.98}$$

We can choose d_0 such that for every $0 < d < d_0$ we have

$$\left(\left[\left(\frac{u_{1-\alpha/2}}{d}\right)^{2/(1+\gamma)}\right]^\circ + 1\right) d^2 \le l \tag{4.99}$$

and

$$2d^2 \le l \tag{4.100}$$

which implies that

$$\begin{aligned} &P(N_2(d)d^2 > l) \\ \le P&\left\{\left(\left[\frac{(\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1})^2}{d^2}\right]^\circ + 1\right) d^2 > l\right\} + P\left\{\left(\left[\frac{k(\xi^A_m(\alpha)\hat{\sigma}_m)^2}{d^2}\right]^\circ + 1\right) d^2 > l\right\} \\ &= P_{1,l}(d) + P_{2,l}(d). \end{aligned} \tag{4.101}$$

Let us first deal with the second probability.

$$\begin{aligned} P_{2,l}(d) &\le P\left\{\xi^A_m(\alpha)\hat{\sigma}_m > \sqrt{\frac{l-d^2}{k}}\right\} \le P\left\{\xi^A_m(\alpha)\hat{\sigma}_m > \sqrt{\frac{l}{2k}}\right\} \\ = P\left\{\mathcal{L}^A_m\left(\sqrt{\frac{l}{2k\hat{\sigma}_m}}\right) < 1-\alpha\right\} &= P\left\{P^A_m\left(|\sqrt{m}(M^*_m - M_m)| \le \sqrt{\frac{l}{2k}}\right) < 1-\alpha\right\} \\ &= P\left\{P^*_m\left(|\sqrt{m}(M^*_m - M_m)| > \sqrt{\frac{l}{2k}}\right) > \alpha\right\} \end{aligned} \tag{4.102}$$

Using Lemma 4.3.1 we have, for all $0 < d < d_0$,

$$\begin{aligned} \sum_{l=1}^{\infty} P_{2,l}(d) &= \sum_{l=1}^{\infty} P\left\{P^*_m\left(|\sqrt{m}(M^*_m - M_m)| > \sqrt{\frac{l}{2k}}\right) > \alpha\right\} \\ &= \sum_{l=1}^{c_3} P\left\{P^*_m\left(|\sqrt{m}(M^*_m - M_m)| > \sqrt{\frac{l}{2k}}\right) > \alpha\right\} \\ &+ \sum_{l=c_3+1}^{\sqrt{m}c_4} P\left\{P^*_m\left(|\sqrt{m}(M^*_m - M_m)| > \sqrt{\frac{l}{2k}}\right) > \alpha\right\} \\ &+ \sum_{l=\sqrt{m}c_4}^{\infty} P\left\{P^*_m\left(|\sqrt{m}(M^*_m - M_m)| > \sqrt{\frac{l}{2k}}\right) > \alpha\right\}, \end{aligned} \tag{4.103}$$

where c_3 and c_4 are given by (4.67) and (4.68), respectively. The first term on the right hand side of (4.103) is clearly finite (smaller than c_3), the finiteness of the second term follows from

Lemma 4.3.1. It remains to investigate the properties of the third term on the right hand side of (4.103). Using Lemma 4.3.2, we get that

$$\sum_{l=\sqrt{m}c_4}^{\infty} P\left\{P_m^*\left(|\sqrt{m}(M_m^* - M_m)| > \sqrt{\frac{l}{2k}}\right) > \alpha\right\}$$

$$\leq \sum_{l=\sqrt{m}c_4}^{\infty} 3\left[4F\left(\theta - \sqrt{\frac{2l}{km}} - h\right)\left(1 - F\left(\theta - \sqrt{\frac{2l}{km}} - h\right)\right)\right]^{,} \tag{4.104}$$

where h is given by (4.9). Notice that, for $t > 0$,

$$[F(\theta - t)(1 - F(\theta - t))]' = f(\theta - t)[2F(\theta - t) - 1] \leq 0.$$

Hence, it follows that the term inside the sum in (4.104) is non-increasing in l. This allows us to show the finiteness of the sum in (4.104) by showing the finiteness of the following integral for which we have

$$\int\limits_{c_5 m^{1/2}}^{\infty} \left\{4F\left(\theta - t\left(\frac{x}{m}\right)^{1/2} - h\right)\left[1 - F\left(\theta - t\left(\frac{x}{m}\right)^{1/2} - h\right)\right]\right\}^m dx$$

$$= \frac{2m}{t^2} \int\limits_{tc_5^{1/2} m^{1/4}}^{\infty} u\left\{4F\left(\theta - u - h\right)\left[1 - F\left(\theta - u - h\right)\right]\right\}^m du$$

$$\leq \frac{2}{t^2}(4c_l^*)^{1/l} m \left\{4F\left(\theta - h\right)\left[1 - F\left(\theta - h\right)\right]\right\}^{m-1-b}$$

$$\times \int\limits_{tc_5^{1/2} m^{1/4}}^{\infty} \left\{4F\left(\theta - u - h\right)\left[1 - F\left(\theta - u - h\right)\right]\right\}^b du, \tag{4.105}$$

where c_l^* is given by (4.6) and b is any number satisfying $b > 1/l$. Notice that $F(\theta - h) < 1/2$ which implies that

$$4F\left(\theta - h\right)\left[1 - F\left(\theta - h\right)\right] < 1.$$

It follows that

$$m\left\{4F\left(\theta - h\right)\left[1 - F\left(\theta - h\right)\right]\right\}^{m-1-b}$$

is uniformly bounded in m and tends to 0 as $m \to \infty$.

Finally,

$$\int\limits_{tc_5^{1/2} m^{1/4}}^{\infty} \left\{4F\left(\theta - u - h\right)\left[1 - F\left(\theta - u - h\right)\right]\right\}^b du < \infty$$

by condition (4.7). Hence, the sum in (4.104) is finite. It remains to investigate the convergence of

$$\sum_{l=1}^{\infty} \sup_{0<d<d_0} P_{1,l}(d).$$

Notice that, by definition, $N_1(d) \geq m(d)$ for all $d > 0$ which immediately implies that

$$\sup_{0<d<d_0} P\left(\xi^A_{N_1(d)}(\alpha)\hat{\sigma}_{N_1(d)} > \sqrt{\frac{l}{2}}\right) < \sup_{0<d<d_0} P\left(\xi^A_{m(d)}(\alpha)\hat{\sigma}_m > \sqrt{\frac{l}{2}}\right). \tag{4.106}$$

Notice that the term on the left hand side is $P_{1,l}(d)$ and the term on the right hand side is exactly $P_{2,l}(d)$, where k is set to 1. The convergence of the term on the right hand side of (4.106) has already been established. This proves the desired uniform integrability result and concludes the proof of the part (iv) of the theorem.

□

4.4.2 Alternative Procedure

We can simplify the procedure by omitting the denominator $\hat{\lambda}'_F(\theta)$ of the estimate of the asymptotic standard deviation $\hat{\sigma}_n$, and by standardizing the estimates only with the term S_n,

$$S_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \psi^2(X_i - M_n)} = \hat{\lambda}'_F(\theta)\hat{\sigma}_n,$$

see the formulas (4.30) and (4.38).

The advantage of this procedure would be its simplicity and greater speed of calculation. The disadvantage is that we cannot define, without knowledge of the value of the $\lambda'(F,\theta)$, similar procedure based only on quantiles of normal distribution. Thus, the method considered in this section cannot be compared with the method based only on normal critical points as directly as the methods considered in previous sections.

We define the standardized bootstrap critical point $\zeta^A_m(\alpha)$ as $1-\alpha$ quantile of the (centered and standardized) conditional distribution of

$$\sqrt{m}\frac{|M^*_m - M_m|}{S_m}. \tag{4.107}$$

The studentized bootstrap critical point $\zeta^U_m(\alpha)$ is defined as $1-\alpha$ quantile of the (centered and studentized) conditional distribution of

$$\sqrt{m}\frac{|M^*_m - M_m|}{S^*_m}. \tag{4.108}$$

These critical points can be used to define the following three-stage procedure analogously as in (4.81). In the first stage we draw

$$m = m(d) = \max\left\{2, \left[\left(\frac{u_{1-\alpha/2}}{d}\right)^{2/(1+\gamma)}\right]^\circ + 1\right\} \tag{4.109}$$

observations.

The intermediate sample size $N^s_1(d)$ based on the standardized bootstrap critical points is then given as

$$N^s_1(d) = \max\left\{m, \left[k\left(\frac{\zeta^A_m(\alpha)S_m}{d}\right)^2\right]^\circ + 1\right\}, \tag{4.110}$$

where $0 < k < 1$ is the parameter controlling the sample size in this stage.

Finally, we draw next $N_2^s(d) - N_1^s(d)$ observations, where

$$N_2^s(d) = \max\left\{N_1^s(d), \left[\left(\frac{\zeta^A_{N_1^s}(\alpha) S_{N_1^s}}{d}\right)^2\right]^\circ + 1\right\} \tag{4.111}$$

and obtain the $1-\alpha$ confidence interval $(M_{N_2^s} - d, M_{N_2^s} + d)$ based on the standardized bootstrap critical points. This procedure is not so complicated as the three-stage procedure defined in (4.81), because it doesn't involve estimation of the term $\lambda'_F(\theta)$.

In the same way, we can construct a fixed-width confidence intervals based on studentized bootstrap critical points.

From comparison of the definitions (4.107) of $\zeta^A_m(\alpha)$ and (4.77) of $\xi^A_m(\alpha)$ we obtain the following:

$$1-\alpha = P^*\left(\sqrt{m}\frac{|M^*_m - M_m|}{S_m} < \zeta^A_m(\alpha)\right) = P^*\left(\sqrt{m}\frac{|M^*_m - M_m|}{\hat{\sigma}_m} < \xi^A_m(\alpha)\right). \tag{4.112}$$

Recall that $\hat{\sigma}_m = S_m/\hat{\lambda}'_F(\theta)$. Thus

$$\zeta^A_m(\alpha) = \xi^A_m(\alpha)/\hat{\lambda}'_F(\theta). \tag{4.113}$$

It follows that the sample sizes $N_1(d)$ and $N_2(d)$ given by formulas (4.80) and (4.81), respectively, are equal to the sample sizes $N_1^s(d)$ and $N_2^s(d)$ which are given by formulas (4.110) and (4.111). The sequential procedure (4.81) based on critical points $\xi^A_n(\alpha)$ gives exactly the same results as the procedure (4.111) which is based on critical points $\zeta^A_n(\alpha)$ and which is computationally simpler.

The critical points $\zeta^A_m(\alpha)$ can be calculated more easily than the critical points $\xi^A_m(\alpha)$ because we do not have to calculate any estimate of $\lambda'_F(\theta)$ any more. Notice also that the estimate of $\lambda'_F(\theta)$ does not appear in the calculation of the stopping times in formulas (4.110) and (4.111). However, keeping in mind the formula (4.113), we see that the asymptotic properties of $\zeta^A_m(\alpha)$ still depend on the asymptotic behaviour of $\hat{\lambda}'_F(\theta)$.

The asymptotic properties of the three-stage procedure (4.111) are simple corollary of (4.113) and we state them in the following Theorem 4.4.2.

THEOREM **4.4.2** *Suppose that the conditions of Section 4.1 hold. Then the three-stage procedure based on bootstrapping M-estimators defined by (4.111) has the following asymptotic properties:*

$$(i) \qquad \lim_{d\to 0+} N_2^s(d) = \infty \qquad [P]\ a.s., \tag{4.114}$$

$$(ii) \qquad \lim_{d\to 0+} \frac{N_2^s(d)}{c_M(d)} = 1 \qquad [P]\ a.s., \tag{4.115}$$

$$(iii) \qquad \lim_{d\to 0+} P(M_{N_2^s(d)} - d < \theta < M_{N_2^s(d)} + d) = 1-\alpha, \tag{4.116}$$

$$(iv) \qquad \lim_{d\to 0+} E\left(\frac{N_2^s(d)}{c_M(d)}\right) = 1 \tag{4.117}$$

PROOF: The proof is corollary of (4.113) and of Theorem 4.4.1.

□

4.5 More Asymptotics for the Procedure Based on Sample Mean

In this section, we want to investigate the asymptotic behaviour of the three-stage procedure based on sample mean which was proposed by Aerts and Gijbels (1993) and which we described in Chapter 1 in Subsection 1.1.4. We will start with calculating the expected value of N_2 — the calculations will follow the proof of Theorem 6.3.1 in Ghosh, Mukhopadhyay, and Sen (1997). The speed of convergence of the bootstrap critical points, using the results of Hall (1992), will be established in Subsection 4.5.2. Finally, In Subsection 4.5.3 we derive the most important result in this section, the asymptotic distribution of $\sqrt{N_2}$.

The results obtained for the procedure based on the sample mean will be further generalized in Section 4.6 where we will investigate the asymptotic properties of the robust procedure.

4.5.1 Expected Value of N_2 — Procedure Based on Sample Mean

Here, we will calculate the expected value of the stopping time for the three-stage procedure based on sample mean and the critical points of normal distribution. We have already established in the proof of Theorem 4.4.1 that there exists d_0 such that for all $d > d_0$ we have

$$N_2 = \left[\frac{u^2_{1-\alpha/2}\hat{\sigma}^2_{N_1}}{d^2}\right]^\circ + 1.$$

Therefore it suffices to investigate the following expression

$$\begin{aligned}
E\left\{\left[\frac{u^2_{1-\alpha/2}\hat{\sigma}^2_{N_1}}{d^2}\right]^\circ + 1\right\} &= E\left\{\frac{u^2_{1-\alpha/2}\hat{\sigma}^2_{N_1}}{d^2} + 1 + \left[\frac{u^2_{1-\alpha/2}\hat{\sigma}^2_{N_1}}{d^2}\right]^\circ - \frac{u^2_{1-\alpha/2}\hat{\sigma}^2_{N_1}}{d^2}\right\} \\
&= \frac{u^2_{1-\alpha/2}}{d^2}E\hat{\sigma}^2_{N_1} + 1 + E\left\{\left[\frac{u^2_{1-\alpha/2}\hat{\sigma}^2_{N_1}}{d^2}\right]^\circ - \frac{u^2_{1-\alpha/2}\hat{\sigma}^2_{N_1}}{d^2}\right\} \\
&= \frac{u^2_{1-\alpha/2}}{d^2}E\frac{1}{N_1}\sum_{i=1}^{N_1}[(X_i-\theta)^2 - (\bar{X}_{N_1}-\theta)^2] + 1 + E\left\{\left[\frac{u^2_{1-\alpha/2}\hat{\sigma}^2_{N_1}}{d^2}\right]^\circ - \frac{u^2_{1-\alpha/2}\hat{\sigma}^2_{N_1}}{d^2}\right\}
\end{aligned}$$

We have that

$$\lim_{d\to 0} E\left\{1 + \left[\frac{u^2_{1-\alpha/2}\hat{\sigma}^2_{N_1}}{d^2}\right]^\circ - \frac{u^2_{1-\alpha/2}\hat{\sigma}^2_{N_1}}{d^2}\right\} = 1/2,$$

see Ghosh, Mukhopadhyay, and Sen (1997), proof of Theorem 6.3.2. For the remaining part, notice that using Taylor expansion of N_1^{-1} around $[kc(d)]^{-1}$ we have

$$N_1^{-1} = [kc(d)]^{-1} - [kc(d)]^{-1}\left(\frac{\hat{\sigma}^2_m}{\sigma^2} - 1\right) + o_P(d^2). \tag{4.118}$$

Hence,

$$\begin{aligned}
E\frac{1}{N_1}\sum_{i=1}^{N_1}(X_i-\theta)^2 &= \left(E\frac{1}{N_1}\sum_{i=1}^{m(d)}(X_i-\theta)^2 + E\left\{\left[\frac{1}{N_1}\sum_{i=m(d)+1}^{N_1}(X_i-\theta)^2\right]|X_1,\ldots,X_m\right\}\right) \\
&= E\left\{\frac{1}{N_1}\sum_{i=1}^{m}(X_i-\theta)^2 + \frac{N_1-m}{N_1}\sigma^2\right\} = \sigma^2 + \sigma^2 E\left\{\frac{1}{N_1}\sum_{i=1}^{m}\frac{(X_i-\theta)^2}{\sigma^2} - \frac{m}{N_1}\right\}
\end{aligned}$$

$$= \sigma^2 + \sigma^2 E\left\{[kc(d)]^{-1}\left(2 - \frac{\hat{\sigma}_m^2}{\sigma^2}\right)\sum_{i=1}^{m}\left(\frac{(X_i-\theta)^2}{\sigma^2} - 1\right) + o(d^2)\right\}$$

$$= \sigma^2 + \frac{\sigma^2}{kc(d)}E\left\{\left(2 - \frac{\hat{\sigma}_m^2}{\sigma^2}\right) m\left(\frac{\hat{\sigma}_m^2}{\sigma^2} - 1\right)\right\} + o(d^2)$$

$$= \sigma^2 + \frac{m}{kc(d)}E(\sigma^2 + (\sigma^2 - \hat{\sigma}_m^2))(\hat{\sigma}_m^2 - \sigma^2) + o(d^2) = \sigma^2 - \frac{mE(\hat{\sigma}_m^2 - \sigma^2)^2}{kc(d)} + o(d^2). \quad (4.119)$$

Notice that by Taylor expansion we have

$$N_1^{-2} = [kc(d)]^{-2} - 2[kc(d)]^{-2}\left(\frac{\hat{\sigma}_m^2}{\sigma} - 1\right) + o_P(d^2).$$

Hence, denoting by $\bar{X}_{m+1:N_1}$ the sample mean of $X_{m+1},\ldots,X_{N_1}$, i.e., $\bar{X}_{m+1:N_1} = [1/(N_1 - m)]\sum_{i=m+1}^{N_1} X_i$, we may write

$$E\left\{\frac{1}{N_1}\sum_{i=1}^{N_1}(\bar{X}_{N_1} - \theta)^2\right\} = E(\bar{X}_{N_1} - \theta)^2 = E\left\{\frac{m}{N_1}(\bar{X}_m - \theta) + \frac{N_1 - m}{N_1}(\bar{X}_{m+1:N_1} - \theta)\right\}^2$$

$$= E\left(E\left\{\frac{m}{N_1}(\bar{X}_m - \theta) + \frac{N_1 - m}{N_1}(\bar{X}_{m+1:N_1} - \theta)\right\}^2 |X_1,\ldots,X_m\right)$$

$$= E\left(\frac{1}{N_1^2}E\left\{m^2(\bar{X}_m - \theta)^2 + 2m(N_1 - m)(\bar{X}_m - \theta)(\bar{X}_{m+1:N_1} - \theta)\right.\right.$$

$$\left.\left. + (N_1 - m)^2(\bar{X}_{m+1:N_1} - \theta)^2\right\} |X_1,\ldots,X_m\right)$$

$$= E\left\{\frac{1}{N_1^2}\left[m^2(\bar{X}_m - \theta)^2 + (N_1 - m)\sigma^2\right]\right\} = E\left\{\frac{\sigma^2}{N_1} + \frac{1}{N_1^2}\left[m^2(\bar{X}_m - \theta)^2 - m\sigma^2\right]\right\}$$

$$= E\left\{\sigma^2[kc(d)]^{-1}\left(2 - \frac{\hat{\sigma}_m^2}{\sigma^2}\right) + [kc(d)]^{-2}\left(3 - 2\frac{\hat{\sigma}_m^2}{\sigma}\right)\left[m^2(\bar{X}_m - \theta)^2 - m\sigma^2\right] + o(d^2)\right\}$$

$$= \frac{\sigma^2}{kc(d)} + o(d^2). \quad (4.120)$$

Combining (4.119) and (4.120) yields that

$$E\frac{1}{N_1}\sum_{i=1}^{N_1}(X_i - \bar{X}_{N_1})^2 = \sigma^2 - \frac{\sigma^2 + mE(\hat{\sigma}_m^2 - \sigma^2)^2}{kc(d)} + o(d^2).$$

and it follows that

$$EN_2 = \frac{u_{1-\alpha/2}^2}{d^2}\left[\sigma^2 - \frac{\sigma^2 + mE(\hat{\sigma}_m^2 - \sigma^2)^2}{kc(d)}\right] + \frac{1}{2} + o(1)$$

$$= c(d) - \frac{1 + mE\left(\hat{\sigma}_m^2/\sigma^2 - 1\right)^2}{k} + \frac{1}{2} + o(1). \quad (4.121)$$

THEOREM **4.5.1** *Assume that X_i's are i.i.d. random variables with variance $0 < \sigma^2 < \infty$. Then we have for the three-stage procedure defined by (1.22)–(1.24) that*

$$EN_2 = c(d) - \frac{1 + mE\left(\hat{\sigma}_m^2/\sigma^2 - 1\right)^2}{k} + \frac{1}{2} + o(1).$$

PROOF: The proof is given earlier in this subsection.

□

4.5.2 Bootstrap Critical Points — Procedure Based on Sample Mean

The asymptotic behaviour of the bootstrap critical points can be established using the Edgeworth expansions (see Theorems A.3.2 and A.3.3 in the Appendix). The Edgeworth expansion for the distribution of the sample mean (Hall 1992) can be written in the following way.

Assume that $X_1, \dots, X_n$ are i.i.d. random variables such that $E|X_i|^{\nu+2} < \infty$ and the characteristic function $\chi(t)$ of X_i satisfies Cramer's condition, i.e.,

$$\lim_{t\to\infty} \sup |\chi(t)| < 1. \tag{4.122}$$

Then

$$P\left\{\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \le x\right\} = \Phi(x) + \sum_{j=1}^{\nu} n^{-j/2}\pi_j(x)\varphi(x) + o(n^{-\nu/2}) \tag{4.123}$$

uniformly in x where π_j is polynomial of degree of $3j-1$ with coefficients depending on moments of X_i up to order $j+2$.

The Edgeworth expansion for the bootstrap distribution (together with the rate of convergence) has been established in Theorems 5.1 and 5.2 in Hall (1992). We present here the versions of Hall's theorems which are useful in our situation. We denote by $\bar{X}_n$ and S_n^2 the sample mean and the sample variance of $X_1, \dots, X_n$, respectively.

THEOREM **4.5.2** *Let $\lambda > 0$ be given, and let $l = l(\lambda)$ denote a sufficiently large positive number. Assume that $E|X|^l < \infty$ and that Cramer's condition holds. Let $\hat{\pi}_j$ be the polynomials obtained from π_j in (4.123) by replacing population moments by the corresponding sample moments. Then there exists a constant C such that*

$$P\left\{\sup_{-\infty<x<\infty} \left|P^*\left(\sqrt{n}\frac{\bar{X}_n^* - \bar{X}_n}{S_n} \le x\right) - \Phi(x) - \sum_{j=1}^{\nu} n^{-j/2}\hat{\pi}_j(x)\varphi(x)\right| > Cn^{-(\nu+1)/2}\right\} = O(n^{-\lambda}) \tag{4.124}$$

and, $[P]$ almost surely,

$$\sup_{-\infty<x<\infty} \left|P^*\left(\sqrt{n}\frac{\bar{X}_n^* - \bar{X}_n}{S_n} \le x\right) - \Phi(x) - \sum_{j=1}^{\nu} n^{-j/2}\hat{\pi}_j(x)\varphi(x)\right| = O(n^{-(\nu+1)/2}). \tag{4.125}$$

PROOF: The first part of the theorem is (simplification of) Theorem 5.1 in Hall (1992), the second part follows from the first part.

□

Theorem 4.5.2 can be restated in terms of the bootstrap critical points defined in Section 4.4 as follows.

THEOREM **4.5.3** *Under the conditions of Theorem 4.5.2, and for each $\delta, \lambda > 0$, there exist constants $C, \varepsilon > 0$ such that*

$$P\left\{\sup_{n^{-\varepsilon}\leq\alpha\leq 1-n^{-\varepsilon}} |\xi_n^A(\alpha) - u_{1-\alpha/2} - \sum_{j=1}^{\nu} n^{-j/2}\hat{\pi}_j(u_{1-\alpha/2})| > Cn^{-(\nu+1)/2}\right\} = O(n^{-\lambda}) \quad (4.126)$$

and, $[P]$ almost surely,

$$\sup_{n^{-\varepsilon}\leq\alpha\leq 1-n^{-\varepsilon}} |\xi_n^A(\alpha) - u_{1-\alpha/2} - \sum_{j=1}^{\nu} n^{-j/2}\hat{\pi}_j(u_{1-\alpha/2})| = O(n^{-(\nu+1)/2}) \quad (4.127)$$

PROOF: See Theorem 5.2 in Hall (1992).

□

We want to show that Theorem 4.5.3 is valid also when we replace the index n by random variable (stopping time). For this purpose, we present the following Lemma 4.5.1.

LEMMA **4.5.1** *Assume that X_n is a sequence of random variables such that*

$$X_n = X + O\left(\frac{1}{n}\right) \qquad [P] \text{ a.s.}$$

and that U_n is integer valued sequence of random variables with the property

$$\frac{U_n}{n} \to 1 \qquad [P] \text{ a.s. as } n \to \infty.$$

Then also

$$X_{U_n} = X + O\left(\frac{1}{n}\right) \qquad [P] \text{ a.s.}$$

PROOF: The almost sure convergence of the sequence X_n is equivalent to (Serfling 1980)

$$\lim_{n\to\infty} P\left(|X_m - X| < \frac{\varepsilon}{m}, \forall m \geq n\right) = 1. \quad (4.128)$$

Similarly,

$$\lim_{n\to\infty} P\left(|\frac{U_m}{m} - 1| < \delta, \forall m \geq n\right) = 1. \quad (4.129)$$

In order to finish the proof of the lemma, we have to show that

$$\lim_{n\to\infty} P\left(|X_{U_m} - X| < \frac{\varepsilon}{m}, \forall m \geq n\right) = 1.$$

We have

$$\begin{aligned} &\liminf_{n\to\infty} P\left(|X_{U_m} - X| < \frac{\varepsilon}{m}, \forall m \geq n\right) \\ &= \liminf_{n\to\infty} \left[P\left(\left\{|X_{U_m} - X| < \frac{\varepsilon}{m}, \forall m \geq n\right\} \cap \{|U_m - m| \leq m\delta, \forall m \geq n\}\right)\right. \\ &\quad \left. + P\left(\left\{|X_{U_m} - X| < \frac{\varepsilon}{m}, \forall m \geq n\right\} \cap \{|U_m - m| > m\delta, \forall m \geq n\}\right)\right] \end{aligned}$$

$$\geq \liminf_{n\to\infty} \left[P\left(\left\{ |X_{U_m} - X| < \frac{\varepsilon}{m}, \forall m \geq n \right\} \mid \{|U_m - m| \leq m\delta, \forall m \geq n\} \right) \right.$$

$$\left. \times P(|U_m - m| \leq m\delta, \forall m \geq n) \right]$$

$$\geq \liminf_{n\to\infty} \left[P\left(|X_m - X| < \frac{\varepsilon}{m + m\delta}, \forall m \geq n - m\delta \right) P(|U_m - m| \leq m\delta, \forall m \geq n) \right]$$

$$= \liminf_{n\to\infty} \left[P\left(|X_m - X| < \frac{\varepsilon/(1+\delta)}{m}, \forall m \geq n/(1+\delta) \right) P\left(|\frac{U_m}{m} - 1| \leq \delta, \forall m \geq n \right) \right] = 1,$$

as $n \to \infty$, by (4.128) and (4.129).

□

Now we are ready to establish the rate of convergence of the bootstrap critical points based on random number of observations.Recall that $N_1(d)$ denotes the intermediate sample size and that $\xi_n^A(\alpha)$ denotes the bootstrap critical points (1.29).

THEOREM **4.5.4** *Assume that X_i's are i.i.d. random variables with variance $0 < \sigma^2 < \infty$. Consider the three-stage procedure based on bootstrap for sample mean described in Subsection 1.1.4 and assume that the assumptions of Theorem 4.5.2 are satisfied. Then we have*

$$\xi^A_{N_1(d)}(\alpha) = \xi_{[kc(d)]^\circ+1}(\alpha) + O(d^2).$$

PROOF: The theorem follows immediately from the property *(ii)* in Theorem 1.1.6 restated in terms of N_1, Theorem 4.5.3, and Lemma 4.5.1.

□

4.5.3 Asymptotic Distribution of N_2 — Procedure Based on Sample Mean

The next two theorems concern the asymptotic properties of the procedure of Aerts and Gijbels (1993) described in Subsection 1.1.4.

THEOREM **4.5.5** *Assume that X_i's are i.i.d. random variables with variance $0 < \sigma^2 < \infty$ and such that $EX_i^4 < \infty$. Then, as $d \to 0+$,*

$$\sqrt{N_1} \left[\frac{1}{N_1} \sum_{i=1}^{N_1} (X_i - \bar{X}_{N_1})^2 - \sigma^2 \right] \xrightarrow{\mathcal{D}} N\left(0, E(X_i - EX_i)^4 - \sigma^4\right).$$

PROOF: It is a consequence of Anscombe Theorem A.4.2.

□

The main assertion in this section is the following Theorem 4.5.6 which gives the asymptotic distribution of the square root of the stopping time for the procedure based on sample mean.

THEOREM **4.5.6** *Consider the three-stage procedure based on a sample mean described in Subsection 1.1.4. Under the assumptions of Theorems 4.5.5 and 4.5.4, we have that, as $d \to 0+$,*

$$\sqrt{N_2(d)} - \sqrt{c(d)} \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{4k}\left[\frac{E(X_i - EX_i)^4}{\sigma^4} - 1\right]\right).$$

PROOF: Using the definition of $N_2(d)$ in Section 1.1.4, we can rewrite $N_2(d)$ equivalently in the following way

$$N_2(d) = \left(\frac{\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1}}{d}\right)^2 + \left[\left(\frac{\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1}}{d}\right)^2 - \left(\frac{\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1}}{d}\right)^2 + 1\right]^\circ$$

$$+ \left\{\max(N_1(d), m(d)) - \left[\left(\frac{\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1}}{d}\right)^2\right]^\circ\right\} I\left[\{N_1(d) > N_2(d)\} \cup \{m(d) > N_2(d)\}\right]. \tag{4.130}$$

The convergence in distribution

$$\left[\left(\frac{\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1}}{d}\right)^2\right]^\circ - \left(\frac{\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1}}{d}\right)^2 + 1 \xrightarrow{\mathcal{D}} U(0,1) \qquad \text{as } d \to 0+, \tag{4.131}$$

where $U(0,1)$ denotes the Uniform distribution on interval $(0,1)$, has been established e.g. in Ghosh, Mukhopadhyay, and Sen (1997) or Hall (1981). Similarly as in the proof of part (ii) of Theorem 4.4.1, we can also show that as $d \to 0+$, $[P]$ almost surely,

$$\left\{\max(N_1(d), m(d)) - \left[\left(\frac{\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1}}{d}\right)^2\right]^\circ\right\} I\left[\{N_1(d) > N_2(d)\} \cup \{m(d) > N_2(d)\}\right] \to 0. \tag{4.132}$$

From (4.130), (4.131), and (4.132) it is easy to see that

$$\sqrt{N_2} = \frac{\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1}}{d} + O_P(1),$$

and therefore it remains to find out the asymptotic distribution of the term

$$\frac{\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1}}{d}. \tag{4.133}$$

Using simple algebra, Theorem 4.5.4, and Theorem 4.5.5 we have the following

$$P\left\{\frac{\xi^A_{N_1}(\alpha)\hat{\sigma}_{N_1}}{d} - \frac{u_{1-\alpha/2}\sigma}{d} < x\right\}$$

$$P\left\{\frac{(\xi^A_{N_1} - u_{1-\alpha/2} + u_{1-\alpha/2})(\alpha)\hat{\sigma}_{N_1}}{d} - \frac{u_{1-\alpha/2}\sigma}{d} < x\right\} = P\left\{u_{1-\alpha/2}\frac{\hat{\sigma}_{N_1} - \sigma}{d} < x + o(1)\right\}$$

$$= P\left\{u_{1-\alpha/2}\frac{\sqrt{N_1}\,(\hat{\sigma}_{N_1} - \sigma)}{\sqrt{N_1}d} < x + o(1)\right\} = P\left\{\frac{\sqrt{N_1}\,(\hat{\sigma}_{N_1} - \sigma)}{\sqrt{k}\sigma}\frac{(\hat{\sigma}_{N_1} + \sigma)}{(\hat{\sigma}_{N_1} + \sigma)} < x + o(1)\right\}$$

$$= P\left\{\frac{\sqrt{N_1}\,(\hat{\sigma}^2_{N_1} - \sigma^2)}{2\sqrt{k}\sigma^2} < x + o(1)\right\} = \Phi\left(2x\sqrt{k}\sigma^2/\sqrt{E(X_i - EX_i)^4 - \sigma^4}\right) + o(1)$$

It follows that
$$\frac{\xi^A_{N_1}(\alpha)\hat\sigma_{N_1}}{d} \xrightarrow{\mathcal{D}} N\left(\sqrt{c(d)}, \frac{E(X_i - EX_i)^4 - \sigma^4}{4k\sigma^4}\right).$$

□

Corollary **4.5.1** *Under the assumptions of Theorem 4.5.6, for the three-stage procedure based on sample mean, we have that*
$$\sqrt{N_2(d)} = \sqrt{c(d)} + O_P(1).$$

4.6 More Asymptotics for the Robust Procedure

In this section, we will investigate the asymptotics for the stopping time N_2 for the robust three stage procedure. We will use similar tools as in the previous section, where we established the asymptotic distribution of N_2 for the three-stage procedure based on sample mean.

Some properties of the random variable $N_2(d)$ were already established in Theorem 4.4.1. Now, we will concentrate on the asymptotic distribution of $N_2(d)$. Similarly as in the previous section it follows that it is sufficient to investigate the asymptotic behaviour of
$$\frac{\xi^A_{N_1}\hat\sigma_{N_1}}{d}.$$

In the following subsection, we will establish the asymptotic distribution of the estimate $\hat\sigma^2_{N_1}$, see formula (4.38), of the asymptotic variance of the M-estimator.

4.6.1 Distribution of the Variance Estimator — Robust Procedure

The asymptotic distribution of the estimator $\hat\sigma^2_{N_1}$ can be established using results of Jurečková and Sen (1981). We start with investigating the asymptotic properties of the numerator S_n^2 of $\hat\sigma^2_n$,
$$S_n^2 = \frac{1}{n}\sum_{i=1}^n \psi^2(X_i - M_n),$$
see formulas (4.30) and (4.38).

Lemma **4.6.1** *Under the assumptions of Section 4.1, we have for every* $0 < \varepsilon_1 < \varepsilon_2 < \infty$ *that*
$$\max_{n\varepsilon_1 \le m \le n\varepsilon_2} \sqrt{m}\left|\frac{1}{m}\sum_{i=1}^m \psi^2(X_i - M_m) - \frac{1}{m}\sum_{i=1}^m \psi^2(X_i - \theta)\right| \xrightarrow{P} 0 \qquad \text{as } n \to \infty$$

Proof: See Lemma 3.1 in Jurečková and Sen (1981).

□

In order to establish the asymptotic distribution of S_{N_1}, it is helpful to introduce the notion of the uniform continuity in probability, see Definition A.4.1. Theorem A.4.1 allows us to investigate the properties of randomly stopped sequences of random variables.

The uniform continuity in probability for the sequence $\sqrt{n}S_n^2$ is established in the following Lemma 4.6.2. This lemma allows us to apply Theorem A.4.1 in our situation.

Lemma **4.6.2** *Under the assumptions of Section 4.1, the sequences of random variables*

$$\sqrt{n}\left[Z_n^2 - E\psi^2(X_i-\theta)\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\psi^2(X_i-\theta) - E\psi^2(X_i-\theta)\right] \tag{4.134}$$

and

$$\sqrt{n}\left[S_n^2 - E\psi^2(X_i-\theta)\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\psi^2(X_i-M_n) - E\psi^2(X_i-\theta)\right] \tag{4.135}$$

are uniformly continuous in probability (u.c.i.p.).

Proof: In view of Definition A.4.1, we have to show that for each $\varepsilon > 0$ there exists $\delta > 0$ such that for all $n \geq 1$ we have

$$P\left\{\max_{n<m\leq n(1+\delta)}\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left[\psi^2(X_i-\theta) - E\psi^2(X_i-\theta)\right]\right.\right.$$

$$\left.\left. -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\psi^2(X_i-\theta) - E\psi^2(X_i-\theta)\right]\right| > \varepsilon\right\} < \varepsilon. \tag{4.136}$$

For the sake of simplicity, let's use the notation

$$U_i = \psi^2(X_i-\theta) - E\psi^2(X_i-\theta) = \psi^2(X_i-\theta) - E\psi^2(X_1-\theta).$$

Then we can bound the probability on the left-hand side of (4.136) as follows:

$$P\left\{\max_{n<m\leq n(1+\delta)}\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m}U_i - \frac{1}{\sqrt{n}}\sum_{j=1}^{n}U_j\right| > \varepsilon\right\}$$

$$\leq P\left\{\max_{n<m\leq n(1+\delta)}\sqrt{\frac{m}{n}}\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m}U_i - \frac{1}{\sqrt{n}}\sum_{j=1}^{n}U_j\right| > \varepsilon\right\}$$

$$= P\left\{\max_{n<m\leq n(1+\delta)}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{m}U_i - \sqrt{\frac{m}{n}}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}U_j\right| > \varepsilon\right\}$$

$$= P\left\{\max_{n<m\leq n(1+\delta)}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i + \frac{1}{\sqrt{n}}\sum_{i=n+1}^{m}U_i - \sqrt{\frac{m}{n}}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}U_j\right| > \varepsilon\right\}$$

$$\leq P\left\{\max_{n<m\leq n(1+\delta)}\left(\left[\sqrt{\frac{m}{n}}-1\right]\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}U_j\right| + \left|\frac{1}{\sqrt{n}}\sum_{i=n+1}^{m}U_i\right|\right) > \varepsilon\right\}$$

$$\leq P\left\{\max_{n<m\leq n(1+\delta)}\left[\sqrt{\frac{m}{n}}-1\right]\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}U_j\right| > \varepsilon/2\right\} + P\left\{\max_{n<m\leq n(1+\delta)}\left|\frac{1}{\sqrt{n}}\sum_{i=n+1}^{m}U_i\right| > \varepsilon/2\right\} \tag{4.137}$$

Using Tchebyschev's inequality (see Lemma A.2.3) we have for the first probability on the right-hand side of (4.137) that

$$P\left\{\max_{n<m\le n(1+\delta)}\left[\sqrt{\frac{m}{n}}-1\right]\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}U_j\right|>\varepsilon/2\right\}\le P\left\{\frac{\sqrt{n(1+\delta)}-\sqrt{n}}{\sqrt{n}}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}U_j\right|>\varepsilon/2\right\}$$

$$=P\left\{\left|\sum_{j=1}^{n}U_j\right|>\frac{\varepsilon\sqrt{n}}{2\left(\sqrt{1+\delta}-1\right)}\right\}\le\frac{4}{\varepsilon^2}\left(\sqrt{1+\delta}-1\right)^2 Var\,\psi^2(X_i-\theta)<\varepsilon/2 \tag{4.138}$$

if we choose $\delta>0$ sufficiently small. It remains to investigate the second probability on the right-hand side of (4.137). Using Kolmogorov inequality (see Lemma A.2.4) we obtain that

$$P\left\{\max_{n<m\le n(1+\delta)}\left|\frac{1}{\sqrt{n}}\sum_{i=n+1}^{m}U_i\right|>\varepsilon/2\right\}=P\left\{\max_{1<k\le n\delta}\left|\sum_{j=1}^{k}U_{n+j}\right|>\sqrt{n}\varepsilon/2\right\}$$

$$\le\frac{4}{\varepsilon^2 n}n\delta\,Var\,\psi^2(X_1-\theta)=\frac{4\delta}{\varepsilon}Var\,\psi^2(X_1-\theta)<\varepsilon/2 \tag{4.139}$$

for $\delta>0$ small enough. Combining (4.137) with (4.138) and (4.139) proves the first part of the theorem.

The uniform continuity in probability of the sequence $\sqrt{n}\left[S_n^2-E\psi^2(X_i-\theta)\right]$ follows immediately from the uniform continuity in probability of the sequence $\sqrt{n}\left[Z_n^2-E\psi^2(X_i-\theta)\right]$ and from Lemma 4.6.1.

□

LEMMA 4.6.3 *Under the assumptions of Section 4.1 we have for the intermediate sample size $N_1(d)$ defined in (4.80) that*

$$\lim_{d\to 0+}\frac{N_1(d)}{kc_M(d)}=1\qquad [P]\ a.s.,$$

where $c_M(d)$ denotes the asymptotically optimal sample size (1.44) and where k is the parameter controling the intermediate sample size $N_1(d)$.

PROOF: The proof is exactly the same as the proof of the equivalent property for $N_2(d)$, see Theorem 4.4.1 and the inequalities in (4.92).

□

LEMMA 4.6.4 *Assume that the conditions of Section 4.1 hold and that N_1 is the intermediate sample size defined in (4.80). Then*

$$\frac{1}{\sqrt{N_1(d)}}\sum_{i=1}^{N_1(d)}\left(\psi^2(X_i-\theta)-\int\psi^2(x-\theta)dF(x)\right)$$

$$\xrightarrow{\mathcal{D}} N\left(0,\int\psi^4(x-\theta)dF(x)-\left[\int\psi^2(x-\theta)dF(x)\right]^2\right)\qquad as\ d\to 0+\,.$$

PROOF: The lemma follows immediately from Anscombe Theorem A.4.2 and from Lemma 4.6.3.
□

THEOREM **4.6.1** *Assume that the conditions of Section 4.1 hold and that N_1 is the intermediate sample size defined in (4.80). Then*

$$\frac{1}{\sqrt{N_1(d)}}\sum_{i=1}^{N_1(d)}\left(\psi^2(X_i - M_{N_1}) - \int \psi^2(x-\theta)dF(x)\right)$$

$$\xrightarrow{\mathcal{D}} N\left(0, \int \psi^4(x-\theta)dF(x) - \left[\int \psi^2(x-\theta)dF(x)\right]^2\right) \qquad \text{as } d \to 0+ .$$

PROOF: Lemma 4.6.1 implies that the asymptotic distribution of

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\psi^2(X_i - M_n) - \int \psi^2(x-\theta)dF(x)\right) \tag{4.140}$$

is the same as the asymptotic distribution of

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\psi^2(X_i - \theta) - \int \psi^2(x-\theta)dF(x)\right)$$

which is $N\left(0, \int \psi^4(x-\theta)dF(x)\right)$, see proof of Lemma 4.6.4. Hence, we have the limit distribution of the random variables in (4.140) for non-random number of observations. The theorem now follows from Theorem A.4.1 using Lemma 4.6.3 and from the second part of Lemma 4.6.2.
□

LEMMA **4.6.5** *Under the assumptions of Section 4.1 we have, for every $\varepsilon > 0$, that there exists $\delta > 0$ and n_0 such that*

$$P\left\{\max_{n \le m \le n(1+\delta)} |\sqrt{n}(M_n - M_m)| > \varepsilon\right\} < \varepsilon$$

for all $n \ge 0$.

PROOF: See Jurečková and Sen (1981), proof of Lemma 3.1.
□

Remark 5 *Notice that Lemma 4.6.5 immediately implies that the sequence $\sqrt{n}(M_n - \theta)$ is uniformly continuous in probability.*

We have to investigate also the properties of the denominator of the estimate $\hat{\sigma}_n^2$ of the asymptotic variance of the M-estimator which is defined in (4.38). We denote

$$D_n(M_n) = \sqrt{n}[\lambda_{F_n}(M_n - tn^{-1/2}) - \lambda_{F_n}(M_n + tn^{-1/2})]. \tag{4.141}$$

Notice that the denominator of $\hat{\sigma}_n^2$ is then $[D_n(M_n)/2t]^2$.

LEMMA **4.6.6** *Let's assume that the assumptions of Section 4.1 hold. Then for every $\varepsilon > 0$ there exists $\delta_0 > 0$ and n_0 such that for all $0 < \delta < \delta_0$ and for all $n > n_0$ we have*

$$P\left\{\sup_{-K\log n<t<K\log n} n^{-1/4} \left| \sum_{i=n}^{n(1+\delta)} \left[\psi(X_i-\theta-tn^{-1/2})-\psi(X_i-\theta)-\lambda_F(\theta+tn^{-1/2})\right]\right| > \varepsilon\right\} < \varepsilon$$

and, under the additional assumption that $\psi_2(.) \equiv 0$ and that $\psi_1(.)$ has a bounded derivative $\psi_1'(.)$ inside the interval $(-h,h)$, also

$$P\left\{\sup_{-K\log n<t<K\log n} \left| \sum_{i=n}^{n(1+\delta)} \left[\psi(X_i-\theta-tn^{-1/2})-\psi(X_i-\theta)-\lambda_F(\theta+tn^{-1/2})\right]\right| > \varepsilon\right\} < \varepsilon.$$

PROOF: The proof of the first part is very similar to the proof of Lemma 4.2.1. The only difference is that we can control one more parameter δ which allows us to make the righthand side of (4.23) arbitrarily small.

The second part of the lemma can be proven, using Taylor expansion and Hoeffding inequality A.2.5, in the same way as Lemma 4.6.2.

□

Remark 6 *Lemma 4.6.6 yields that the sequence*

$$n^{-1/4}\sum_{i=n}^{n(1+\delta)} \left[\psi(X_i-\theta-tn^{-1/2})-\psi(X_i-\theta)-\lambda_F(\theta+tn^{-1/2})\right]$$

is u.c.i.p. and that also the sequence

$$\sum_{i=n}^{n(1+\delta)} \left[\psi(X_i-\theta-tn^{-1/2})-\psi(X_i-\theta)-\lambda_F(\theta+tn^{-1/2})\right],$$

is u.c.i.p. under the additional assumption that $\psi_2 \equiv 0$ and that $\psi_1(.)$ has a bounded derivative $\psi_1'(.)$ inside the interval $(-h,h)$.

Remark 7 *Notice that Lemma 4.2.2 allows us to replace θ by M_n in Lemma 4.6.6. Therefore, we get also the uniform continuity in probability for the sequence $n^{1/4}D_n(M_n)$ (and $n^{1/2}D_n(M_n)$ if $\psi_2 \equiv 0$).*

In order to establish the asymptotic normality for the sequence $D_n(\theta)$, defined in (4.141), we have to introduce additional assumptions.

J1: Recall that the step component ψ_2 can be written as

$$\psi_2(x) = \beta_j \quad \text{for } a_{j-1} < x < a_j, \tag{4.142}$$

where $j = 1,\ldots,m+1$, $a_0 = -h$, $a_{m+1} = h$. Assume that at least two β_j's are different and that $f'(.)$ is bounded in neighbourhoods of $a_0,\ldots,a_{m+1}$.

J2: Assume that $\psi_2 \equiv 0$ and that $\psi_1(.)$ has two bounded derivatives in the interval $(-h, h)$.

LEMMA **4.6.7** *Let's assume that the assumptions of Section 4.1 hold.*
(i) Assume that the condition J1 is satisfied. Then

$$n^{-1/4} \sum_{i=1}^{n} \left\{ \psi(X_i - \theta - tn^{-1/2}) - \psi(X_i - \theta) - t\lambda_F'(\theta) \right\} \xrightarrow{\mathcal{D}} N\left(0, t\sum_{i=1}^{m+1} (\beta_j - \beta_{j-1})^2 f(\alpha_j)\right)$$

(ii) Assuming that the condition J2 holds, we have that

$$\sum_{i=1}^{n} \left\{ \psi(X_i - \theta - tn^{-1/2}) - \psi(X_i - \theta) - t\lambda_F'(\theta) \right\}$$
$$\xrightarrow{\mathcal{D}} N\left(0, t\left\{ \int [\psi'(x-\theta)]^2 dF(x) - \left[\int \psi'(x-\theta) dF(x)\right]^2 \right\}\right).$$

PROOF: See Theorem 2.2 and the corollary of Theorem 2.3 in Jurečková (1980).

□

LEMMA **4.6.8** *Assume that the assumptions of Section 4.1 hold.*
(i) Assume that the condition J1 is satisfied. Recall that $N_1(d)$ is the intermediate sample size given by (4.80). Then we have that

$$N_1^{-1/4} \sum_{i=1}^{N_1} \left\{ \psi(X_i - \theta - tN_1^{-1/2}) - \psi(X_i - \theta) - t\lambda_F'(\theta) \right\} \xrightarrow{\mathcal{D}} N\left(0, t\sum_{i=1}^{m+1} (\beta_j - \beta_{j-1})^2 f(\alpha_j)\right)$$

(ii) Assuming that the condition J2 is satisfied, we have that

$$\sum_{i=1}^{N_1} \left\{ \psi(X_i - \theta - tN_1^{-1/2}) - \psi(X_i - \theta) - t\lambda_F'(\theta) \right\}$$
$$\xrightarrow{\mathcal{D}} N\left(0, t\left\{ \int [\psi'(x-\theta)]^2 dF(x) - \left[\int \psi'(x-\theta) dF(x)\right]^2 \right\}\right).$$

PROOF: It follows from Theorem A.4.1 using Lemma 4.6.7 and Remark 6.

□

From the results of Jurečková (1980) it follows that, if the score function is not continuous, the order of convergence of the estimator (4.38) of the asymptotic variance of the M-estimator is $n^{-1/4}$. Some alternative estimators were considered e.g. by Dodge and Jurečková (1995), but the rate of convergence $n^{-1/4}$ was not improved. It would be possible to estimate the asymptotic variance of the M-estimator also from the length of the confidence interval which can be obtained similarly as in Section 1.2.2 (Jurečková and Sen 1978). Unfortunatelly, the rate of convergence of this estimator is also $n^{-1/4}$ as follows e.g. from the results of Aerts (1988).

This means that the M-estimators with the discontinuous score function have some undesireable properties in the context of the three-stage procedures.

In the following, we will restrict ourselves only to the M-estimators with smooth score functions.

The next theorem gives us a better estimator for $\lambda_F'(\theta)$.

THEOREM **4.6.2** *Assume that the conditions of Section 4.1 hold with the jump component $\psi_2(x) \equiv 0$. Assume that the continuous component $\psi_1(x)$ has bounded derivative inside the interval $(-h, h)$, where h is given by (4.9). Then we have*

$$\frac{1}{N_1}\sum_{i=1}^{N_1}\psi'(X_i - M_{N_1}) \to \lambda'_F(\theta) \qquad [P]\ a.s. \tag{4.143}$$

and

$$\sqrt{N_1}\left[\frac{1}{N_1}\sum_{i=1}^{N_1}\psi'(X_i - M_{N_1}) - \lambda'_F(\theta)\right] \to N\left(0, E\left[\psi'(X_i-\theta)\right]^2 - \left[E\psi'(X_i-\theta)\right]^2\right) \tag{4.144}$$

PROOF: The proof of the first part follows from the Strong Law of Large Numbers, the second part can be proven e.g. by the Anscombe Theorem A.4.2.

□

4.6.2 Bootstrap Critical Points — Robust Procedure

The rate of convergence of bootstrap critical points for M-estimators has been investigated only under assumptions of differentiability of the score function $\psi(.)$. Some results can be found e.g. in Lahiri (1992).

Let's assume (in addition to assumptions of Section 4.1) that

L1: there exists a Borel set $C \subset \Re$ such that $P_\theta(X \in C) = 1$ for all $\theta \in \Theta$ and the function $\psi(x-t)$ has continuous ν-th order derivatives in t for $1 \le \nu \le s$ at each $(x,t) \in C \times \Theta$ for some integer $s \ge 3$.

L2: $E|\psi^{(\nu)}|^s < \infty$ for $0 \le |\nu| \le s-1$, and there exists an $\varepsilon > 0$ such that

$$E\left(\sup_{|t-\theta|<\varepsilon}|\psi^{(s)}|^s\right) < \infty.$$

THEOREM **4.6.3** *Assume that the assumptions of Section 4.1 are satisfied and that also the assumptions L1 and L2 hold with $s = 3$ and Cramer's condition is satisfied. Then, $[P]$ a.s.,*

$$\sup_{-\infty<x<\infty}|P_n^*\left(\sqrt{n}\frac{M_n^*-M_n}{\hat\sigma} < x\right) - P\left(\sqrt{n}\frac{M_n-\theta}{\sigma} < x\right)| = o(n^{-1/2})$$

PROOF: The theorem is proven in more general form in Lahiri (1992). Notice that Lahiri's condition (2.3) is satisfied automatically if we define B as intervals $(-\infty, x]$.

□

4.6.3 Asymptotic Distribution of N_2 — Robust Procedure

The asymptotic distribution of $\sqrt{N_2}$ for the robust procedure can be derived similarly as in Theorem 4.5.6 for the procedure based on sample mean. The only difference is that we have to take into account also the estimate of $\lambda'_F(\theta)$ which makes the situation a little more complicated and which also increases the variance of the limit distribution of $\sqrt{N_2}$ in the following theorem.

THEOREM **4.6.4** *Let the assumptions of Theorems 4.6.2 and 4.6.3 be satisfied. Then we have for the robust three-stage procedure based on standardized bootstrap critical points that*

$$\sqrt{N_2(d)} - \sqrt{c_M(d)} \xrightarrow{\mathcal{D}}$$

$$N\left(0, \frac{1}{k}\left\{\frac{\int \psi^4(x-\theta)dF(x)}{4\left[\int \psi^2(x-\theta)dF(x)\right]^2} - \frac{\int \psi^2(x-\theta)\psi'(\theta)dF(x)}{\lambda_F'(\theta)\int \psi^2(x-\theta)dF(x)} + \frac{\int [\psi'(X_i-\theta)]^2 dF(x)}{[\lambda_F'(\theta)^2]} - \frac{1}{4}\right\}\right)$$

as $d \to 0+$.

PROOF: Using Theorem 4.6.3 with Lemma 4.5.1, Theorem 4.6.1, and Theorem 4.6.2, we have the following

$$\sqrt{N_2(d)} - \sqrt{c_M(d)} = \frac{\xi^A_{N_1}\sqrt{\frac{1}{N_1}\sum_{i=1}^{N_1}\psi^2(X_i - M_{N_1})}}{d\frac{1}{N_1}\sum_{i=1}^{N_1}\psi'(X_i - M_{N_1})} - \frac{u_{1-\alpha/2}\sqrt{\int_{-\infty}^{\infty}\psi^2(x-\theta)dF(x)}}{d\lambda_F'(\theta)}$$

$$= \frac{u_{1-\alpha/2}}{d}\left\{\frac{\sqrt{\frac{1}{N_1}\sum_{i=1}^{N_1}\psi^2(X_i - M_{N_1})}}{\frac{1}{N_1}\sum_{i=1}^{N_1}\psi'(X_i - M_{N_1})} - \frac{\sqrt{\int_{-\infty}^{\infty}\psi^2(x-\theta)dF(x)}}{\lambda_F'(\theta)}\right\} + o_P(1)$$

$$= \frac{u_{1-\alpha/2}}{d}\left\{\frac{1}{2\sqrt{\int_{-\infty}^{\infty}\psi^2(x-\theta)dF(x)}\lambda_F'(\theta))}\left[\frac{1}{N_1}\sum_{i=1}^{N_1}\psi^2(X_i - M_{N_1}) - \int_{-\infty}^{\infty}\psi^2(x-\theta)dF(x)\right]\right.$$

$$\left. - \left[\frac{\sqrt{\int_{-\infty}^{\infty}\psi^2(x-\theta)dF(x)}}{\left[\lambda_F'(\theta)\right]^2}\right]\left[\frac{1}{N_1}\sum_{i=1}^{N_1}\psi'(X_i - M_{N_1}) - \lambda_F'(\theta)\right]\right\} + o_P(1)$$

$$= \sqrt{\frac{N_1}{k}}\left\{\frac{1}{2\int_{-\infty}^{\infty}\psi^2(x-\theta)dF(x)}\left[\frac{1}{N_1}\sum_{i=1}^{N_1}\psi^2(X_i - M_{N_1}) - \int_{-\infty}^{\infty}\psi^2(x-\theta)dF(x)\right]\right.$$

$$\left. - \frac{1}{\lambda_F'(\theta)}\left[\frac{1}{N_1}\sum_{i=1}^{N_1}\psi'(X_i - M_{N_1}) - \lambda_F'(\theta)\right]\right\} + o_P(1)$$

$$\xrightarrow{\mathcal{D}} N\left(0, \frac{1}{k}\left\{\frac{\int \psi^4(x-\theta)dF(x) - \left[\int \psi^2(x-\theta)dF(x)\right]^2}{4\left[\int \psi^2(x-\theta)dF(x)\right]^2}\right.\right.$$

$$\left.\left. - \frac{\int \psi^2(x-\theta)\psi'(\theta)dF(x) - \lambda_F'(\theta)\int \psi^2(x-\theta)dF(x)}{\lambda_F'(\theta)\int \psi^2(x-\theta)dF(x)} + \frac{\int [\psi'(X_i-\theta)]^2 dF(x) - [\lambda_F'(\theta)]^2}{[\lambda_F'(\theta)^2]}\right\}\right)$$

$$\equiv N\left(0, \frac{1}{k}\left\{\frac{\int \psi^4(x-\theta)dF(x)}{4\left[\int \psi^2(x-\theta)dF(x)\right]^2} - \frac{\int \psi^2(x-\theta)\psi'(\theta)dF(x)}{\lambda_F'(\theta)\int \psi^2(x-\theta)dF(x)} + \frac{\int [\psi'(X_i-\theta)]^2 dF(x)}{[\lambda_F'(\theta)^2]} - \frac{1}{4}\right\}\right)$$

□

If we compare the asymptotic distribution of $\sqrt{N_2}$ for the procedure based on sample mean given in Theorem 4.5.6 with the asymptotic distribution of $\sqrt{N_2}$ for the procedure based on M-estimators provided by Theorem 4.6.4, we see that the formulas agree nicely — the only difference is the variance of the estimate of $\lambda'_F(\theta)$ which increases the asymptotic variance of $\sqrt{N_2}$ for the robust procedure.

We see that the asymptotic variance of the asymptotic normal distribution in Theorem 4.6.4 depends both on the score function $\psi(.)$ and on the distribution function of the observations $F(.)$. It would be very interesting to find the score function which minimizes the asymptotic variance of $\sqrt{N_2} - \sqrt{c_M(d)}$ for a given distribution function. It might be possible to use some type of adaptive estimators — however, this area remains open for future research.

Chapter 5

Simulations

In this chapter, we will try to compare the methods based on bootstrap with the method based on normal approximation via simulation. The hope is that the bootstrap methods will give better results, especially for small sample sizes.

The advantage of the methods based on bootstrap is closer approximation of the unknown critical points of the distribution of the M-estimator. Via simulation, for a known distribution of the random variables, we can get some idea about the real critical points.

We decided to use Huber's $\psi_h(.)$ function which is (for $h > 0$) defined as

$$\psi_h(x) = \begin{cases} -h & \text{if } x < -h \\ x & \text{if } |x| \leq h \\ h & \text{if } x > h \end{cases} \tag{5.1}$$

This score function has been derived by Huber (1981) as a score function which minimizes the worst possible variance which can be obtained for ε-contaminated Normal distribution. The value of h should be chosen accordingly to the level of contamination ε and it can be obtained as a solution of the equation

$$2\Phi(h) - 1 + \frac{2\varphi(h)}{h} = \frac{1}{1-\varepsilon}, \tag{5.2}$$

where $\Phi(\cdot)$ and $\varphi(\cdot)$ denote the distribution and density function of the Normal distribution, respectively (Antoch and Vorlíčková 1992).

We give the values of ε for $0 < h < 3$ in Table B.1 in Appendix B.

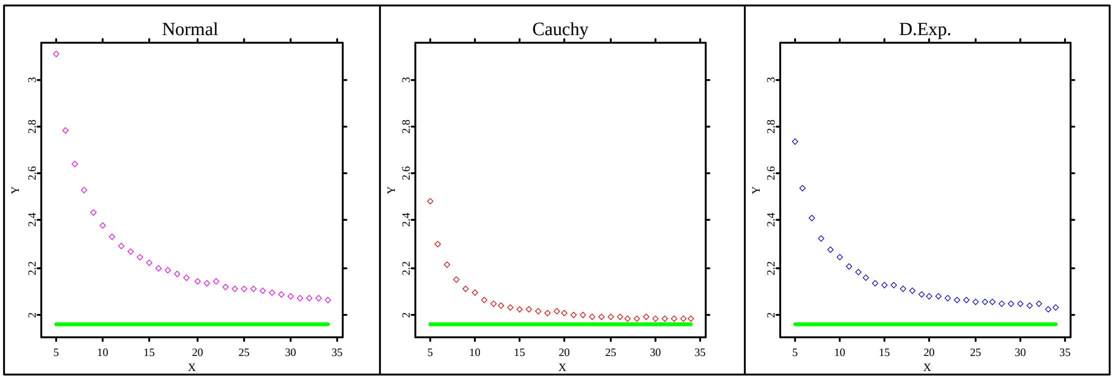
In Section 5.1, we will try to approximate the exact critical points of the distribution of the M-estimator for smaller sample sizes. In Sections 5.2–5.5, we will present the results of simulations for Normal, Cauchy, Double Exponential, and contaminated Normal distributions. In particular, we will concentrate on the Cauchy distribution which will be investigated in Section 5.3.

All simulations presented in this chapter were programmed in GNU Fortran. All graphics was created in XploRe (Härdle, Klinke, and Müller 2000).

5.1 Simulations of Exact Critical Points

In order to investigate the behaviour of our procedure for small sample sizes, we will try to investigate the magnitude of the error caused by using the asymptotic normal critical points instead of the unknown exact critical points of the M-estimator.

Figure 5.1: *Simulated critical points for Huber's $\psi(.)$ with parameter $h = 1.5$ for Normal, Cauchy and Double Exponential distribution, $n = 5, \ldots, 30$.*



We tried to find the correct critical points by simulation. The results for $n = 10, \ldots, 30$ for Normal, Cauchy and Double Exponential distributions are presented in Figure 5.1. The values plotted on each graph were obtained in the following way:

1. we simulated large number (200.000) random samples of the desired size $(10, \ldots, 30)$ from the distribution under consideration (Normal, Cauchy, Double Exponential).

2. we calculated the M-estimate for every generated sample and we find out the 95% (empirical) critical point of the distribution of $\sqrt{n}|M_n - \theta|/\hat{\sigma}_n$.

3. we plot the obtained "simulated critical points" in the graph together with the green line at the bottom of the pictures which corresponds to the limiting $u_{0.975}$ quantile of the standard Normal distribution.

The plots show the dependence of the simulated critical points on the sample size.

We see that the critical points converge to $u_{0.975}$ as the sample size increases. The convergence to the critical point of the asymptotic normal distribution seems to be fastest for the Cauchy distribution.

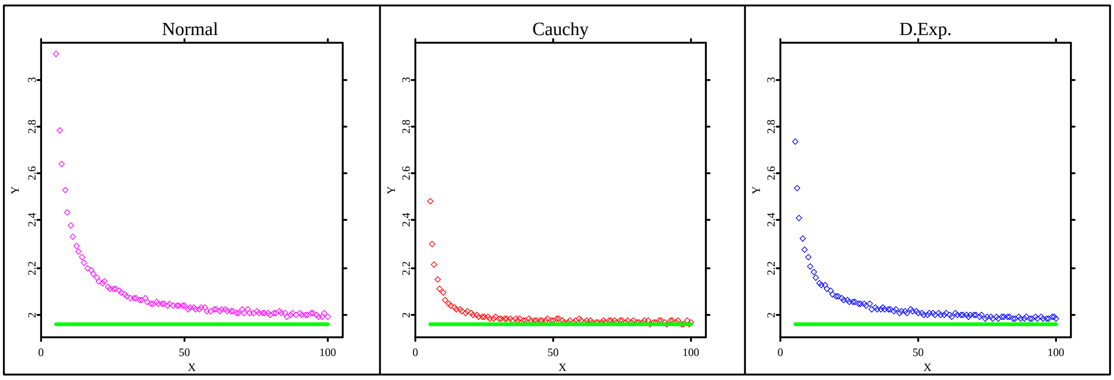
On Figure 5.2, we plotted exactly the same quantities as on Figure 5.1, but for sample sizes between 5 and 100. Also this plot suggests that the normal approximation works very well for Cauchy distribution, the convergence to the normal critical point for the Normal and Double Exponential distributions seems to be a bit slower.

5.2 Simulations for Normal Distribution

For the simulations presented in this section, we simulated data from standard Normal distribution $N(0, 1)$. We denote by $\Phi(\cdot)$ and $\varphi(\cdot)$ the distribution function and the density function of $N(0, 1)$, respectively.

The asymptotically optimal stopping time for normally distributed data and for the sequential procedure based on M-estimators with Huber's score function $\psi_h(x) = \max\{-h, \min(x, h)\}$,

Figure 5.2: *Simulated critical points for Huber's $\psi(.)$ with parameter $h = 1.5$ for Normal, Cauchy and Double Exponential distribution, $n = 5, \ldots, 100$.*



see also formula (5.1), depends on

$$\sigma^2(\psi_h, \Phi) = \frac{\int \psi_h^2(x) d\Phi(x)}{\int \psi'(x) d\Phi(x)}.$$

Straightforward calculations lead that

$$\int_{-\infty}^{\infty} \psi_h^2(x) d\Phi(x) = 1 - 2h\psi(h) + (2h^2 - 2)(1 - \Phi(h)) \tag{5.3}$$

and

$$\int_{-\infty}^{\infty} \psi_h'(x) d\Phi(x) = 2\Phi(h) - 1. \tag{5.4}$$

This gives that the asymptotically optimal stopping time for observations coming from standard Normal distribution can be calculated as

$$c_M(d) = \left(\frac{u_{1-\alpha/2}}{d}\right)^2 \frac{1 - 2\varphi(h) + (h^2 - 2)(1 - \Phi(h))}{(2\Phi(h) - 1)^2}. \tag{5.5}$$

We carried out the simulations for the desired lenght of interval equal to 1, 0.6, and 0.2, i.e., for $d = 0.5$, $d = 0.3$, and $d = 0.1$. The tuning parameters γ and k for the three-stage procedure were chosen as $\gamma = 1/3$ and $k = 1/2$. The calculations were repeated 1000 times. The results of simulations are presented in Tables 5.1 and 5.2 and in Figures 5.3–5.6. Table 5.1 and Figures 5.3 and 5.5 show the results of simulations for the robust three-stage procedure described in Section 4.4. Table 5.2 and Figures 5.4 and 5.6 display the results of simulations for the "alternative procedure" which has been proposed in Subsection 4.4.2. In each table we compare the results for the methods based on normal critical points, standardized and studentized critical points.

Table 5.1: *Normal distribution, Huber's ψ with $h = 1.5$, $\gamma = 1/3$, $k = 1/2$*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	16	N	15.71	14	0.923
		A	18.86	17	0.934
		U	23.94	21	0.964
0.30	45	N	39.95	37	0.920
		A	48.65	44	0.932
		U	55.88	52	0.945
0.10	399	N	392.74	391.5	0.927
		A	485.46	457	0.948
		U	490.42	465.5	0.943

Table 5.2: *Normal distribution, Huber's ψ with $h = 1.5$, $\gamma = 1/3$, $k = 1/2$, alternative method*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	16	N	15.57	14	0.922
		A	18.28	16	0.944
		U	21.83	19	0.965
0.30	45	N	40.73	38	0.912
		A	49.93	45	0.925
		U	55.39	51	0.930
0.10	399	N	391.71	391	0.920
		A	483.46	453	0.943
		U	487.44	453	0.942

Figure 5.3: N_2 *for Normal distribution.*

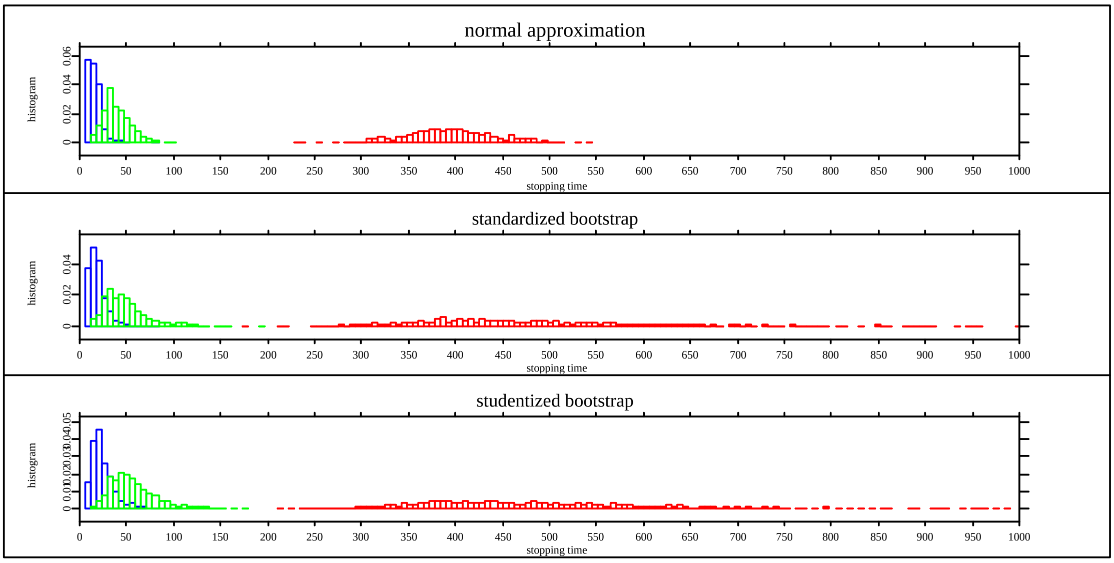


Figure 5.4: N_2 *for Normal distribution, alternative method.*

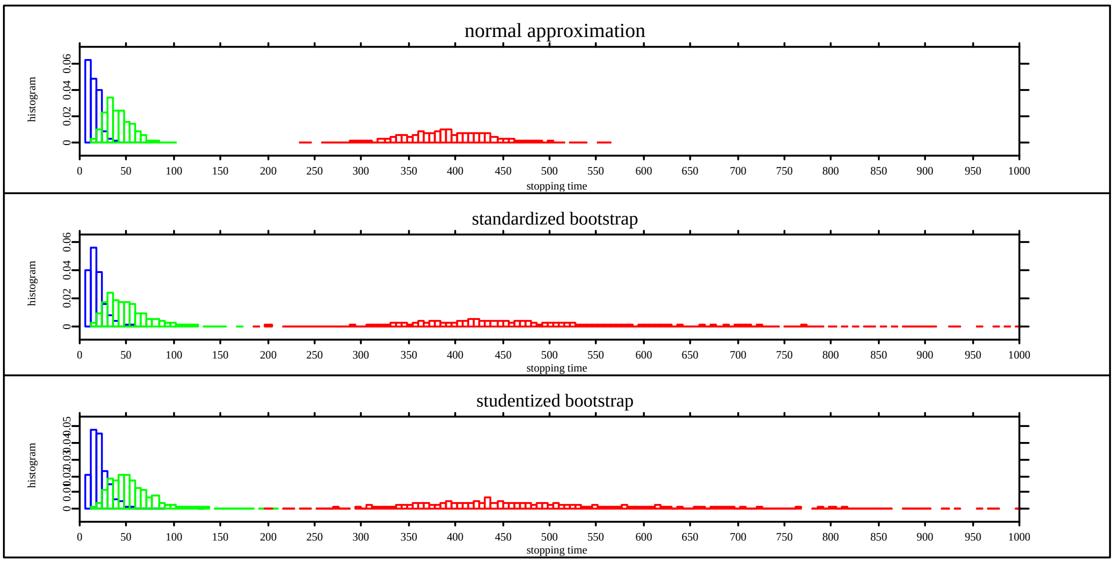


Figure 5.5: $\sqrt{N_2}$ *for Normal distribution.*

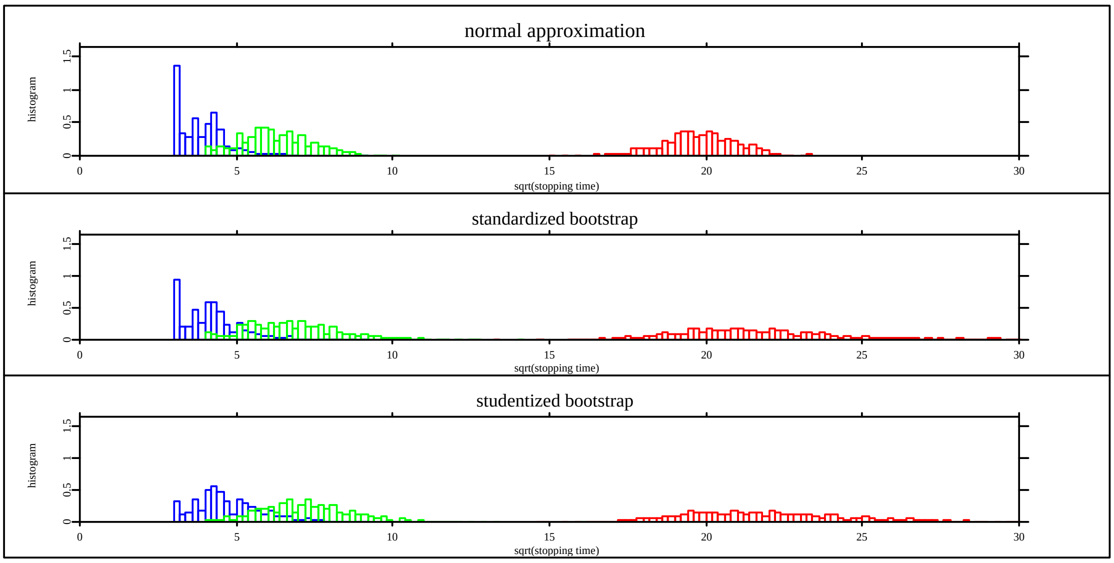
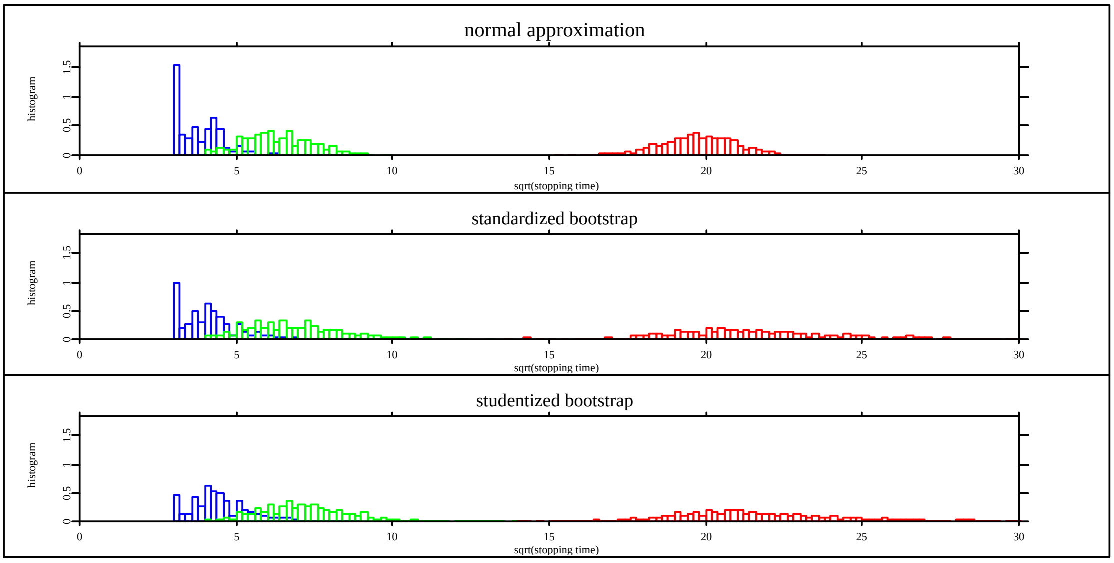


Figure 5.6: $\sqrt{N_2}$ *for Normal distribution, alternative method.*



Notice that the procedures based on normal critical points are exactly the same and that the procedures based on standardized bootstrap critical points are equivalent in all tables and figures.

In each table, the first column contains the value of d which is one half of the desired length $2d$ of the confidence interval. The (asymptotically) optimal number of observations for the standard Normal distribution (5.5) is given in the second column. The third column specifies the critical points which were used for calculations. "N" denotes the normal critical points, "A" stands for the standardized bootstrap critical points, and "U" denotes the studentized bootstrap critical points. In the fourth, fifth and sixth column, we present the mean, median and the coverage probability estimated from 1000 simulations.

We carried out the simulations for the desired lengths of confidence interval equal to $2d = 1$, $2d = 0.6$, and $2d = 0.2$. As expected, $d = 0.5$ typically leads to a smaller sample size. Comparing the values of mean and median for all three methods with the optimal sample size given in the tables, we see that the method "N" tends to underestimate the optimal sample size. The coverage probability of 95% was reached only for the method based on bootstrap critical points for $d = 0.5$, other coverage probabilities lie below the value 0.95. Notice that in all cases, the coverage probabilites for the bootstrap based methods are closer to 0.95 than the method based on normal critical points.

In Figures 5.3 and 5.4, you can see the histograms of the stopping times. The first display on each figure shows histograms for the normal critical points, the second display shows histograms for the procedure based on the standardized bootstrap critical points and on the third display we plot the histograms for the procedure based on studentized bootstrap. The blue, green and red histograms at each display correspond to the stopping times for procedures giving fixed-width confidence intervals of lengths $2d = 1$, $2d = 0.6$, and $2d = 0.2$, respectively. We can clearly see the dependence of the mean (and also of the variance) on the desired length of the confidence interval.

From the theory in Chapter 4 (Theorem 4.6.4) it follows that $\sqrt{N_2}$ has asymptotic distribution whose variance does not depend on d. Therefore we display also the histograms of $\sqrt{N_2}$ on Figures 5.5 and 5.6.

The difference in the magnitude of the stopping times between these three methods is clearly visible for $d = 0.1$.

Also the coverage probabilities and stopping times given in Tables 5.1 and 5.2 suggest that the method based on normal approximation tends to stop too early. It seems that the methods based on bootstrap tend to correct this feature. On the other hand, the graphics in Figures 5.3–5.6 suggests that the variance of N_2 for the methods based on bootstrap is larger than for the method based on normal critical points.

From the point of view of the coverage probability, the best results, especially for smaller sample sizes, are provided by the studentized bootstrap. However, the coverage probabilities obtained for the method based on standardized bootstrap are not much worse.

We conclude that our method works well for the normal distribution and that its results correspond to the theoretical values. One possible reason for the bigger variance of N_2 for the bootstrap based methods might be insufficient starting sample size $m(d)$. This phenomenon will be further investigated in the next section which concerns Cauchy distribution.

More results of simulations for the Huber's $\psi_h(.)$ function with $h = 1$ and $h = 0.5$ can be found in Appendix C. The problem of automatic choice of h remains one of the open questions.

Table 5.3: *Cauchy distribution, Huber's* ψ *with* $h = 1.5$, $\gamma = 1/3$, $k = 1/2$

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	46	N	51.59	42	0.939
		A	91.71	54	0.954
		U	109.25	50	0.959
0.30	128	N	128.56	126	0.960
		A	144.43	128.5	0.963
		U	145.56	129	0.962
0.10	1150	N	1142.3	1146.5	0.959
		A	1053.7	1047.5	0.944
		U	1055.8	1055	0.945

5.3 Cauchy Distribution

Even more interesting are simulations for the Cauchy distribution. In this case, the methods based on sample mean can not be used at all, because Cauchy distribution has heavy tails. The possibility to use M-estimators in this situation is therefore of great importance.

We simulated observations from the Cauchy distribution given by density

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \qquad x \in \Re. \tag{5.6}$$

Straightforward calculations lead that

$$\int_{-\infty}^{\infty} \psi_h'(x)dF(x) = \frac{2}{\pi}\operatorname{arctg} h$$

and

$$\int \psi_h^2(x)dF(x) = h^2 + \frac{2}{\pi}[h - (h^2+1)\operatorname{arctg} h].$$

This leads that the asymptotically optimal stopping time is equal to

$$c_M(d) = \Big(\frac{u_{1-\alpha/2}}{d}\Big)^2 \frac{h^2 + 2/\pi[h - (h^2+1)\operatorname{arctg} h]}{(2/\pi \operatorname{arctg} h)^2} \tag{5.7}$$

The results of the simulations are are given in Tables 5.3 and 5.4 and in Figures 5.7—5.10. Similarly as in the previous section, the plots contain the histograms of the stopping times for the three-stage procedure based on normal critical points, standardized bootstrap critical points, and studentized bootstrap critical points displayed on the original and square root scale.

From Figures 5.7–5.10, it seems that here the asymptotics works better than for the Normal distribution in the previous section. One possible reason is that the confidence intervals of the same width require approximately two times more observations for the Cauchy than for the standard Normal distribution, see Tables B.4–B.9.

The coverage probabilities in Tables 5.3 and 5.4 seem to lie very close to the desired value 0.95. The only value which seems to be a bit lower is the coverage probability for the method based on normal critical points for $d = 0.5$ in both tables.

Table 5.4: *Cauchy distribution, Huber's* ψ *with* $h = 1.5$, $\gamma = 1/3$, $k = 1/2$, *alternative method*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	46	N	53.32	42	0.931
		A	97.62	53	0.950
		U	104.09	52	0.952
0.30	128	N	130.49	124	0.947
		A	143.26	127	0.952
		U	139.74	125	0.956
0.10	1150	N	1146.9	1142	0.967
		A	1056.6	1046	0.949
		U	1053.8	1055	0.949

Figure 5.7: N_2 *for Cauchy distribution.*

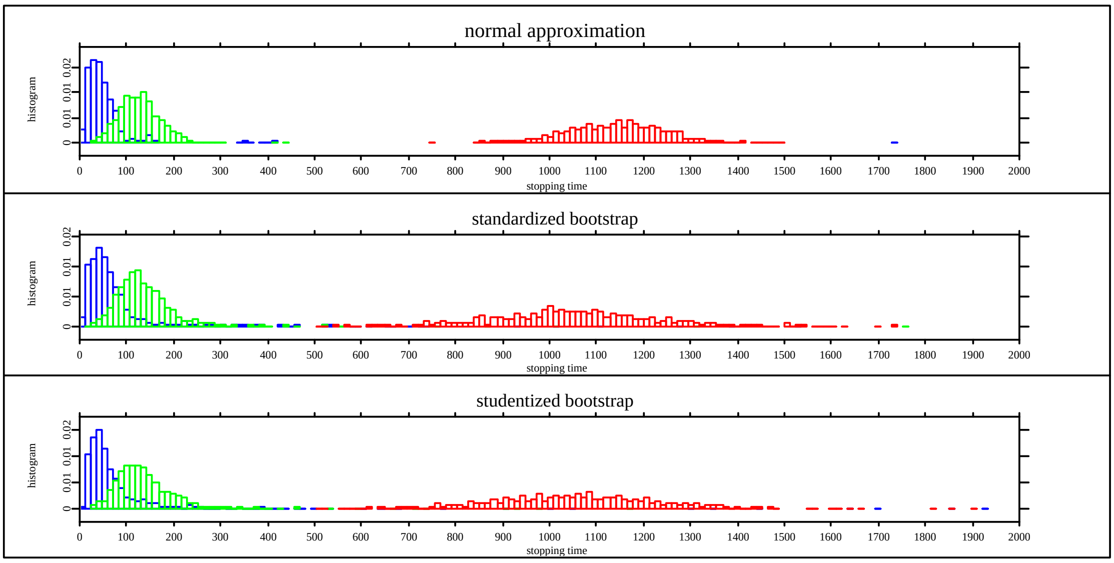


Figure 5.8: N_2 *for Cauchy distribution, alternative method.*

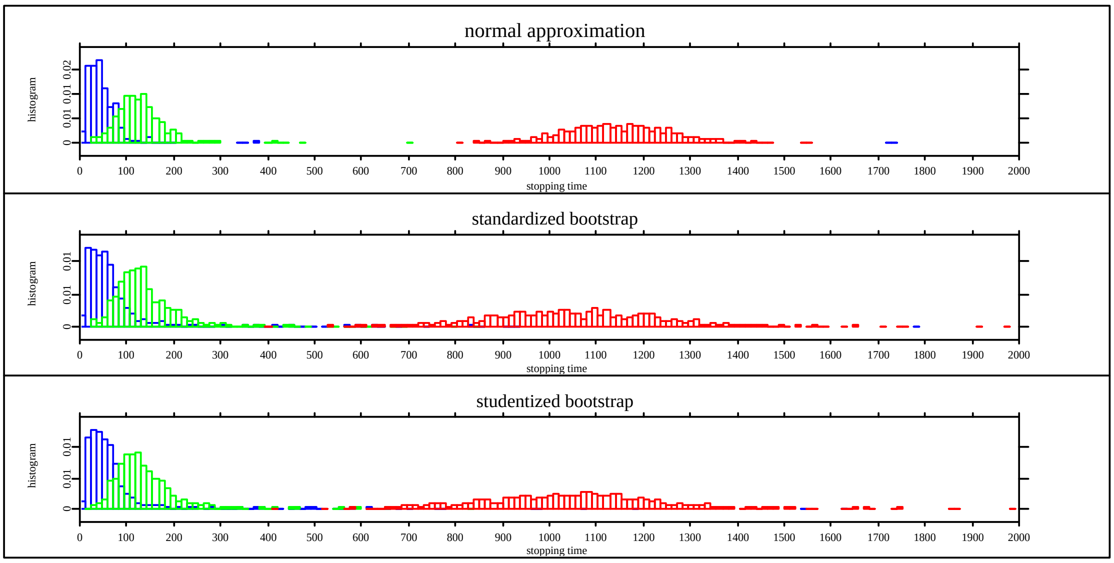


Figure 5.9: $\sqrt{N_2}$ *for Cauchy distribution.*

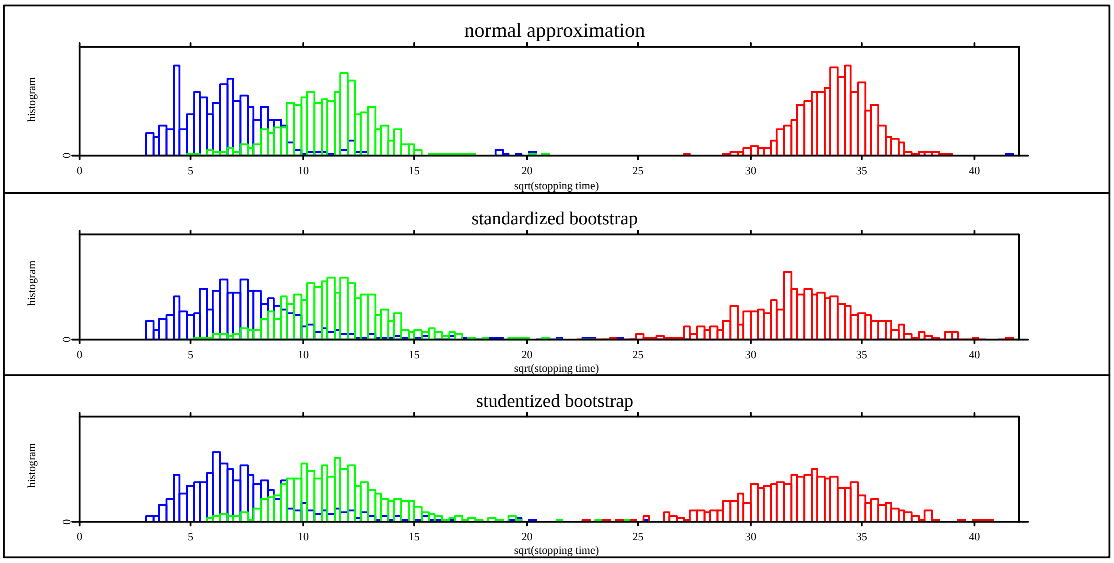
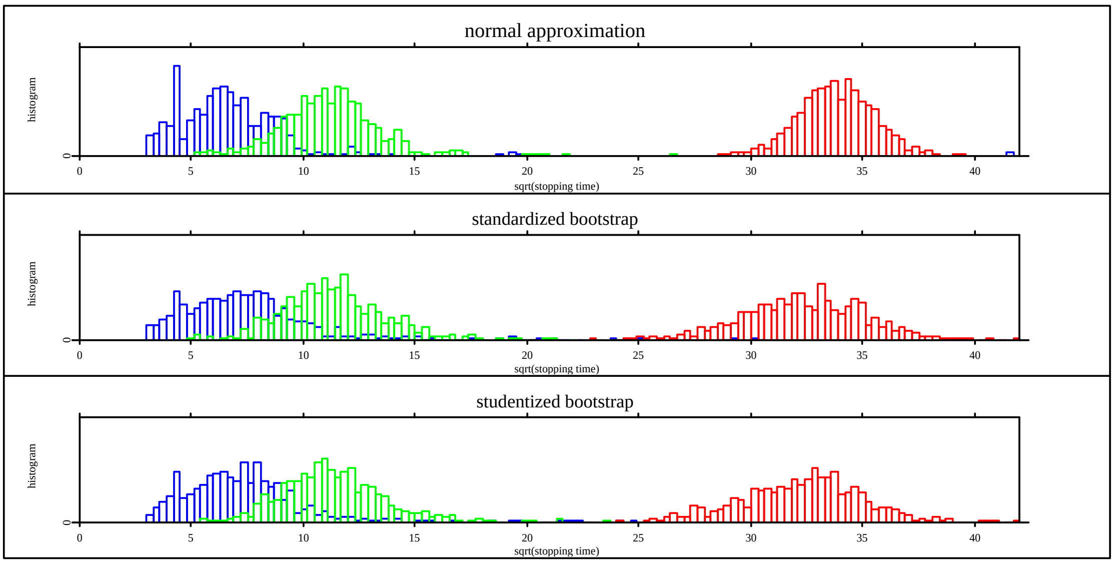


Figure 5.10: $\sqrt{N_2}$ *for Cauchy distribution, alternative method.*



Looking at the values of mean and median, we can notice one unpleasant feature of the bootstrap based method. The huge difference between the values of the mean and median (the mean is approximately two times higher than the median) suggests that the distribution of N_2 might be skewed. The reason for this behaviour seems to be too small starting sample size $m(d)$. The methods based on bootstrap tend to overestimate the critical points dramatically if $m(d)$ is less than 20. We tried to increase the starting sample size and we modified the formula (4.79) as follows:

$$m = m(d) = \max\left\{20, \left[\left(\frac{u_{1-\alpha/2}}{d}\right)^{2/(1+\gamma)}\right]^{\circ} + 1\right\}.$$

This guarantees that the starting sample size is so big that the bootstrap procedure has chance to work well. The results of simulations using this starting sample size are given in Table 5.5 and in Figure 5.11. Indeed, we can see that the difference between the mean and median is now much smaller.

In Tables 5.5–5.7 and Figures 5.11–5.13 we tried to pursue another interesting phenomenon. From Theorem 4.6.4 we know that the variance of the asymptotic normal distribution depends on the parameter k. We can try to decrease the asymptotic variance of $\sqrt{N_2}$ by increasing the value of k. In Figures 5.11–5.13, we can indeed see a sligth decrease in the variance of the $\sqrt{N_2}$. At the same time it seems that the histograms shift slightly to the right. This is confirmed by the values in Tables 5.5–5.7. We see that both mean and median of N_2 increase with increasing k. Unfortunatelly, this is true also for the coverage probabilities. Choosing $k = 0.9$ means that our procedure overestimates the necessary sample size — this effect is more dramatic for longer confidence intervals which results in smaller sample sizes.

We conclude that our method works well also for the Cauchy distribution. We can see that the bootstrap critical points approximate the correct distribution of the M-estimator better than the normal approximation. We recommend to choose the starting sample size to be at least 20

Table 5.5: *Cauchy distribution, Huber's ψ with $h = 1.5$, $\gamma = 1/3$, $k = 0.5$, $m_0 \geq 20$*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	46	N	45.39	40	0.944
		A	51.95	45	0.959
		U	49.68	43	0.952
0.30	128	N	126.54	123	0.940
		A	134.34	126	0.950
		U	133.66	124	0.953
0.10	1150	N	1144.9	1147	0.955
		A	1060.2	1049	0.946
		U	1056.2	1043	0.948

Table 5.6: *Cauchy distribution, Huber's ψ with $h = 1.5$, $\gamma = 1/3$, $k = 0.7$, $m_0 \geq 20$*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	46	N	47.89	45	0.951
		A	54.83	49	0.957
		U	53.89	47	0.965
0.30	128	N	139.71	130	0.960
		A	160.50	138.5	0.961
		U	164.49	133	0.957
0.10	1150	N	1151.5	1144	0.959
		A	1080.4	1066	0.948
		U	1075.0	1067.5	0.938

Table 5.7: *Cauchy distribution, Huber's ψ with $h = 1.5$, $\gamma = 1/3$, $k = 0.9$, $m_0 \geq 20$*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	46	N	53.35	49	0.961
		A	66.66	54	0.968
		U	63.30	53	0.967
0.30	128	N	155.44	140	0.968
		A	185.94	153	0.975
		U	183.36	146	0.971
0.10	1150	N	1220.4	1177	0.963
		A	1185.6	1139	0.955
		U	1181.6	1134	0.956

Figure 5.11: $\sqrt{N_2}$ *for Cauchy distribution*, $k = 0.5$, $m_0 \geq 20$.

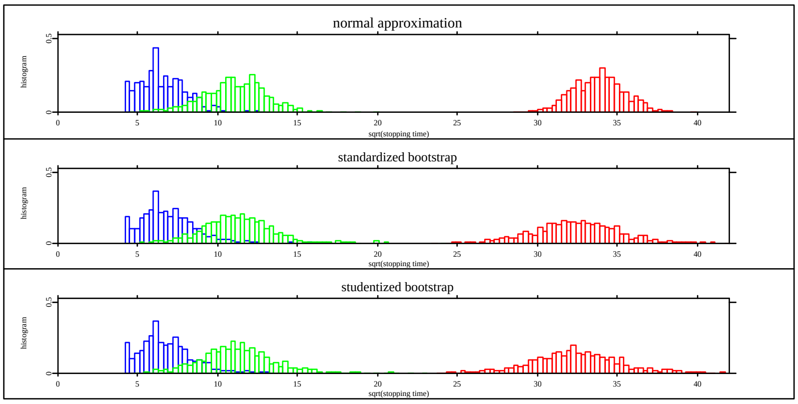


Figure 5.12: $\sqrt{N_2}$ *for Cauchy distribution*, $k = 0.7$, $m_0 \geq 20$.

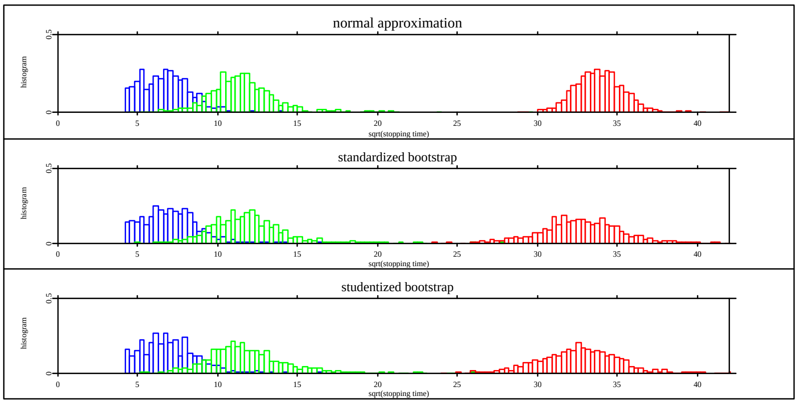


Figure 5.13: $\sqrt{N_2}$ *for Cauchy distribution*, $k = 0.9$, $m_0 \geq 20$.

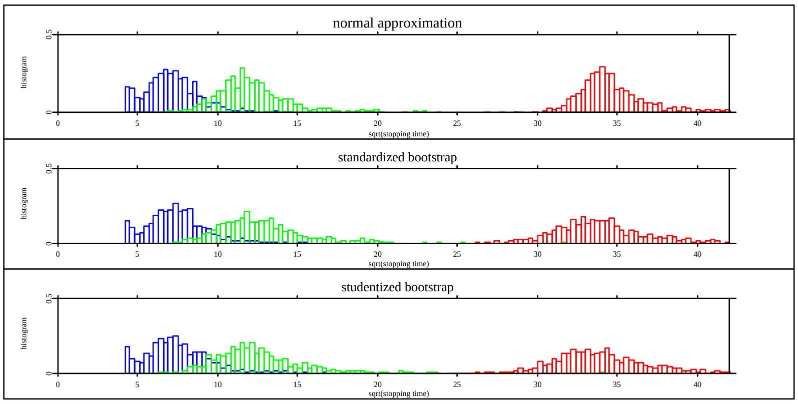


Table 5.8: *Double Exponential distribution, Huber's ψ with $h = 1.5$, $\gamma = 1/3$, $k = 1/2$*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	23	N	22.29	19	0.930
		A	25.00	21	0.957
		U	34.22	26	0.962
0.30	63	N	58.76	56	0.933
		A	61.18	59	0.942
		U	68.81	63	0.949
0.10	563	N	555.31	552.5	0.965
		A	504.81	496	0.958
		U	505.49	497	0.961

and we recommend to choose $k = 0.5$, or at most $k = 0.7$. Higher values should be chosen only for very small lengths of confidence intervals, where the choice of the value of k does not seem to have negative impact on the coverage probability of the resulting confidence interval.

Another simulations from the Cauchy distribution with different choice of h in the Huber's score function are presented in Appendix C. These simulations for $h = 1$ and $h = 0.5$ are of special interest here because the Tables B.4–B.9 suggest that by choosing smaller value of h we can decrease the final sample size N_2. This is completely different than for the Normal distribution where we can decrease the final sample size by increasing the value of h.

5.4 Double Exponential Distribution

In this section, we simulated observations from the Double Exponential distribution which is given by the density

$$f(x) = \frac{1}{2}\exp\{-|x|\}, \qquad x \in \Re. \tag{5.8}$$

Straightforward calculations lead that

$$\int_{-\infty}^{\infty} \psi_h'(x)dF(x) = 1 - \exp\{-h\}$$

and

$$\int \psi_h^2(x)dF(x) = 2 - (2+2h)\exp\{-h\}.$$

This leads the asymptotically optimal stopping time

$$c_M(d) = \left(\frac{u_{1-\alpha/2}}{d}\right)^2 \frac{2-(2+2h)\exp\{-h\}}{(1-\exp\{-h\})^2} \tag{5.9}$$

The values of the optimal stopping time are given in Appendix B in Tables B.4–B.9. Notice that, similarly as for the Cauchy distribution, the final sample size is decreasing with decreasing value of h.

Table 5.9: *Double Exponential, Huber's ψ with $h = 1.5$, $\gamma = 1/3$, $k = 1/2$, alternative method*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	23	N	21.76	19	0.921
		A	24.00	20	0.932
		U	26.48	23	0.957
0.30	63	N	58.23	55	0.932
		A	60.82	57.5	0.939
		U	62.99	59.5	0.947
0.10	563	N	559.40	561	0.964
		A	507.81	499	0.956
		U	509.16	501	0.955

Figure 5.14: N_2 *for Double Exponential distribution.*

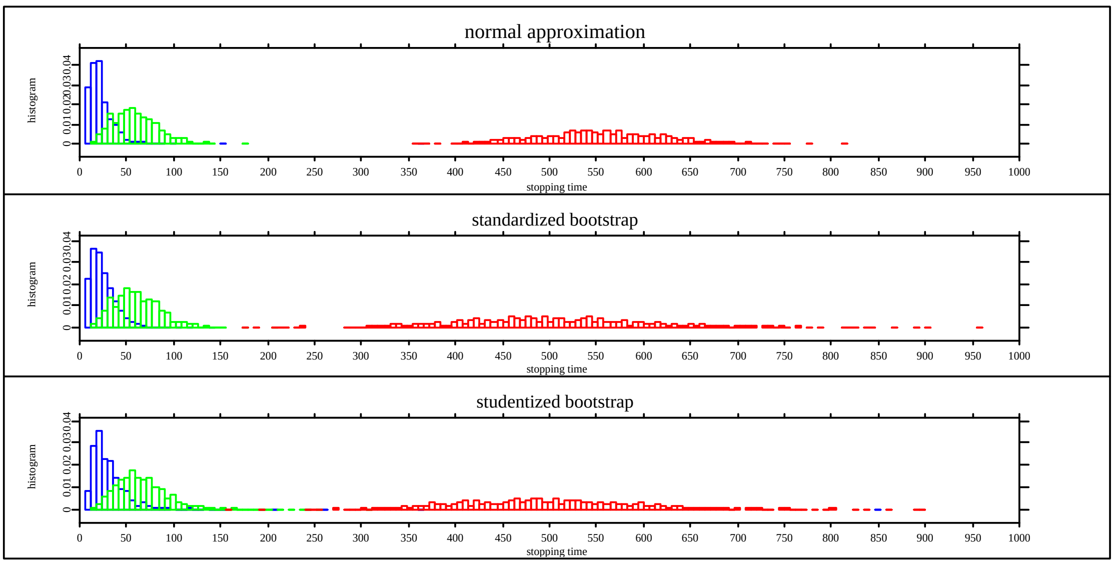


Figure 5.15: N_2 *for Double Exponential, alternative method.*

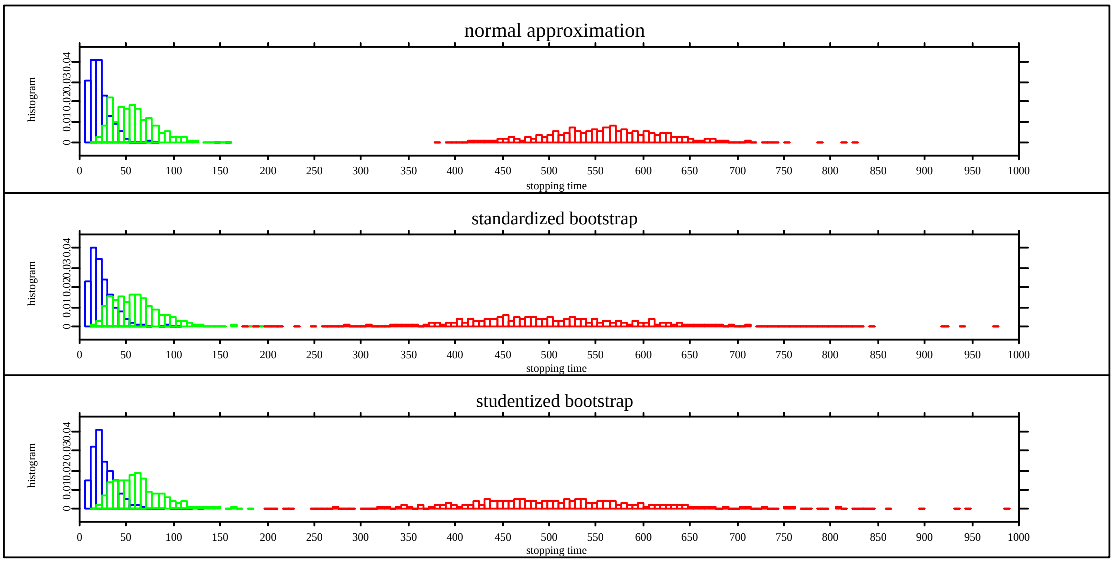


Figure 5.16: $\sqrt{N_2}$ *for Double Exponential distribution.*

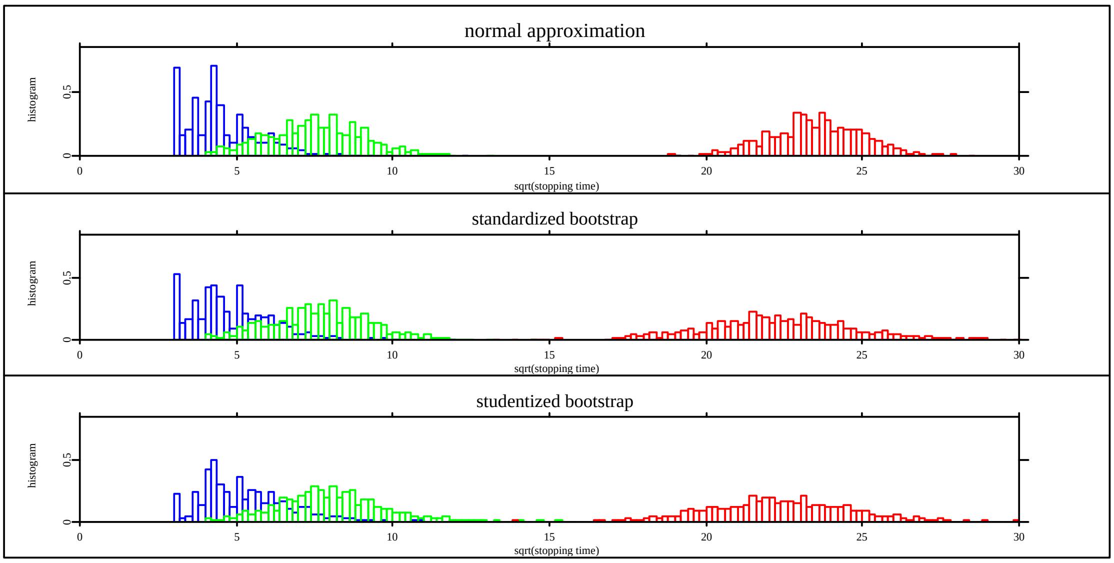
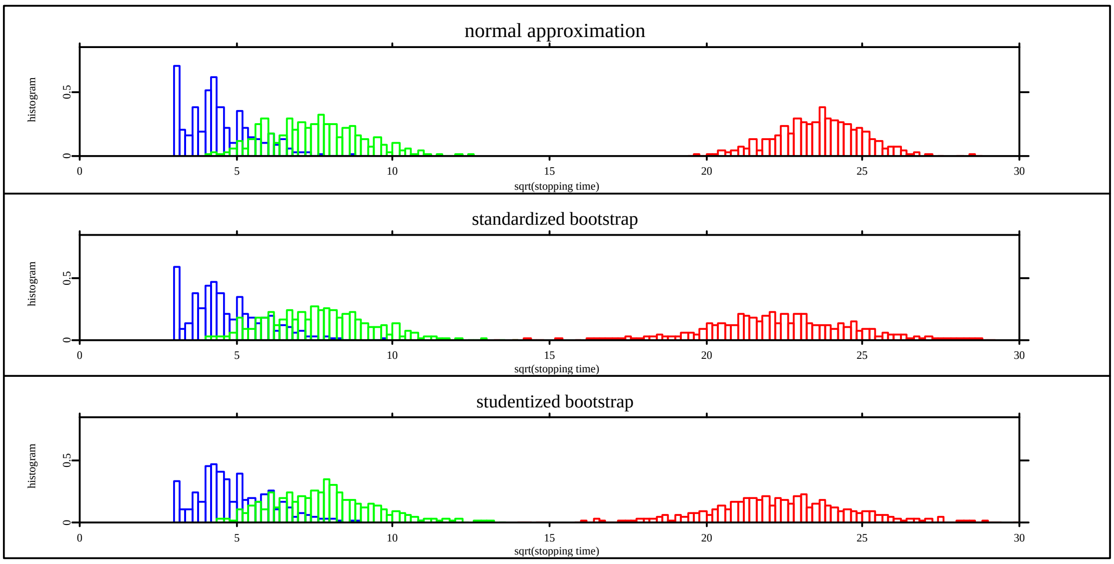


Figure 5.17: $\sqrt{N_2}$ *for Double Exponential, alternative method.*



The results of simulations for the Double Exponential distribution are given in Tables 5.8 and 5.9 and graphically displayed in Figures 5.14–5.17.

From the tables, it seems that the methods based on bootstrap critical points are more reliable than the method based on the normal critical points. In all cases, the coverage probabilities for the bootstrap based methods lie closer to the 0.95. The method based on normal approximation seems to underestimate the optimal sample size for $d = 0.5$ and $d = 0.3$ and to overestimate it for $d = 0.1$.

The histograms on Figures 5.14–5.17 suggest that (for $d = 0.1$) the procedures based on bootstrap tend to stop earlier than the procedure based on normal critical points, even though the coverage probabilities are better for the bootstrap based procedures. The higher variance of N_2 for the methods based on bootstrap seems to be due to the variation in the bootstrap critical points.

For Double Exponential distribution, the methods based on bootstrap give better results than the method based on normal critical points.

5.5 Mixture of Two Normal Distributions

In this section we simulated observations from a mixture of two Normal distributions. We draw observations from distribution $N(0, 1)$ with probability p and observations from the distribution $N(0, \sigma^2)$ with probability $1 - p$.

Straightforward calculations lead, in the same way as in the previous sections, that for the distribution $N(0,\sigma^2)$, which has distribution function $\Phi(x/\sigma)$,

$$\int_{-\infty}^{\infty} \psi_h^2(x)d\Phi(x/\sigma) = \sigma^2 - 2h\sigma\psi(h/\sigma) + (2h^2 - 2\sigma^2)(1-\Phi(h/\sigma)) \tag{5.10}$$

and

$$\int_{-\infty}^{\infty} \psi_h'(x)d\Phi(x) = 2\Phi(h/\sigma) - 1. \tag{5.11}$$

Denoting the distribution function of the mixture of the normal distributions by $F(.)$, we have that

$$\int_{-\infty}^{\infty} \psi_h^2(x)dF(x)$$

$$= p(1-2h\psi(h) + (2h^2-2)(1-\Phi(h))) + (1-p)(\sigma^2 - 2h\sigma\psi(h/\sigma) + (2h^2-2\sigma^2)(1-\Phi(h/\sigma))) \tag{5.12}$$

and

$$\int_{-\infty}^{\infty} \psi_h'(x)d\Phi(x) = p(2\Phi(h)-1) + (1-p)(2\Phi(h/\sigma)-1). \tag{5.13}$$

This gives that the asymptotically optimal stopping time can be calculated as

$$c_M(d) = \left(\frac{u_{1-\alpha/2}}{d}\right)^2$$

$$\times \frac{p(1-2h\psi(h) + (2h^2-2)(1-\Phi(h))) + (1-p)(\sigma^2 - 2h\sigma\psi(h/\sigma) + (2h^2-2\sigma^2)(1-\Phi(h/\sigma)))}{[p(2\Phi(h)-1) + (1-p)(2\Phi(h/\sigma)-1)]^2}. \tag{5.14}$$

In our simulations, we have drawn observations from the distribution with distribution function

$$F(x) = 0.95\Phi(x) + 0.05\Phi(x/3),$$

i.e., observations coming from $N(0,1)$ were contaminated by 5% of observations which came from $N(0,3)$.

The asymptotically optimal sample sizes for this situation are given in Appendix B in Tables B.4–B.9.

The results of simulations are given, in the same format as before, in Tables 5.10 and 5.11. The histograms of N_2 and $\sqrt{N_2}$ are plotted in Figures 5.18–5.21.

From the tables, we see that the coverage probabilities of the method based on normal approximation lie below the desired value, whereas the coverage probabilities of the bootstrap based methods lie in all cases closer to 0.95. From the histograms on the figures we can see that the methods based on bootstrap tend to stop later.

The behaviour of the method based on normal approximations is getting even worse if we decrease the value of h in the score function as can be seen from the tables and figures in Appendix C.

We conclude that also for contaminated Normal distribution, the methods based on bootstrap work better than the method based on normal critical points and can be recommended.

Table 5.10: *Mixture of Normal distributions, Huber's ψ with $h = 1.5$, $\gamma = 1/3$, $k = 1/2$*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	18	N	17.21	16	0.925
		A	20.33	18	0.950
		U	26.12	22	0.957
0.30	50	N	45.77	44	0.905
		A	54.95	51	0.916
		U	62.52	58	0.934
0.10	446	N	438.15	438	0.924
		A	526.74	498	0.945
		U	532.04	502.5	0.944

Table 5.11: *Mixture of Normal Distributions, Huber's ψ with $h = 1.5$, $\gamma = 1/3$, $k = 1/2$, alternative method*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	18	N	17.24	16	0.925
		A	20.06	18	0.939
		U	23.55	20	0.949
0.30	50	N	45.54	44	0.914
		A	53.18	49	0.928
		U	58.45	54	0.935
0.10	446	N	442.91	443	0.927
		A	528.00	501	0.940
		U	533.63	505	0.941

Figure 5.18: *N_2 for Mixture of Normal Distributions.*

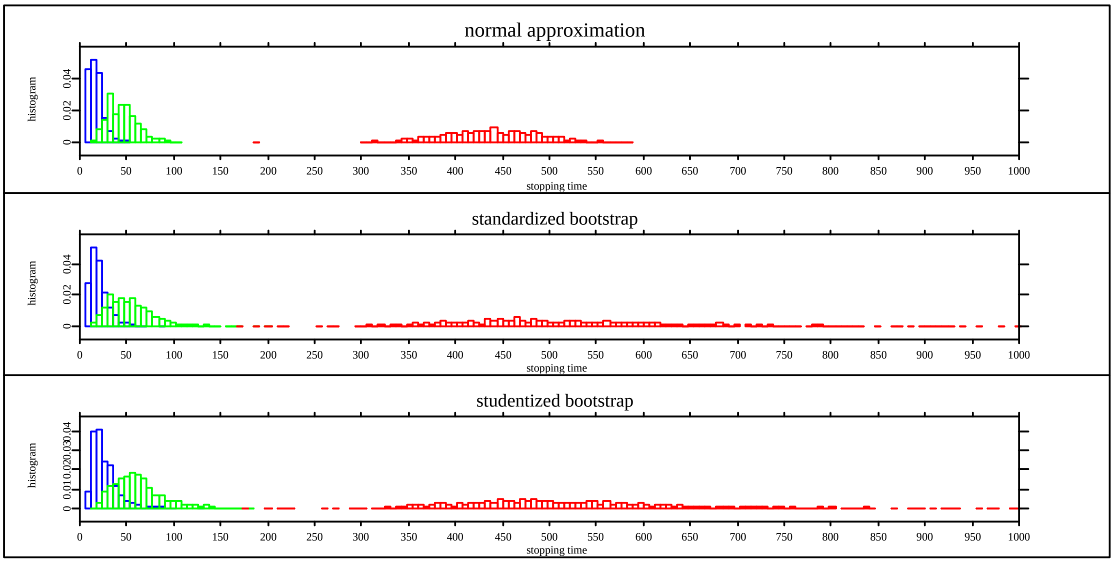


Figure 5.19: *N_2 for Mixture of Normal Distributions, alternative method.*

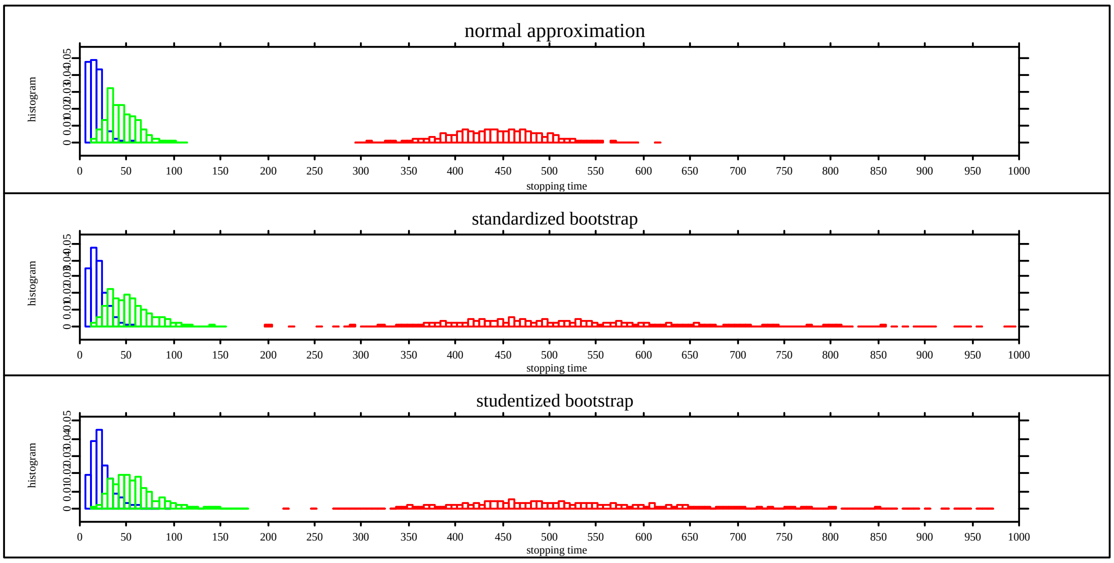


Figure 5.20: $\sqrt{N_2}$ *for Mixture of Normal Distributions.*

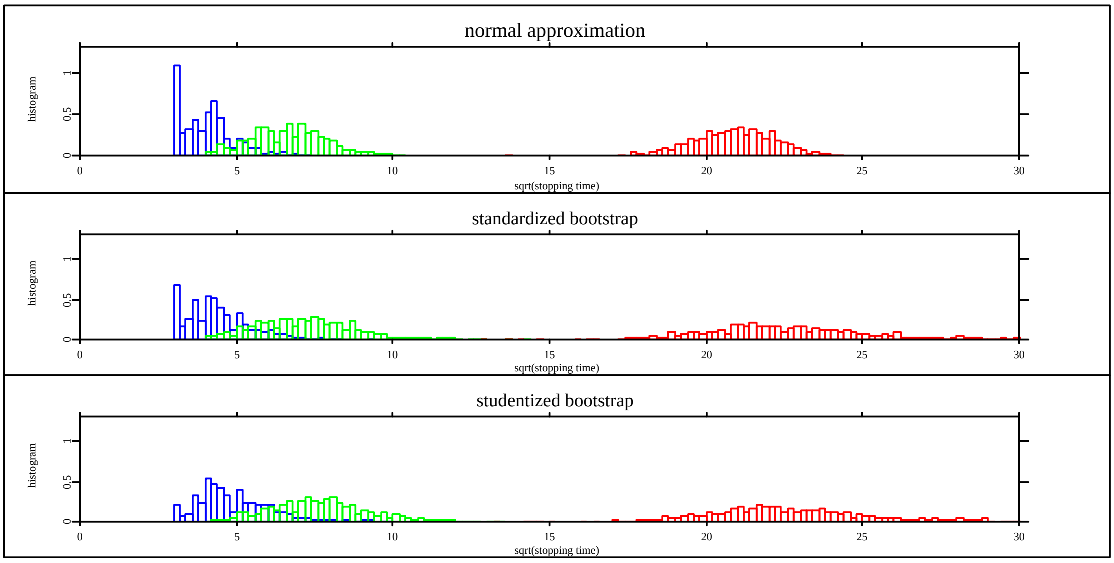
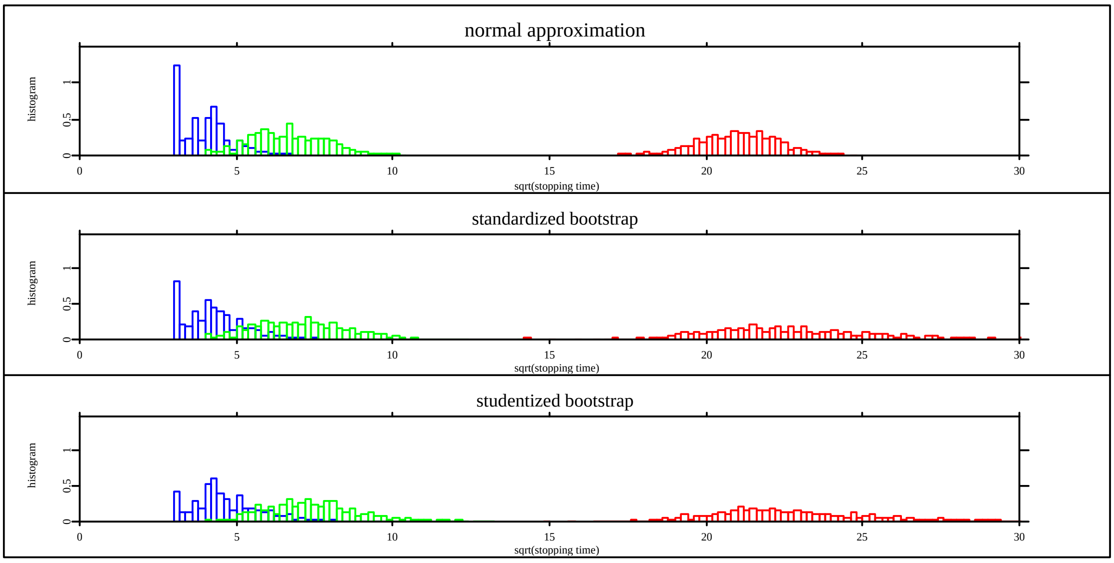


Figure 5.21: $\sqrt{N_2}$ *for Mixture of Normal Distributions, alternative method.*



Chapter 6

Conclusions and Open Questions

The procedure investigated in the previous chapters seems to work well. Unfortunatelly, it can be used only for the simple case of parameters of location. Some possibilities of adapting the procedure to some more general and more complicated situations are outlined in this chapter.

6.1 Linear Model

In this section, we will consider the general linear model

$$\mathbf{Y}_n = \mathbf{X}_n \underset{\sim}{\Theta} + \underset{\sim}{\varepsilon}_n, \tag{6.1}$$

where $\mathbf{Y}_n$ is the vector of observations of the dependent variable, $\mathbf{X}_n$ is the $(n \times p)$ design matrix with $rank(\mathbf{X}_n) = p < n$, $\underset{\sim}{\Theta}$ is the $p-$dimensional vector of the (unknown) regression coefficients, and σ is dispersion parameter of the vector $\underset{\sim}{\varepsilon}_n$ of random errors ε_i which are i.i.d. random variables with common distribution function $F(.)$.

Least squares estimate $\hat{\underset{\sim}{\Theta}}_n$ of the vector of parameters Θ can be obtained as

$$\hat{\underset{\sim}{\Theta}}_n = (\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'\mathbf{Y}_n. \tag{6.2}$$

The variance σ^2 of the random errors ε_i can be estimated by

$$\hat{\sigma}_n^2 = \frac{1}{n}\mathbf{Y}_n' \left\{ I_n - \mathbf{X}_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n' \right\} \mathbf{Y}_n. \tag{6.3}$$

Under the assumptions of normality and assuming that the matrix $\mathbf{X}_n'\mathbf{X}_n$ is not singular, we obtain that

$$\sqrt{n}\left(\hat{\underset{\sim}{\Theta}}_n - \Theta\right) \sim N\left(0, \sigma(\mathbf{X}_n'\mathbf{X}_n)^{-1}\right),$$

and

$$\frac{(n-p)\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2_{n-p},$$

and $\hat{\underset{\sim}{\Theta}}_n$ is independent of $\hat{\sigma}_n^2$. Thus the $1-\alpha$ confidence ellipsoid for Θ is

$$L_n(\alpha) = \left\{ x : \frac{(\hat{\underset{\sim}{\Theta}}_n - x)'(\mathbf{X}_n'\mathbf{X}_n)(\hat{\underset{\sim}{\Theta}}_n - x)}{p\hat{\sigma}_n^2} \leq F_{\alpha;p,n-p} \right\}. \tag{6.4}$$

The maximum diameter of $L_n(\alpha)$ is equal to

$$\frac{2\sqrt{pF_{\alpha;p,n-p}\hat{\sigma}_n}}{\sqrt{n}}\sqrt{\lambda_n},$$

where λ_n is the maximum eigenvalue of $n(\mathbf{X}_n'\mathbf{X}_n)^{-1}$. Fixing $0 < d < \infty$, we can use the following ellipsoidal confidence region for $\underset{\sim}{\Theta}$:

$$R_n(d) = \left\{x : n^{-1}(\hat{\underset{\sim}{\Theta}}_n - x)'(\mathbf{X}_n'\mathbf{X}_n)(\hat{\underset{\sim}{\Theta}}_n - x) \leq d^2\right\}. \tag{6.5}$$

The coverage probability of $R_n(d)$ is easily seen to be equal to

$$P\{R_n(d) \ni \underset{\sim}{\Theta}\} = P\left\{\chi_p^2 \leq \frac{d^2 n}{\sigma^2}\right\}.$$

Equating this probability to $1-\alpha$ yields the optimal fixed sample size $c(d)$ as

$$c(d) = \frac{\chi_p^2(\alpha)\sigma^2}{d^2}, \tag{6.6}$$

where $\chi_p^2(\alpha)$ denotes the critical value of the χ_p^2 distribution, i.e. $P(\chi_p^2 \leq \chi_p^2(\alpha)) = 1-\alpha$.

The asymptotically optimal fixed sample size $c(d)$ given by (6.6) depends on the unknown quantity σ^2. Similarly as for the parameter of location, we need a sequential procedure in order to obtain a confidence region of type (6.5) with prescribed coverage probability $1-\alpha$. To achieve this goal, the following three-stage procedure was suggested by Mukhopadhyay and Abid (1986).

In the first stage, we fix parameter $\gamma > 0$ which controls the starting sample size, and we draw

$$m = m(d) = \max\left\{p+1, \left[\left(\frac{\chi_p^2(\alpha)}{d^2}\right)^{1/1+\gamma}\right]^\circ\right\} \tag{6.7}$$

observations. These observations are used to calculate the estimate $\hat{\sigma}_m^2$ of the parameter σ^2 which we use to determine the number $N_1(d) - m(d)$ of observations drawn in the second stage of the sequential procedure,

$$N_1(d) = \max\left\{m, \left[k\frac{\chi_p^2(\alpha)\hat{\sigma}_m^2}{d^2}\right]^\circ\right\}, \tag{6.8}$$

where $0 < k < 1$ is the parameter controlling the sample size. In the third stage we draw $N_2(d) - N_1(d)$ observations, where $N_2(d)$ is given as

$$N_2(d) = \max\left\{N_1(d), \left[\frac{\chi_p^2(\alpha)\hat{\sigma}_{N_1(d)}^2}{d^2} + 3k^{-1} - \frac{1}{2}[p - \chi_p^2(\alpha)]k^{-1} - \frac{1}{2}\right]^\circ\right\}. \tag{6.9}$$

This is used to construct the ellipsoidal confidence region $R_{N_2}(d)$ given by formula (6.5). The following properties were proved by Mukhopadhyay and Abid (1986).

THEOREM **6.1.1** *Under the assumptions of normality and independence of the random errors ε_i, we have for the three-stage procedure defined by (6.7)—(6.9) that*

$$(i) \qquad P(R_{N_2}(d) \ni \theta) = 1 - \alpha + o(d^2); \tag{6.10}$$

$$(ii) \qquad EN(d) = c(d) + k^{-1}\left(1 - \frac{1}{2}\left[p - \chi_p^2(\alpha)\right]\right) + o(1). \tag{6.11}$$

PROOF: See Mukhopadhyay and Abid (1986) or Ghosh, Mukhopadhyay, and Sen (1997).

6.1.1 Bootstrap Critical Points

The motivation for using bootstrap critical points instead of the critical points of the asymptotic χ^2 distribution is the same as the motivation for using bootstrap critical points instead of the asymptotic normal critical points in the estimation of the parameter of location in the previous chapter.

The χ^2 approximation of the true distribution of the estimates $\hat{\underset{\sim}{\Theta}}_n$ does not have to be accurate if the distribution of errors is not normal or if the sample size is small. In these situations, it makes sense to consider the bootstrap approximation of the distribution of the estimates. The bootstrap based on residuals was proposed by Efron (1979). Other, more sophisticated, approaches to the bootstrap of regression coefficients can be found e.g. in the monograph Shao and Tu (1995).

Let the unknown vector of the regression coefficients $\underset{\sim}{\Theta}$ be estimated by the least squares estimator $\hat{\underset{\sim}{\Theta}}_n$, see (6.2). Denote by $\mathbf{r}_n = (r_{1n}, r_{2n}, \ldots, r_{nn})'$ the vector of residuals, i.e.,

$$\mathbf{r}_n = \mathbf{Y}_n - \mathbf{X}_n \hat{\underset{\sim}{\Theta}}_n. \tag{6.12}$$

Define the empirical distribution function of the centered residuals

$$F_{C,n}(x) = \frac{1}{n} \sum_{i=1}^{n} I(r_i - \bar{r}_n < x). \tag{6.13}$$

The true distribution of the estimate $\hat{\underset{\sim}{\Theta}}_n$ is now approximated by the conditional distribution of the bootstrap estimate $\hat{\underset{\sim}{\Theta}}{}^*_n$, which is obtained as

$$\hat{\underset{\sim}{\Theta}}{}^*_n = (\mathbf{X}_n' \mathbf{X}_n)^{-1} \mathbf{X}_n' \mathbf{Y}_n^*, \tag{6.14}$$

where the vector of observations was obtained as

$$\mathbf{Y}_n^* = \mathbf{X}_n \underset{\sim}{\Theta} + \underset{\sim}{\varepsilon}{}^*_n, \tag{6.15}$$

with $\underset{\sim}{\varepsilon}{}^*_n$ denoting the vector of i.i.d. random variables with common distribution function $F_{C,n}$ given by formula (6.13). It is easy to see that

$$E^*(\hat{\underset{\sim}{\Theta}}{}^*_n) = \hat{\underset{\sim}{\Theta}}_n,$$

and

$$Var^*(\hat{\underset{\sim}{\Theta}}{}^*_n) = \hat{\sigma}_n^2 (\mathbf{X}_n' \mathbf{X}_n)^{-1},$$

where $\hat{\sigma}_n^2 = Var^*(\varepsilon_i^*) = n^{-1} \sum_{i=1}^{n} (r_i - \bar{r}_n)^2$. The asymptotic properties of $\hat{\underset{\sim}{\Theta}}{}^*_n$ will be stated in Theorem 6.1.2, but before that we have to introduce Mallows' distance (Mallows 1972). The definition and basic properties can be found also in Bickel and Freedman (1981) and Shao and Tu (1995). For two distributions $\mathcal{H}$ and $\mathcal{G}$, their Mallows' distance is

$$\tilde{\rho}_r(\mathcal{H}, \mathcal{G}) = \inf_{\mathcal{T}_{X,Y}} (E\|X - Y\|^r)^{1/r},$$

where $\mathcal{T}_{X,Y}$ is the collection of all possible joint distributions (X, Y) whose marginal distributions are $\mathcal{H}$ and $\mathcal{G}$, respectively.

The Mallows' distance has some appealing properties and is often used in order to establish the consistency of a bootstrap estimator. One of basic results claims that $\tilde{\rho}_r(\mathcal{G}_n, \mathcal{G}) \to 0$ implies the convergence in distribution $\mathcal{G}_n \xrightarrow{\mathcal{D}} \mathcal{G}$.

THEOREM **6.1.2** *Assume that and that the ε_i's are i.i.d. random variables and that also*

$$(\mathbf{X}_n'\mathbf{X}_n) \to \infty, \qquad h_{max} = \max_{1\le i\le n} x_i'(\mathbf{X}'\mathbf{X})x_i \to 0.$$

Then

$$\tilde{\rho}_2(\mathcal{L}_n^*, \mathcal{L}_n) \to 0 \qquad [P] \ a.s., \tag{6.16}$$

where the symbols $\mathcal{L}_n$ and $\mathcal{L}_n^*$ denote the distribution of $(\hat{\underset{\sim}{\Theta}}_n - \underset{\sim}{\Theta})'(\mathbf{X}_n'\mathbf{X}_n)(\hat{\underset{\sim}{\Theta}}_n - \underset{\sim}{\Theta})$ and the bootstrap distribution of $(\hat{\underset{\sim}{\Theta}}_n^* - \hat{\underset{\sim}{\Theta}}_n)'(\mathbf{X}_n'\mathbf{X}_n)(\hat{\underset{\sim}{\Theta}}_n^* - \hat{\underset{\sim}{\Theta}}_n)$, respectively.

PROOF: See the proof of Theorem 7.6 in Shao and Tu (1995).

□

6.1.2 Three-stage Sequential Procedure Based on Bootstrap

In the first stage we draw

$$m = m(d) = \max\left\{ p+1, \left[\left(\frac{\chi_p^2(\alpha)}{d^2} \right)^{1/1+\gamma} \right]^{\circ} \right\} \tag{6.17}$$

observations.

In the next stages of the sequential procedure, we suggest to replace the asymptotic critical points $\chi_p^2(\alpha)$ by the bootstrap critical points. Using the critical points of

$$(\hat{\underset{\sim}{\Theta}}_n^* - \hat{\underset{\sim}{\Theta}}_n)' \frac{(\mathbf{X}_n'\mathbf{X}_n)}{\hat{\sigma}_n^2} (\hat{\underset{\sim}{\Theta}}_n^* - \hat{\underset{\sim}{\Theta}}_n),$$

we get the standardized bootstrap critical points $\xi_m^{A2}(\alpha)$. By using critical points of

$$(\hat{\underset{\sim}{\Theta}}_n^* - \hat{\underset{\sim}{\Theta}}_n)' \frac{(\mathbf{X}_n'\mathbf{X}_n)}{\hat{\sigma}_n^{2*}} (\hat{\underset{\sim}{\Theta}}_n^* - \hat{\underset{\sim}{\Theta}}_n),$$

we obtain the studentized bootstrap critical points $\xi_m^{U2}(\alpha)$.

The intermediate sample size is thus given as

$$N_1(d) = \max\left\{ m, \left[k \left(\frac{\xi^A(\alpha)\hat{\sigma}_m}{d} \right)^2 \right]^{\circ} \right\}, \tag{6.18}$$

where $0 < k < 1$ is the parameter controlling the sample size and where $\xi^*(\alpha)$ denotes the bootstrap critical point. The final sample size $N_2(d)$ is given as

$$N_2(d) = \max\left\{ N_1(d), \left[\left(\frac{\xi^A(\alpha)\hat{\sigma}_{N_1(d)}}{d} \right)^2 \right]^{\circ} \right\}. \tag{6.19}$$

These observations are used to construct the ellipsoidal confidence region $R_{N_2}(d)$ given by formula (6.5).

THEOREM **6.1.3** *Assume that and that the ε_i's are i.i.d. random variables with zero mean and variance $0 < \sigma < \infty$ and that also*

$$(\mathbf{X}_n'\mathbf{X}_n) \to \infty, \qquad h_{max} = \max_{1\le i\le n} x_i'(\mathbf{X}'\mathbf{X})x_i \to 0,$$

and there exists n_0 such that for all $n \ge n_0$ we have

$$\operatorname{rank} \frac{1}{n}(\mathbf{X}_n'\mathbf{X}_n) = p.$$

Then we have for the three-stage procedure defined by (6.17)—(6.19) the following:

$$(i) \qquad \lim_{d\to 0+} N_2(d) = \infty \qquad [P] \ a.s., \tag{6.20}$$

$$(ii) \qquad \lim_{d\to 0+} \frac{N_2(d)}{c(d)} = 1 \qquad [P] \ a.s., \tag{6.21}$$

$$(iii) \qquad \lim_{d\to 0+} P(R_{N_2}(d) \ni \Theta) = 1 - \alpha, \tag{6.22}$$

$$(iv) \qquad \lim_{d\to 0+} E\left(\frac{N_2(d)}{c(d)}\right) = 1, \ \textit{if there exists } \delta > 0 \textit{ such that } E\varepsilon_i^{2+\delta} < \infty. \tag{6.23}$$

PROOF: Denote by $\mathcal{R}_n^A(x)$ the distribution function of the standardized bootstrap statistics. By Theorem 6.1.2 we have that

$$\lim_{n\to\infty} \mathcal{R}_n^A(x) = K_p^2(x) \qquad [P] \ a.s., \tag{6.24}$$

where $K_p^2(x)$ denotes the distribution function of the χ_p^2 distribution.

Recall that $\xi_n^{A2}(\alpha)$ denotes the $1-\alpha$ quantile of the distribution of

$$(\hat{\underset{\sim}{\Theta}}_n^* - \hat{\underset{\sim}{\Theta}}_n)' \frac{(\mathbf{X}_n'\mathbf{X}_n)}{\hat{\sigma}_n^2} (\hat{\underset{\sim}{\Theta}}_n^* - \hat{\underset{\sim}{\Theta}}_n).$$

By the strict monotonicity and continuity of $\chi_p^2(x)$ and by e.g. Lemma 1.5.6 in Serfling (1980) we have for all $t \in (0,1)$

$$\lim_{n\to\infty} \xi_n^{A2}(t) = \chi_p^2(1-t) \qquad [P] \ a.s. \tag{6.25}$$

where the symbol $\chi_p^2(1-t)$ denotes the $1-t$ quantile of the χ_p^2 distribution.

Notice that $\lim_{d\to 0+} m(d) = \infty$ and therefore

$$\lim_{d\to 0+} \xi_{m(d)}^{A2}(\alpha) = \chi_p^2(1-\alpha) \qquad [P] \ a.s. \tag{6.26}$$

This, together with consistence of $\hat{\sigma}_{m(d)}^2$ and the definition of $N_1(d)$ implies that

$$\lim_{d\to 0+} N_1(d) = \infty \qquad [P] \ a.s. \tag{6.27}$$

which in turn implies that also

$$\lim_{d\to 0+} N_2(d) = \infty \qquad [P] \ a.s. \tag{6.28}$$

The definition of the stopping time $N_2(d)$ implies following inequalities.

$$\frac{\xi_{N_1}^{A2}(\alpha)\hat{\sigma}_{N_1}^2}{d^2} < N_2(d) \leq \frac{\xi_{N_1}^{A2}(\alpha)\hat{\sigma^2}_{N_1}}{d^2} + 1 + N_1(d) I\left[\xi_{N_1}^{A2}(\alpha)\hat{\sigma}_{N_1}^2 \leq k\xi_m^{A2}(\alpha)\hat{\sigma}_m^2 + md^2 + d^2\right] \tag{6.29}$$

Similarly, as in the proof of Theorem 4.4.1, it can be shown that

$$\lim_{d\to 0} I\left[\xi_{N_1}^{A2}(\alpha)\hat{\sigma}_{N_1}^2 \leq k\xi_m^{A2}(\alpha)\hat{\sigma}_m^2 + md^2 + d^2\right] = 0 \qquad [P]\ a.s. \tag{6.30}$$

Combining (6.29) and (6.30) proves part (ii).

Part (iii) follows from the Slutzky Theorem, Anscombe Theorem and part (ii).

In order to establish part (iv) of the theorem, it suffices to verify the uniform integrability of the set $\{N_2(d)d^2\}_{d>0}$. Similarly, as in the proof of the Theorem 4.4.1, it is sufficient to verify the convergence of the series

$$\sum_{l=1}^{\infty} \sup_{0<d<d_0} P\{N_2(d)d^2 > l\} \tag{6.31}$$

for some $d_0 > 0$.

We can choose d_0 such that for every $0 < d < d_0$ we have

$$\left(\left[\left(\frac{u_{1-\alpha/2}}{d}\right)^{2/(1+\gamma)}\right]^{\circ} + 1\right) d^2 \leq l \tag{6.32}$$

and

$$2d^2 \leq l \tag{6.33}$$

which implies that

$$\begin{aligned} &P(N_2(d)d^2 > l) \\ &\leq P\left\{\left(\left[\frac{(\xi_{N_1}^A(\alpha)\hat{\sigma}_{N_1})^2}{d^2}\right]^{\circ} + 1\right) d^2 > l\right\} + P\left\{\left(\left[\frac{k(\xi_m^A(\alpha)\hat{\sigma}_m)^2}{d^2}\right]^{\circ} + 1\right) d^2 > l\right\} \\ &= P_{1,l}(d) + P_{2,l}(d). \end{aligned} \tag{6.34}$$

Let us first deal with the second probability.

$$\begin{aligned} P_{2,l}(d) &\leq P\left\{\xi_m^A(\alpha)\hat{\sigma}_m > \sqrt{\frac{l-d^2}{k}}\right\} \leq P\left\{\xi_m^A(\alpha)\hat{\sigma}_m > \sqrt{\frac{l}{2k}}\right\} \\ &= P\left\{P^*\left((\hat{\underset{\sim}{\Theta}}{}_m^* - \hat{\underset{\sim}{\Theta}}{}_m)'(\mathbf{X}_m'\mathbf{X}_m)(\hat{\underset{\sim}{\Theta}}{}_m^* - \hat{\underset{\sim}{\Theta}}{}_m) > \frac{l}{2k}\right) > \alpha\right\} \\ &\leq P\left\{\frac{2k}{l}E^*\left((\hat{\underset{\sim}{\Theta}}{}_m^* - \hat{\underset{\sim}{\Theta}}{}_m)'(\mathbf{X}_m'\mathbf{X}_m)(\hat{\underset{\sim}{\Theta}}{}_m^* - \hat{\underset{\sim}{\Theta}}{}_m)\right) > \alpha\right\} \\ &= P\left\{\frac{2k}{l}\hat{\sigma}_m^2 > \alpha\right\} \leq \left(\frac{2k}{\alpha l}\right)^q E\hat{\sigma}_m^{2q}. \end{aligned} \tag{6.35}$$

Choosing $q > 1$ (while keeping $2q < 2 + \delta$) guarantees the convergence of the series

$$\sum_{l=1}^{\infty} \sup_{0<d<d_0} P_{2,l}(d).$$

The convergence of the series containing terms $P_{2,l}(d)$ follows in the same way as in the proof of Theorem 4.4.1.

□

6.2 Robust Regression

The results of Chapter 4 can be also extended to the models of robust regression. The procedure leading to confidence regions with fixed maximal diameter would be very similar to the procedure for the Least Squares estimates described in the previous Section 6.1.

The existing results on the bootstrap on M-estimators of the parameters of linear model, see Lahiri (1992) or Karabulut and Lahiri (1997), require very strong conditions. One of the challenges in this area would be to find some weaker conditions which would be still sufficient for establishing the Edgeworth expansions which are needed for the proof of the desired asymptotic properties of the three-stage sequential procedure.

6.3 GM-statistics

Another exciting possibility of developing the ideas of Chapter 4 is to generalize the procedure for the generalized M-estimators, the so-called GM-statistics, which were first mentioned by Serfling (1984).

Recall that the M-estimators of location are defined as a solution of the equation

$$\sum_{i=1}^{n} \psi(X_i - t) = 0.$$

In the definition of GM-statistics, the observations X_i are replaced by the U-statistics, i.e., the GM-statistics is defined as a solution of the equation

$$\sum_{\{i_1,\ldots,i_m\} \in C_m^n} \psi\left(h(X_{i_1}, \ldots, X_{i_m}) - t\right) = 0,$$

where $C_m^n = \{(i_1, \ldots, i_m) \in N^n | 1 \leq j_1 < \cdots < j_m \leq n\}$ and where $h(.)$ is the kernel of degree m of the U-statistics.

Clearly, the GM-statistics cover both the U-statistics and M-estimators. An example of GM-statistics is the generalized Hodges-Lehmann estimator

$$\mathrm{med}_{1 \leq i_1 < \cdots < i_m \leq n} \frac{1}{m}(X_{i_1} + \cdots + X_{i_m}).$$

A systematic account of (fully sequential) fixed-width confidence intervals for GM-statistics is given in Aerts (1988).

6.4 Multi-stage Procedures with Lower Bound for Variance

Mukhopadhyay and Duggan (1997) suggested to improve the asymptotic properties of the two-stage procedure by adding the assumption that there exists a lower band for the variance σ^2 of the observations, i.e.,

$$0 < \sigma_L^2 < \sigma^2.$$

The procedure goes as follows. In the first stage we draw

$$m(d) = \max\left\{2, \left[\frac{\sigma_L^2 u_{1-\alpha/2}^2}{d^2}\right]^\circ + 1\right\}$$

observations. In the second stage, we add $N(d) - m(d)$ observations, where $N(d)$ is given by (1.11). Mukhopadhyay and Duggan (1997) showed that this two-stage procedure is second order efficient.

A drawback of this procedure might be the assumption concerning the lower bound σ_L^2. It is clearly desirable to choose σ_L^2 as close to the real value of σ^2 as possible. On the other hand, we have to be very careful that we do not choose σ_L^2 too large as we might easily overestimate the optimal sample size. The solution of this dilemma can be found by means of three-stage procedure.

We keep the assumption that $\sigma_L^2 > \sigma^2$ and we draw

$$m(d) = \max\left\{2, \left[\frac{\sigma_L^2 u_{1-\alpha/2}^2}{d^2}\right]^\circ + 1\right\}$$

observations in the first stage as before. In the second stage, we draw $N_1(d) - m(d)$ observations,

$$N_1(d) = \max\left\{m(d), \left[k\frac{\hat{\sigma}_m^2 u_{1-\alpha/2}^2}{d^2}\right]^\circ + 1\right\},$$

where $0 < k < 1$ is the parameter controlling the intermediate sample size. In the third stage, we draw additional $N_2(d) - N_1(d)$ observations,

$$N_2(d) = \max\left\{N_1(d), \left[k\frac{\hat{\sigma}_m^2 u_{1-\alpha/2}^2}{d^2}\right]^\circ + 1\right\}.$$

With this three-stage procedure, we can choose σ_L^2 very far from σ^2 and still obtain good results.

Another advantage of this procedure is that we do not have to take care about the tuning parameter γ which controls the sample size in the first stage. This parameter is replaced by the more natural condition $\sigma^2 > \sigma_L^2$.

The generalization of this procedure to the procedure based on bootstrap critical points and to M-estimators is straightforward and all theoretical results of Chapter 4 are still valid.

Appendix A

Useful Theorems and Lemmas

In this appendix we collect basic probability inequalities and theorems which were used in the proofs. Most of them can be found in any textbook concerning theory of probability, see e.g. Feller (1966), Loève (1977), Serfling (1980), or Štěpán (1987).

A.1 Basic Inequalities

LEMMA **A.1.1** *(c_r inequality) $E|X+Y|^r \leq c_r E|X|^r + c_r E|Y|^r$, where $c_r = 1$ or 2^{r-1} according as $r \leq 1$ or $r \geq 1$.*

LEMMA **A.1.2** *(Hölder inequality) $E|XY| \leq E^{1/r}|X|^r E^{1/s}|Y|^s$, where $r > 1$ and $1/r + 1/s = 1$.*

LEMMA **A.1.3** *(Minkowski inequality) If $r \geq 1$, then*

$$E^{1/r}|X+Z|^r \leq E^{1/r}|X|^r + E^{1/r}|Z|^r$$

LEMMA **A.1.4** *(Schwarz inequality) $E^2|XY| \leq E|X|^2 E|Y|^2$.*

LEMMA **A.1.5** *(Jensen inequality) If g is convex and EX finite, then $g(EX) \leq Eg(X)$.*

A.2 Probability Inequalities

LEMMA **A.2.1** *(Basic inequality) Let X be an arbitrary random variables and let g on $\Re$ be a nonnegative Borel function.*

If g is even and nondecreasing on $[0, +\infty)$, then for every $a \geq 0$

$$\frac{Eg(X) - g(a)}{a.s.\sup g(X)} \leq P[|X| \geq a] \leq \frac{Eg(X)}{g(a)}.$$

If g is nondecreasing on $\Re$, then the middle term is replaced by $P[X \geq a]$, where a is an arbitrary number.

LEMMA **A.2.2** *(Markov inequality)*

$$P[|X| \geq a] \leq \frac{E|X|^r}{a^r}.$$

LEMMA **A.2.3** *(Tchebyschev inequality)*

$$P[|X| \geq a] \leq \frac{E|X|^2}{a^2}.$$

LEMMA **A.2.4** *(Kolmogorov inequality) If $X_1, \ldots, X_n$ are independent and such that $EX_i^2 < \infty$, then for every $\varepsilon > 0$*

$$P\left(\max_{1\leq k\leq n} |S_k| \geq \varepsilon\right) \leq \varepsilon^{-2} \sum_{k=1}^{n} Var\ X_k, \tag{A.1}$$

where $S_k = \sum_{j=1}^{k}(X_j - EX_j)$.

LEMMA **A.2.5** *(Hoeffding inequality) If $X_1, \ldots, X_n$ are independent and $a_i < X_i < b_i$, then for $t > 0$*

$$P\left(\bar{X} - \mu \geq t\right) \leq \exp\left\{-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right\} \tag{A.2}$$

LEMMA **A.2.6** *(Dvoretzky, Kiefer, and Wolfowitz) Let F and F_n be the population and sample distribution function of a sequence of iid random variables $\{X_i\}$ and let D_n denote the Kolmogorov-Smirnov distance, i.e.*

$$D_n = \sup_{x\in\Re} |F_n(x) - F(x)|.$$

Then there exists a finite positive constant C (not depending on F) such that

$$P(D_n > d) \leq C \exp\{-2nd^2\}, \qquad d > 0, \tag{A.3}$$

for all $n = 1, 2, \ldots$.

A.3 CLT

THEOREM **A.3.1** *(Lindeberg–Feller) Let $\{X_i\}$ be independent with means $\{\mu_i\}$, finite variances $\{\sigma_i\}$, and distribution functions $\{F_i\}$. Suppose that $B_n^2 = \sum_1^n \sigma_i$ satisfies*

$$\frac{\sigma_n^2}{B_n^2} \to 0, \quad B_n \to \infty, \quad as\ n \to \infty. \tag{A.4}$$

Then, as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathcal{D}} N\left(\frac{1}{n}\sum_{i=1}^{n}\mu_i, \frac{1}{n^2}B_n^2\right) \tag{A.5}$$

if and only if the Lindeberg condition

$$\frac{\sum_{i=1}^{n}\int_{|t-\mu_i|>\varepsilon B_n}(t-\mu_i)^2dF_i(t)}{B_n^2}\to 0,\ \text{as } n\to\infty,\ \forall\varepsilon>0, \tag{A.6}$$

is satisfied.

PROOF: See Serfling (1980).

□

THEOREM **A.3.2** *(Edgeworth expansion) If F is not a lattice distribution and if the third moment μ_3 exists then*

$$F_n(x)-\Phi(x)-\frac{\mu_3}{6\sigma^3\sqrt{n}}(1-x^2)\varphi(x)=o(n^{-1/2})$$

uniformly for all x, where $F_n(.)$ is the distribution function of $\sqrt{n}\sum_{i=1}^{n}(X_i-\mu)/\sigma$ and where X_i are iid with mean μ and variance σ^2.

PROOF: See Feller (1966).

THEOREM **A.3.3** *(Edgeworth expansion) If Cramer's condition (4.122) holds and if the third moment μ_3 and the fourth moment μ_4 exist then*

$$F_n(x)-\Phi(x)-\frac{\mu_3}{6\sigma^3\sqrt{n}}(1-x^2)\varphi(x)+\left(\frac{\mu_3^2}{72\sigma_2^6}(x^2-1)+\frac{\mu_4-3\sigma^4}{24\sigma^4}(x^3-3x)\right)n^{-1}\varphi'(x)=o(n^{-1})$$

uniformly for all x, where $F_n(.)$ is the distribution function of $\sqrt{n}\sum_{i=1}^{n}(X_i-\mu)/\sigma$ and where X_i are iid with mean μ and variance σ^2.

PROOF: See Feller (1966).

□

THEOREM **A.3.4** *(Berry–Esséen) Let $\{X_i\}$ be iid random variables with mean μ and variance $\sigma^2>0$. Then*

$$\sup_t|G_n(t)-\Phi(t)|\le\frac{33}{4}\frac{E|X_1-\mu|^3}{\sigma^3n^{1/2}},\qquad \text{for all } n,$$

where $G_n(.)$ is the distribution function of the normalized sum

$$\frac{\sum_{i=1}^{n}(X_i-E\sum_{i=1}^{n}X_i)}{\sqrt{Var\sum_{i=1}^{n}X_i}}$$

PROOF: See Theorem 1.9.5 in Serfling (1980).

□

A.4 Random CLT

DEFINITION **A.4.1** *We say that $Y_1, Y_2, \ldots$ are uniformly continuous in probability if and only if for every $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$P\left(\max_{0<k<n\delta} |Y_{n+k} - Y_n| \geq \varepsilon\right) < \varepsilon \tag{A.7}$$

for all $n \geq 1$.

THEOREM **A.4.1** *Suppose that $Y_1, Y_2, \ldots$ are uniformly continuous in probability. Let N_a, $(a > 0)$ be a positive integer-valued random variable for which N_a/a converges in probability to a finite positive constant k. Define n_a as the integer part of ak. Then*

$$Y_{N_a} - Y_{n_a} \xrightarrow{P} 0 \qquad \text{as } n \to \infty.$$

If in addition Y_n converges in distribution to a random variable Y as $n \to \infty$ then also Y_{N_a} converges in distribution to Y as $n \to \infty$.

PROOF: See Theorem 2.7.1 in Ghosh, Mukhopadhyay, and Sen (1997).

□

THEOREM **A.4.2** *(Anscombe) Let $\{X_i\}$ be iid random variables with mean μ and finite variance $\sigma^2 > 0$. Let $\{\nu_n\}$ be a sequence of integer-valued random variables and $\{a_n\}$ a sequence of positive constants tending to ∞, such that*

$$\frac{\nu_m}{a_n} \xrightarrow{P} c$$

for some positive constant c. Then

$$\frac{\sum_{i=1}^{\nu_n}(X_i - \mu)}{\sqrt{\nu_n}} \xrightarrow{D} N(0, \sigma^2).$$

PROOF: See Theorem 1.9.4 in Serfling (1980).

□

Appendix B

Tables

Table B.1: Values of h for Huber's $\psi_h(.)$ score function and percentage of contamination.

h	ε	h	ε	h	ε	h	ε	h	ε	h	ε
0.00	1.000000	0.50	0.441714	1.00	0.142845	1.50	0.037629	2.00	0.008418	2.50	0.001600
0.01	0.987467	0.51	0.432995	1.01	0.139328	1.51	0.036576	2.01	0.008156	2.51	0.001545
0.02	0.974939	0.52	0.424395	1.02	0.135887	1.52	0.035551	2.02	0.007903	2.52	0.001492
0.03	0.962417	0.53	0.415915	1.03	0.132520	1.53	0.034553	2.03	0.007656	2.53	0.001441
0.04	0.949908	0.54	0.407554	1.04	0.129227	1.54	0.033580	2.04	0.007417	2.54	0.001391
0.05	0.937413	0.55	0.399313	1.05	0.126006	1.55	0.032633	2.05	0.007184	2.55	0.001343
0.06	0.924936	0.56	0.391192	1.06	0.122856	1.56	0.031710	2.06	0.006959	2.56	0.001296
0.07	0.912482	0.57	0.383190	1.07	0.119775	1.57	0.030812	2.07	0.006740	2.57	0.001251
0.08	0.900055	0.58	0.375307	1.08	0.116763	1.58	0.029937	2.08	0.006528	2.58	0.001207
0.09	0.887656	0.59	0.367544	1.09	0.113819	1.59	0.029085	2.09	0.006322	2.59	0.001165
0.10	0.875292	0.60	0.359900	1.10	0.110940	1.60	0.028256	2.10	0.006122	2.60	0.001125
0.11	0.862964	0.61	0.352374	1.11	0.108126	1.61	0.027448	2.11	0.005928	2.61	0.001085
0.12	0.850676	0.62	0.344966	1.12	0.105376	1.62	0.026663	2.12	0.005739	2.62	0.001047
0.13	0.838433	0.63	0.337676	1.13	0.102688	1.63	0.025898	2.13	0.005557	2.63	0.001010
0.14	0.826236	0.64	0.330504	1.14	0.100061	1.64	0.025153	2.14	0.005379	2.64	0.000975
0.15	0.814091	0.65	0.323448	1.15	0.097495	1.65	0.024428	2.15	0.005207	2.65	0.000940
0.16	0.801999	0.66	0.316508	1.16	0.094988	1.66	0.023723	2.16	0.005041	2.66	0.000907
0.17	0.789964	0.67	0.309683	1.17	0.092538	1.67	0.023036	2.17	0.004879	2.67	0.000875
0.18	0.777991	0.68	0.302973	1.18	0.090146	1.68	0.022369	2.18	0.004722	2.68	0.000844
0.19	0.766080	0.69	0.296377	1.19	0.087809	1.69	0.021719	2.19	0.004570	2.69	0.000813
0.20	0.754236	0.70	0.289894	1.20	0.085527	1.70	0.021086	2.20	0.004422	2.70	0.000784
0.21	0.742462	0.71	0.283524	1.21	0.083298	1.71	0.020471	2.21	0.004279	2.71	0.000756
0.22	0.730761	0.72	0.277265	1.22	0.081122	1.72	0.019873	2.22	0.004141	2.72	0.000729
0.23	0.719134	0.73	0.271116	1.23	0.078997	1.73	0.019291	2.23	0.004006	2.73	0.000703
0.24	0.707586	0.74	0.265077	1.24	0.076922	1.74	0.018697	2.24	0.003876	2.74	0.000678
0.25	0.696119	0.75	0.259146	1.25	0.074897	1.75	0.018147	2.25	0.003749	2.75	0.000653
0.26	0.684735	0.76	0.253323	1.26	0.072921	1.76	0.017611	2.26	0.003627	2.76	0.000630
0.27	0.673437	0.77	0.247607	1.27	0.070991	1.77	0.017090	2.27	0.003508	2.77	0.000607
0.28	0.662228	0.78	0.241995	1.28	0.069108	1.78	0.016584	2.28	0.003393	2.78	0.000585
0.29	0.651109	0.79	0.236489	1.29	0.067271	1.79	0.016091	2.29	0.003281	2.79	0.000564
0.30	0.640083	0.80	0.231085	1.30	0.065478	1.80	0.015612	2.30	0.003173	2.80	0.000543
0.31	0.629152	0.81	0.225784	1.31	0.063728	1.81	0.015147	2.31	0.003069	2.81	0.000523
0.32	0.618318	0.82	0.220583	1.32	0.062022	1.82	0.014694	2.32	0.002967	2.82	0.000504
0.33	0.607583	0.83	0.215483	1.33	0.060357	1.83	0.014254	2.33	0.002869	2.83	0.000486
0.34	0.596949	0.84	0.210481	1.34	0.058732	1.84	0.013826	2.34	0.002773	2.84	0.000468
0.35	0.586418	0.85	0.205576	1.35	0.057148	1.85	0.013410	2.35	0.002681	2.85	0.000451
0.36	0.575991	0.86	0.200768	1.36	0.055603	1.86	0.013006	2.36	0.002592	2.86	0.000434
0.37	0.565671	0.87	0.196054	1.37	0.054096	1.87	0.012614	2.37	0.002505	2.87	0.000418
0.38	0.555458	0.88	0.191434	1.38	0.052627	1.88	0.012232	2.38	0.002421	2.88	0.000403
0.39	0.545354	0.89	0.186907	1.39	0.051194	1.89	0.011861	2.39	0.002340	2.89	0.000388
0.40	0.535360	0.90	0.182471	1.40	0.049797	1.90	0.011501	2.40	0.002261	2.90	0.000373
0.41	0.525478	0.91	0.178125	1.41	0.048435	1.91	0.011151	2.41	0.002185	2.91	0.000359
0.42	0.515708	0.92	0.173867	1.42	0.047107	1.92	0.010810	2.42	0.002112	2.92	0.000346
0.43	0.506052	0.93	0.169697	1.43	0.045813	1.93	0.010480	2.43	0.002040	2.93	0.000333
0.44	0.496511	0.94	0.165613	1.44	0.044551	1.94	0.010159	2.44	0.001971	2.94	0.000321
0.45	0.487085	0.95	0.161614	1.45	0.043322	1.95	0.009847	2.45	0.001904	2.95	0.000309
0.46	0.477776	0.96	0.157698	1.46	0.042123	1.96	0.009544	2.46	0.001839	2.96	0.000297
0.47	0.468583	0.97	0.153864	1.47	0.040956	1.97	0.009250	2.47	0.001777	2.97	0.000286
0.48	0.459509	0.98	0.150112	1.48	0.039818	1.98	0.008965	2.48	0.001716	2.98	0.000275
0.49	0.450552	0.99	0.146439	1.49	0.038709	1.99	0.008687	2.49	0.001657	2.99	0.000265

Table B.2: Values of starting sample size for $\alpha = 0.05$ and different values of γ.

$d\backslash\gamma$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
1	4	4	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2
0.95	4	4	4	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2
0.9	5	4	4	4	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2
0.85	5	5	4	4	4	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2
0.8	6	5	4	4	4	4	3	3	3	3	3	3	3	3	3	2	2	2	2	2
0.75	6	5	5	4	4	4	4	3	3	3	3	3	3	3	3	3	3	2	2	2
0.7	7	6	5	5	4	4	4	4	3	3	3	3	3	3	3	3	3	3	3	2
0.65	8	7	6	5	5	4	4	4	4	4	3	3	3	3	3	3	3	3	3	3
0.6	9	8	7	6	5	5	5	4	4	4	4	3	3	3	3	3	3	3	3	3
0.55	11	9	8	7	6	5	5	5	4	4	4	4	4	3	3	3	3	3	3	3
0.5	12	10	9	8	7	6	5	5	5	4	4	4	4	4	3	3	3	3	3	3
0.49	13	11	9	8	7	6	6	5	5	4	4	4	4	4	4	3	3	3	3	3
0.48	13	11	9	8	7	6	6	5	5	5	4	4	4	4	4	3	3	3	3	3
0.47	14	11	9	8	7	6	6	5	5	5	4	4	4	4	4	3	3	3	3	3
0.46	14	12	10	8	7	7	6	6	5	5	4	4	4	4	4	4	3	3	3	3
0.45	15	12	10	9	8	7	6	6	5	5	5	4	4	4	4	4	3	3	3	3
0.44	16	13	10	9	8	7	6	6	5	5	5	4	4	4	4	4	4	3	3	3
0.43	16	13	11	9	8	7	6	6	5	5	5	4	4	4	4	4	4	3	3	3
0.42	17	14	11	10	8	7	7	6	6	5	5	5	4	4	4	4	4	4	3	3
0.41	18	14	12	10	9	8	7	6	6	5	5	5	4	4	4	4	4	4	3	3
0.4	18	15	12	10	9	8	7	6	6	5	5	5	4	4	4	4	4	4	3	3
0.39	19	15	12	11	9	8	7	7	6	6	5	5	5	4	4	4	4	4	4	3
0.38	20	16	13	11	9	8	7	7	6	6	5	5	5	4	4	4	4	4	4	3
0.37	21	17	13	11	10	9	8	7	6	6	5	5	5	5	4	4	4	4	4	4
0.36	22	17	14	12	10	9	8	7	6	6	6	5	5	5	4	4	4	4	4	4
0.35	23	18	15	12	10	9	8	7	7	6	6	5	5	5	4	4	4	4	4	4
0.34	25	19	15	13	11	9	8	8	7	6	6	5	5	5	5	4	4	4	4	4
0.33	26	20	16	13	11	10	9	8	7	6	6	6	5	5	5	4	4	4	4	4
0.32	27	21	17	14	12	10	9	8	7	7	6	6	5	5	5	5	4	4	4	4
0.31	29	22	18	14	12	11	9	8	7	7	6	6	5	5	5	5	4	4	4	4
0.3	31	23	18	15	13	11	10	9	8	7	6	6	6	5	5	5	5	4	4	4
0.29	33	25	19	16	13	11	10	9	8	7	7	6	6	5	5	5	5	4	4	4
0.28	35	26	20	17	14	12	10	9	8	7	7	6	6	6	5	5	5	5	4	4
0.27	37	28	22	17	15	12	11	10	9	8	7	7	6	6	5	5	5	5	4	4
0.26	40	29	23	18	15	13	11	10	9	8	7	7	6	6	6	5	5	5	5	4
0.25	43	31	24	19	16	14	12	10	9	8	8	7	6	6	6	5	5	5	5	4
0.24	46	34	26	21	17	14	12	11	10	9	8	7	7	6	6	6	5	5	5	5
0.23	50	36	28	22	18	15	13	11	10	9	8	8	7	6	6	6	5	5	5	5
0.22	54	39	29	23	19	16	14	12	10	9	9	8	7	7	6	6	6	5	5	5
0.21	59	42	32	25	20	17	14	12	11	10	9	8	7	7	6	6	6	5	5	5
0.2	64	45	34	27	21	18	15	13	12	10	9	8	8	7	7	6	6	6	5	5
0.19	70	49	37	29	23	19	16	14	12	11	10	9	8	7	7	7	6	6	5	5
0.18	77	54	40	31	25	20	17	15	13	11	10	9	8	8	7	7	6	6	6	5
0.17	86	59	44	33	27	22	18	16	14	12	11	10	9	8	8	7	7	6	6	6
0.16	96	66	48	36	29	23	20	17	14	13	11	10	9	9	8	7	7	6	6	6
0.15	107	73	53	40	31	25	21	18	15	14	12	11	10	9	8	8	7	7	6	6
0.14	122	82	58	44	34	28	23	19	17	14	13	12	10	10	9	8	8	7	7	6
0.13	139	93	65	49	38	30	25	21	18	16	14	12	11	10	9	9	8	7	7	7
0.12	161	106	74	55	42	33	27	23	19	17	15	13	12	11	10	9	8	8	7	7
0.11	189	122	85	62	47	37	30	25	21	18	16	14	13	12	11	10	9	8	8	7
0.1	224	143	98	71	53	42	34	28	23	20	18	15	14	12	11	10	10	9	8	8
0.09	271	170	115	82	61	48	38	31	26	22	19	17	15	14	12	11	10	10	9	8
0.08	336	207	138	97	72	55	44	35	29	25	22	19	17	15	13	12	11	10	10	9
0.07	428	259	169	117	86	65	51	41	34	28	24	21	19	17	15	13	12	11	10	10
0.06	567	334	214	146	105	79	61	49	40	33	28	24	21	19	17	15	14	13	12	11
0.05	789	453	283	189	134	99	75	59	48	40	33	29	25	22	19	17	16	14	13	12
0.04	1184	657	399	260	180	130	98	76	61	49	41	35	30	26	23	20	18	17	15	14
0.03	1997	1060	621	392	264	186	137	104	82	66	54	45	38	33	29	25	23	20	18	17
0.02	4173	2084	1158	700	452	309	221	164	125	98	79	65	54	46	40	35	30	27	24	22
0.01	14714	6614	3362	1883	1139	734	498	353	259	196	153	122	99	82	69	58	50	44	39	34

Table B.3: Values of starting sample size for $\alpha = 0.01$ and different values of γ.

$d\backslash\gamma$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
1	6	5	5	4	4	4	4	3	3	3	3	3	3	3	3	3	3	2	2	2
0.95	7	6	5	5	4	4	4	4	3	3	3	3	3	3	3	3	3	3	2	2
0.9	7	6	6	5	5	4	4	4	4	3	3	3	3	3	3	3	3	3	3	3
0.85	8	7	6	5	5	4	4	4	4	4	3	3	3	3	3	3	3	3	3	3
0.8	9	8	7	6	5	5	4	4	4	4	4	3	3	3	3	3	3	3	3	3
0.75	10	8	7	6	6	5	5	4	4	4	4	4	3	3	3	3	3	3	3	3
0.7	11	9	8	7	6	6	5	5	4	4	4	4	4	3	3	3	3	3	3	3
0.65	13	10	9	8	7	6	6	5	5	4	4	4	4	4	4	3	3	3	3	3
0.6	15	12	10	9	7	7	6	6	5	5	5	4	4	4	4	4	3	3	3	3
0.55	17	14	11	10	8	7	7	6	6	5	5	5	4	4	4	4	4	4	3	3
0.5	20	16	13	11	9	8	7	7	6	6	5	5	5	4	4	4	4	4	4	3
0.49	21	16	13	11	10	8	8	7	6	6	5	5	5	4	4	4	4	4	4	4
0.48	22	17	14	12	10	9	8	7	6	6	5	5	5	5	4	4	4	4	4	4
0.47	23	18	14	12	10	9	8	7	6	6	6	5	5	5	4	4	4	4	4	4
0.46	23	18	15	12	10	9	8	7	7	6	6	5	5	5	4	4	4	4	4	4
0.45	24	19	15	13	11	9	8	7	7	6	6	5	5	5	5	4	4	4	4	4
0.44	25	20	16	13	11	10	8	8	7	6	6	5	5	5	5	4	4	4	4	4
0.43	26	20	16	13	11	10	9	8	7	6	6	6	5	5	5	4	4	4	4	4
0.42	28	21	17	14	12	10	9	8	7	7	6	6	5	5	5	5	4	4	4	4
0.41	29	22	17	14	12	10	9	8	7	7	6	6	5	5	5	5	4	4	4	4
0.4	30	23	18	15	12	11	9	8	8	7	6	6	6	5	5	5	4	4	4	4
0.39	31	24	19	15	13	11	10	9	8	7	7	6	6	5	5	5	5	4	4	4
0.38	33	25	19	16	13	11	10	9	8	7	7	6	6	5	5	5	5	4	4	4
0.37	35	26	20	16	14	12	10	9	8	7	7	6	6	6	5	5	5	4	4	4
0.36	36	27	21	17	14	12	11	9	8	8	7	6	6	6	5	5	5	5	4	4
0.35	38	28	22	18	15	13	11	10	9	8	7	7	6	6	5	5	5	5	4	4
0.34	40	30	23	19	15	13	11	10	9	8	7	7	6	6	6	5	5	5	5	4
0.33	42	31	24	19	16	14	12	10	9	8	8	7	6	6	6	5	5	5	5	4
0.32	45	33	25	20	17	14	12	11	9	9	8	7	7	6	6	5	5	5	5	5
0.31	47	35	26	21	17	15	13	11	10	9	8	7	7	6	6	6	5	5	5	5
0.3	50	37	28	22	18	15	13	11	10	9	8	8	7	7	6	6	5	5	5	5
0.29	54	39	29	23	19	16	14	12	10	9	9	8	7	7	6	6	6	5	5	5
0.28	57	41	31	24	20	17	14	12	11	10	9	8	7	7	6	6	6	5	5	5
0.27	61	43	33	26	21	17	15	13	11	10	9	8	8	7	7	6	6	6	5	5
0.26	65	46	35	27	22	18	15	13	12	10	9	9	8	7	7	6	6	6	5	5
0.25	70	49	37	28	23	19	16	14	12	11	10	9	8	7	7	7	6	6	5	5
0.24	75	53	39	30	24	20	17	14	13	11	10	9	8	8	7	7	6	6	6	5
0.23	81	57	42	32	26	21	18	15	13	12	10	9	9	8	7	7	6	6	6	6
0.22	88	61	45	34	27	22	19	16	14	12	11	10	9	8	8	7	7	6	6	6
0.21	96	66	48	36	29	23	20	17	14	13	11	10	9	9	8	7	7	6	6	6
0.2	105	71	51	39	31	25	21	18	15	13	12	11	10	9	8	8	7	7	6	6
0.19	115	78	56	42	33	27	22	19	16	14	12	11	10	9	9	8	7	7	7	6
0.18	127	85	60	45	35	28	23	20	17	15	13	12	11	10	9	8	8	7	7	6
0.17	141	93	66	49	38	30	25	21	18	16	14	12	11	10	9	9	8	7	7	7
0.16	157	103	72	53	41	33	27	22	19	17	15	13	12	11	10	9	8	8	7	7
0.15	176	115	80	59	45	35	29	24	20	18	15	14	12	11	10	9	9	8	8	7
0.14	200	129	89	65	49	39	31	26	22	19	17	15	13	12	11	10	9	9	8	7
0.13	229	146	99	72	54	42	34	28	24	20	18	16	14	13	11	10	10	9	8	8
0.12	264	166	112	80	60	47	37	31	26	22	19	17	15	13	12	11	10	9	9	8
0.11	310	192	128	91	67	52	41	34	28	24	21	18	16	14	13	12	11	10	9	9
0.1	368	225	149	104	77	59	46	37	31	26	23	20	17	15	14	13	12	11	10	9
0.09	446	268	175	121	88	67	52	42	35	29	25	22	19	17	15	14	12	11	11	10
0.08	552	326	209	143	103	77	60	48	39	33	28	24	21	19	17	15	14	12	11	11
0.07	703	408	257	173	123	91	70	55	45	37	31	27	23	21	18	17	15	14	13	12
0.06	931	527	326	216	151	110	84	66	53	43	36	31	27	23	21	19	17	15	14	13
0.05	1297	714	431	280	192	139	104	80	64	52	43	37	31	27	24	21	19	17	16	14
0.04	1945	1035	607	384	259	183	135	103	81	65	53	45	38	33	28	25	22	20	18	17
0.03	3281	1672	945	579	379	262	189	141	109	86	70	58	49	41	36	31	28	25	22	20
0.02	6858	3285	1762	1034	651	434	304	221	167	129	103	83	69	58	49	42	37	33	29	26
0.01	24182	10428	5119	2781	1639	1032	687	478	345	258	198	156	125	103	85	72	62	53	46	41

Table B.4: Asymptotically optimal sample size for $\alpha = 0.05$ for Huber's $\psi_{0.5}(.)$ score function.

d	N(0,1)	Cauchy	Double Exponential	0.95 N(0,1) + 0.05 N(0,9)
1	5	9	5	6
0.95	6	10	5	6
0.9	6	11	6	7
0.85	7	13	7	8
0.8	8	14	7	9
0.75	9	16	8	10
0.7	10	18	10	11
0.65	12	21	11	13
0.6	14	25	13	15
0.55	17	30	15	18
0.5	20	36	18	21
0.49	21	37	19	22
0.48	22	39	20	23
0.47	22	40	21	24
0.46	23	42	22	25
0.45	24	44	23	26
0.44	26	46	24	28
0.43	27	48	25	29
0.42	28	50	26	30
0.41	29	53	27	32
0.4	31	55	28	33
0.39	32	58	30	35
0.38	34	61	32	37
0.37	36	65	33	39
0.36	38	68	35	41
0.35	40	72	37	43
0.34	42	77	39	46
0.33	45	81	42	49
0.32	48	86	44	52
0.31	51	92	47	55
0.3	54	98	50	59
0.29	58	105	54	63
0.28	62	113	58	67
0.27	67	121	62	72
0.26	72	131	67	78
0.25	78	141	72	84
0.24	85	153	78	92
0.23	92	167	85	100
0.22	101	182	93	109
0.21	110	200	102	119
0.2	122	220	112	132
0.19	135	244	125	146
0.18	150	272	139	162
0.17	168	305	155	182
0.16	190	344	175	205
0.15	216	391	199	234
0.14	248	449	229	268
0.13	287	521	265	311
0.12	337	611	311	365
0.11	401	727	370	434
0.1	485	879	448	525
0.09	599	1086	553	648
0.08	758	1374	700	820
0.07	990	1794	914	1071
0.06	1348	2442	1244	1457
0.05	1940	3516	1791	2098
0.04	3032	5494	2798	3278
0.03	5389	9767	4974	5827
0.02	12125	21975	11192	13109
0.01	48500	87898	44765	52435

Table B.5: Asymptotically optimal sample size for $\alpha = 0.05$ for Huber's $\psi_1(.)$ score function.

d	N(0,1)	Cauchy	Double Exponential	0.95 N(0,1) + 0.05 N(0,9)
1	5	10	6	5
0.95	5	11	6	6
0.9	6	13	7	6
0.85	6	14	8	7
0.8	7	16	8	8
0.75	8	18	10	9
0.7	9	20	11	10
0.65	11	24	13	12
0.6	12	28	15	13
0.55	15	33	17	16
0.5	18	40	21	19
0.49	18	41	22	20
0.48	19	43	23	21
0.47	20	45	24	22
0.46	21	47	25	23
0.45	22	49	26	24
0.44	22	51	27	25
0.43	24	53	28	26
0.42	25	56	29	27
0.41	26	59	31	28
0.4	27	62	32	30
0.39	28	65	34	31
0.38	30	68	36	33
0.37	32	72	38	35
0.36	33	76	40	36
0.35	35	80	42	39
0.34	37	85	44	41
0.33	40	90	47	43
0.32	42	96	50	46
0.31	45	102	53	49
0.3	48	109	57	52
0.29	51	117	61	56
0.28	55	125	65	60
0.27	59	135	70	64
0.26	63	145	76	69
0.25	69	157	82	75
0.24	74	170	89	81
0.23	81	185	97	89
0.22	88	203	105	97
0.21	97	222	116	106
0.2	107	245	128	117
0.19	118	271	141	130
0.18	132	302	157	144
0.17	148	339	176	162
0.16	167	383	199	183
0.15	190	435	226	208
0.14	218	500	260	238
0.13	252	579	301	276
0.12	296	680	353	324
0.11	352	809	420	386
0.1	426	979	509	467
0.09	526	1208	628	576
0.08	665	1529	794	729
0.07	869	1997	1037	952
0.06	1182	2718	1412	1296
0.05	1702	3913	2033	1866
0.04	2659	6114	3176	2916
0.03	4726	10870	5646	5183
0.02	10634	24456	12702	11661
0.01	42533	97822	50808	46641

Table B.6: Asymptotically optimal sample size for $\alpha = 0.05$ for Huber's $\psi_{1.5}(.)$ score function.

d	N(0,1)	Cauchy	Double Exponential	0.95 N(0,1) + 0.05 N(0,9)
1	4	12	6	5
0.95	5	13	7	5
0.9	5	15	7	6
0.85	6	16	8	7
0.8	7	18	9	7
0.75	8	21	11	8
0.7	9	24	12	10
0.65	10	28	14	11
0.6	12	32	16	13
0.55	14	39	19	15
0.5	16	46	23	18
0.49	17	48	24	19
0.48	18	50	25	20
0.47	19	53	26	21
0.46	19	55	27	22
0.45	20	57	28	23
0.44	21	60	30	24
0.43	22	63	31	25
0.42	23	66	32	26
0.41	24	69	34	27
0.4	25	72	36	28
0.39	27	76	38	30
0.38	28	80	39	31
0.37	30	84	42	33
0.36	31	89	44	35
0.35	33	94	46	37
0.34	35	100	49	39
0.33	37	106	52	41
0.32	39	113	55	44
0.31	42	120	59	47
0.3	45	128	63	50
0.29	48	137	67	53
0.28	51	147	72	57
0.27	55	158	78	62
0.26	59	171	84	66
0.25	64	184	91	72
0.24	70	200	98	78
0.23	76	218	107	85
0.22	83	238	117	93
0.21	91	261	128	102
0.2	100	288	141	112
0.19	111	319	156	124
0.18	123	355	174	138
0.17	138	398	195	155
0.16	156	450	220	175
0.15	178	511	251	199
0.14	204	587	288	228
0.13	236	681	334	264
0.12	277	799	391	310
0.11	330	951	466	369
0.1	399	1150	563	446
0.09	492	1420	695	551
0.08	623	1797	880	697
0.07	813	2347	1149	910
0.06	1107	3194	1564	1238
0.05	1594	4599	2252	1783
0.04	2490	7186	3519	2785
0.03	4427	12774	6255	4952
0.02	9959	28741	14073	11140
0.01	39836	114964	56289	44560

Table B.7: Asymptotically optimal sample size for $\alpha = 0.01$ for Huber's $\psi_{0.5}(.)$ score function.

d	N(0,1)	Cauchy	Double Exponential	0.95 N(0,1) + 0.05 N(0,9)
1	9	16	8	10
0.95	10	17	9	11
0.9	11	19	10	12
0.85	12	22	11	13
0.8	14	24	13	15
0.75	15	27	14	17
0.7	18	31	16	19
0.65	20	36	19	22
0.6	24	43	22	26
0.55	28	51	26	30
0.5	34	61	31	37
0.49	35	64	33	38
0.48	37	66	34	40
0.47	38	69	36	41
0.46	40	72	37	43
0.45	42	75	39	45
0.44	44	79	40	47
0.43	46	83	42	49
0.42	48	87	44	52
0.41	50	91	46	54
0.4	53	95	49	57
0.39	56	100	51	60
0.38	59	106	54	63
0.37	62	111	57	67
0.36	65	118	60	70
0.35	69	124	64	74
0.34	73	132	67	79
0.33	77	140	71	84
0.32	82	149	76	89
0.31	88	158	81	95
0.3	94	169	86	101
0.29	100	181	92	108
0.28	107	194	99	116
0.27	115	209	107	125
0.26	124	225	115	134
0.25	135	243	124	145
0.24	146	264	135	158
0.23	159	287	147	172
0.22	174	314	160	188
0.21	190	345	176	206
0.2	210	380	194	227
0.19	233	421	215	251
0.18	259	469	239	280
0.17	290	526	268	314
0.16	328	594	303	354
0.15	373	675	344	403
0.14	428	775	395	463
0.13	496	899	458	536
0.12	582	1055	537	629
0.11	693	1255	639	749
0.1	838	1519	774	906
0.09	1035	1875	955	1119
0.08	1309	2373	1209	1416
0.07	1710	3099	1578	1849
0.06	2327	4218	2148	2516
0.05	3351	6073	3093	3623
0.04	5236	9489	4833	5661
0.03	9308	16869	8591	10063
0.02	20942	37954	19329	22642
0.01	83768	151816	77316	90565

Table B.8: Asymptotically optimal sample size for $\alpha = 0.01$ for Huber's $\psi_1(.)$ score function.

d	N(0,1)	Cauchy	Double Exponential	0.95 N(0,1) + 0.05 N(0,9)
1	8	17	9	9
0.95	9	19	10	9
0.9	10	21	11	10
0.85	11	24	13	12
0.8	12	27	14	13
0.75	14	31	16	15
0.7	15	35	18	17
0.65	18	40	21	20
0.6	21	47	25	23
0.55	25	56	30	27
0.5	30	68	36	33
0.49	31	71	37	34
0.48	32	74	39	35
0.47	34	77	40	37
0.46	35	80	42	39
0.45	37	84	44	40
0.44	38	88	46	42
0.43	40	92	48	44
0.42	42	96	50	46
0.41	44	101	53	48
0.4	46	106	55	51
0.39	49	112	58	53
0.38	51	118	61	56
0.37	54	124	65	59
0.36	57	131	68	63
0.35	60	138	72	66
0.34	64	147	76	70
0.33	68	156	81	74
0.32	72	165	86	79
0.31	77	176	92	84
0.3	82	188	98	90
0.29	88	201	105	96
0.28	94	216	112	103
0.27	101	232	121	111
0.26	109	250	130	120
0.25	118	271	141	129
0.24	128	294	153	140
0.23	139	320	166	153
0.22	152	350	182	167
0.21	167	384	199	183
0.2	184	423	220	202
0.19	204	469	244	224
0.18	227	522	271	249
0.17	255	585	304	279
0.16	287	660	343	315
0.15	327	751	391	359
0.14	375	863	448	412
0.13	435	1000	520	477
0.12	511	1174	610	560
0.11	608	1397	726	666
0.1	735	1690	878	806
0.09	907	2086	1084	995
0.08	1148	2640	1372	1259
0.07	1500	3449	1791	1645
0.06	2041	4694	2438	2238
0.05	2939	6759	3511	3223
0.04	4592	10560	5485	5035
0.03	8163	18773	9751	8951
0.02	18366	42239	21939	20140
0.01	73462	168956	87754	80557

Table B.9: Asymptotically optimal sample size for $\alpha = 0.01$ for Huber's $\psi_{1.5}(.)$ score function.

d	N(0,1)	Cauchy	Double Exponential	0.95 N(0,1) + 0.05 N(0,9)
1	7	20	10	8
0.95	8	23	11	9
0.9	9	25	13	10
0.85	10	28	14	11
0.8	11	32	16	13
0.75	13	36	18	14
0.7	15	41	20	16
0.65	17	47	24	19
0.6	20	56	28	22
0.55	23	66	33	26
0.5	28	80	39	31
0.49	29	83	41	33
0.48	30	87	43	34
0.47	32	90	45	35
0.46	33	94	46	37
0.45	34	99	49	39
0.44	36	103	51	40
0.43	38	108	53	42
0.42	40	113	56	44
0.41	41	119	58	46
0.4	44	125	61	49
0.39	46	131	64	51
0.38	48	138	68	54
0.37	51	146	72	57
0.36	54	154	76	60
0.35	57	163	80	63
0.34	60	172	85	67
0.33	64	183	90	71
0.32	68	194	95	76
0.31	72	207	102	81
0.3	77	221	109	86
0.29	82	237	116	92
0.28	88	254	125	99
0.27	95	273	134	106
0.26	102	294	144	114
0.25	111	318	156	124
0.24	120	345	169	134
0.23	131	376	184	146
0.22	143	411	201	160
0.21	157	451	221	175
0.2	173	497	244	193
0.19	191	551	270	214
0.18	213	613	301	238
0.17	239	688	337	267
0.16	269	776	380	301
0.15	306	883	433	343
0.14	352	1014	497	393
0.13	408	1175	576	456
0.12	478	1379	676	535
0.11	569	1642	804	637
0.1	689	1986	973	770
0.09	850	2452	1201	951
0.08	1076	3103	1520	1203
0.07	1405	4053	1985	1571
0.06	1912	5516	2701	2138
0.05	2753	7943	3889	3079
0.04	4301	12411	6077	4811
0.03	7645	22063	10803	8552
0.02	17201	49641	24306	19241
0.01	68804	198562	97222	76962

Appendix C

More Simulations

Table C.1: *Normal distribution, Huber's ψ with $h = 1.0$, $\gamma = 1/3$, $k = 1/2$*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	18	N	17.97	15	0.940
		A	19.77	17	0.959
		U	36.94	22	0.967
0.30	48	N	44.33	41	0.910
		A	51.13	47	0.925
		U	65.00	57	0.945
0.10	426	N	419.70	421	0.940
		A	497.47	475.5	0.949
		U	509.99	487	0.956

Table C.2: *Normal distribution, Huber's ψ with $h = 1.0$, $\gamma = 1/3$, $k = 1/2$, alternative method*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	18	N	16.77	14	0.924
		A	19.42	17	0.947
		U	22.03	19	0.948
0.30	48	N	43.84	40	0.922
		A	50.49	45	0.934
		U	54.76	50	0.951
0.10	426	N	417.51	417	0.932
		A	501.13	477	0.947
		U	504.90	480	0.948

Figure C.1: N_2 *for Normal distribution.*

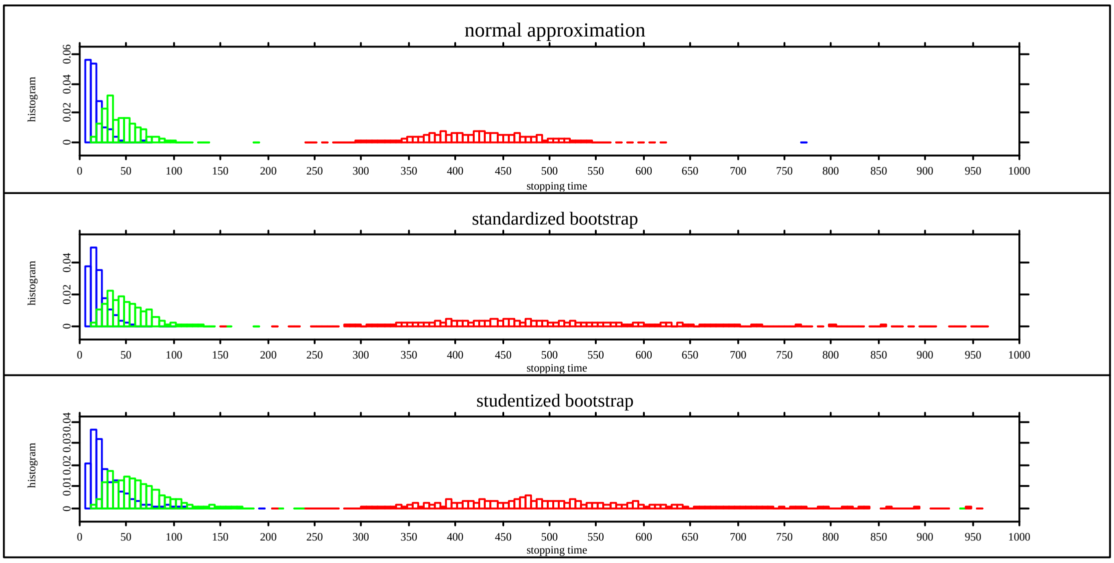


Figure C.2: N_2 *for Normal distribution, alternative method.*

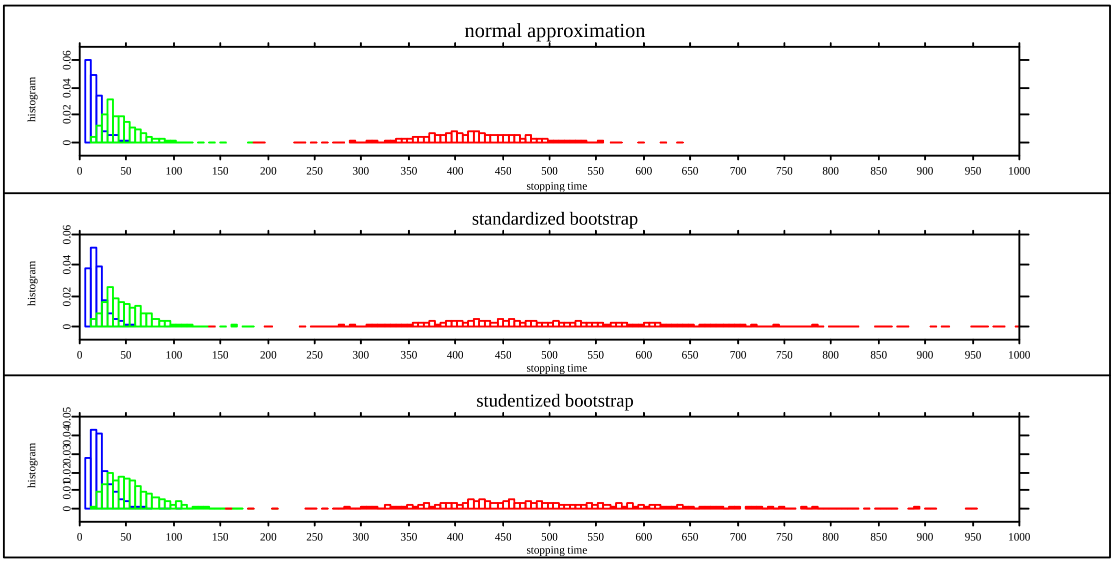


Figure C.3: $\sqrt{N_2}$ *for Normal distribution.*

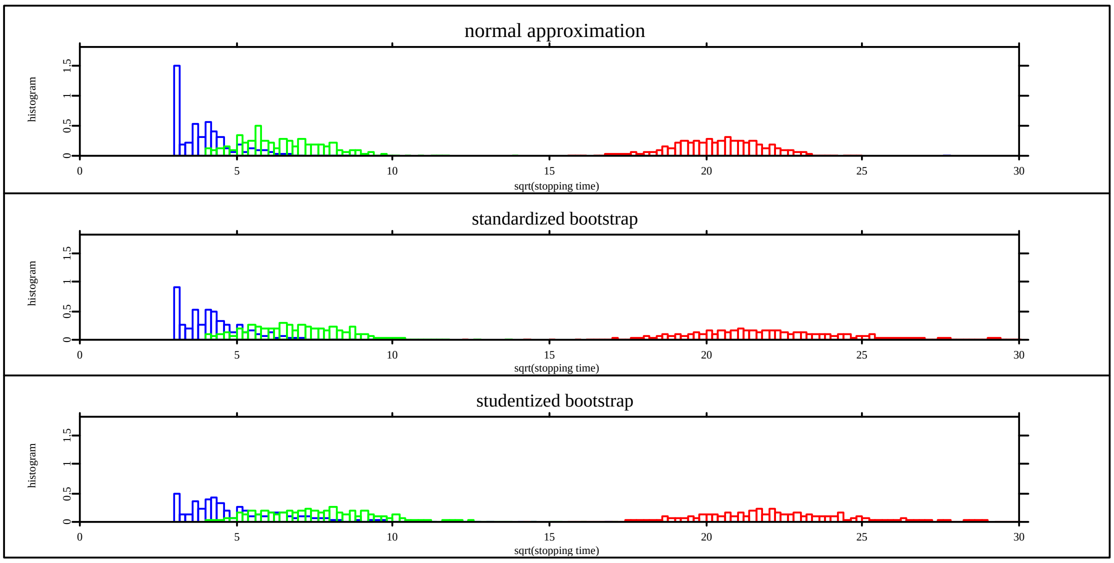


Figure C.4: $\sqrt{N_2}$ *for Normal distribution, alternative method.*

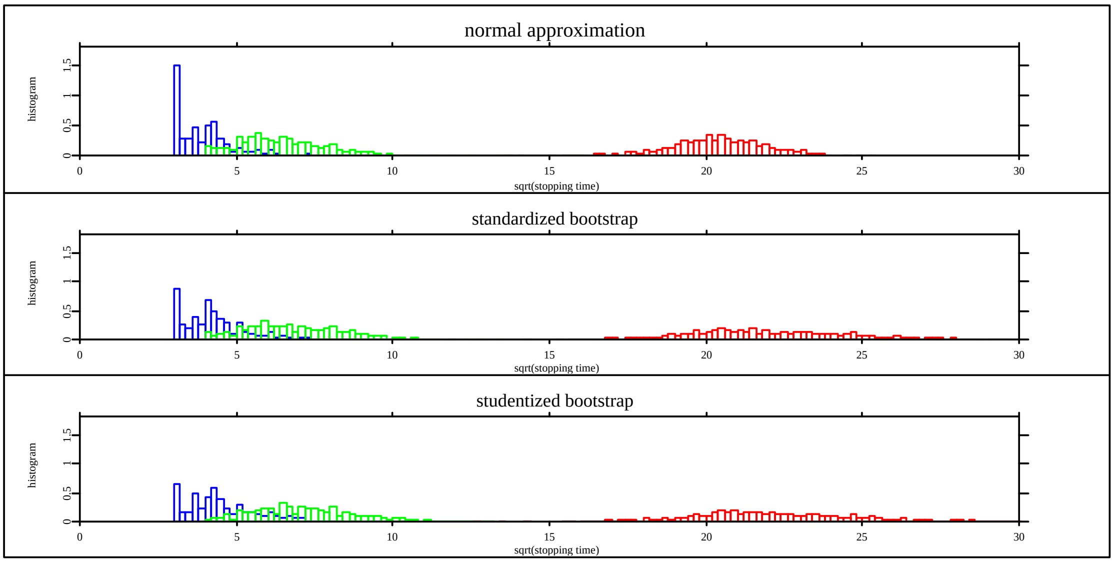


Table C.3: *Cauchy distribution, Huber's ψ with $h = 1.0$, $\gamma = 1/3$, $k = 1/2$*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	40	N	55.01	35	0.931
		A	80.38	49.5	0.954
		U	187.26	48	0.950
0.30	109	N	115.13	101	0.954
		A	136.44	119	0.955
		U	165.85	112.5	0.954
0.10	979	N	970.43	964	0.958
		A	970.39	957	0.951
		U	970.74	953	0.958

Table C.4: *Cauchy distribution, Huber's ψ with $h = 1.0$, $\gamma = 1/3$, $k = 1/2$, alternative method*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	40	N	52.61	34	0.928
		A	95.75	49	0.945
		U	116.46	49.5	0.948
0.30	109	N	112.38	103.5	0.946
		A	142.73	115	0.963
		U	140.76	114	0.953
0.10	979	N	977.49	975	0.956
		A	969.21	951	0.957
		U	970.93	959	0.957

Figure C.5: N_2 *for Cauchy distribution.*

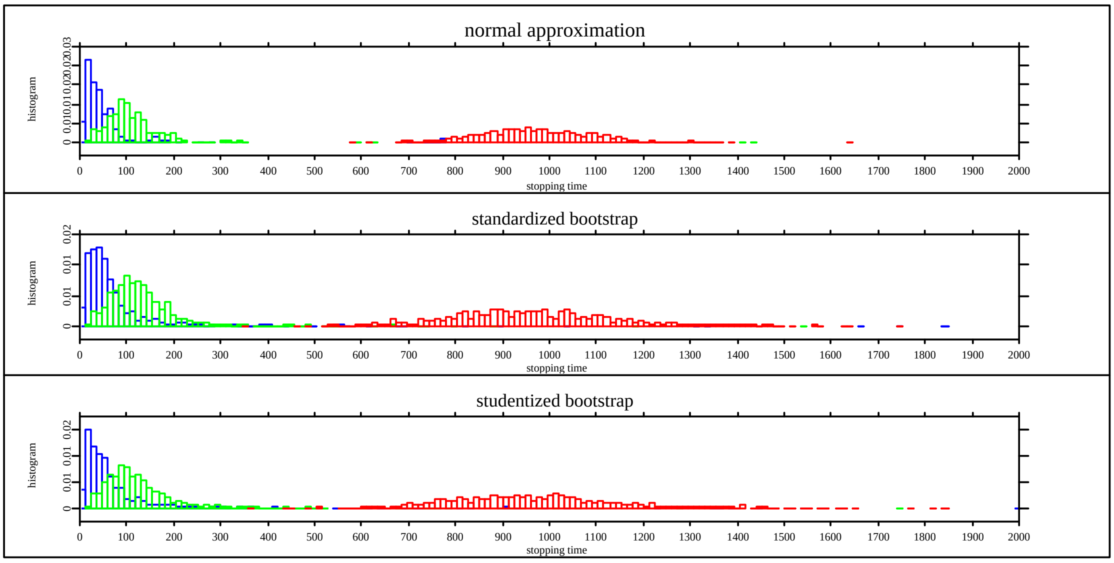


Figure C.6: N_2 *for Cauchy distribution, alternative method.*

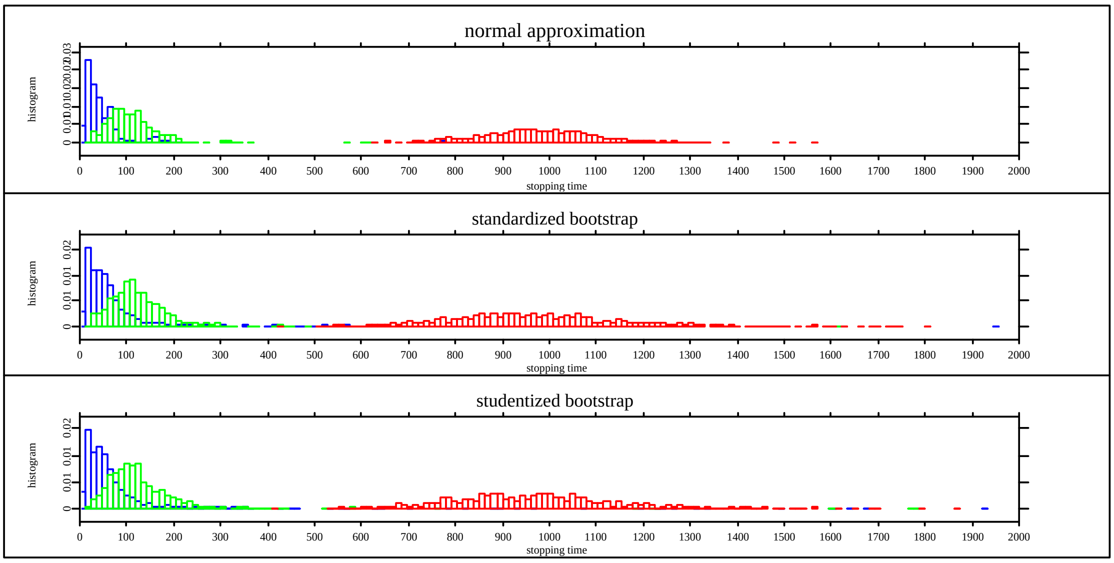


Figure C.7: $\sqrt{N_2}$ *for Cauchy distribution.*

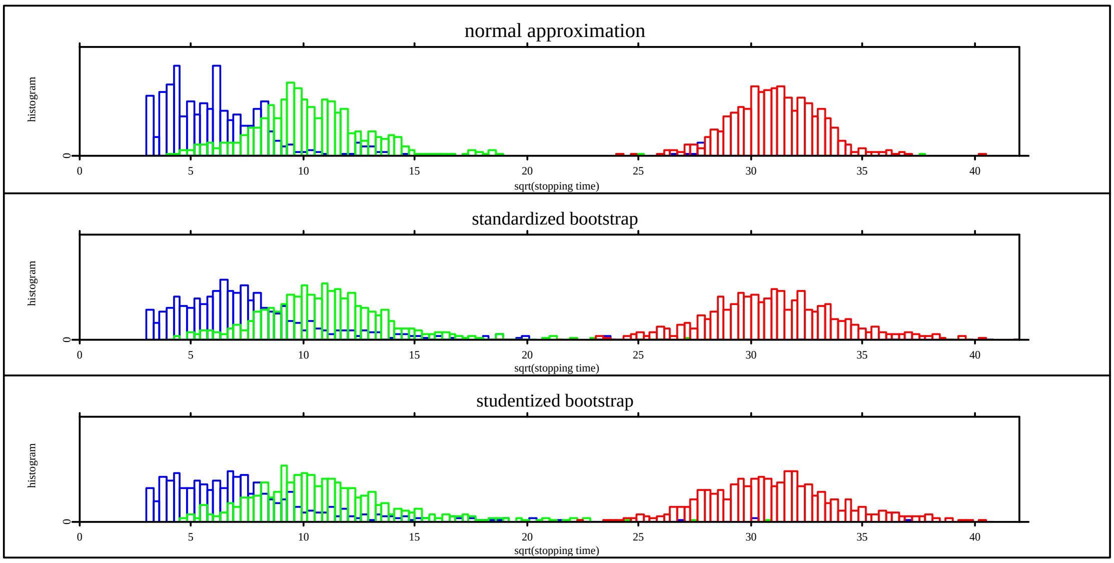


Figure C.8: $\sqrt{N_2}$ *for Cauchy distribution, alternative method.*

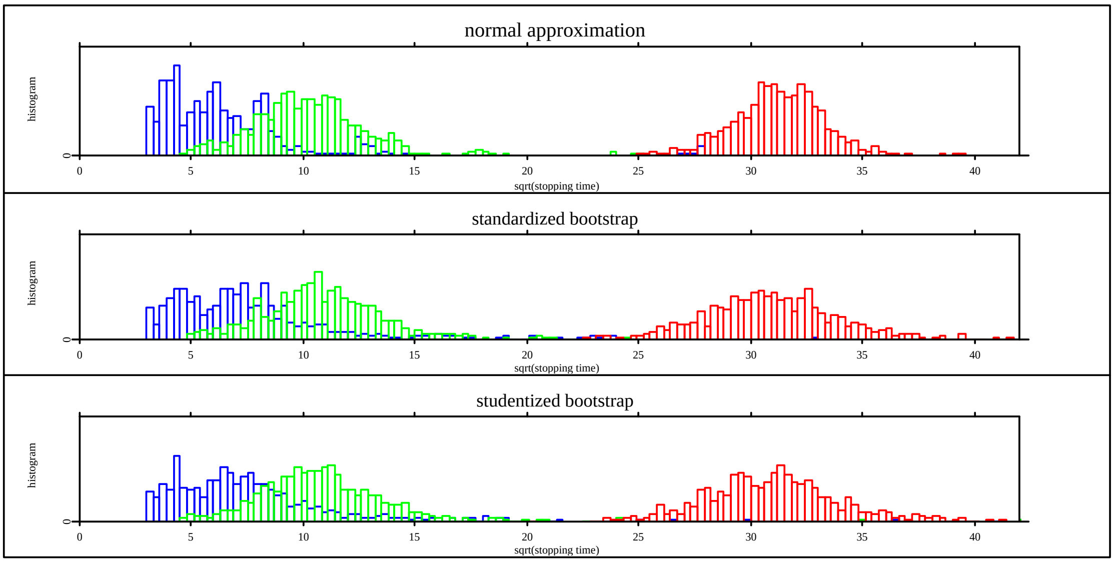


Table C.5: *Double Exponential distribution, Huber's ψ with $h = 1.0$, $\gamma = 1/3$, $k = 1/2$*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	21	N	20.98	17	0.926
		A	24.53	19.5	0.950
		U	37.05	23	0.958
0.30	57	N	54.61	51	0.934
		A	59.64	56	0.938
		U	70.18	61	0.968
0.10	509	N	502.06	501	0.954
		A	479.00	473	0.946
		U	481.67	476	0.940

Table C.6: *Double Exponential, Huber's ψ with $h = 1.0$, $\gamma = 1/3$, $k = 1/2$, alternative method*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	21	N	20.97	17	0.930
		A	24.19	20	0.948
		U	25.39	20	0.951
0.30	57	N	54.27	50	0.935
		A	58.27	54	0.944
		U	59.49	54	0.959
0.10	509	N	501.69	502	0.959
		A	478.76	475.5	0.956
		U	479.31	477	0.955

Figure C.9: N_2 *for Double Exponential distribution.*

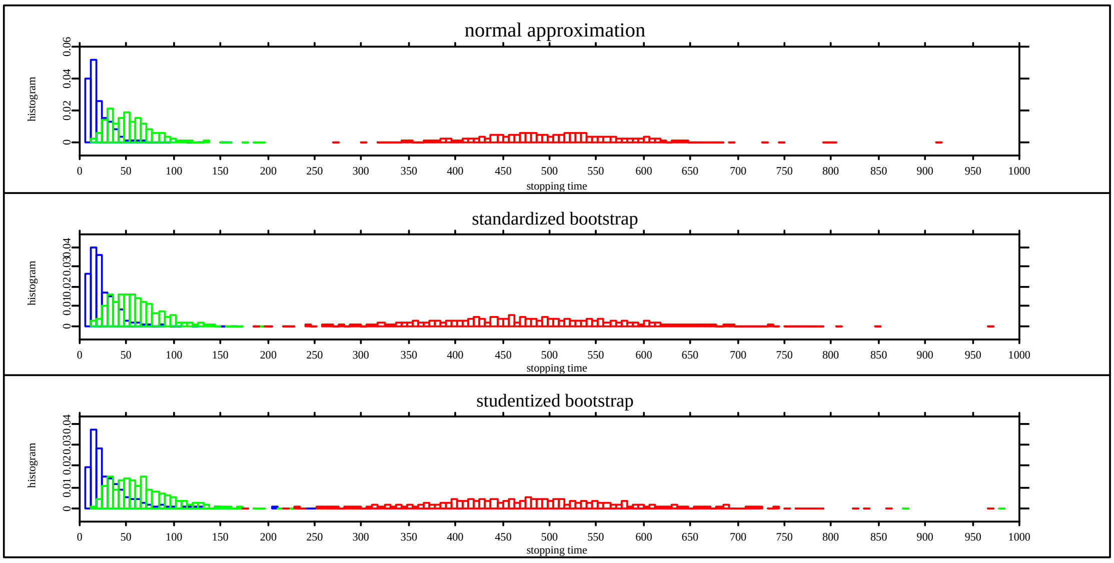


Figure C.10: N_2 *for Double Exponential, alternative method.*

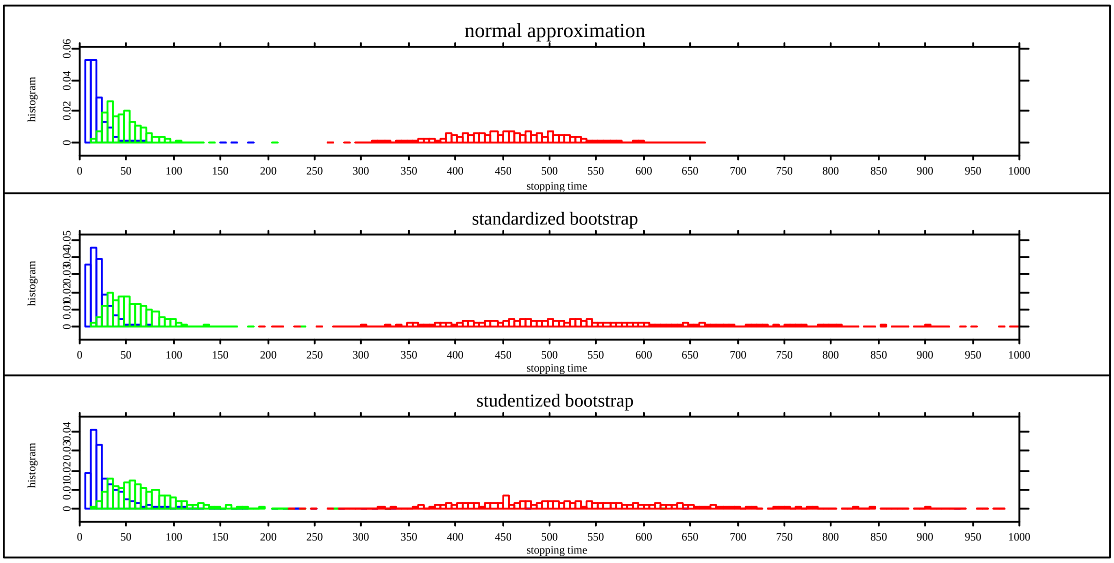


Figure C.11: $\sqrt{N_2}$ *for Double Exponential distribution.*

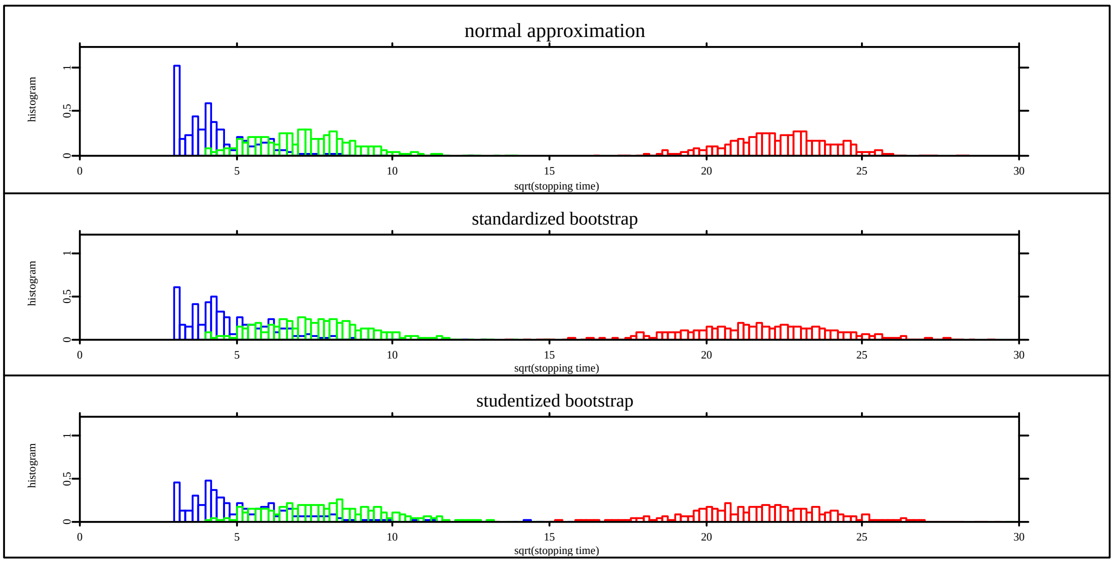


Figure C.12: $\sqrt{N_2}$ *for Double Exponential, alternative method.*

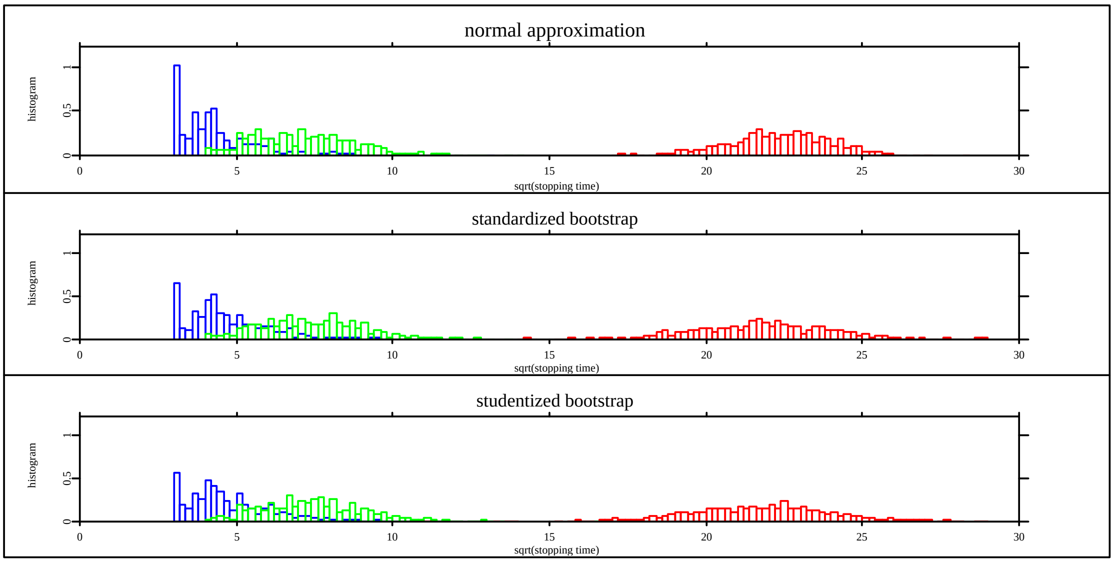


Table C.7: *Mixture of Normal distributions, Huber's ψ with $h = 1.0$, $\gamma = 1/3$, $k = 1/2$*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	19	N	18.36	15	0.906
		A	20.68	18	0.928
		U	36.27	21	0.941
0.30	52	N	47.63	44	0.908
		A	55.86	52	0.928
		U	69.80	61	0.940
0.10	467	N	455.83	454.5	0.932
		A	537.42	513	0.939
		U	546.98	520	0.952

Table C.8: *Mixture of Normal Distributions, Huber's ψ with $h = 1.0$, $\gamma = 1/3$, $k = 1/2$, alternative method*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	19	N	18.86	15	0.919
		A	20.96	18	0.933
		U	23.91	19	0.934
0.30	52	N	48.15	44	0.890
		A	55.04	50	0.911
		U	59.22	55	0.914
0.10	467	N	462.19	461	0.916
		A	546.32	515	0.932
		U	551.86	519	0.939

Figure C.13: N_2 *for Mixture of Normal Distributions.*

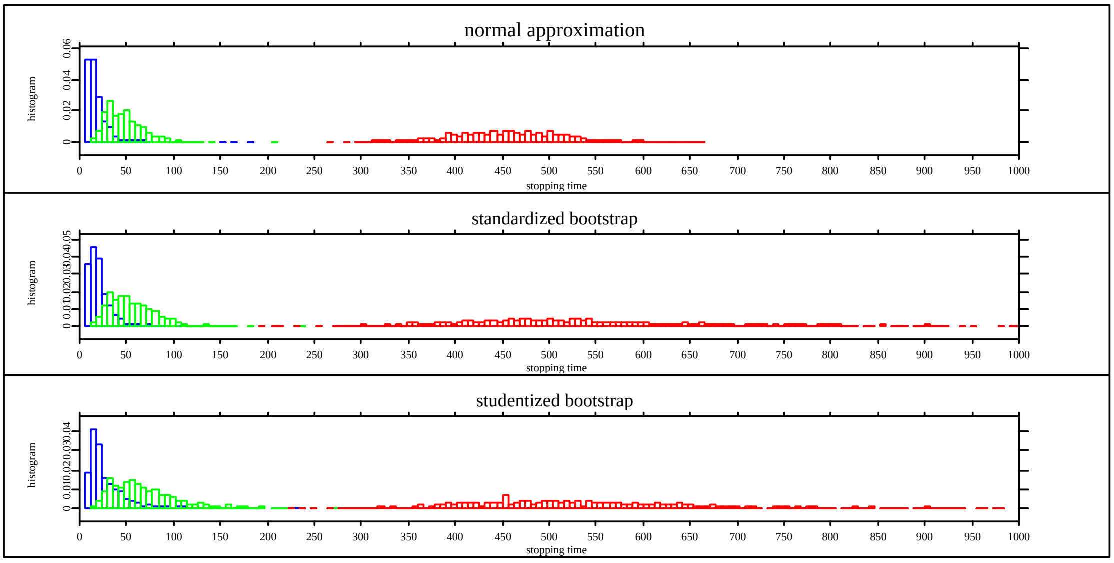


Figure C.14: N_2 *for Mixture of Normal Distributions, alternative method.*

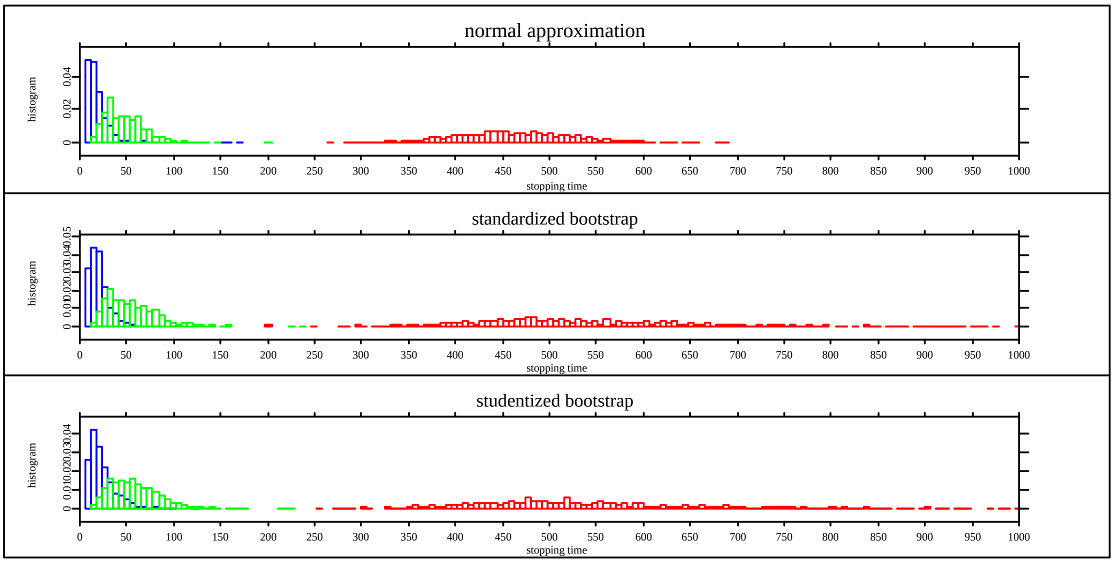


Figure C.15: $\sqrt{N_2}$ *for Mixture of Normal Distributions.*

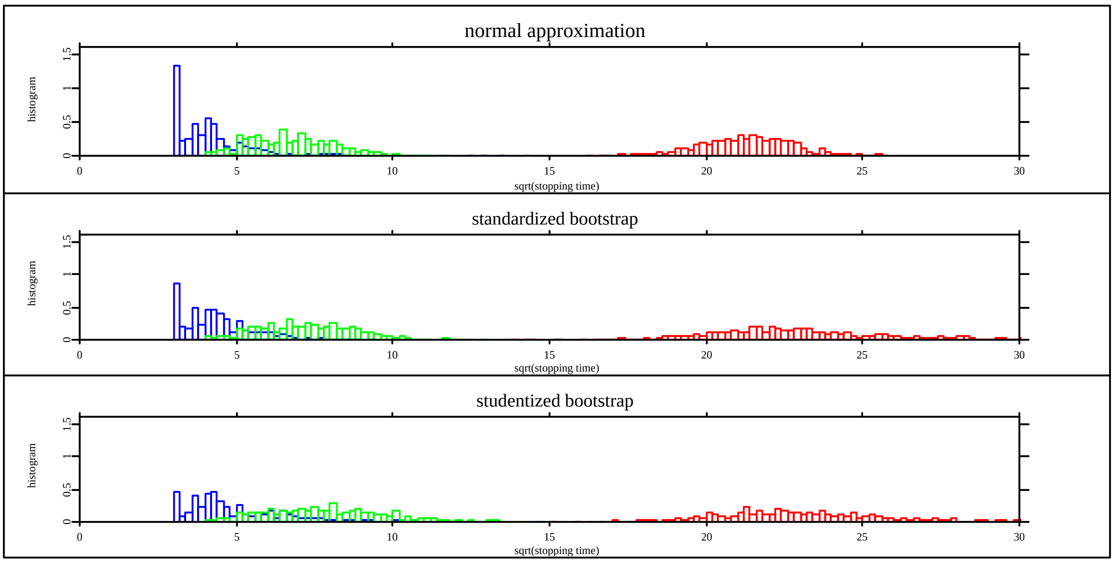


Figure C.16: $\sqrt{N_2}$ *for Mixture of Normal Distributions, alternative method.*

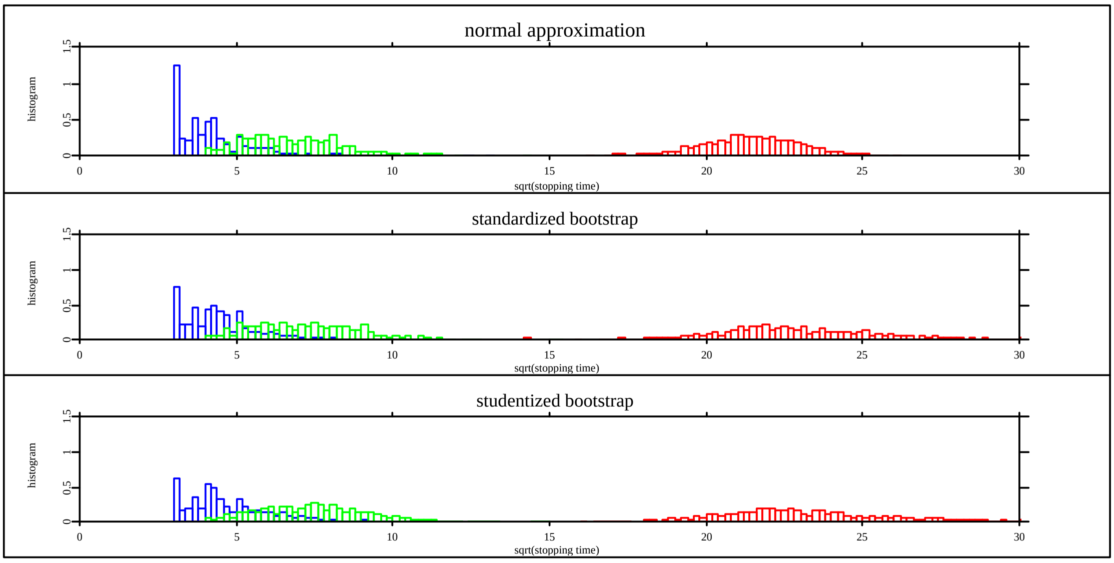


Table C.9: *Normal distribution, Huber's ψ with $h = 0.5$, $\gamma = 1/3$, $k = 1/2$*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	20	N	21.61	15	0.913
		A	23.50	19	0.939
		U	91.96	29	0.954
0.30	54	N	50.90	39	0.890
		A	56.10	51	0.911
		U	111.61	68	0.924
0.10	485	N	473.55	467	0.935
		A	531.38	516	0.950
		U	551.61	535	0.951

Table C.10: *Normal distribution, Huber's ψ with $h = 0.5$, $\gamma = 1/3$, $k = 1/2$, alternative method*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	20	N	22.017	16	0.906
		A	48.05	18	0.924
		U	27.18	20	0.941
0.30	54	N	52.48	39	0.894
		A	57.81	49.5	0.918
		U	57.71	51	0.923
0.10	485	N	476.39	475	0.920
		A	532.88	519	0.936
		U	539.15	525	0.927

Figure C.17: *N_2 for Normal distribution.*

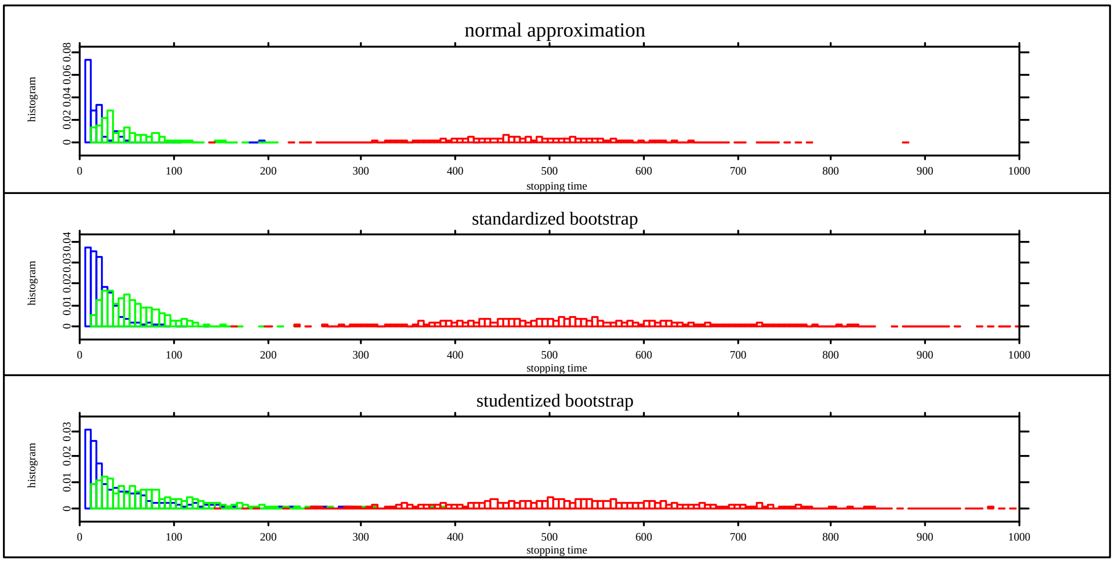


Figure C.18: *N_2 for Normal distribution, alternative method.*

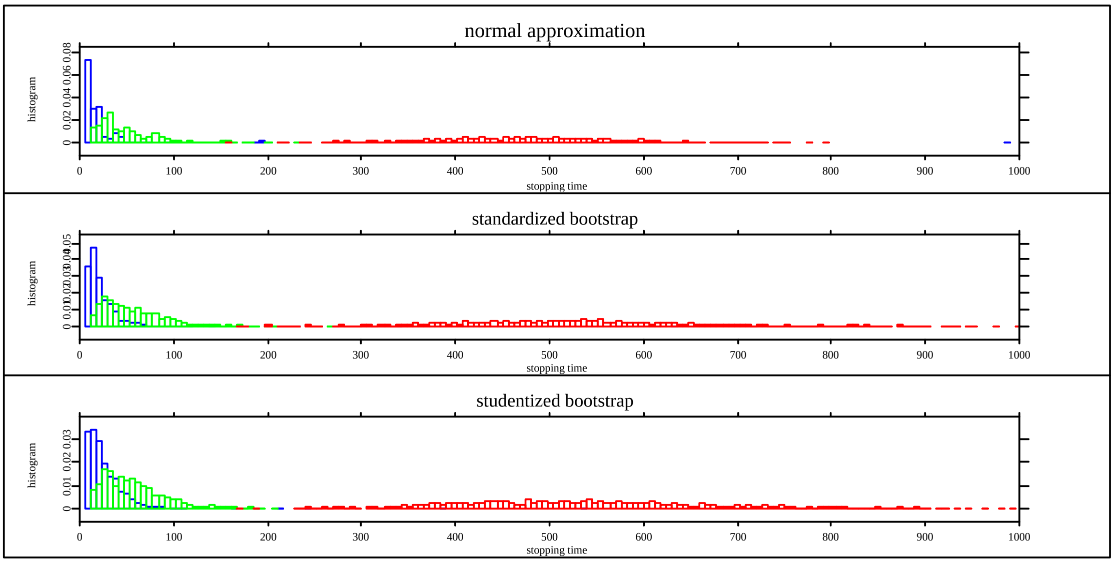


Figure C.19: $\sqrt{N_2}$ *for Normal distribution.*

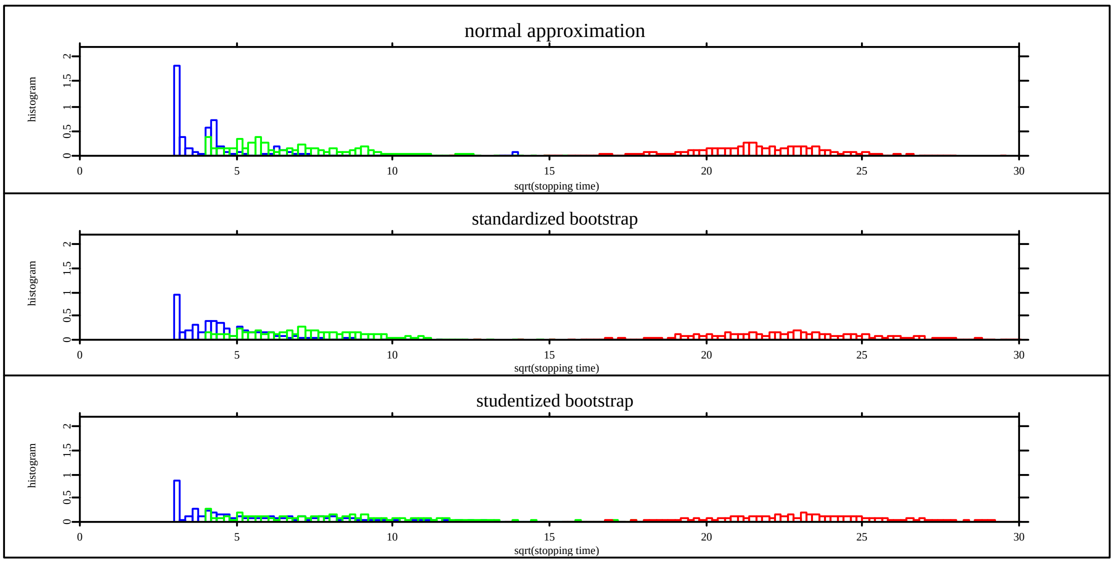


Figure C.20: $\sqrt{N_2}$ *for Normal distribution, alternative method.*

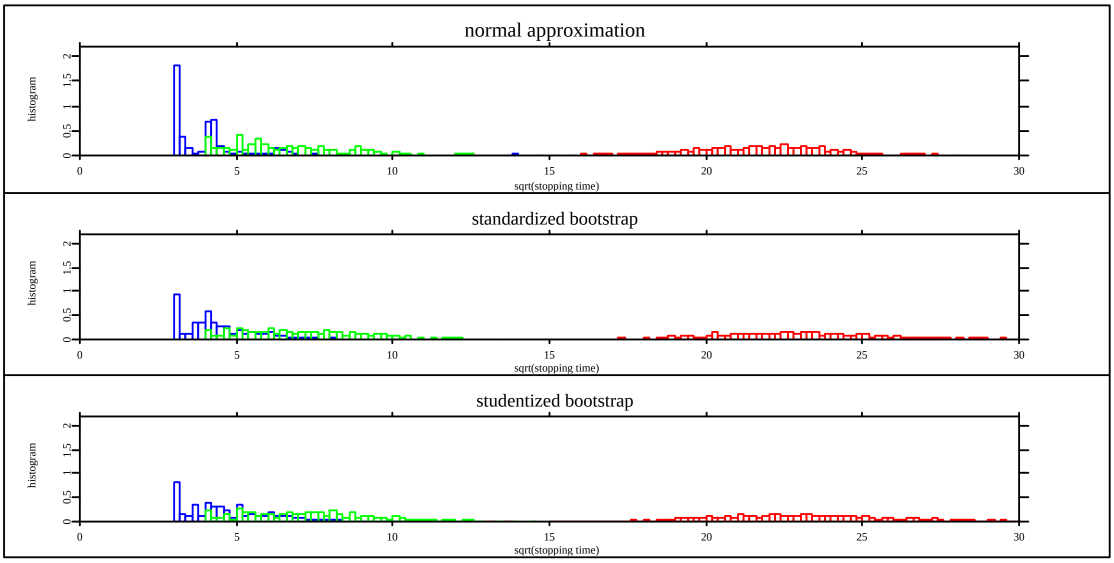


Table C.11: *Cauchy distribution, Huber's ψ with $h = 0.5$, $\gamma = 1/3$, $k = 1/2$*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	36	N	53.15	23	0.897
		A	98.49	52.5	0.962
		U	331.28	73	0.961
0.30	98	N	118.42	89	0.919
		A	138.38	119	0.955
		U	296.09	130	0.941
0.10	879	N	871.97	864.5	0.946
		A	919.59	902.5	0.961
		U	941.52	908.5	0.960

Table C.12: *Cauchy distribution, Huber's ψ with $h = 0.5$, $\gamma = 1/3$, $k = 1/2$, alternative method*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	36	N	52.32	20	0.906
		A	144.15	47	0.945
		U	123.10	52	0.946
0.30	98	N	117.58	93	0.929
		A	135.66	117.5	0.954
		U	158.12	123	0.952
0.10	879	N	874.82	867	0.944
		A	910.63	883.5	0.938
		U	919.08	904	0.942

Figure C.21: N_2 *for Cauchy distribution.*

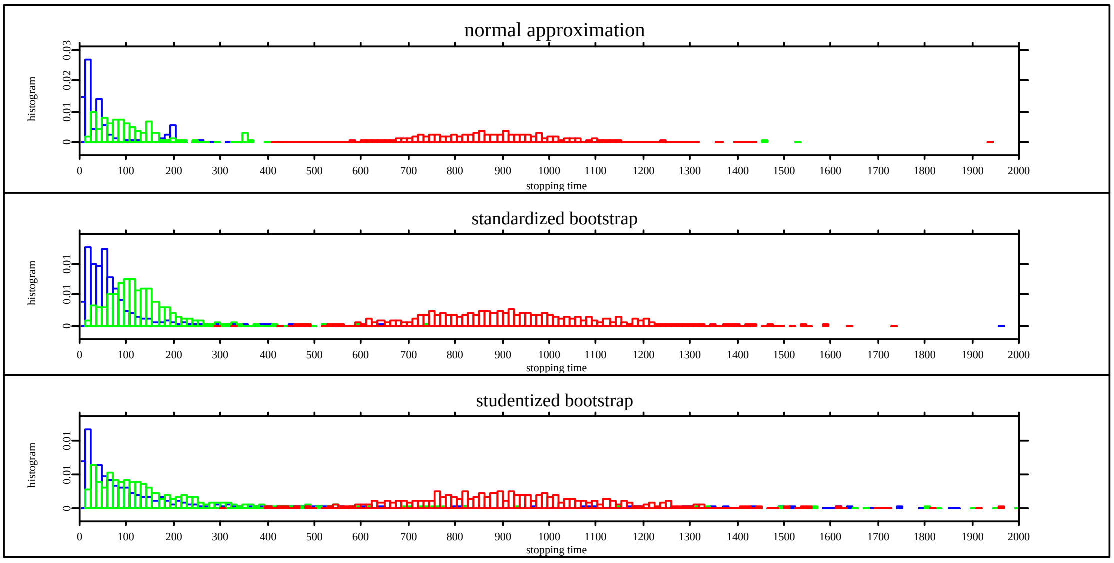


Figure C.22: N_2 *for Cauchy distribution, alternative method.*

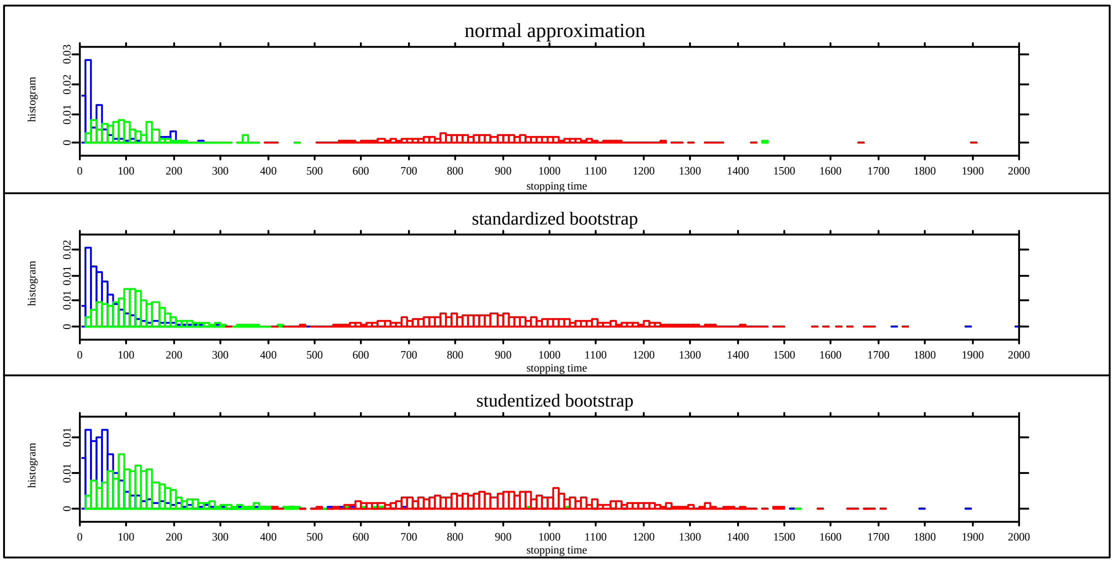


Figure C.23: $\sqrt{N_2}$ *for Cauchy distribution.*

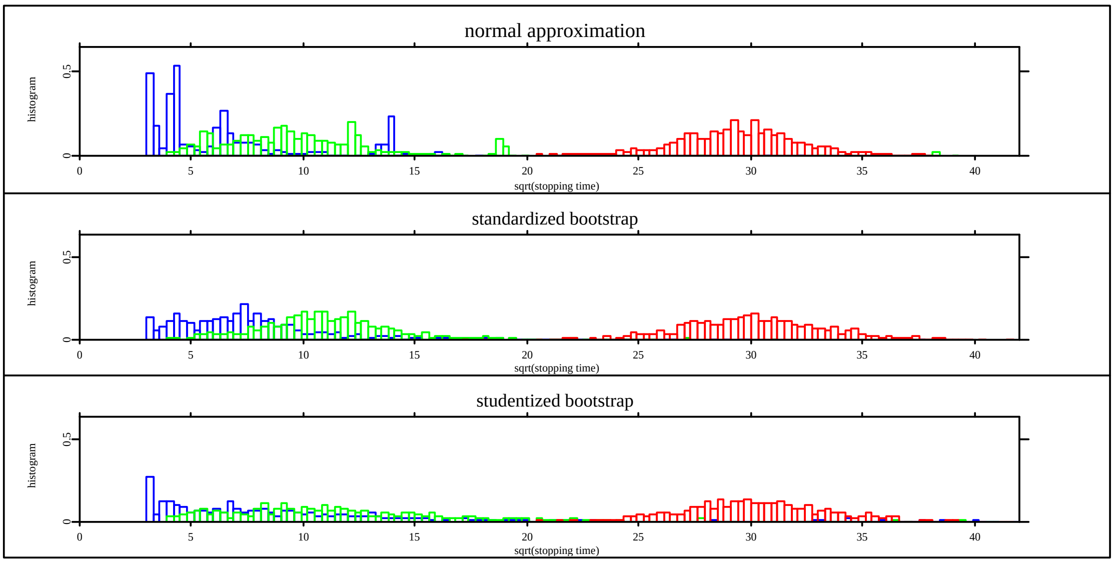


Figure C.24: $\sqrt{N_2}$ *for Cauchy distribution, alternative method.*

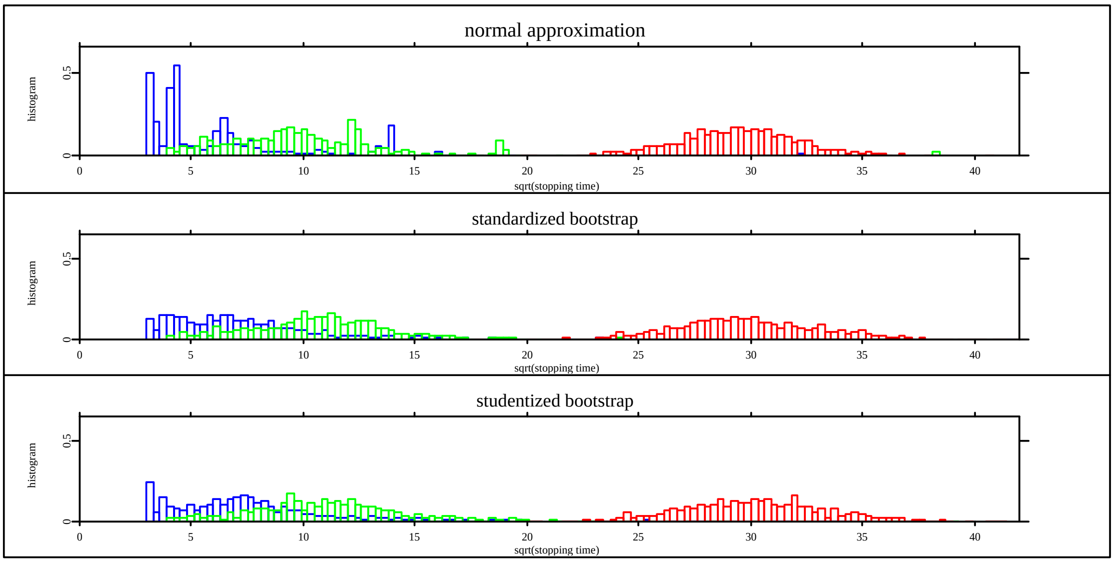


Table C.13: *Double Exponential distribution, Huber's ψ with $h = 0.5$, $\gamma = 1/3$, $k = 1/2$*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	18	N	23.281	16	0.921
		A	27.67	21	0.951
		U	94.75	26	0.963
0.30	50	N	54.145	41	0.931
		A	60.945	54	0.947
		U	109.1	59	0.947
0.10	448	N	437.8	435	0.946
		A	445.61	435	0.946
		U	453.68	447	0.947

Table C.14: *Double Exponential, Huber's ψ with $h = 0.5$, $\gamma = 1/3$, $k = 1/2$, alternative method*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	18	N	23.10	16	0.933
		A	37.57	20	0.949
		U	36.59	22	0.959
0.30	50	N	55.52	42	0.943
		A	63.29	54	0.954
		U	63.23	53	0.950
0.10	448	N	440.46	433	0.943
		A	447.35	438	0.952
		U	444.90	434	0.949

Figure C.25: N_2 *for Double Exponential distribution.*

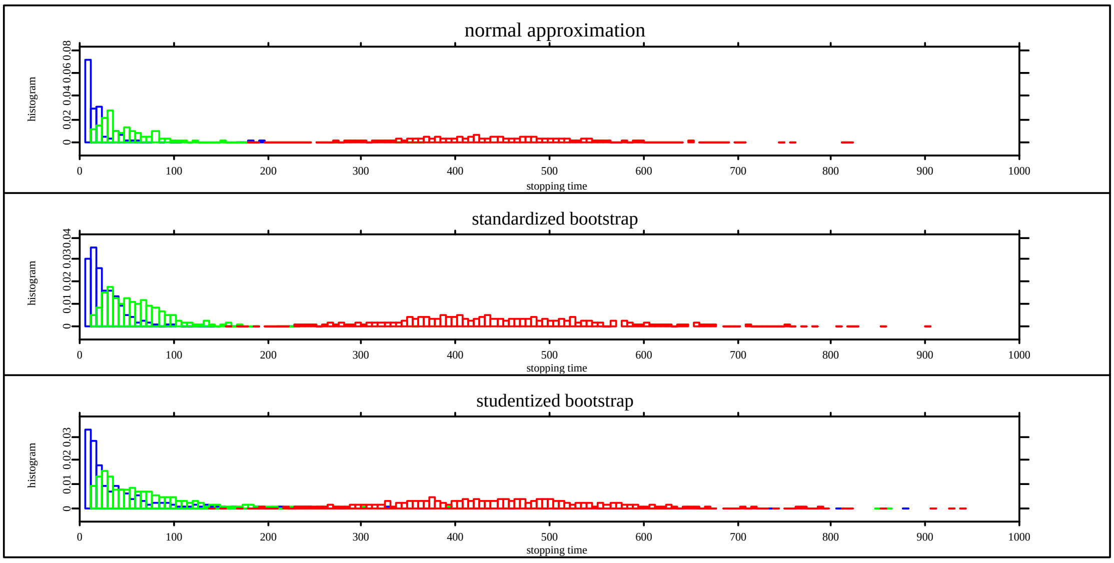


Figure C.26: N_2 *for Double Exponential, alternative method.*

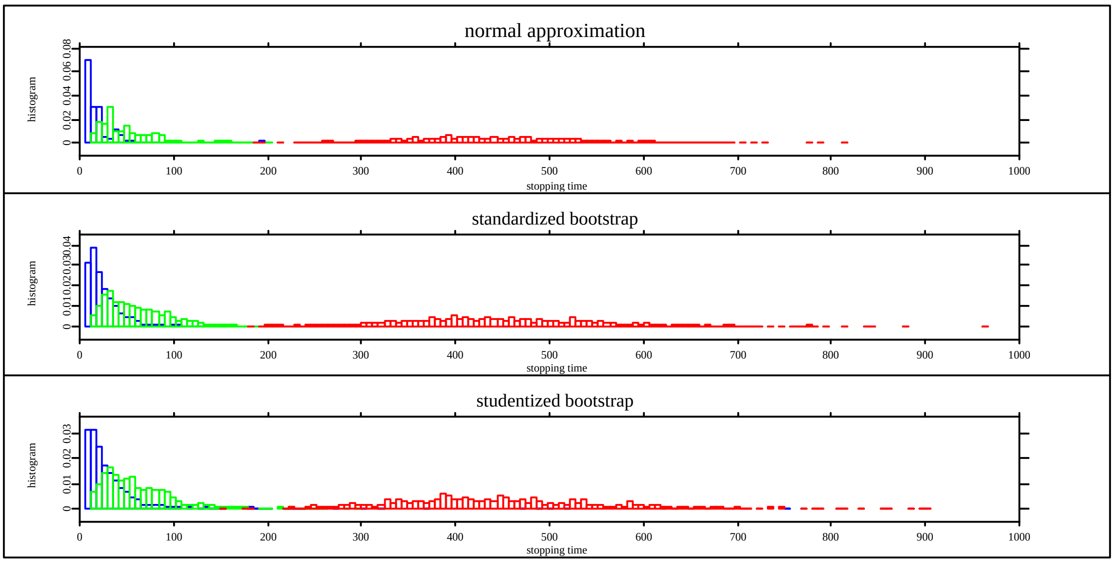


Figure C.27: $\sqrt{N_2}$ *for Double Exponential distribution.*

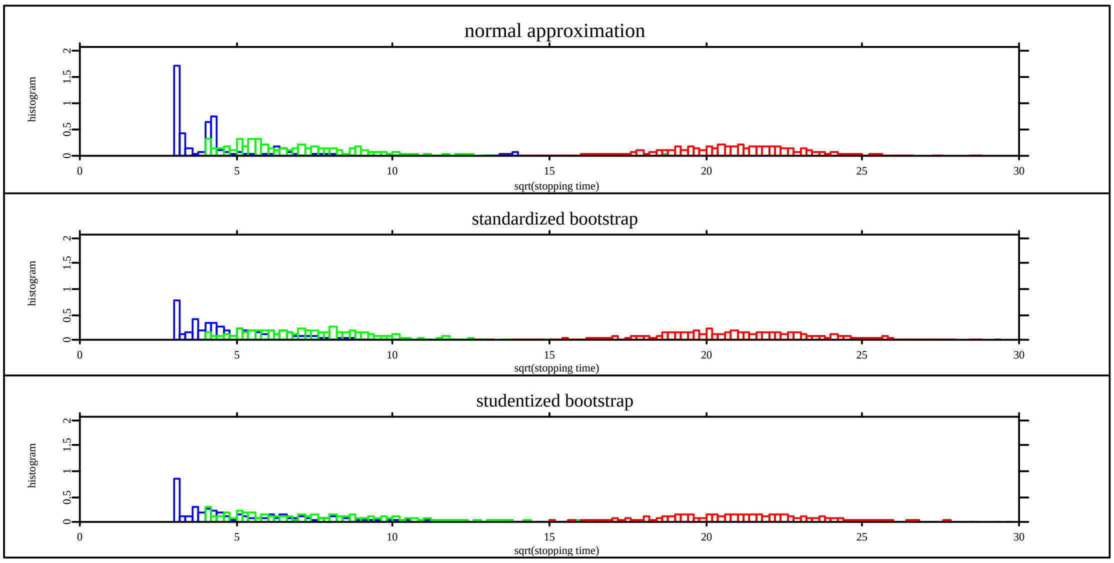


Figure C.28: $\sqrt{N_2}$ *for Double Exponential, alternative method.*

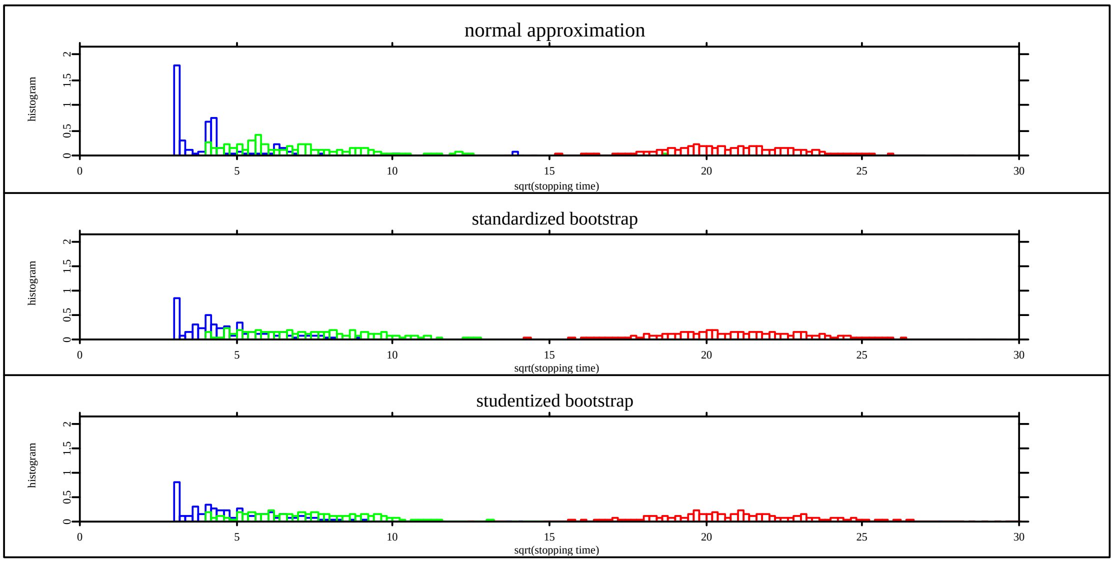


Table C.15: *Mixture of Normal distributions, Huber's ψ with $h = 0.5$, $\gamma = 1/3$, $k = 1/2$*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	21	N	22.77	16	0.911
		A	25.02	20	0.933
		U	86.71	33.5	0.947
0.30	59	N	59.62	48	0.904
		A	62.90	57.5	0.935
		U	133.08	78	0.945
0.10	525	N	518.39	517.5	0.939
		A	584.12	561	0.952
		U	609.25	589.5	0.954

Table C.16: *Mixture of Normal Distributions, Huber's ψ with $h = 0.5$, $\gamma = 1/3$, $k = 1/2$, alternative method*

d	$c_M(d)$	critical points	mean	median	coverage probability
0.50	21	N	22.20	16	0.915
		A	34.77	19	0.941
		U	31.82	22.5	0.952
0.30	59	N	57.94	43	0.889
		A	63.85	56	0.910
		U	63.58	58	0.926
0.10	525	N	517.77	517	0.928
		A	572.82	552	0.945
		U	580.50	554.5	0.945

Figure C.29: N_2 *for Mixture of Normal Distributions.*

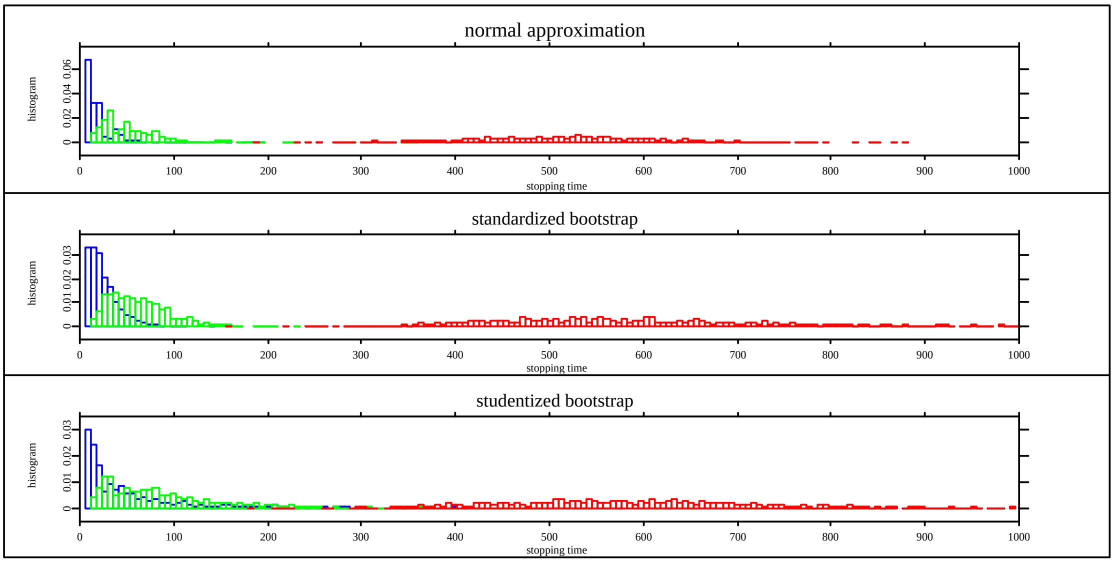


Figure C.30: N_2 *for Mixture of Normal Distributions, alternative method.*

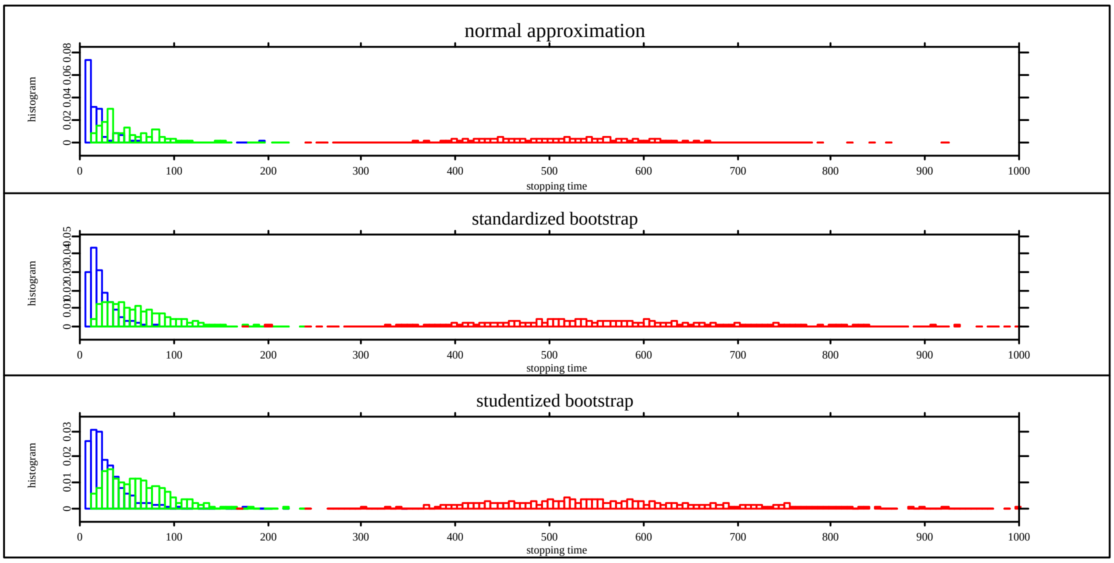


Figure C.31: $\sqrt{N_2}$ *for Mixture of Normal Distributions.*

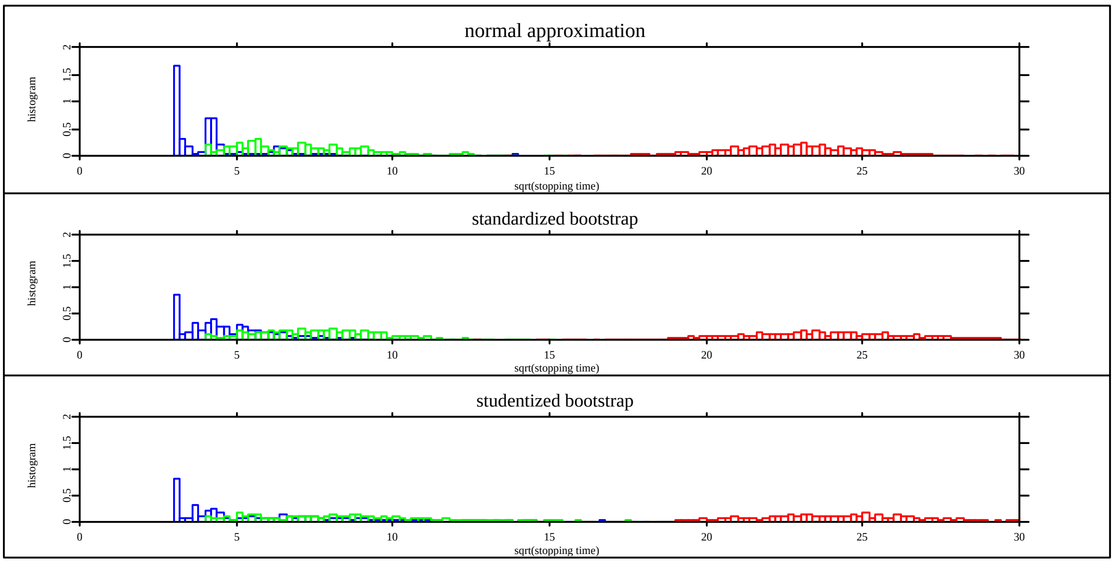
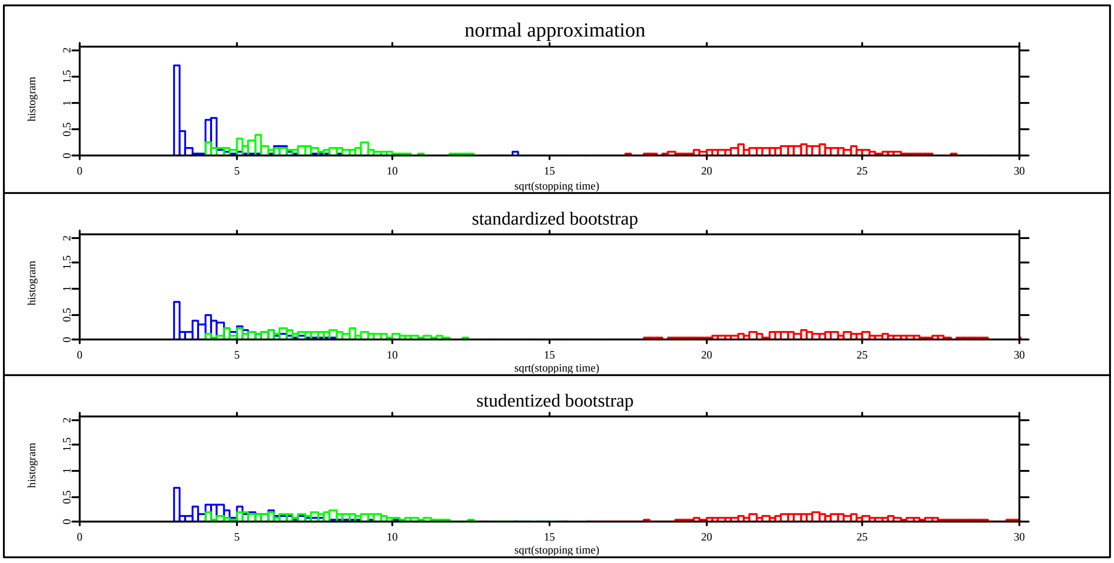


Figure C.32: $\sqrt{N_2}$ *for Mixture of Normal Distributions, alternative method.*



Bibliography

Aerts, M. (1988). *Asymptotic Distribution Theory for Random U- and GM-statistics with Application in Sequential Estimation*, PhD. thesis, Limburgs Universitair Centrum Diepenbeek

Aerts, M. & Gijbels, I. (1993). A three-stage procedure based on bootstrap critical points. *Sequential Analysis,* **12(2)**: 93–113

Anděl, J. (1985). *Matematická statistika*, SNTL, Praha

Antoch, J. & Vorlíčková, D. (1992). *Vybrané metody statistické analýzy dat*, Academia, Praha

Anscombe, F.J. (1952). Large sample theory of sequential estimation. *Proc. Cambridge Philos. Soc.,* **48**: 600–607

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

Dodge, Y. &Jurečková, J. (1995). Estimation of quantile density function based on regression quantiles. *Statistics & Probability Letters,* **23**: 73–78

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

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications, Volume II*, Wiley, New York

Ghosh, M. & Mukhopadhyay, N. (1981). Consistency and asymptotic efficiency of two-stage and sequential estimation procedures. *Sankhya Ser. A,* **43**: 220–227

Ghosh, M., Mukhopadhyay, N. & Sen, P.K. (1997). *Sequential Analysis*, Wiley, New York

Ghosh, B.K. & Sen, P.K. (1991). *Handbook of Sequential Analysis*, Dekker, New York

Härdle, W., Klinke, S., & Müller, M. (2000). *XploRe, Academic Edition*, Springer-Verlag, Berlin

Hall, P. (1981). Asymptotic theory of triple sampling for sequential estimation of a mean. *Ann. Statist.,* **9**: 1229–1238

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York

Hampel, F.R., Ronchetti, E.R., Rousseeuw, P.J. & Stahel, W.A. (1990). *Robust Statistics - The Approach Based on Influence Functions*, Wiley, New York

Hlávka, Z. (1999). Robust sequential procedure based on bootstrap. In *Eleventh European Young Statisticians Meeting Proceedings, Marly-le-Roi, August 24–28,1999*, pages 85–95, INRA, Paris

Hlávka, Z. (2000). Some aspects of fixed-width confidence intervals for Cauchy distribution. Discussion Paper, SFB 373, Berlin, submitted

Hlávka, Z. (2000). Asymptotic properties of robust three-stage procedure based on bootstrap for M-estimators. Discussion Paper, SFB 373, Berlin, submitted

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Jour. Amer. Statist. Assoc.,* **58**: 13–29

Huber, P. (1981). *Robust Statistics*, Wiley, New York

Jurečková, J. (1980). Asymptotic representation of M-estimators of location. *Opert. Statist., Ser. Stat.,* **11**: 61–73

Jurečková, J. & Sen, P.K. (1978). Bounded-length sequential confidence intervals for regression and location parameters. *The Second Prague Symposiium on Asymptotic Statistics*, Charles University, Prague

Jurečková, J. & Sen, P.K. (1981). Sequential procedures based on M-estimators with discontinuous score functions. *Jour. Statist. Plan. Infer.*, **5**: 253–266

Jurečková, J. & Sen, P.K. (1982). M-estimators and L-estimators of location: uniform integrability and asymptotic risk-efficient sequential versions. *Commun. Statist., -Sequential Analysis*, **1(1)**: 27–56

Jurečková, J. & Sen, P.K. (1995). *Robust Statistical Procedures: Asymptotics and Interrelations*, Wiley, New York

Karabulut, I. & Lahiri, S.N. (1997). Two-term edgeworth expansion for M-estimators of a linear regression parameter without Cramér-type conditions and an application to the bootstrap. *J. Austral. Math. Soc. (Series A),* **62**: 361–370

Lahiri, S.N. (1992). Bootstrapping M-estimators of a multiple linear regression parameter. *Ann. Statist.,* **20**: 1548–1570

Lahiri, S.N. (1992). On bootstrapping M-estimators. *Sankhyā A,* **54**: 157–170

Loève, M. (1977). *Probability Theory I*, Springer-Verlag, New York

MacCracken, D. (1985). *Computerpraxis mit FORTRAN 77 in Naturwissenschaft und Technik*, Carl Hanser Verlag, München

Mallows, C.L. (1972). A note on asymptotic joint normality. *Ann. Math. Statist.,* **39**: 755–771

Mukhopadhyay, N. (1980). A consistent and asymptotically efficient two-stage procedure to construct fixed-width confidence interval for the mean. *Metrika,* **24**: 281–284

Mukhopadhyay, N. & Abid, A.D. (1986). On fixed-size confidence regions for the regression parameters. *Metron,* **44**: 297–306

Mukhopadhyay, N. & Duggan, W.T. (1997). Can a two-stage procedure enjoy second-order properties? . *Sankhyā Ser. A,* **59**: 435–448

Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. (1992). *Numerical Recipes in FORTRAN, The Art of Scientific Computing, Second Edition*, Cambridge University Press

Rektorys, K. (1988). *Přehled užité matematiky*, SNTL, Praha

Rieder, H. (1994). *Robust Asymptotic Statistics*, Springer-Verlag, New York

Searle, S.R. (1982). *Matrix Algebra Useful for Statisticians*, Wiley, New York

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York

Serfling, R.J. (1984). Generalized L-, M-, and R-statistics. *Ann. Math. Statist.,* **12**: 76–86

Shao, J. & Tu, D. (1995). *The Jacknife and Bootstrap*, Springer-Verlag, New York

Shorack, G.R. (1982). Bootstrapping robust regression. *Commun. in Statist., -Theor. Meth.,* **11(9)**: 961–972

Stein, C. (1945). A two sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.*, **17**: 498–499

Stein, C. (1949). Some problems in sequential estimation (abstract). *Econometrica,* **17**: 77–78

Štěpán, J. (1987). *Teorie Pravděpodobnosti*, Academia, Praha

Swanepoel, J.W.H, van Wijk, J.W.J. &Venter, J.H. (1984). Fixed-width confidence intervals based on bootstrap procedures. *Commun. in Statist., -Seq. Analys.*, **2**: 289–310