Robust Estimation of the Location Parameter from a Two-Parameter Exponential Distribution

JÖRG PAWLITSCHKO[*]

Department of Statistics
University of Dortmund
D-44221 Dortmund

Abstract

This paper deals with the problem of estimating the location parameter of a two-parameter exponential distribution in case of contaminated data. Since in this case the sample minimum is an extremely unreliable estimator, robust alternatives are necessary. We investigate two types of estimators closer. The first type is based on a simple relation for the median. The second type has originally been suggested for Type-II-censored samples, but it also has good robustness properties. We discuss the breakdown properties of the two types of estimators and compare their performance for various patterns of data contamination in an extensive simulation study.

1 Introduction

In this paper we consider the problem of estimating the location parameter θ of a two-parameter exponential distribution $Exp(\theta, \nu)$ with density

$$f_{\theta,\nu}(t) \;=\; \frac{1}{\nu}\;\exp\Big(-\frac{t-\theta}{\nu}\Big), \qquad t > \theta.$$

The two-parameter exponential distribution provides a simple but nevertheless useful model for the analysis of lifetimes, especially when investigating reliability of technical equipment.

[*]pawl@statistik.uni-dortmund.de

In this context, the parameter θ can be interpreted as a guarantee time. When conducting a reliability experiment some of the items being tested may fail quite early. A common way to treat such observations is to consider the recorded times to failure not as trustworthy and to ignore them until the experiment has been running for a certain time. This proceeding leads to left censored lifetimes. Alternatively one may consider the complete data but take account of those spurious lifetimes by application of robust statistical methods. As we will see, on the one hand this approach leads to estimators already known from the analysis of censored data which now however can be investigated with respect to their robustness to violation of the exponential model. On the other hand, application of a wider class of robust methods offers a variety of new estimators for the location parameter.

Let $\underline{x}_N = (x_1, \ldots, x_N)'$ be a sample of size N with assumed parent distribution $Exp(\theta, \nu)$. With $x_{1:N} \leq \cdots \leq x_{N:N}$ we denote the corresponding ordered values. A common estimator for θ is the sample minimum $\hat{\theta}_{\min} = x_{1:N}$. This estimator is the maximum likelihood estimator for θ and it is known to be strongly consistent with convergence order N. However, it is extremely non-robust. Not only that it has low finite-sample breakdown point $1/(N+1)$, a single outlier that falls below the true θ determines this estimator completely and will inevitably lead to false conclusions. Therefore, robust alternatives to $\hat{\theta}_{\min}$ are necessary.

In this paper we investigate two approaches closer. One is based on a simple representation of the median of an $Exp(\theta, \nu)$-distributed random variable, the other makes use of ideas that have already successfully been applied in the case of censored data. We introduce these two types of estimators in Section 2 and discuss their breakdown behavior. Section 3 contains the results of an extensive simulation study that has been conducted to investigate the performance of these estimators in different situations of data contamination.

2 Robust Estimators

For an $Exp(\theta, \nu)$-distributed random variable X its median is given by

$$\mathsf{Med}(X) \;=\; \theta + \nu \, \ln 2. \tag{1}$$

Hence, if $S_N(\underline{x}_N)$ is a robust estimator of the scale parameter ν, a robust estimator of θ can be obtained by setting

$$\hat{\theta}_S \;=\; \mathsf{Med}(\underline{x}_N) - S_N(\underline{x}_N) \, \ln 2. \tag{2}$$

Estimators of this type have been suggested by Rousseeuw and Croux (1993). In the following, we will call such an estimator a Median-Scale- (MS-) estimator. Of course, there are many possible choices for S_N in (2). We concentrate on some which have explicit representations so that they are unique and simple to calculate and have breakdown point equal to 1/2. Further, all these choices render the corresponding MS-estimator affine equivariant.

(i) Median absolute deviation from the median:

$$MAD_N(\underline{x}_N) \;=\; a\,\mathsf{Med}_i\big|x_i - \mathsf{Med}(\underline{x}_N)\big|.$$

(ii) RCS-estimator (Rousseeuw and Croux, 1993):

$$RCS_N(\underline{x}_N) \;=\; b\,\mathsf{Med}_i\,\big(\mathsf{Med}_j\,\big|x_i - x_j\big|\big).$$

(iii) RCQ-estimator (Rousseeuw and Croux, 1993):

$$RCQ_N(\underline{x}_N) \;=\; c\,\big\{|x_i - x_j|;\; i,j \in \{1,\dots,N\},\, i<j\big\}_{l:N\,(N-1)},$$

where

$$l \;=\; \left\lceil \frac{N\,(N-1)}{8} \right\rceil.$$

(iv) Length of the shortest half sample (Rousseeuw and Leroy, 1988):

$$LSH_N(\underline{x}_N) \;=\; d \min_{i=1,\dots,N-h} \big(x_{h+i:N} - x_{i:N}\big),$$

where $h = \lfloor N/2 \rfloor$.

For any distribution from a location-scale family the constants in (i) – (iv) can be chosen such that the respective estimator is Fisher-consistent for the scale parameter. In case of an exponential distribution one has to set $a = 2.0781$, $b = 1.6982$, $c = 3.476$, $d = 1.4427$.

A different type of estimator can be found by using results from the analysis of censored data. In case of a doubly Type-II-censored sample the smallest and largest values are not observable. Suppose that only $x_{r:N} \le \cdots \le x_{N-s:N}$ are known for some fixed $1 < r < N-s < N$. In this situation the BLUE of θ is well-known to equal

$$\hat{\theta}_{r,s} \;=\; x_{r:N} - \frac{\sum_{i=N-r+1}^{N} 1/i}{N-r-s} \left(\sum_{i=r}^{N-s} x_{i:N} + s\,x_{N-s:N} - (N-r+1)\,x_{r:N} \right), \qquad (3)$$

cf. e.g. Balakrishnan and Sandhu (1995). In case of a complete sample, this estimator can be interpreted as a special kind of L-estimator giving weight zero to the $r-1$ smallest and s largest observations. If the proportion of outliers in the complete sample is not too large, then it can be expected that for appropriately chosen values of r and s this L-estimator will have good robustness properties.

In the following, we discuss the breakdown behavior of the location estimators introduced so far. Since in case of an exponential distribution the location parameter θ does not characterize the center of the distribution, it is useful to distinguish between two types of breakdown.

Let T_N denote any estimator of θ which is invariant under permutations of the observations. Then we define finite-sample breakdown points of T_N as follows:

$$\begin{aligned}
b_+(T_N, \underline{x}_N) &= \frac{1}{N} \min\big\{m : \sup_{y_1,\dots,y_m \in \mathbb{R}} T_N(x_1,\dots,x_{N-m},y_1,\dots,y_m) = \infty\big\}, \\
b_-(T_N, \underline{x}_N) &= \frac{1}{N} \min\big\{m : \inf_{y_1,\dots,y_m \in \mathbb{R}} T_N(x_1,\dots,x_{N-m},y_1,\dots,y_m) = -\infty\big\}, \\
b(T_N, \underline{x}_N) &= \min\big\{b_-(T_N, \underline{x}_N), b_+(T_N, \underline{x}_N)\big\}.
\end{aligned}$$

These breakdown points are replacement breakdown points in the sense of Donohoe and Huber (1983). Further, both b_+ and b_- can be seen as explosion breakdown points describing the proneness to positive explosion in case of b_+ and to negative explosion in case of b_-. Since θ is the lower boundary of the support of $Exp(\theta, \nu)$, the second type of breakdown poses the more severe problem. Obviously, for the sample minimum one has $b_+(\hat{\theta}_{\min}, \underline{x}_N) = 1$, however $b_-(\hat{\theta}_{\min}, \underline{x}_N) = 1/(N+1)$ which indicates that this estimator is not very reliable.

Proposition 1

(a) For the L-estimator $\hat{\theta}_{r,s}$ the finite-sample breakdown points are given by

$$b_+(\hat{\theta}_{r,s}, \underline{x}_N) = \frac{N-r+1}{N}, \qquad b_-(\hat{\theta}_{r,s}, \underline{x}_N) = \frac{\min\{r, s+1\}}{N}.$$

(b) For all four choices of S_N, the finite-sample breakdown points of the corresponding MS-estimators are given by

$$b_+(\hat{\theta}_S, \underline{x}_N) = \frac{1}{N}\left\lfloor \frac{N+1}{2} \right\rfloor, \qquad b_-(\hat{\theta}_S, \underline{x}_N) = \frac{1}{N}\left\lfloor \frac{N+1}{2} \right\rfloor.$$

Proof. For proving (a), note that an L-estimator of θ can also be written as

$$\hat{\theta}_{r,s}(\underline{x}_N) = \sum_{i=r}^{N-s} w_{[i:N]}\, x_{i:N}$$

with weights

$$\begin{aligned}
w_{[r:N]} &= 1 + \frac{(N-r)\ \sum_{i=N-r+1}^{N} 1/i}{N-r-s}, \\
w_{[i:N]} &= -\frac{\sum_{i=N-r+1}^{N} 1/i}{N-r-s} \qquad \text{for } i = r+1, \dots, N-s-1,
\end{aligned}$$

$$w_{[N-s:N]} \;=\; -\frac{(s+1)\,\sum_{i=N-r+1}^{N} 1/i}{N-r-s}.$$

Note that $w_{[r:N]} > 0$, $w_{[i:N]} < 0$ for $i = r+1, \ldots, N-s$, and

$$\sum_{i=r+1}^{N-s} w_{[i:N]} \;=\; 1 - w_{[r:N]}. \tag{4}$$

Now breakdown of $\hat{\theta}_{r,s}$ with respect to positive explosion is attained by replacing $N - r + 1$ observations in $\underline{x}_N$ by the same number y and letting $y \to \infty$. Then for the new sample $\underline{\tilde{x}}_N$, also $\tilde{x}_{r:N} \to \infty$ and from (4) it is clear that $\hat{\theta}_{r,s}(\underline{\tilde{x}}_N) \to \infty$. Note that no smaller number of replacements leads to positive explosion, since this would require $\tilde{x}_{i:N} \to -\infty$ for some $i > r$ which is only possible if also $\tilde{x}_{r:N} \to -\infty$ with at least the same rate of divergence.

Breakdown of $\hat{\theta}_{r,s}$ due to negative explosion can be achieved in two different ways: (i) replacing r observations in $\underline{x}_N$ by the same number y and letting $y \to -\infty$ and (ii) replacing $s+1$ observations by y and letting $y \to \infty$. The smaller number of replacements yields the corresponding breakdown point.

Part (b) can easily be derived from the explosion breakdown points of the standardized median and the respective scale estimators which (with the exception of LSH_N) have been calculated in Gather and Schultze (1999). $\square$

From Proposition 1 one sees that L-estimators with optimal breakdown points can be obtained by choosing $r = (N+1)/2$, $s = (N-3)/2$ or $r = (N-1)/2$, $s = (N-1)/2$ if N is odd and $r = N/2$, $s = N/2 - 1$ if N is even, however, the variances of these optimal L-estimators are much larger than that of the four MS-estimators making them not recommendable.

3 Finite-sample behavior of the location estimators under contamination – a simulation study

To compare the performance of the location estimators in case that not all observations in $\underline{x}_N$ come from the same two-parameter exponential distribution we conducted an extensive simulation study. To create different situations of "bad" data we considered a certain contamination model, namely a slippage model of Ferguson-type. For given sample size N and η_0, $\eta_1 \geq 0$, let $k_0 = \lfloor N\eta_0 \rfloor$, $k_1 = \lfloor N\eta_1 \rfloor$, and $n = N - k_0 - k_1$. We do not require $\eta_0 + \eta_1 \leq 1/2$ however both should not exceed the fraction of "good" observations. Then samples were generated which contained

- $n = N - k_0 - k_1$ observations from $Exp(\theta, 1)$; $\theta > 0$ (the "good" observations),
- k_0 observations from $Exp(0, 1)$ (location slippage),

and

- k_1 observations from $Exp(\theta, b)$, $b \gg 1$ (scale slippage).

We first investigate the more interesting case that no scale slippage is present. Figures 1–3 give bias and standard deviation of L-estimators with $s = 0$ and different values of r. Note that in this case the choice of $r = 1$ corresponds to the bias-corrected sample minimum. The location slippage amounts to $\theta = 0.5$, 1.5, 2.5. Each simulation run consisted of 5000 repetitions. The figures refer to a sample size of $N = 50$, results for other sample sizes turned out to be quite similar.

The simulations lead to the following results: If $\eta_0 > 0$ then all L-estimators have negative bias as could have been expected. The ability of coping with contaminated observations depends of the location difference. In case of small θ L-estimators can cope with a fraction η_0 of location-contaminants that is higher than r/N (see Figure 1). This effect disappears with θ becoming larger. The L-estimator with $r = \lfloor (N + 1)/2 \rfloor$ is resistant against any location-

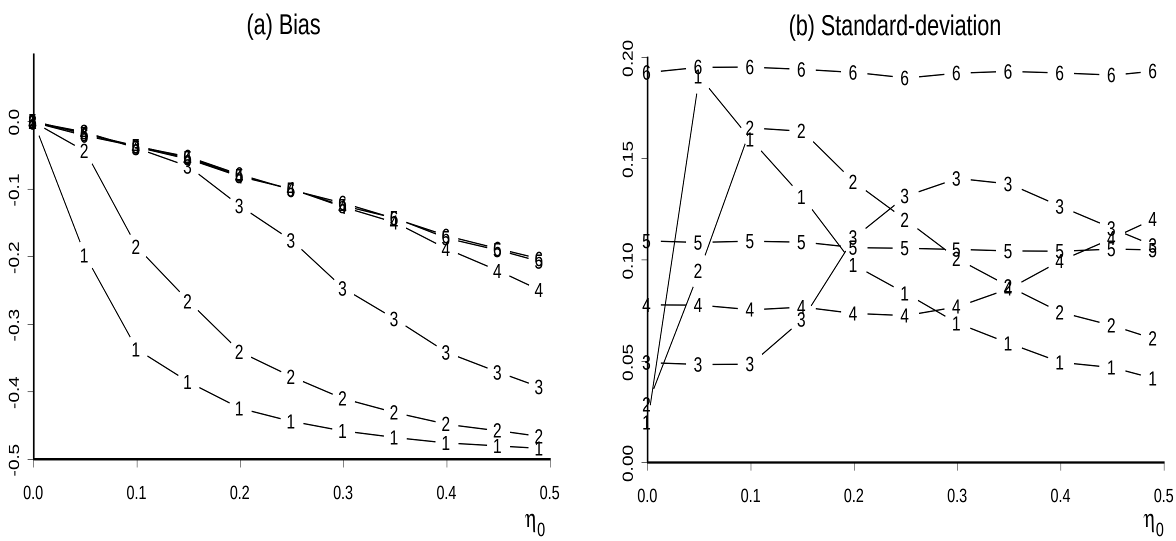

Figure 1: Bias and standard deviation of L-estimators with $N = 50$, $\theta = 0.5$; – 1 – : $r = 1$, – 2 – : $r = 2$, – 3 – : $r = 5$, – 4 – : $r = 10$, – 5 – : $r = 15$, – 6 – : $r = 25$.

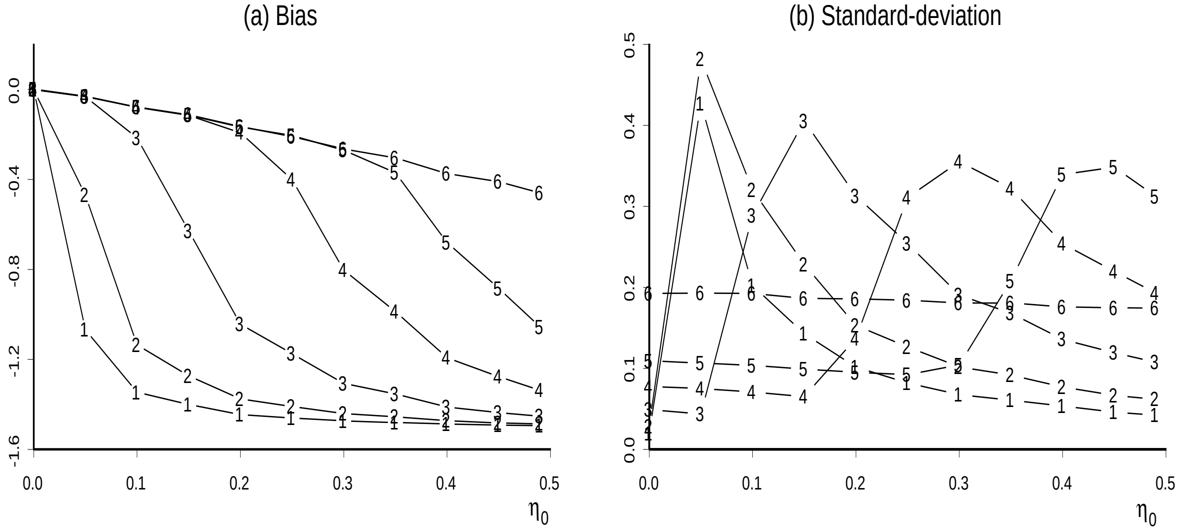

Figure 2: Bias and standard deviation of L-estimators with $N = 50$, $\theta = 1.5$; $-\,1\,-:\ r = 1$, $-\,2\,-:\ r = 2$, $-\,3\,-:\ r = 5$, $-\,4\,-:\ r = 10$, $-\,5\,-:\ r = 15$, $-\,6\,-:\ r = 25$.

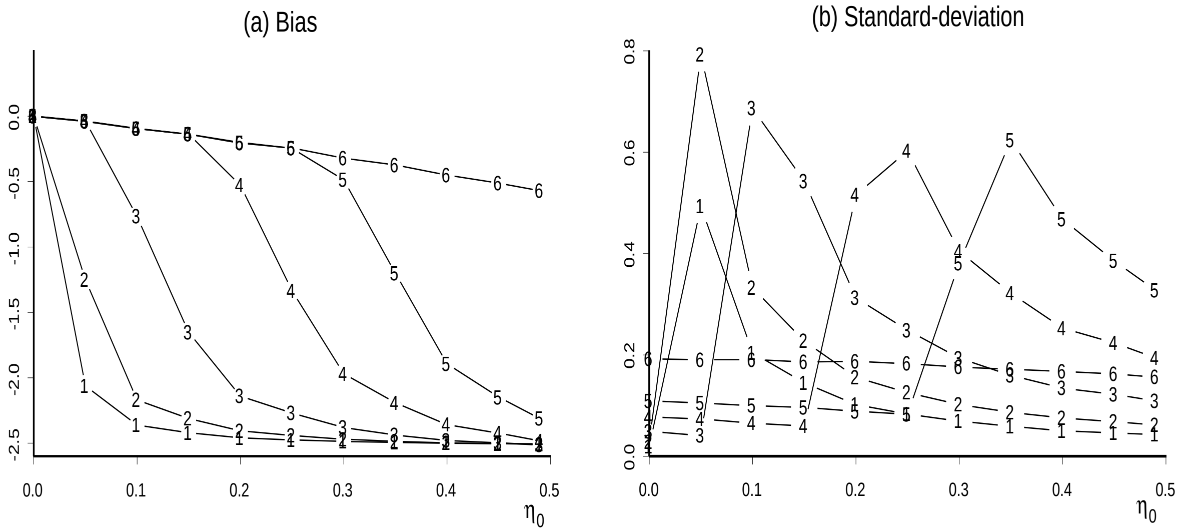

Figure 3: Bias and standard deviation of L-estimators with $N = 50$, $\theta = 2.5$; $-\,1\,-:\ r = 1$, $-\,2\,-:\ r = 2$, $-\,3\,-:\ r = 5$, $-\,4\,-:\ r = 10$, $-\,5\,-:\ r = 15$, $-\,6\,-:\ r = 25$.

contamination, however it has larger standard deviation than any of the MS-estimators (see below) and therefore cannot be recommended.

Figures 4–6 show comparable results for MS-estimators, when no scale contamination is present. The corresponding results for the sample minimum are displayed as well.

The simulation results for the MS-estimators lead not to a clear recommendation. Again, if $\eta_0 > 0$ then all MS-estimators considered here have negative bias. Further, the MAD-MS-estimator performs best with respect to bias with the exception of the case that both η_0 and θ are large and it has small standard-deviation if the fraction of location contaminants η_0 is not too large. The LSH-MS-estimator has always high (absolute) bias and standard-deviation and therefore is not recommendable. The RCQ-MS- and RCS-MS-estimators have similar standard deviation, the former however mostly has the larger (absolute) bias which for small η_0 even exceeds that of the LSH-MS-estimator. As a final conclusion one may state that within the class of MS-estimators the RCS-MS-estimator performs best if η_0 approaches 1/2, otherwise the MAD-MS-estimator has the edge.

When being compared with L-estimators, one finds that MS-estimators give better protection against a great variety of different contamination situations. However, if one has knowledge about amount and form of possible contamination, then an L-estimator with appropriately chosen r and $s = 0$ is preferable.

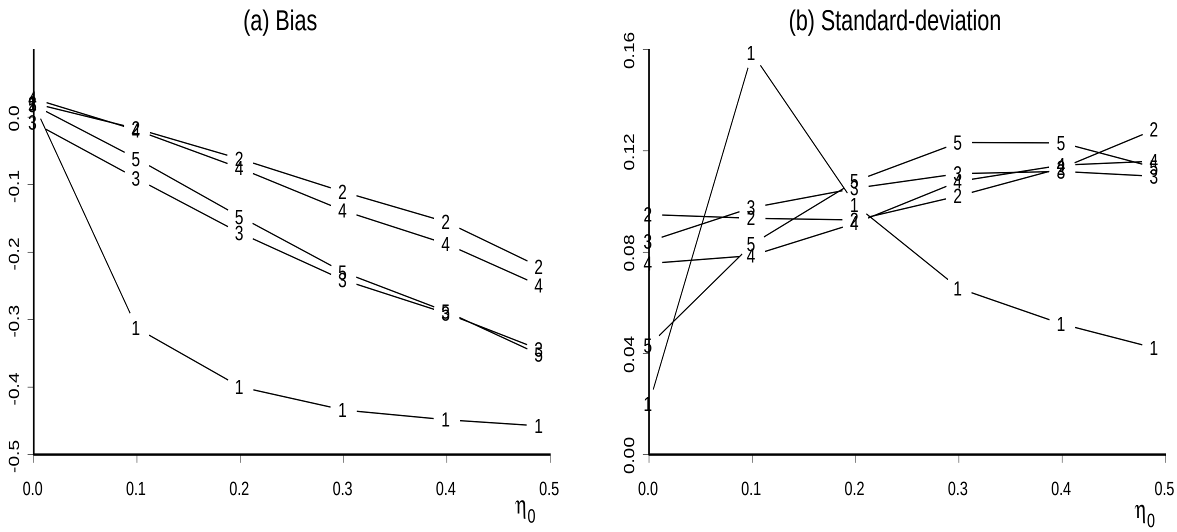

Figure 4: Bias and standard deviation of MS-estimators with $N = 50$, $\theta = 0.5$, no scale contamination; – 1 – : $\hat{\theta}_{\min}$; MS-estimators based on – 2 – : MAD, – 3 – : RCQ, – 4 – : RCS, – 5 – : LSH.

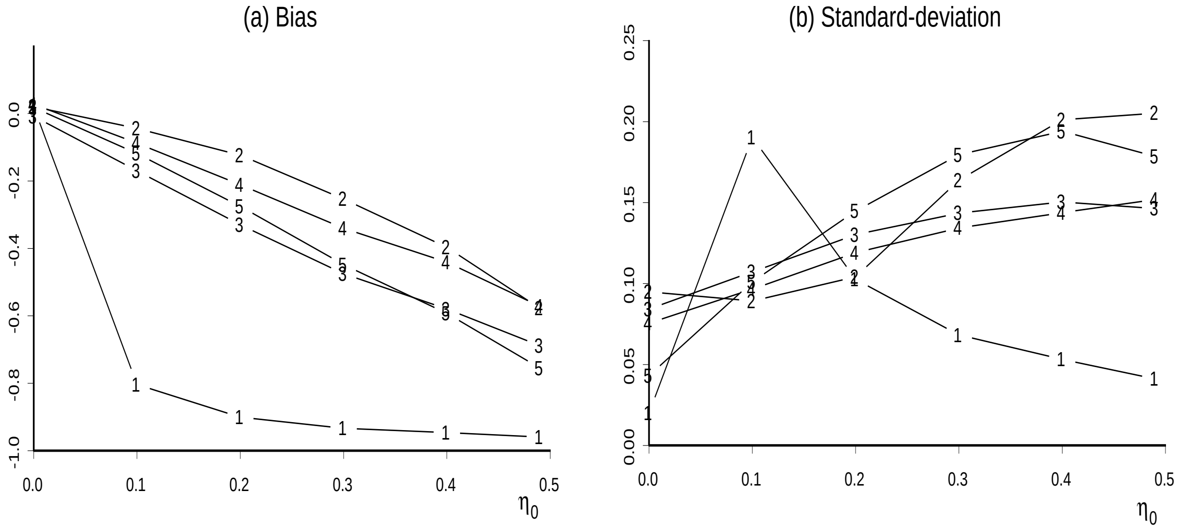

Figure 5: Bias and standard deviation of MS-estimators with $N = 50$, $\theta = 1$, no scale contamination ; $-\ 1\ -$: $\hat{\theta}_{\min}$; MS-estimators based on $-\ 2\ -$: MAD, $-\ 3\ -$: RCQ, $-\ 4\ -$: RCS, $-\ 5\ -$: LSH.

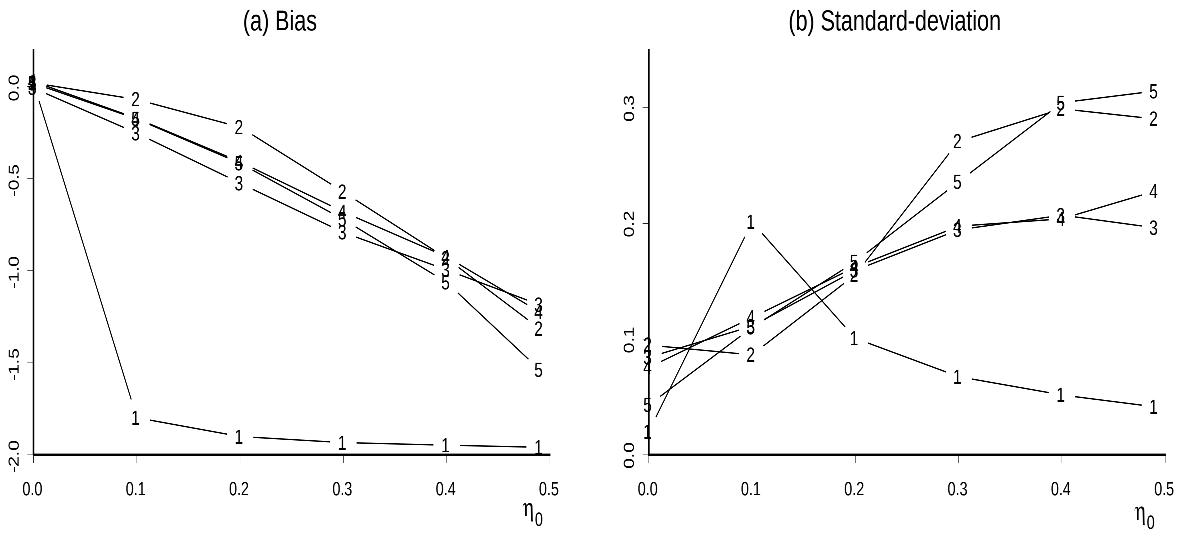

Figure 6: Bias and standard deviation of MS-estimators with $N = 50$, $\theta = 2$, no scale contamination ; $-\ 1\ -$: $\hat{\theta}_{\min}$; MS-estimators based on $-\ 2\ -$: MAD, $-\ 3\ -$: RCQ, $-\ 4\ -$: RCS, $-\ 5\ -$: LSH.

Figure 7 shows further results for L-estimators in case of additional scale contamination. Location slippage is fixed at $\theta = 2.5$, $\eta_0 = 0.1$, the parameter b characterizing the scale slippage has been set to $b = 16$.

As it could be expected, now also the choice of s should carefully be made. If s is chosen to small, then with growing r the additional scale contamination leads to a negative bias of the L-estimator, which can be very substantial if η_1 is large. The bias-problem can be avoided if s is chosen as large as possible, however, as already mentioned, if also r is large the resulting breakdown-optimal L-estimator has the drawback of a large variance.

Interestingly, if r is chosen such that it exceeds $N\,\eta_0$ only by a small amount, then the choice of s has no effect on the performance of the resulting L-estimator.

With the exception of the RCQ-MS-estimator, the performance of the MS-estimators is not much affected by additional scale contamination. The former tends to be biased downwards with increasing η_1 and is therefore not recommendable in this case. Again, best results are obtained with the MAD- and the RCS-MS-estimator, the former mostly having the smaller bias, the latter having the smaller variance.

Figures 8–10 show simulation results for bias and standard-deviation of the MAD-MS- and RCS-MS-estimator together with related results for some selected L-estimators in the special case that the parameters θ and b are set to $\theta = 2$ and $b = 16$, respectively. As in the case of no scale contamination, the performance of the MS-estimators is more stable over different contamination situations than that of the L-estimators. Again, for fixed η_0 and η_1 it is nearly always possible to choose certain values for r and s such that the resulting L-estimator shows better results than its competitors. The same L-estimator, however, may lead to very bad results if the fractions of contaminated observations change.

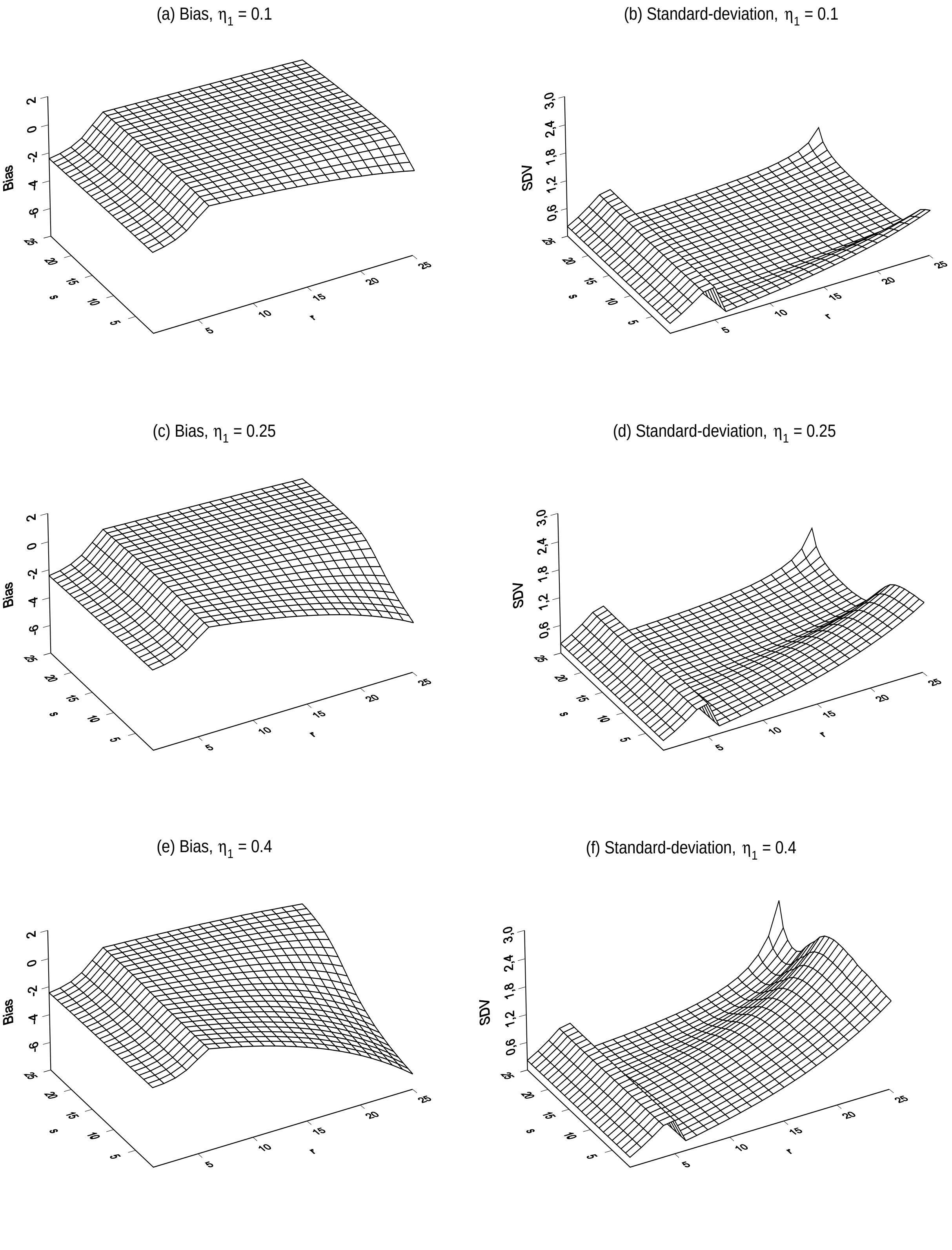

Figure 7: Bias and standard deviation of L-estimators with $N = 50$, $\theta = 2.5$, 10% location contamination for selected values of η_1

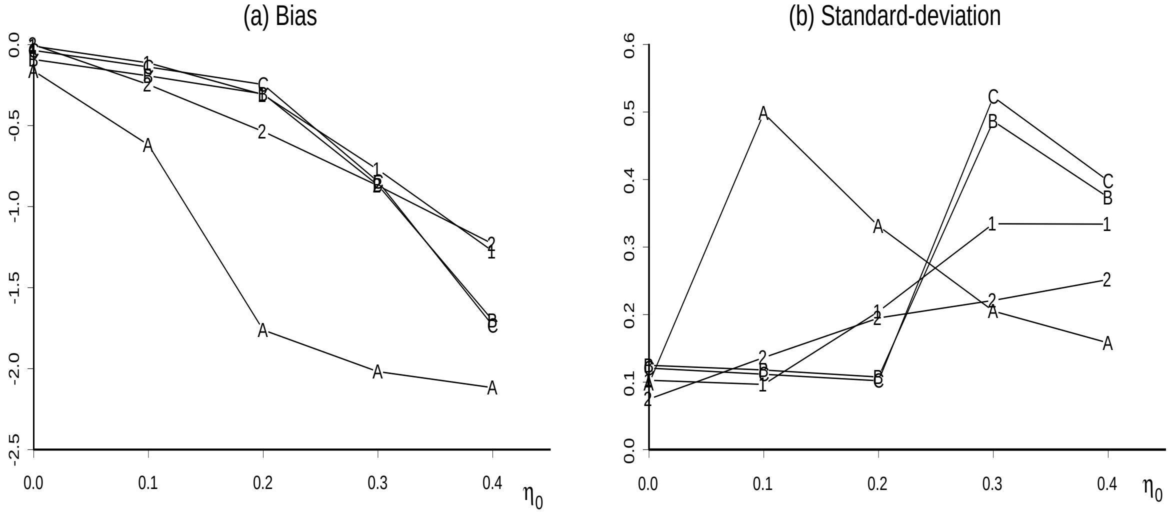

Figure 8: Bias and standard-deviation of some L- and MS-estimators with $N = 50$, $\theta = 2$, $b = 16$, $\eta_1 = 0.1$; MS-estimators based on – 1 – : MAD, – 2 – : RCS; L-estimators with – A – : $r = 5$, $s = 0$, – B – : $r = 13$, $s = 5$, – C – : $r = 13$, $s = 13$.

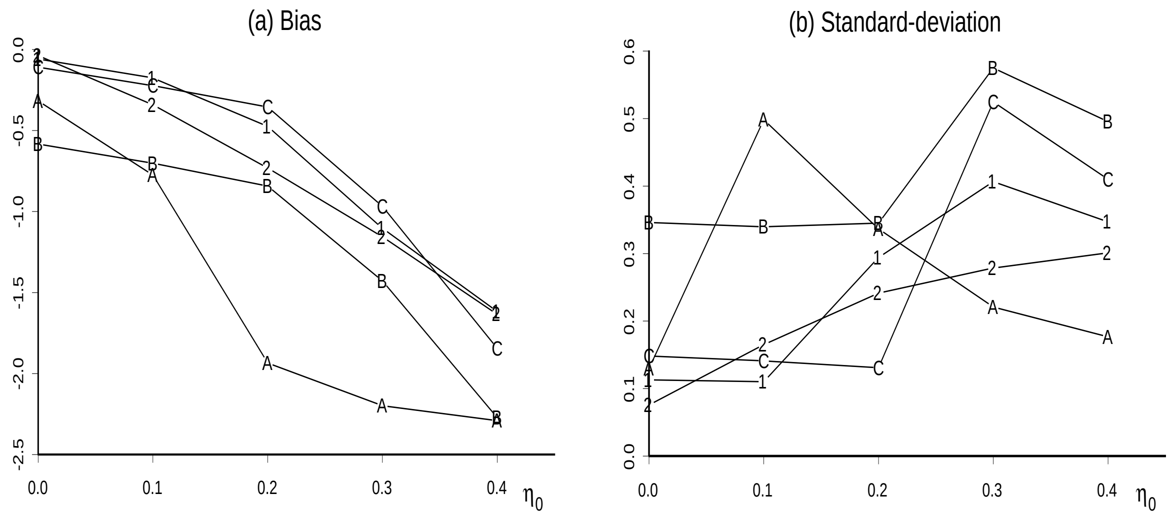

Figure 9: Bias and standard-deviation of some L- and MS-estimators with $N = 50$, $\theta = 2$, $b = 16$, $\eta_1 = 0.2$; MS-estimators based on – 1 – : MAD, – 2 – : RCS; L-estimators with – A – : $r = 5$, $s = 0$, – B – : $r = 13$, $s = 5$, – C – : $r = 13$, $s = 13$.

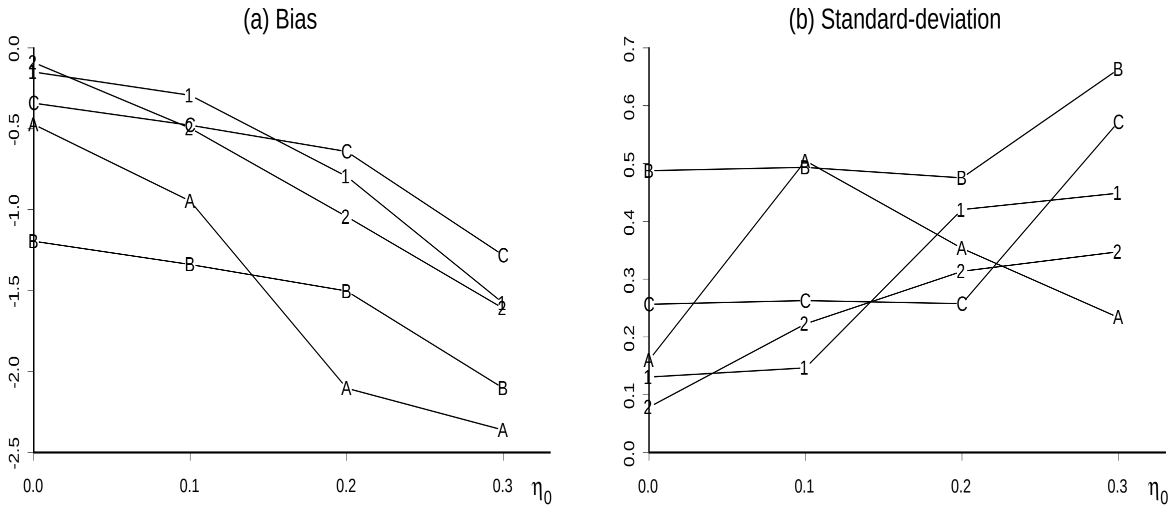

Figure 10: Bias and standard-deviation of some L- and MS-estimators with $N = 50$, $\theta = 2$, $b = 16$, $\eta_1 = 0.3$; MS-estimators based on – 1 – : MAD, – 2 – : RCS; L-estimators with – A – : $r = 5$, $s = 0$, – B – : $r = 13$, $s = 5$, – C – : $r = 13$, $s = 13$.

4 Example and Conclusion

Consider the following sample of $N = 15$ observations which are assumed to come i.i.d. from a two-parameter exponential distribution:

1.38,	11.31,	13.46,	15.01,	16.00,	17.49,	17.54,	17.89,
19.89,	23.07,	25.53,	32.44,	36.16,	40.61,	49.72.	

Figure 11 contains QQ-plots for checking the assumption of exponentiality, where in plot (a) the population quantiles are estimated non-robustly and in plot (b) their estimation is based on the RCS-estimator of the scale and the corresponding MS-estimator of the location parameter. Note that from (a) one would conclude that the assumption of a two-parameter exponential distribution is questionable. In plot (b), the smallest observation clearly sticks out as outlying and the remaining points roughly lie on the main diagonal thus supporting the assumption of exponentiality for the main body of the data.

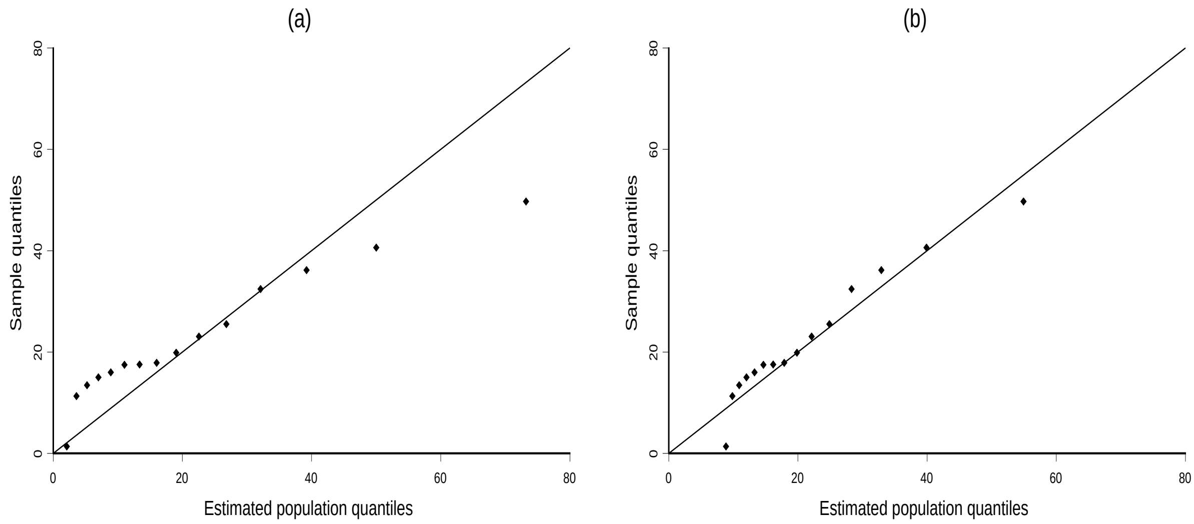

Figure 11: QQ-plots for the example: Parameters estimated with (a) sample mean and minimum, (b) RCS- and RCS-MS-estimator.

Table 1 contains the values of the four MS-estimators discussed so far as well as the values of some selected L-estimators.

MS-estimator based on	
MAD	10.4247
RCS	8.4059
RCQ	6.1291
LSH	9.3064

L-estimator		$s =$		
		0	1	2
	1	-0.1286	-0.1979	-0.2799
	2	9.4242	9.3716	9.3070
	3	10.7734	10.7074	10.6238
$r =$	4	11.4541	11.3699	11.2605
	5	11.3213	11.1949	11.0273
	6	11.7517	11.5918	11.3726
	7	9.6564	9.3116	8.8322

Table 1: Estimators of the location parameter for the example

Indeed, the sample has been generated as a random sample of size 14 from an $Exp(10, 16)$-distribution with an additional lower outlier. Therefore, all robust estimators, maybe with exception of the RCQ-MS-estimator, behave satisfactorily.

In this paper we have discussed several robust estimators of the location parameter of a two-dimensional exponential distribution. We have seen that estimators of the median-scale type provide good protection against different situations of location and/or scale contamination of the regular data. For a special contamination situation one might find an L-estimator which performs better with respect to its bias and variance, this location estimator, however, might fail in other situations. An important application of robust location estimators will be their implementation in outlier-identification procedures. Since mostly the fraction and type of outliers in a given sample will be unknown, for this application the use of MS-estimators is suggested. We will address this topic in further research.

Acknowledgement

Support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

Balakrishnan, N. and R.A. Sandhu (1995). Linear estimation under censoring and inference, in: N. Balakrishnan, and A.P. Basu (eds.), *The Exponential Distribution – Theory, Methods and Applications*, Gordon and Breach, Amsterdam, 53–72.

Donoho, D.L. and P.J. Huber (1983). The notion of breakdown point, in: P.J. Bickel, K.A. Doksum, and J.L. Hodges, Jr. (eds.), *A Festschrift for Erich Lehmann*, Wadsworth, Belmont, 157–184.

Gather, U. and V. Schultze (1999). Robust estimation of scale of an exponential distribution. Statistica Neerlandica, 53, 327–341.

Rousseeuw, P.J. and C. Croux (1993). Alternatives to the median absolute deviation. J. Amer. Statist. Assoc., 88, 1273–1283.

Rousseeuw, P.J. and A.M. Leroy (1988). A robust scale estimator based on the length of the shortest half. Statistica Neerlandica, 42, 103–116.