Aust. N. Z. J. Stat. 2017 doi: 10.1111/j.1467-842X.XXX

INVESTIGATION OF THE PERFORMANCE OF TRIMMED ESTIMATORS OF LIFE TIME DISTRIBUTIONS WITH CENSORING

BRENTON R. CLARKE[1]* ALEXANDRA HÖLLER[2] CHRISTINE H. MÜLLER[3] AND KARURU WAMAHIU[4]

Murdoch University and Technische Universität Dortmund

Summary

For the lifetime (or negative) exponential distribution, the trimmed likelihood estimator has been shown to be explicit in the form of a β-trimmed mean which is representable as an estimating functional that is both weakly continuous and Fréchet differentiable and hence qualitatively robust at the parametric model. It also has high efficiency at the model. The robustness is in contrast to the maximum likelihood estimator (MLE) involving the usual mean which is not robust to contamination in the upper tail of the distribution. When there is known right censoring, it may be perceived that the MLE which is the most asymptotically efficient estimator may be protected from the effects of "outliers" due to censoring. We demonstrate that this is not the case generally, and in fact, based on the functional form of the estimators, suggest a hybrid defined estimator that incorporates the best features of both the MLE and the β-trimmed mean. Additionally, we study the pure trimmed likelihood estimator for censored data and show that it can be easily calculated and that not always the censored observations are trimmed. The different trimmed estimators are compared by a simulation study.

Key words: Efficiency, Trimmed likelihood, Life time distributions, β-trimmed mean

1. Introduction

It is folklore that the classical maximum likelihood estimator (MLE) is non-robust. In fact it is well observed in the robustness literature that the MLE based on heavy tailed distributions such as the Student t-distribution on $\nu < +\infty$ degrees of freedom is in fact both efficient and robust. For example, Clarke (1983) shows in Section 7 that the estimator of location and scale of the Cauchy distribution, which is a special case of the Student t-distribution with $\nu = 1$, has a smooth bounded influence function. This means that the estimating functional is weakly continuous, Fréchet differentiable and, as the estimator is the MLE, it is efficient with asymptotic variance being the inverse of the Fisher Information.

For lifetime exponential distributions it is well understood that the classical maximum likelihood estimator is known not to be robust, however, for lifetime distributions with censoring we see bounded influence of outlying values due to fixed and known right censoring. This raises the question of whether or not the MLE is robust in this setting and

*Author to whom correspondence should be addressed.
[1]Brenton R Clarke, Mathematics and Statistics, School of Engineering and Information Technology, Murdoch University, 90 South Street, Murdoch, Western Australia, 6150. email: b.clarke@murdoch.edu.au
[2]Alexandra Höller, Fakultät Statistik, Technische Universität Dortmund, 44221 Dortmund, Germany. email: alexandra.hoeller@tu-dortmund.de [3]Christine H Müller, Fakultät Statistik, Technische Universität Dortmund, 44221 Dortmund, Germany. email: cmueller@statistik.tu-dortmund.de [4]Karuru Wamahiu, School of Veterinary and Life Sciences, Environmental and Conservation Sciences, Murdoch University, 90 South Street, Murdoch, Western Australia, 6150. email: k.wamahiu@murdoch.edu.au

whether one should use alternative estimates. This is important in both medical research and reliability studies. See Collett (2003), and James (1986) more generally, for the development and theory of classical estimation in the presence of censoring, particularly in medical applications. Engineering and reliability applications often include industrial examples where experiments are often being censored at a particular time point C which is fixed before the experiments. In particular this situation as well as random noninformative censoring is considered here in the context of robust estimation.

Trimmed likelihood estimation is a relatively new approach to robust estimation and was introduced in separate developments by Bednarski & Clarke (1993), Vandev & Neykov (1993) and Hadi & Luceño (1997). Trimmed likelihood estimators extend the least median of squares estimators and the least trimmed squares estimators of Rousseeuw (1984) and Rousseeuw & Leroy (1987) by replacing the likelihood functions of the normal distribution by likelihood functions of other distributions. Müller & Neykov (2003) applied them for generalized linear models and other applications can be found for example in Neykov, Filzmoser, Dimova *et al.* (2007), Cheng & Biswas (2008), Neykov, Filzmoser & Neytchev (2014). Clarke, Gamble & Bednarski (2000) (see also Clarke, Gamble & Bednarski 2011) showed that the functional form of the trimmed likelihood estimator for the (negative) exponential distribution is equivalent to that for the β-trimmed mean studied at length in Staudte & Sheather (1990). Hence in the case of the exponential distribution, the trimmed likelihood estimator is an explicit estimator. Given the trimming and outlier proportion β is fixed and known, the estimator has relatively high efficiency at the model exponential distribution and can be more efficient than estimators based on quantiles as demonstrated in Staudte & Sheather (1990). Moreover the estimator functional is both weakly continuous and Fréchet differentiable at the model. This, for example, guarantees both consistency and asymptotic normality of the estimator. It is also qualitatively robust in a sense due to Hampel (1971). The trimmed likelihood estimator for the exponential distribution was also extended to accelerated lifetime experiments and to general regression setups with the exponential distribution by Müller, Szugat, Celik *et al.* (2016) who derived the influence function for this situation.

In a parallel development Ahmed, Volodin & Hussein (2005) investigate a weighted trimmed likelihood for the exponential distribution which leads to a weighted mean, where the weights depend on a significance level and the data. This leads to subsequent trimming of large observations before calculating the mean of the remainder. This is not the same form as the β-trimmed mean of Staudte & Sheather (1990) which re-scales the trimmed mean, with largest observations trimmed, based on a constant that depends on the proportion of observations trimmed. This re-scaling gives a consistent estimator at the model parametric family. Nevertheless both approaches make use of a trimmed mean.

In this paper, we consider trimmed estimators for censored data. Recently, Farcomeni & Viviani (2011) proposed a trimmed estimator for censored data in the Cox regression model by maximizing a trimmed partial likelihood via a Metropolis-type algorithm. However, here we will provide foundations for trimming censored data with the exponential distribution. In this case, the MLE can be written down explicitly. We show in Section 4 that this leads to a very simple method to calculate a trimmed likelihood estimator for this case. From the uncensored case where always the largest observations are trimmed with the β-trimmed mean, it could be expected that the largest observations, which are often the censored observations, are trimmed with a trimmed likelihood as well. However, we show that this is not always the case.

Nevertheless the trimmed likelihood estimator, like the MLE, has the drawback of a strong bias in some situations. Therefore we introduce additionally in Section 5 two natural hybrid estimators based on the MLE and the β-trimmed mean for the case of a fixed censoring constant C. These estimators are based on the idea that the trimmed estimator could be the MLE when the trimming proportion is not larger than the proportion of censored observations.

Before we consider the different types of trimmed estimators, Section 2 repeats known results for the maximum likelihood estimator and Section 3 offers known results for the β-trimmed mean. Section 6 provides a comparison of the different estimators via simulation.

2. The Maximum Likelihood Estimator

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent times until the occurrence of an event of interest and let $C_1, C_2, \ldots, C_n$ be a sequence of independent censoring variables, which are additionally independent of $X_1, X_2, \ldots, X_n$. Then what we observe is

$$Z_i = \min(X_i, C_i), \quad \text{for} \quad i = 1, 2, \ldots, n.$$

Now, suppose that events X_i have a probability distribution, $F_\theta(x)$, with density function $f_\theta(x)$ then the survivor function is

$$S_\theta(x) = 1 - F_\theta(x) = P(X_i \ > \ x)$$

where the distribution function $F_\theta(x) = 1 - e^{-x/\theta}$ and density $f_\theta(x) = \theta^{-1} e^{-x/\theta}$, for $x > 0, \theta > 0$. Therefore, the hazard function is $\lambda_\theta = f_\theta(x)/S_\theta(x) = \theta^{-1}$. If we denote

$$\begin{aligned}\Delta_i &= \begin{cases} 1 & \text{if } X_i \ < \ C_i \text{ (uncensored)} \\ 0 & \text{if } X_i \ \geq \ C_i \text{ (censored)} \end{cases} \\ &= I\{X_i \ < \ C_i\}\end{aligned}$$

where I is the indicator variable. Since the censoring variables are independent of $X_1, X_2, \ldots, X_n$, we can assume that they are independent of the parameter θ as well so that their distributions do not influence the likelihood function. The realizations of X_i, C_i, Z_i, Δ_i are denoted by x_i, c_i, z_i, δ_i. We find that what is recorded is (z_i, δ_i) for $i = 1, 2 \ldots n$, which leads to a likelihood function given by

$$\begin{aligned}L(\theta) &= \prod_{i=1}^n f_\theta(z_i)^{\delta_i} S_\theta(z_i)^{1-\delta_i} \\ &= \prod_{i=1}^n \left(\frac{f_\theta(z_i)}{S_\theta(z_i)}\right)^{\delta_i} S_\theta(z_i) \\ &= \prod_{i=1}^n \lambda_\theta(z_i)^{\delta_i} S_\theta(z_i) \\ &= \prod_{i=1}^n \left(\frac{1}{\theta}\right)^{\delta_i} e^{-z_i/\theta}.\end{aligned}$$

Now, taking the log of the likelihood

$$\log L(\theta) = -\sum_{i=1}^{n} \delta_i \log(\theta) - \frac{1}{\theta}\sum_{i=1}^{n} z_i$$

and applying the partial derivative and equating to zero

$$\frac{\partial}{\partial \theta} \log L(\theta) = -\frac{1}{\theta}\sum_{i=1}^{n} \delta_i + \frac{1}{\theta^2}\sum_{i=1}^{n} z_i \ = \ 0$$

leads to the maximum likelihood estimator (MLE) for θ

$$\hat{\theta} = \frac{\sum_{i=1}^{n} z_i}{\sum_{i=1}^{n} \delta_i} = \frac{\sum_{i=1}^{n} z_i}{n_{uc}}. \tag{1}$$

Here, n_{uc} is the total number of uncensored observations.

A special case is where all censoring variables are equal and given by a fixed constant C. Then, we notice that

$$\sum_{i=1}^{n} z_i = \sum_{i=1}^{n} x_i \ I(x_i < \ C) + (n - n_{uc})C$$

and substituting this into the MLE gives

$$\begin{aligned}\hat{\theta} &= \frac{\sum_{i=1}^{n} x_i \ I(x_i < \ C) + (n - n_{uc})C}{n_{uc}} \\ &= \frac{n \int_{[0,C)} x dF_n(x) + (n - n_{uc})C}{n_{uc}}.\end{aligned}$$

It can also be noticed that, the denominator, $n_{uc} = \#\{x_i < \ C\} = \lim_{x \uparrow C} nF_n(x)$, so that now the MLE

$$\hat{\theta} = \frac{n \int_{[0,C)} x dF_n(x) + (n - \lim_{x \uparrow C} nF_n(x))C}{\lim_{x \uparrow C} nF_n(x)}. \tag{2}$$

Here F_n is the empirical distribution function. We shall denote the estimator as $\hat{\theta}_{\text{mle}} \equiv \hat{\theta}_{\text{mle}}(F_n)$.

3. The β-Trimmed Mean

Staudte & Sheather (1990) gave the form for the β-trimmed mean

$$\begin{aligned}T_\beta[F] &= \mathbb{E}[X|X < x_{1-\beta}] \\ &= \frac{1}{1-\beta}\int_0^{F^{-1}(1-\beta)} x dF(x) \\ &= \mu_\beta\end{aligned}$$

where X has exponential distribution here, i.e. $F = F_\theta$. Now, noting that we are only concerned with trimming on the right, the estimator is then defined by the conditional distribution X

$$P(X < x | X < x_{1-\beta}) = \frac{1 - e^{-(x/\theta)}}{1-\beta}, 0 < x < x_{1-\beta}$$

which has density function

$$f_{\theta|X<x_{1-\beta}}(x) = \frac{1}{\theta(1-\beta)} e^{-(x/\theta)}, 0 < x < x_{1-\beta}.$$

The β-trimmed mean is then given by evaluating F at the empirical distribution F_n. Then the estimator for μ_β is achieved by replacing F with F_n

$$T_\beta[F_n] = \bar{x}_{n,\beta} = \frac{1}{r}\sum_{i=1}^{r} x_{(m)}$$

where $r = n - [n\beta]$ is the number of observations remaining after $[n\beta]$ have been trimmed, and $x_{(m)}$ is the *m*th order statistic. It follows that the consistent estimator for θ at the exponential distribution is (see Clarke, Gamble & Bednarski 2000, 2011)

$$\hat{\theta}_\beta = \frac{1-\beta}{1-\beta+\beta\ln\beta}\bar{x}_{n,\beta}. \tag{3}$$

4. The Trimmed Likelihood Estimator

The loglikelihood function L is given by $L(\theta) = \sum_{i=1}^{n} l_i(\theta)$ where

$$l_i(\theta) = -\delta_i \log(\theta) - \frac{1}{\theta} z_i, \quad i = 1, \dots, n,$$

are the individual loglikelihood functions. Let be $l_{(1)}(\theta) \le l_{(2)}(\theta) \le \dots \le l_{(n)}(\theta)$ the ordered individual loglikelihood functions where the ordering depends on θ. The h-trimmed likelihood estimator $\hat{\theta}_h$ is given by (see Müller & Neykov 2003)

$$\hat{\theta}_h = \arg\max_\theta \sum_{i=h+1}^{n} l_{(i)}(\theta)$$

where the h smallest individual loglikelihood functions are trimmed.

If there are no censored data, i.e. $\delta_i = 1$ for all $i = 1, \dots, n$, then the ordering $l_{(1)}(\theta) \le \dots \le l_{(n)}(\theta)$ is independent of θ so that the largest observations z_i must be trimmed to maximize $\sum_{i=h+1}^{n} l_{(i)}(\theta)$. Hence in this case, the h-trimmed likelihood estimator is the β-trimmed mean with $h = n - r = [n\beta]$.

If there are censored data then the ordering of $l_{(1)}(\theta) \le \dots \le l_{(n)}(\theta)$ depends on θ. Nevertheless, the h-trimmed likelihood estimator depends on the ordered observations as well and we have to consider at most $h+1$ cases, namely that $0, 1, \dots, h$ of the censored observations are trimmed. To see this, set $m = n_{uc} = \sum_{i=1}^{n} \delta_i$ for the number of uncensored observations and assume without loss of generality

$$\begin{aligned} &z_1 = x_1 \le z_2 = x_2 \le \ldots \le z_m = x_m, \\ &z_{m+1} = c_{m+1} \le z_{m+2} = c_{m+2} \le \ldots \le z_n = c_n. \end{aligned} \tag{4}$$

Subsets of the sample $\{z_1, \ldots, z_n\}$ can be defined by index sets $I \subset \{1, \ldots, n\}$ so that $z(I) = \{z_i;\ i \in I\}$. If $\hat{\theta}(z(I))$ denotes the maximum likelihood estimator for the subsample $z(I)$, which can be explicitely calculated according to Section 2, and $\sharp I$ denotes the number of elements in I then the h-trimmed likelihood estimator is also given by

$$\hat{\theta}_h = \hat{\theta}(z(I_*)) \quad \text{where} \quad I_* = \arg\max \left\{ \sum_{i \in I} l_i(\hat{\theta}(z(I)));\ I \text{ with } \sharp I = n - h \right\}.$$

Using this representation of the h-trimmed likelihood estimator, $\binom{n}{h}$ index sets I have to be considered which is usually too much for applications. However, here only at most $h+1$ index sets must be considered. For

$$k \in \mathcal{K} = \{\max\{0, m-h\}, \max\{0, m-h\}+1, \ldots, \min\{m, n-h\}-1, \min\{m, n-h\}\}$$

define the special index sets

$$\begin{aligned} &I_k = \{h+1, \ldots, n\} \quad \text{for} \quad k = 0 \quad \text{if} \quad m - h \le 0, \\ &I_k = \{1, \ldots, k, m+1, \ldots, m+n-h-k\} \\ &\qquad\qquad \text{for} \quad 0 < \max\{0, m-h\} < k < \min\{m, n-h\} < n-h, \\ &I_k = \{1, \ldots, n-h\} \quad \text{for} \quad k = n-h \quad \text{if} \quad n - h \le m. \end{aligned}$$

Note that $\mathcal{K}$ has only $h+1$ elements if $0 \le m-h$ and $m \le n-h$. Otherwise it has less than $h+1$ elements since the number m of uncensored observations is less than h or the number $n-m$ of censored observations is less than h. The index set I_0 includes no uncensored observations so that all uncensored observations are trimmed which is only possible if $h \ge m$. The set I_{n-h} includes no censored observations so that all censored observations are trimmed which is only possible if $h \ge n-m$. Generally, the set I_k includes k uncensored observations and $n-h-k$ censored observations.

Now consider an arbitrary index set I with $\sharp I = n-h$ and k indices less than $m+1$ so that the corresponding subsample contains k uncensored observations $z_{j(1)} = x_{j(1)} \le \ldots \le z_{j(k)} = x_{j(k)}$ and $n-h-k$ censored observations $z_{j(k+1)} = c_{j(k+1)} \le \ldots \le z_{j(n-h)} = c_{j(n-h)}$. The ordering of the observations given by (4) provides

$$z_1 \le z_{j(1)}, \ldots, z_k \le z_{j(k)},\ z_{m+1} \le z_{j(k+1)}, \ldots, z_{m+n-h-k} \le z_{j(n-h)},$$

so that for all $\theta > 0$

$$\begin{aligned} \sum_{i \in I} l_i(\theta) &= \sum_{i=1}^{k} \left(-\log(\theta) - \frac{1}{\theta} z_{j(i)} \right) - \sum_{i=k+1}^{n-h} \frac{1}{\theta} z_{j(i)} \\ &\le \sum_{i=1}^{k} \left(-\log(\theta) - \frac{1}{\theta} z_i \right) - \sum_{i=m+1}^{m+n-h-k} \frac{1}{\theta} z_i \\ &= \sum_{i \in I_k} l_i(\theta) \le \sum_{i \in I_k} l_i(\hat{\theta}(z(I_k))). \end{aligned}$$

Hence it is enough to consider only the index sets I_k with $k \in \mathcal{K}$ and the following Theorem holds.

Theorem 1. *The h-trimmed likelihood estimator $\hat{\theta}_h$ is given by*

$$\hat{\theta}_h = \hat{\theta}(z(I_{k_*})) \quad \textit{where} \quad k_* = \arg\max \left\{ \sum_{i \in I_k} l_i \left(\hat{\theta}(z(I_k)) \right); \ k \in \mathcal{K} \right\}$$

and $\hat{\theta}(z(I_k)) = \frac{1}{k} \left(\sum_{i=1}^{k} z_i + \sum_{i=m+1}^{m+n-h-k} z_i \right)$ *is the maximum likelihood estimator for the subsample given by I_k.*

The following simple example with fixed censoring constant shows that not always $k_* = \min\{m, n-h\}$ holds so that even in this case not always all censored observations are trimmed.

Example 1. 10 observations were simulated with an exponential distribution with $\theta = 10$ and the fixed censoring constant $C = 10$ was used. This led to the following ordered observations

$$1.397953, 1.457067, 1.47046, 4.360686, 5.396828, 7.551818, 9.565675, 10, 10, 10$$

so that 3 observations were censored. If $h = 2$ then $k_* = 5$ so that only the two largest uncensored observations are trimmed. The 2-trimmed estimator is then $\hat{\theta}_2 = 8.816599$ while the maximum likelihood estimator for the whole sample is given by 8.742927.

If θ is less than 1 or close to 1 then it is more likely that the censored observations are trimmed.

5. The Proposed Hybrid Estimators

Here we assume a fixed censoring constant C. Then the β-trimmed mean, $\hat{\theta}_\beta$, can be used provided that $F_n^{-1}(1-\beta) < C$. Now replacing F_n by $F_{\hat{\theta}_\beta}$ in (2) allows us to consider the pseudo MLE, $\hat{\theta}_{\text{pmle}}$, which is given by

$$\hat{\theta}_{pmle} = \frac{n \int_0^C x dF_{\hat{\theta}_\beta}(x) + (n - nF_{\hat{\theta}_\beta}(C))C}{nF_{\hat{\theta}_\beta}(C)}. \quad (5)$$

5.1. Proposal 1

The β-trimmed mean, $\hat{\theta}_\beta$, makes no sense if $F_n^{-1}(1-\beta) \geq C$, so that we suggest to just use $\hat{\theta}_{\text{mle}}$ in this case. Hence the new trimmed estimator can be of the form

$$\hat{\theta}_{\text{p1}} = \begin{cases} \hat{\theta}_{\text{pmle}} & \text{if } F_n^{-1}(1-\beta) < C, \\ \hat{\theta}_{\text{mle}} & \text{if } F_n^{-1}(1-\beta) \geq C. \end{cases} \quad (6)$$

Note that $F_n^{-1}(1-\beta) \geq C$ means $F_n^{-1}(1-\beta) = C$ since $F_n(C) = 1$. Now, evaluating the integral from (5)

$$\int_0^C x dF_{\hat{\theta}_\beta}(x) = \int_0^C x \frac{1}{\hat{\theta}_\beta} e^{-x/\hat{\theta}_\beta} dx$$
$$= \hat{\theta}_\beta - (C + \hat{\theta}_\beta) e^{-C/\hat{\theta}_\beta} \quad (7)$$

and we get

$$\hat{\theta}_{\text{pmle}} = \frac{n(\hat{\theta}_\beta - (C+\hat{\theta}_\beta)e^{-C/\hat{\theta}_\beta}) + (n - nF_{\hat{\theta}_\beta}(C))C}{nF_{\hat{\theta}_\beta}(C)}.$$

Then substitute $F_{\hat{\theta}_\beta}(C) = 1 - e^{-C/\hat{\theta}_\beta}$ into the equation

$$\begin{aligned}
\hat{\theta}_{\text{pmle}} &= \frac{n\hat{\theta}_\beta - n(C+\hat{\theta}_\beta)e^{-C/\hat{\theta}_\beta} + (n - n(1-e^{-C/\hat{\theta}_\beta}))C}{n(1-e^{-C/\hat{\theta}_\beta})} \\
&= \frac{n\hat{\theta}_\beta - nCe^{-C/\hat{\theta}_\beta} - n\hat{\theta}_\beta e^{-C/\hat{\theta}_\beta} + nCe^{-C/\hat{\theta}_\beta}}{n(1-e^{-C/\hat{\theta}_\beta})} \\
&= \frac{n\hat{\theta}_\beta(1-e^{-C/\hat{\theta}_\beta})}{n(1-e^{-C/\hat{\theta}_\beta})} \\
&= \hat{\theta}_\beta
\end{aligned}$$

which is the consistent trimmed mean estimator in (3). Hence, the proposal for the new censored data estimator is defined by

$$\hat{\theta}_{\text{p1}} = \begin{cases} \hat{\theta}_\beta & \text{if } F_n^{-1}(1-\beta) < C, \\ \hat{\theta}_{\text{mle}} & \text{if } F_n^{-1}(1-\beta) \geq C. \end{cases} \tag{8}$$

5.2. Proposal 2

A second possibility is to recognize that we can expand the integral, from (2)

$$\int_{[0,C)} x dF_n(x) = \int_{[0,F_n^{-1}(1-\beta))} x dF_n(x) + \int_{[F_n^{-1}(1-\beta),C)} x dF_n(x)$$

and noticing that $(1-\beta)T_\beta[F_n] = \int_{[0,F_n^{-1}(1-\beta))} x dF_n(x)$ then gives

$$\int_{[0,C)} x dF_n(x) = (1-\beta)T_\beta[F_n] + \int_{[F_n^{-1}(1-\beta),C)} x dF_n(x). \tag{9}$$

The second integral in the expression could be unstable, due to large observations. We therefore replace F_n by $F_{\hat{\theta}_\beta}$ which gives

$$\begin{aligned}
\int_{[F_n^{-1}(1-\beta),C)} x dF_n(x) &\approx \int_{F_n^{-1}(1-\beta)}^{C} x dF_{\hat{\theta}_\beta} \\
&= \int_{F_n^{-1}(1-\beta)}^{C} x \frac{1}{\hat{\theta}_\beta} e^{-x/\hat{\theta}_\beta} dx \\
&= -(C+\hat{\theta}_\beta)e^{-C/\hat{\theta}_\beta} + (F_n^{-1}(1-\beta) + \hat{\theta}_\beta)e^{-F_n^{-1}(1-\beta)/\hat{\theta}_\beta}
\end{aligned}$$

So

$$\int_{[0,C)} x dF_n(x) \approx (1-\beta)T_\beta[F_n] - (C+\hat{\theta}_\beta)e^{-C/\hat{\theta}_\beta} + (F_n^{-1}(1-\beta)+\hat{\theta}_\beta)e^{-F_n^{-1}(1-\beta)/\hat{\theta}_\beta} \equiv \text{Correction}(F_n,\beta,C)$$

Here, we find the resulting pseudo corrected maximum likelihood estimator (pcmle) is then

$$\hat{\theta}_{\text{pcmle}} = \frac{n\text{Correction}(F_n,\beta,C) + (n - \lim_{x\uparrow C} nF_n(x))C}{\lim_{x\uparrow C} nF_n(x)}.$$

Hence, our proposed estimator is defined by

$$\hat{\theta}_{\text{p2}} = \begin{cases} \hat{\theta}_{\text{pclme}} & \text{if } F_n^{-1}(1-\beta) < C, \\ \hat{\theta}_{\text{mle}} & \text{if } F_n^{-1}(1-\beta) \geq C. \end{cases} \tag{10}$$

6. Simulation results

The performances of the proposed estimators, $\hat{\theta}_h$ (TLE), $\hat{\theta}_{\text{p1}}$ (P1), and $\hat{\theta}_{\text{p2}}$ (P2), were compared with that of the maximum likelihood estimator (MLE) at the contamination “model” $F = (1-\epsilon)\,F_{\theta_1} + \epsilon\,F_{\theta_2}$. At this contamination model, a small fraction ϵ of observations were generated with θ_2 satisfying $\theta_2 > \theta_1$ so that these observations are outliers, i.e. they tend to be much larger then the observations of the central model generated with θ_1. We are especially interested in the results attained by the estimators when considering small and large sample sizes and when the censoring constant C ranges from one to large. Hence $10,000$ replicates of size $n = 20, 100, 500$ were generated.

Figure 1 shows the means, the variances, the squared biases, and the mean squared errors (MSE) of the estimates in these simulations for n=20. The biases and the MSE were calculated with respect to θ_1 of the central model. All estimates are strongly biased. However the trimmed likelihood estimator (TLE) underestimates the value of the central model while all other estimators overestimate. Using this small sample size, the MSE of the TLE is the best as soon as the censoring constant is not too small. This changes completely using larger sample sizes as shown in Figure 2. For the larger sample sizes, the TLE provides the worst MSE because it underestimates the parameter of the central model much more than the other estimate overestimate it. The estimator P1 shows the best MSE for these sample sizes if the censoring constant is not too small. It outperforms clearly the MLE which is not the case for the TLE. The estimator P2 is also better than the MLE for larger censoring constants but it is always worse than P1.

Additionally, we modified the proportion of outliers and the parameters of the central model and the contamination. Some of these results are shown in Figure 3. Here the TLE is the best as soon as the trimming proportion β is the same as the contamination proportion ϵ. If this is not the case then P1 is superior. The MLE becomes the worst for large censoring constant.

All scenarios showed that the MSE of the MLE is an increasing function of the censoring constant while the trimmed estimators are reaching a plateau or are decreasing for large

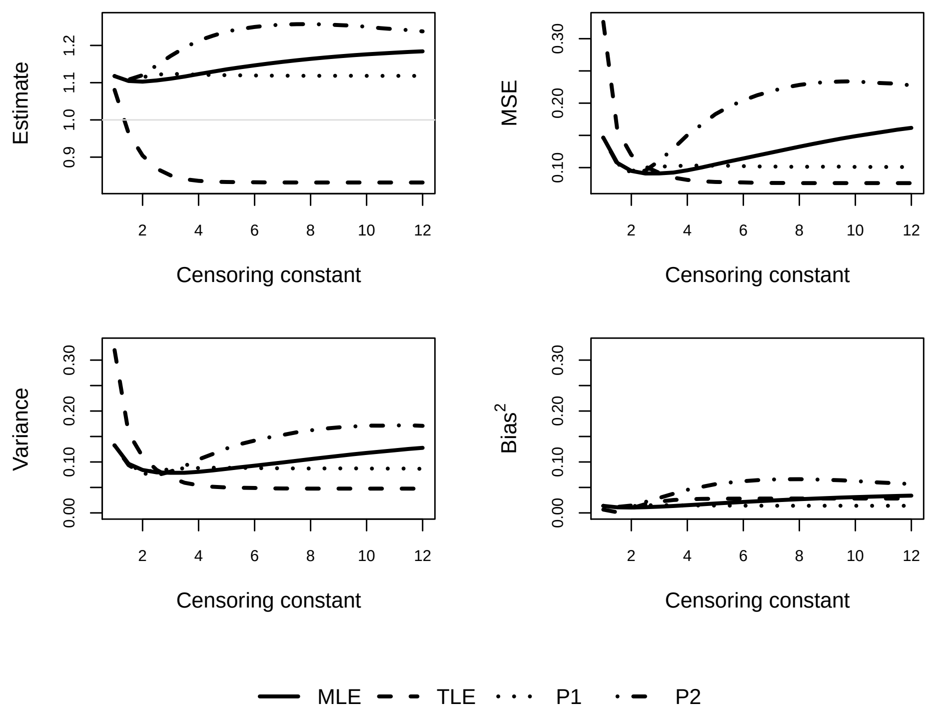

Figure 1. Performance of the estimators for n=20 with $\epsilon = 0.05$, $\theta_1 = 1$, $\theta_2 = 5$, and 10% trimming

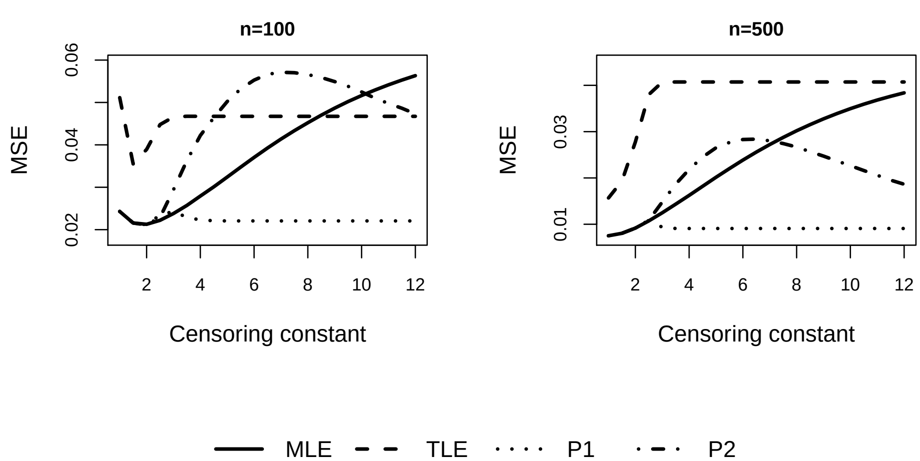

Figure 2. MSE of the estimators for n=100 and n=500 with $\epsilon = 0.05$, $\theta_1 = 1$, $\theta_2 = 5$, and 10% trimming

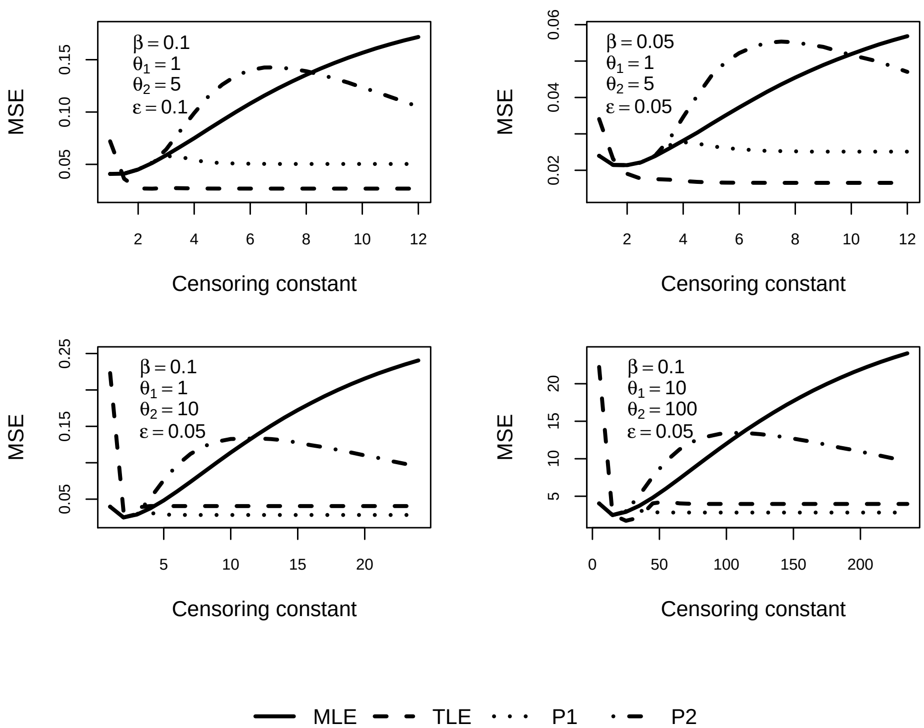

Figure 3. MSE of the estimators for n=100 with different constellations of ϵ, θ_1, θ_2, and the trimming proportion β

censoring constants like P2. Hence the trimmed estimators are much more robust against contamination than the MLE. Moreover, P1 is always better than P2. Hence the only robust competitors are the TLE and the P1 estimator. The P1 estimator outperforms the TLE for larger samples sizes if the proportions of contamination and trimming do not coincide.

Acknowledgement

The research was supported by the DFG Collaborative Research Center SFB 823 *Statistical modeling of nonlinear dynamic processes*.

References

AHMED, E.S., VOLODIN, A.I. & HUSSEIN, A.A. (2005). Robust weighted likelihood estimation of exponential parameters. *IEEE Trans. Reliab.* **54**, 389–395.

BEDNARSKI, T. & CLARKE, B.R. (1993). Trimmed likelihood estimation of location and scale of the normal distribution. *Austral. J. Statist.* **35**, 141–153.

CHENG, T.C. & BISWAS, A. (2008). Maximum trimmed likelihood estimator for multivariate mixed continuous and categorical data. *Computational Statistics & Data Analysis* **52**, 2042–2065.

CLARKE, B.R. (1983). Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations. *Ann. Stat.* **11**, 1196–1205.

CLARKE, B.R., GAMBLE, D.K. & BEDNARSKI, T. (2000). A note on robustness of the β-trimmed mean. *Australian & New Zealand Journal of Statistics* **42**, 113–117.

CLARKE, B.R., GAMBLE, D.K. & BEDNARSKI, T. (2011). On a note on robustness of the β-trimmed mean. *Australian & New Zealand Journal of Statistics* **53**, 481.

COLLETT, D. (2003). *Modelling Survival Data in Medical Research, 2nd ed.* New York: Chapman & Hall/CRC, 2nd edn.

FARCOMENI, A. & VIVIANI, S. (2011). Robust estimation for the cox regression model based on trimming. *Biometrical Journal* **53**, 956–973.

HADI, A.S. & LUCEÑO, A. (1997). Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms. *Computational Statistics & Data Analysis* **25**, 251–272.

HAMPEL, F.R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42**, 1887–1896.

JAMES, I.R. (1986). On estimating equations with censored data. *Biometrika* **73**, 35–42.

MÜLLER, C.H. & NEYKOV, N. (2003). Breakdown points of trimmed likelihood estimators and related estimators in generalized linear models. *Journal of Statistical Planning and Inference* **116**, 503–519.

MÜLLER, C.H., SZUGAT, S., CELIK, N. & CLARKE, B.R. (2016). Influence functions of trimmed likelihood estimators for lifetime experiments. *Statistics* **50**, 505–524.

NEYKOV, N., FILZMOSER, P., DIMOVA, R. & NEYTCHEV, P. (2007). Robust fitting of mixtures using the trimmed likelihood estimator. *Computational Statistics & Data Analysis* **52**, 299–308.

NEYKOV, N., FILZMOSER, P. & NEYTCHEV, P. (2014). Ultrahigh dimensional variable selection through the penalized maximum trimmed likelihood estimator. *Statistical Papers* **55**, 187–207.

NEYKOV, N. & MÜLLER, C.H. (2003). Breakdown point and computation of trimmed likelihood estimators in generalized linear models. In *Developments in Robust Statistics*, eds. R. Dutter, P. Filzmoser, U. Gather & P.J. Rousseeuw. Physica-Verlag, Heidelberg, pp. 277–286.

ROUSSEEUW, P.J. (1984). Least median of squares regression. *Journal of the American Statistical Association* **79**, 871–880.

ROUSSEEUW, P.J. & LEROY, A.M. (1987). *Robust Regression and Outlier Detection.* John Wiley, New York.

STAUDTE, R.G. & SHEATHER, S.J. (1990). *Robust Estimation and Testing.* Wiley Series in Probability and Statistics, New York: Wiley.

VANDEV, D.L. & NEYKOV, N.M. (1993). Robust maximum likelihood in the gaussian case. *in : New Directions in Statistical Data Analysis and Robustness, eds. E. Ronchetti, W.A. Stahel: Birkhäuser, Basel* , 259–264.