A small sample limitation for estimators based on order statistics

by Max Petzold

Department of Statistics, School of Economics and Commercial Law, Göteborg University, Box 660, SE-405 30 GÖTEBORG, Sweden.
Email: Max.Petzold@statistics.gu.se

Key words and phrases: order statistics; robust; sample size limitations.

Order statistics based estimators are often used for small samples where asymptotic theory is improper, and it is therefore important to know the small sample properties and possible limitations. This master thesis focus on the sample size limitations for estimators which are restricted to come up with an estimate within the interval $\left[X_{(1)}^{(n)}, X_{(n)}^{(n)}\right]$ where the order statistics $X_{(1)}^{(n)}$ and $X_{(n)}^{(n)}$ are the smallest and the largest observation respectively in a sample of n observations. Examples of this type of estimators are the standard percentile estimators included in statistical software, e.g. SAS(r) [4] and SPSS(r) [5], which are further described by Petzold [3].

To find out the sample size limitations one first has to notice that the expected values of order statistics are dependent of the sample size.

Lemma 1 *Let $X_{(1)}^{(n)}$ and $X_{(n)}^{(n)}$ denote the smallest and largest, respectively, of n independent and identically distributed continuous random variables. Then*

$$\begin{aligned} E\left[X_{(1)}^{(n+1)}\right] &< E\left[X_{(1)}^{(n)}\right] \quad (a) \\ E\left[X_{(n+1)}^{(n+1)}\right] &> E\left[X_{(n)}^{(n)}\right] \quad (b) \end{aligned}$$

Proof. Consider $F_{X_{(1)}^{n}}(x) = 1 - (S(x))^n$ where $S(x) = 1 - F(x)$ is the survival function. Then, for all such distributions, satisfying the regular condition $xS(x)(1 - S(x)) = 0$ when $x \to \pm\infty$, we have that

$$
\begin{aligned}
& E\left[X_{(1)}^{(n+1)}\right]-E\left[X_{(1)}^{(n)}\right] \\
= & \int_{-\infty}^{\infty} x d F_{X_{(1)}^{(n+1)}}(x)-\int_{-\infty}^{\infty} x d F_{X_{(1)}^{(n)}}(x)=\{\text{integration by parts}\} \\
= & \left[x F_{X_{(1)}^{(n+1)}}(x)\right]_{-\infty}^{\infty}-\left[x F_{X_{(1)}^{(n)}}(x)\right]_{-\infty}^{\infty}+\int_{-\infty}^{\infty} F_{X_{(1)}^{(n)}}(x)\, d x-\int_{-\infty}^{\infty} F_{X_{(1)}^{(n+1)}}(x)\, d x \\
= & \left[x\left(F_{X_{(1)}^{(n+1)}}(x)-F_{X_{(1)}^{(n)}}(x)\right)\right]_{-\infty}^{\infty}+\int_{-\infty}^{\infty}\left[F_{X_{(1)}^{(n)}}(x)-F_{X_{(1)}^{(n+1)}}(x)\right] d x \\
= & \left[x(S(x))^{n}(1-S(x))\right]_{-\infty}^{\infty}+\int_{-\infty}^{\infty}\left[(S(x))^{n+1}-(S(x))^{n}\right] d x \\
= & 0+\int_{-\infty}^{\infty}\left[(S(x))^{n+1}-(S(x))^{n}\right] d x \leq 0, \text { since }(S(x))^{n+1} \leq(S(x))^{n}
\end{aligned}
$$

(b) is shown in a similar way. ■

Consequently, the size of an estimate based on order statistics is dependent on the sample size. It is now clear that, e.g., a percentile estimator which is restricted to come up with an estimate within the interval $\left[X_{(1)}^{(n)}, X_{(n)}^{(n)}\right]$ will be biased for small sample sizes and high or low percentiles.

It is important to find out the sample sizes limitations in each situation. With some knowledge about the true distribution it is possible to calculate the required minimum sample size. Some asymptotic results on moments of order statistics are given by Cramer [1], and results for finite samples from the normal distribution are given by Petzold [2]. The limitations could easily be illustrated using the results in the latter one. Estimating e.g. the $97.5th$ percentile from the standard normal distribution the minimum sample size where $E\left[X_{(n)}\right]$ equals or exceeds the true percentile will be $n = 25$. This sample size limit depends only on the choice of percentile. A higher percentile will demand a larger sample size, and the opposite will be true for low percentiles, i.e. under the $50th$ percentile.

Acknowledgement

I would like to thank my supervisor, Ph.D. Robert Jonsson, for his support and helpful comments during the work with this thesis.

References

[1] Cramér, H. (1957). Mathematical methods of statistics, 7. pr., pp. 575. Princeton: Princeton University Press, 374-376.

[2] Petzold, M. (2000). A note on the first moment of extreme order statistics from the normal distribution. Working paper. Department of Statistics, Göteborg University.

[3] Petzold, M. (2000). A comparison of some commonly used sample percentile definitions. Working paper. Department of Statistics, Göteborg University

[4] SAS(r) (1990). SAS(r) Procedures Guide. SAS Institute Inc. version 6 third edition, 625-626.

[5] SPSS(r) (1997). SPSS(r) Base 7.5 Syntax Reference Guide. SPSS Inc. , 318-319.