Journal of Statistical Research *ISSN 0256 - 422 X*
2012, Vol . 46, No. 1, pp. 31-35
Bangladesh

ON HOW TO FIND THE NORMING CONSTANTS FOR THE MAXIMA OF A FOLDED NORMALLY DISTRIBUTED VARIABLE

KUDAKWASHE MUTANGI
Harare Institute of Technology, Department of Mathematics, Harare, Zimbabwe
Email: kmutangi@gmail.com

FLORENCE MATARISE
Department of Statistics, University of Zimbabwe, Harare, Zimbabwe
Email: matarise@science.uz.ac.zw

ABSTRACT

A general procedure when one is looking for a limiting distribution of $X_n = \max(X_1, \ldots, X_n)$ is to first center X_n by subtracting c_n and then scale by d_n (Mood et al., 1974). This article is focused on finding the norming constants c_n and d_n for the maxima of the folded normal random variable X_n, where $X = |Z|,\ Z \sim N(0, 1)$. We also show that after appropriate normalisation, X_n has a limiting distribution $H(x) = \exp(-\exp(x))$, which is the Gumbel distribution.

Keywords and phrases: normalising, constants, convergence, Gumbel, limiting, folded normal

1 Introduction

The classical theory of extreme values of probability theory deals with the asymptotic distribution theory of the maxima and minima of independent and identically distributed random variables. Suppose H_n and L_n are distribution functions of the maxima and minima respectively, then

$$\begin{aligned} L_n(x) &= P(W_n \leq x) = 1 - [1 - F(x)]^n \\ H_n(x) &= P(Z_n \leq x) = F^n(x), \end{aligned} \tag{1.1}$$

where $W_n = \min(X_1, \ldots, X_n)$ and $Z_n = \max(X_1, \ldots, X_n)$ (Mood et al., 1974), then the values of $H_n(x)$ and $L_n(x)$ cannot be computed from 1.1 and (1.1) due to the sensitivity of u^n to u for large n. This is true if distribution function $F(x)$ is approximated (Leadbetter et al., 1983), where $F(x)$ is varying, linearly normalised extremes $(Z_n - a_n)/b_n$ or $(W_n - c_n)/d_n$ have the same limiting distribution functions $H(x)$ and $L(x)$, respectively.

In this article, let $X = |Z|$ where $Z \sim N(0, 1)$, then X is distributed as a folded normal and is always positive. The maximum $X_n = \max(X_1, \ldots, X_n)$ can be normalised using constants c_n

and d_n (Embrechts et al., 1997; Leadbetter et al., 1983; Ahsanullah and Kirmani, 2006). This article attempts to provide a methodology for finding these constants and also show that the limiting distribution of the linearly normalised extreme from the folded normal is actually the Gumbel distribution $\wedge(x) = \exp(\exp(x))$.

A probability distribution with distribution function $F(x)$ is said to have an extreme value distribution if for each positive integer n, $F(x)$ satisfies the max-stable condition $F^n(x) = F(d_n + c_n x)$ for some constants $\{d_n\}$ and $\{c_n\}$ (Leadbetter and Rootzen, 1988). This means that if we can find c_n and d_n for $X_n = \max(X_1, \ldots, X_n)$ as defined above, we can also find the extreme value distribution of X. We however need the following definition in order to find the limiting distribution.

Definition 1.1. The maximum of a sequence $X_1, \ldots, X_n$ of independent and identically distributed random variables is said to have a limiting distribution if there are constants $\{d_n\}$ and $\{c_n\}$ such that

$$\lim_{n\to\infty} F^n(d_n + c_n x) = H(x) \text{ or equivalently } \lim_{n\to\infty} n[1 - F(d_n + c_n x)] = -\ln[H(x)],$$

where $H(x)$ is the limiting distribution.

It is this definition which will help us to find the limiting distribution of the normalised maxima once the norming constants have been found.

2 Methodology

We begin by introducing the extreme value theorem and some extreme value concepts which are necessary in this article.

2.1 Some important extreme value concepts

Theorem 1 (Extreme Value theorem)**.** *Let $F(x)$ be a distribution function. If there are numbers d_n and $c_n > 0$ such that $F^n(d_n + c_n x) \to G(x)$ as $n \to \infty$, where $G(x)$ is a non-degenerate distribution, then there is a re-scaled version $H(x = G(a + bx)$ of $G(x)$ such that $H(x)$ is one of the three standard extreme value distributions, namely the Gumbel, Fretchet and the Weibull distributions (Leadbetter et al., 1983; Ahsanullah and Nevzorov, 2001).*

The extreme value theorem implies that each probability distribution may be associated with one of the three standard extreme value distributions. This leads to the following definition.

Definition 2.1 (Domain of attraction)**.** Let $H(x)$ be a non-degenerate probability distribution function. Then the domain of attraction of $H(x)$ is the set of all probability distribution functions $F(x)$ such that if $X_1, \ldots, X_n$ are independent and identically distributed (iid) random variables each with probability function $F(x)$ then there exists normalising constants d_n and $c_n > 0$ such that

$$\lim_{n\to\infty} F^n(d_n + c_n x) = H(x)$$

for all x at which $H(x)$ is continuous (Ahsanullah and Kirmani, 2006).

The following theorem is crucial in finding norming constants for all distributions which converge to the three standard extremal types.

Theorem 2. *Let $\{X_n\}$ be an iid sequence. Let $0 \leq \tau \leq \infty$ and suppose that $\{u_n\}$ is a sequence of real continuous numbers such that*

$$n(1 - F(u_n)) \to \tau \ \ as \ \ n \to \infty \tag{2.1}$$

then

$$P\{M_n \leq u_n\} \to e^{-\tau} \ \ as \ \ n \to \infty \tag{2.2}$$

conversely if (2.1) holds for some $\tau, \ 0 \leq \tau \leq \infty$, then so does (2.2) (Ahsanullah and Kirmani, 2006).

Proof. The proof of the above theorem can be found in Leadbetter et al. (1983, page 13) and is thus omitted. □

The norming constants will be found using the above theorems and the relationship which exist between the folded normal and the standard normal distribution. The following theorem gives the norming constants for the maxima of the standard normal and the asymptotic distribution of $M_n = \max(Z_1, \ldots, Z_n)$ where $Z \sim N(0,1)$. We will then use the relationship between the standard normal and the folded normal to find the norming constants for $\max(X_1, \ldots, X_n)$.

Theorem 3. *If $\{Z_n\}$ is an iid standard normal sequence of random variables, then the asymptotic distribution of $M_n = \max(Z_1, \ldots, Z_n)$ is of type 1 (Gumbel). Specifically*

$$P\{c_n^{-1}(M_n - d_n) \leq x\} \to \exp(-\exp(x)), \tag{2.3}$$

where $c_n = (2\ln n)^{-\frac{1}{2}}$ and $d_n = (2\ln n)^{\frac{1}{2}} - \frac{1}{2}(2\ln n)^{-\frac{1}{2}}[\ln(\ln n) + \ln 4\pi + O((\ln n)^{-\frac{1}{2}})]$ (see Leadbetter et al., 1983 and Ahsanullah and Kirmani, 2006)

Proof. The proof of this theorem is straight forward and is thus omitted. □

2.2 Norming constants of the maxima of the folded normal random variable

We begin by finding the cumulative distribution of $X = |Z|, \ F(x)$. Note that $f_X(x) = 2f_Z(x), X > 0$ then it is trivial to show that $F(x) = 2\Phi(x) - 1$, where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution. Thus $1 - F(x) = 2(1 - \Phi(x))$ and it follows that

$$\begin{aligned} -2\ln H(x) &= \lim_{n\to\infty}[2n(1 - \Phi(c_n x + d_n))] \\ &= \lim_{n\to\infty}[n(1 - F(c_n x + d_n))] \end{aligned}$$

but,

$$\begin{aligned} 2\ln H(x) &= -2e^{-x}, \ (\text{since } H(x) = e^{-e^{-x}}) \\ &= -e^{-(x - \ln 2)} = \ln H(cx + d), \end{aligned}$$

where $c = 1$ and $d = -\ln 2$. We can now apply the Fisher Tippet theorem (limit laws of maxima) (see Embrechts et al. (1997), page 121).

Theorem 4 (Fisher Tippet)**.** *Let $\{X_n\}$ be a sequence of iid random variables. If there exist norming constants $c_n > 0,\ d_n \in \mathbb{R}$ and some non-degenerate distribution function $H(x)$ such that*

$$c_n^{-1}(M_n - d_n) \to H(x) \ \text{ as } n \to \infty \tag{2.4}$$

then $H(x)$ belongs to one of the three extremal types.

If the limit law in (2.4) appear as $H(cx + d)$, that is

$$\lim_{n\to\infty} P[c_n^{-1}(M_n - d_n) \leq x] = H(cx + d)$$

then $H(x)$ is also a limit under a simple change of norming constants:

$$\lim_{n\to\infty} P[a_n^{-1}(M_n - b_n) \leq x] = H(x), \ \text{ where } \ a_n = \frac{c_n}{c} \ \text{ and } \ b_n = d_n - \frac{dc_n}{c},$$

in this case

$$a_n = \frac{c_n}{1} = c_n = (2\ln n)^{-\frac{1}{2}} \tag{2.5}$$

and

$$b_n = d_n + c_n \ln 2 = (\ln 2)(2\ln n)^{-\frac{1}{2}} + (2\ln n)^{\frac{1}{2}} - \frac{\ln(\ln n) + \ln 4\pi}{2(2\ln n)^{\frac{1}{2}}} \tag{2.6}$$

The norming constants for the folded normal are given in equations (2.5) and (2.6) above.

Lemma 2.1. *Let $Z \sim N(0, 1)$ and let $X = |Z|$. Let $F(x) = P(X \leq x)$ be the distribution function of X, then the asymptotic distribution of $M_n = \max(X_1, \ldots, X_n)$ is of type 1. Specifically*

$$\lim_{n\to\infty} P[a_n^{-1}(M_n - b_n)] = \exp[-\exp(x)],$$

where $a_n = (2\ln n)^{-\frac{1}{2}}$ and $b_n = (\ln 2)(2\ln n)^{-\frac{1}{2}} + (2\ln n)^{\frac{1}{2}} - \dfrac{\ln(\ln n) + \ln 4\pi}{2(2\ln n)^{\frac{1}{2}}}$.

Proof. The proof follows from the main result and is thus omitted. □

The norming constants can also be given implicitly by

$$a_n = f^{-1}(1 - 1/n), \ \ b_n = a(a_n), \ \text{ where } a(x) = \int_x^\infty \frac{\overline{F}(t)}{\overline{F}(x)} dt$$

(see Embrechts et al. (1997) and Lemma 1.1.6 in Ahsanullah and Kirmani (2006) for the detail). The norming constants given above are not unique and according to Lemma 1.1.7 (Ahsanullah and Kirmani, 2006), any other sequences a_n^* and b_n^* satisfying $\lim_{n\to\infty}(a_n - a_n^*)/b_n = 0$ and $\lim_{n\to\infty} b_n^*/b_n = 0$ mean that $(\max(X_1, \ldots, X_n) - a_n^*)/b_n^*$ will converge to the Gumbel distribution in this case.

3 Conclusion

In this article, we have presented a procedure for finding the norming constants for the maxima of a folded normal variable. Use has been made of the relationship between the normal and the folded normal and the asymptotic theory of extremes. This result is very important in modeling extremal events especially in time series where it is necessary to detect outliers. Chang and Tiao (1983) developed a statistic for detecting the presence of outliers, but the statistic had no limiting distribution and critical values had to be calculated using simulation algorithms for every sample size. One can use the norming constants above and develop a statistic that can detect outliers but whose limiting distribution is the Gumbel distribution.

4 Acknowledgments

The author is indebted to Jimma University, Ethiopia for the use of its computer facilities and libraries for this research.

References

[1] Embrechts, P., Kluppelberg, C. and Mikosh, T (1997). *Modeling extremal events for insurance and finance*. New York. Springer.

[2] Leadbetter, M.R., Lindgren, G. and Rootzen, H (1983). *Extreme s and related properties of random sequences and processes*. New York. Springer.

[3] Leadbetter, M.R. and Rootzen, H (1988). Extremal theory of stochastic processes. *Annals of probability*, 16, 431-478.

[4] Chang, I. and Tiao, G.C. (1983). *Estimation of time series parameters in the presence of outliers*. Technical report 8, Statistics Research Centre, Graduate School of Business, university of Chicago.

[5] Mood, A.M., Graybill, F.A. and Boes, D.C. (1974). *Introduction to the theory of statistics*, 3rd edition, McGraw-Hill Inc.

[6] Ahsanullah, M. and Kirmani, S.N.U.A. (2006). *Topics in extreme values*. Nova Science Publishers, New York.

[7] Ahsanullah, M. and Nevzorov, V.B (2001). *Ordered random variables*. Nova Science Publishers, New York.