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

Jörg Pawlitschko

University of Dortmund, Department of Statistics

D-44221 Dortmund, Germany

pawl@statistik.uni-dortmund.de

1. Introduction

We consider the problem of estimating the location parameter θ of a two-parameter exponential distribution $Exp(\theta, \nu)$ where $\nu > 0$ denotes the respective scale parameter. This distribution is a simple but nevertheless useful tool in lifetime analysis, especially when investigating reliability of technical equipment. In this context, θ can be seen as a guarantee time.

For a sample $\underline{x}_N = (x_1, \ldots, x_N)'$ assumed to come i.i.d. from $Exp(\theta, \nu)$ let $x_{1:N} \leq \cdots \leq x_{N:N}$ denote the corresponding ordered values. A common estimator for θ is the sample minimum $\hat{\theta}_{\min} = x_{1:N}$. This estimator, however, is highly non-robust: A single outlier falling below the true θ may determine $\hat{\theta}_{\min}$ completely and will inevitably lead to false conclusions. Therefore, robust alternatives are necessary.

2. Robust Estimators

In this paper we investigate two types of estimators closer. For the first type, note that the median of an $Exp(\theta, \nu)$-distribution is given by $\theta + \nu \ln 2$. Hence, following a suggestion of Rousseeuw and Croux (1993), a robust estimator of θ can be obtained by setting

$$\hat{\theta}_S \;=\; \mathsf{Med}(\underline{x}_N) - S_N(\underline{x}_N)\;\ln 2 \tag{1}$$

where $S_N(\underline{x}_N)$ is a robust estimator of ν and $\mathsf{Med}(\underline{x}_N)$ is the sample median. We will denote such an estimator as a Median-Scale- (MS-) estimator. There are many possible choices for S_N in (**??**). We discuss some which have explicit representations so that they are unique and have breakdown point of (at least asymptotically) 0.5, cf. Gather and Schultze (1999).

The other type of estimator comes from the analysis of doubly Type-II-censored samples where only $x_{r:N} \leq \cdots \leq x_{N-s:N}$ are observable for some fixed $1 < r < N-s < N$. In this case, the BLUE of θ is well-known to equal

$$\hat{\theta}_{r,s} \;=\; x_{r:N} - \frac{\sum_{i=N-r+1}^{N} 1/i}{N-r-s}\left(\sum_{i=r}^{N-s} x_{i:N} + s\,x_{N-s:N} - (N-r+1)\,x_{r:N}\right). \tag{2}$$

If used in complete samples, $\hat{\theta}_{r,s}$ can be seen as a special L-estimator giving weight zero to the $r-1$ smallest and s largest observations and can be expected to have good robustness properties for appropriately chosen r and s.

3. Results

We first investigate the performance of the two types of estimators with respect to their worst-case behaviour. For this purpose, we adopt the notion of the finite-sample replacement breakdown point as introduced in Donohoe and Huber (1983). For any estimator T_N of θ, this is the minimal fraction of sample elements which need to be replaced by arbitrary ones to make $|T_N|$ grow beyond all limits. It turns out that MS-estimators retain the optimal breakdown properties of the sample median (as an estimator of the corresponding location functional) and the robust scale estimators involved. For an L-estimator $\hat{\theta}_{r,s}$, its breakdown point equals $\min\{r, s+1\}/N$. Further, we have conducted an extensive simulation study to compare the performance of the competing estimators with respect to their bias and standard deviation if the null model (all x_i come i.i.d. from $Exp(\theta, \nu)$) does not hold. To create different situations of "bad" data we consider a certain contamination model, namely a slippage model of Ferguson-type. In the class of MS-estimators, the one based on the S-estimator of Rousseeuw and Croux (1993) performs best if the fraction of contaminants is close to 0.5, otherwise the MS-estimator based on MAD (median deviation from the median) has the edge. When being compared with L-estimators, one finds that MS-estimators give better protection against a greater variety of contamination situations. However, if one has knowledge about the amount and type of possible contamination, then an L-estimator with appropriately chosen r and s is preferable.

REFERENCES

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*, 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 Croux, C. (1993). Alternatives to the median absolute deviation. *J. Amer. Statist. Assoc.*, **88**, 1273-1283.

RÉSUMÉ

Nous considérons l'estimation robuste du paramètre de position d'une distribution exponentielle avec deux paramètres. Nous examinons deux classes d'estimateurs. L'une se base sur une équation pour la médiane. L'autre est motivée par l'estimation des échantillons censurés. Les points de rupture sont calculés et la performance des estimateurs est étudiée dans une étude de simulation.