A note on the first moment of extreme order statistics from the normal distribution

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: asymptotic; bias; Cramér; extreme; normal; order statistic.

Abstract

In this note some well known asymptotic results for moments of order statistics from the normal distribution are treated. The results originates from the work of Cramér. A bias correction for finite sample sizes is proposed for the expected value of the largest observation.

1 Introduction

Let $Y_{(i)}^{(n)}$ denote the i:th value from the bottom in a sample of n i.i.d. observations, such that e.g. $Y_{(n)} \equiv Y_{(n)}^{(n)}$ is the largest observation. The p.d.f. of $Y_{(i)}$ is well known

$$f_{Y_{(i)}}(y) = \frac{n!}{(i-1)!\,(n-i)!} \left(F_Y(y)\right)^{i-1} \left(1 - F_Y(y)\right)^{n-i} f_Y(y) \qquad (1)$$

but could hardly be used to find e.g. expected values of order statistics. To reach an asymptotic result for $Y_{(n)}$, consider the c.d.f. $F_{Y_{(n)}}(y) = \left[F_Y(y)\right]^n$.

Since

$$F_{Y_{(n)}}(y) \rightarrow \begin{cases} 1 & \text{if } F_Y(y) = 1 \\ 0 & \text{if } F_Y(y) < 1 \end{cases}$$

it is obvious that some transformation of $Y_{(n)}$ is needed in order to avoid trivial limiting results. Such transformations of order statistics have been extensively studied by e.g. Fréchet [2] and Gnedenko [3]. Here, some asymptotic results by Cramér [1] for normally distributed observations are presented and a bias correction of the expected value for finite sample sizes is proposed.

2 Asymptotic results for the normal distribution

According to Cramér, for $i \geq n/2$ and $n \rightarrow \infty$

$$Y_{(i)} \approx \mu + \sigma \left[\sqrt{2 \ln n} - \frac{\left(\ln(\ln n) + \ln(4\pi) - 2W_{(i)} \right)}{2\sqrt{2 \ln n}} \right] \tag{2}$$

where $W_{(i)}$ has the density

$$g(w) = \frac{1}{(n-i)!} \exp\left(-(n-i+1)\, w - \exp(-w) \right), -\infty < w < \infty$$

Since $E\left[W_{(n)}\right] = -\psi(1) = 0.57722$ and $V\left[W_{(n)}\right] = \psi'(1) = 1.64493$, see Johnson [4], the largest observation has the moments

$$E\left[Y_{(n)}\right] \approx \mu + \sigma \left[\sqrt{2 \ln n} - \frac{(\ln(\ln n) + \ln(4\pi) - 2 \cdot 0.5772)}{2\sqrt{2 \ln n}} \right] \tag{3}$$

$$V\left[Y_{(n)}\right] \approx \frac{\sigma^2 \cdot 1.64493}{2 \ln n} \tag{4}$$

It is of some interest to study how the results in (2) were obtained. Let $Z = (y - \mu)/\sigma$ and let $Z_{(i)}$ be the corresponding order statistic. Starting from the equation

$$T_{(i)} = n\left[1 - F_Z\left(Z_{(i)}\right)\right] = n\int_{Z_{(i)}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-u^2/2\right) du =$$

$$= \frac{n}{\sqrt{2\pi}} \int_{Z_{(i)}}^{\infty} \frac{1}{u} u \exp\left(-u^2/2\right) du = \frac{n}{\sqrt{2\pi}} \cdot \frac{1}{\left(Z_{(i)} + \delta\right)} \exp\left(-\left(Z_{(i)} + \delta\right)^2/2\right)$$

by the mean value theorem. From this $\left(Z_{(i)} + \delta\right)$ can be solved

$$\left(Z_{(i)} + \delta\right) \approx \sqrt{2\ln n} - \frac{\ln(\ln n) + \ln(4\pi) + 2\ln T_{(i)}}{2\sqrt{2\ln n}}$$

Proceeding to the general case of the normal distribution, replacing $Z_{(i)}$ with $\left(Y_{(i)} - \mu\right)/\sigma$, substituting $W_{(i)}$ for $-\ln T_{(i)}$ and finally choosing $\delta = 0$ we get equation (2). Any order statistic can be obtained in a similar way. Due to symmetry $E\left[Y_{(1)}\right]$ could be found directly from equation (3).

3 A bias correction of the expected value for finite sample sizes

The exact and the approximative value, respectively, of $E\left[Y_{(n)}\right]$ are given in Figure 1 for $Y \sim N(0;1)$ and $n = 5, 6, ..., 101$. The exact value was simulated with the statistical software SAS(r) IML. For all simulations in this paper, the computer clock was used as random seed and 10^6 replicates were taken. The asymptotic result in equation (3) was used as an approximation for finite sample sizes.

As can be seen from Figure 1 and Figure 2 the equation (3) gives a biased result for finite sample sizes. Looking back at the derivatives of equation (3) it can be seen that two arguments are made. Specially the choice of $\delta = 0$

may be put in question. To correct the equation a function $\delta(n)$ was fit to the bias given in Figure 2:

$$\delta(n) = 0.1727 n^{-0.2759}$$

Subtracting this function in equation (3) we get:

$$E\left[Y_{(n)}\right] \approx \mu + \sigma\left[\sqrt{2\ln n} - \frac{(\ln(\ln n) + \ln(4\pi) - 2\cdot 0.5772)}{2\sqrt{2\ln n}} - 0.1727 n^{-0.2759}\right] \tag{5}$$

This correction was tested for some other parameter values in the normal distribution and showed up to compensate the bias effectively. In Table 1 the uncorrected equation (3) and the corrected equation (5) are compared for some normal distributions. The maximal and minimal quota between the bias of the uncorrected equation and the bias of the corrected equation is given for $n = 5, 6, ...101$. As can be read out, the bias of the uncorrected equation is at least 23 times larger than the corresponding bias of the corrected equation. Also the variance in equation (4) is biased and may need to be corrected in some way, but is left in this paper.

Table 1: The maximal and minimal quota between the bias of the uncorrected equation (3) and the bias of the corrected equation (5) for $n = 5, 6, ...101$.

Distribution:	$N(0;1)$	$N(0;5)$	$N(0;20)$
Maximal	10000	13000	7400
Minimal	23	24	27

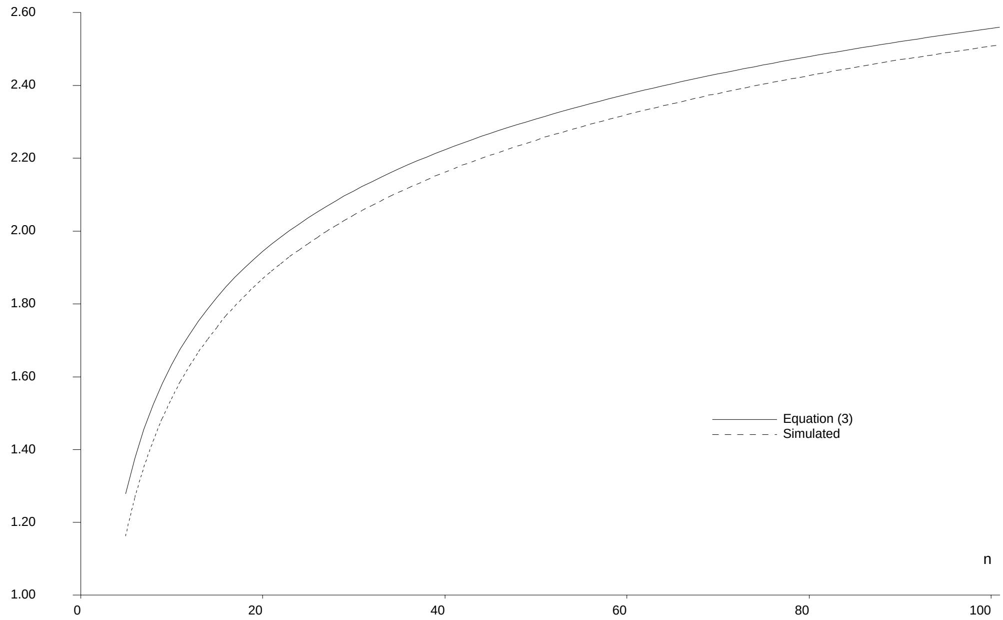

Figure 1: A comparison of the simulated and the approximative $E\left[Y_{(n)}\right]$.

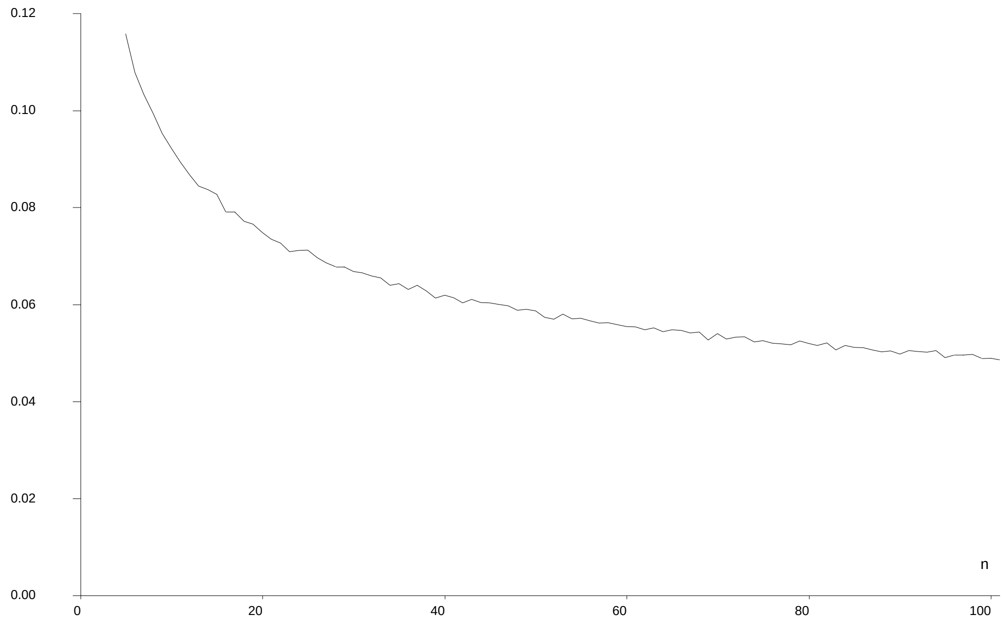

Figure 2: The bias of the approximative $E\left[Y_{(n)}\right]$ in Figure 1.

References

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

[2] Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. Annales de la Société de Mathématique 18, 93-116.

[3] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. Annals of Mathematics 44, 423-453.

[4] Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). Continuous univariate distributions, vol. 1, pp. 756. New York: Wiley, 272-281.