

THE BETA GUMBEL DISTRIBUTION

SARALEES NADARAJAH AND SAMUEL KOTZ

Received 22 March 2004 and in revised form 16 June 2004

The Gumbel distribution is perhaps the most widely applied statistical distribution for problems in engineering. In this paper, we introduce a generalization—referred to as the beta Gumbel distribution—generated from the logit of a beta random variable. We provide a comprehensive treatment of the mathematical properties of this new distribution. We derive the analytical shapes of the corresponding probability density function and the hazard rate function and provide graphical illustrations. We calculate expressions for the n th moment and the asymptotic distribution of the extreme order statistics. We investigate the variation of the skewness and kurtosis measures. We also discuss estimation by the method of maximum likelihood. We hope that this generalization will attract wider applicability in engineering.

1. Introduction

The Gumbel distribution is perhaps the most widely applied statistical distribution for problems in engineering. It is also known as the extreme value distribution of type I. Some of its recent application areas in engineering include flood frequency analysis, network engineering, nuclear engineering, offshore engineering, risk-based engineering, space engineering, software reliability engineering, structural engineering, and wind engineering. A recent book by Kotz and Nadarajah [3], which describes this distribution, lists over fifty applications ranging from accelerated life testing through earthquakes, floods, horse racing, rainfall, queues in supermarkets, sea currents, wind speeds, and track race records (to mention just a few).

In this paper, we propose a generalization of the Gumbel distribution with the hope it will attract wider applicability in engineering. The generalization is motivated by the work of Eugene et al. [1]. If G denotes the cumulative distribution function (cdf) of a random variable, then a generalized class of distributions can be defined by

$$F(x) = I_{G(x)}(a, b) \tag{1.1}$$

for $a > 0$ and $b > 0$, where

$$I_y(a, b) = \frac{B_y(a, b)}{B(a, b)} \quad (1.2)$$

denotes the incomplete beta function ratio, and

$$\begin{aligned} B_y(a, b) &= \int_0^y w^{a-1}(1-w)^{b-1} dw \\ &= y^a \left\{ \frac{1}{a} + \frac{1-b}{1+a}x + \dots + \frac{(1-b) \cdots (n-b)}{n!(a+n)}y^n + \dots \right\} \end{aligned} \quad (1.3)$$

denotes the incomplete beta function. Eugene et al. [1] introduced what is known as the beta normal distribution by taking G to be the cdf of the normal distribution with parameters μ and σ . The properties of this distribution have been studied in more detail by Gupta and Nadarajah [2]. In this paper, we introduce a generalization of the Gumbel distribution—referred to as the beta Gumbel (BG)—by taking G in (1.1) to be the cdf of the Gumbel distribution, that is,

$$G(x) = \exp \left\{ - \exp \left(- \frac{x - \mu}{\sigma} \right) \right\} \quad (1.4)$$

for $-\infty < x < \infty$, $-\infty < \mu < \infty$, and $\sigma > 0$. Thus, the cdf of the BG distribution is given by

$$F(x) = I_{\exp(-u)}(a, b) \quad (1.5)$$

for $-\infty < x < \infty$, $a > 0$, $b > 0$, $-\infty < \mu < \infty$, and $\sigma > 0$, where $u = \exp\{-(x - \mu)/\sigma\}$. The corresponding probability density function (pdf) is

$$f(x) = \frac{u \exp(-au) \{1 - \exp(-u)\}^{b-1}}{\sigma B(a, b)}. \quad (1.6)$$

Using the series representation

$$(1+z)^\alpha = \sum_{j=0}^{\infty} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha-j+1)} \frac{z^j}{j!}, \quad (1.7)$$

(1.6) can be expressed in the mixture form

$$f(x) = \frac{\Gamma(a+b)}{\sigma \Gamma(a)} \sum_{k=0}^{\infty} \frac{(-1)^k u \exp\{-(a+k)u\}}{k! \Gamma(b-k)}. \quad (1.8)$$

As mentioned above, the Gumbel distribution is widely applied in many areas of engineering. The generalization given by (1.6) allows for greater flexibility of its tail. Consider the following problems.

(1) Many hydrological engineering planning, design, and management problems require a detailed knowledge of flood event characteristics, such as flood peak, volume, and duration. Flood frequency analysis often uses the Gumbel distribution to model flood peak values, which provides an assessment of flood events.

(2) Corrosion science has been based mainly upon deterministic approaches, particularly the electrochemical theory of corrosion. Localized corrosion, however, cannot be explained without statistical and stochastic points of view because of the large scatter in data common in the laboratory and the field. Toshio Shibata was the 1996 recipient of the W. R. Whitney Award sponsored by NACE International. In his award lecture at CORROSION/96, Shibata reviewed successful applications of statistical approaches to localized corrosion in engineering data and presented a stochastic theory of pitting corrosion based upon sensitivity analysis of parameters in the stochastic model that rationally could explain statistical distributions of pitting potential and induction time for pit formation. The most successful application in the statistical analysis was found using the Gumbel distribution to estimate the maximum pit depth that will be found in a large-area installation by using a small number of samples with a small area.

(3) Gumbel distribution has also been shown to provide good fits to the time series of the extreme dynamic pressures (i.e., of the squares of the extreme wind speeds).

Each of the problems above is concerned with the tail behavior of one or more variables. Thus, by capturing the tail behavior more accurately, one could obtain improved estimation and prediction. The model given by (1.6) provides one way of doing this.

In the rest of this paper, we provide a comprehensive description of the mathematical properties of (1.6). We identify several particular cases of (1.5). We examine the shape of (1.6) and its associated hazard rate function. We derive formulas for the n th moment and the asymptotic distribution of the extreme order statistics. We also consider estimation issues.

2. Particular cases

The Gumbel distribution is the particular case of (1.5) for $a = 1$ and $b = 1$. Four more particular cases of (1.5) can be identified using special properties of the incomplete beta function ratio. They are

$$F(x) = \sum_{i=a}^n \binom{n}{i} \{1 - \exp(-u)\}^{n-i} \exp(-iu) \tag{2.1}$$

for $b = n - a + 1$ and integer values of a ;

$$F(x) = 1 - \frac{\{1 - \exp(-u)\}^b}{\Gamma(b)} \sum_{i=1}^a \frac{\Gamma(b+i-1)}{\Gamma(i)} \exp\{-(i-1)u\} \tag{2.2}$$

for integer values of a ;

$$F(x) = \frac{\exp(-au)}{\Gamma(a)} \sum_{i=1}^b \frac{\Gamma(a+i-1)}{\Gamma(i)} \{1 - \exp(-u)\}^{i-1} \tag{2.3}$$

for integer values of b ; and

$$F(x) = \frac{2}{\pi} \arctan \sqrt{\frac{\exp(-u)}{1 - \exp(-u)}} \tag{2.4}$$

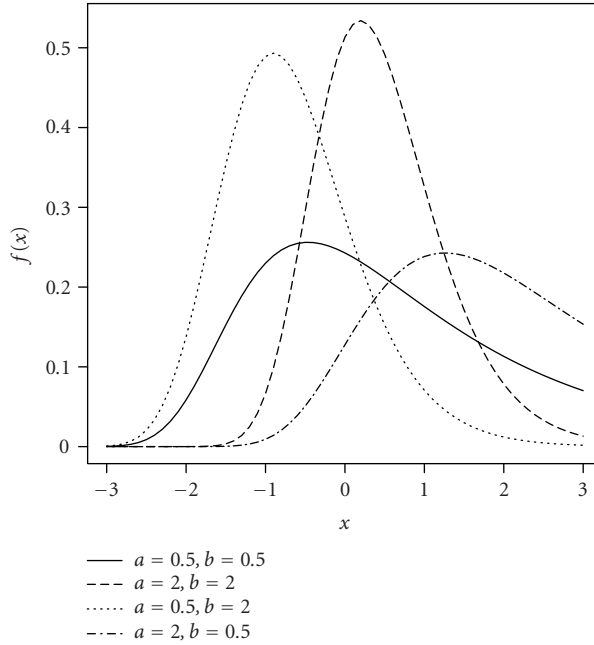


Figure 3.1. The pdf of the beta Gumbel distribution (1.6) for selected values of (a, b) and $\mu = 0$, $\sigma = 1$.

for $a = 1/2$ and $b = 1/2$, where $u = \exp\{-(x - \mu)/\sigma\}$. These four formulas are of special interest by themselves. For example, the last formula corresponds to the arcsine distribution, which arises naturally in statistical communication theory as a model for the amplitude of a periodic signal in thermal noise and the limiting spectral density function of a high-index-angle modulated carrier (see Lee [5, Chapter 6] and Middleton [6, Chapter 14]).

3. Shape

The first derivative of $\log f(x)$ for the BG distribution is

$$\frac{d \log f(x)}{dx} = \frac{au - 1 + \{1 - (a + b - 1)u\} \exp(-u)}{\sigma \{1 - \exp(-u)\}}, \quad (3.1)$$

where $u = \exp\{-(x - \mu)/\sigma\}$. Standard calculations based on this derivative show that $f(x)$ exhibits a single mode at $x = x_0$, where $u_0 = \exp\{-(x_0 - \mu)/\sigma\}$ is the solution of

$$1 - au = \{1 - (a + b - 1)u\} \exp(-u). \quad (3.2)$$

Furthermore, $f(-\infty) = 0$ and $f(\infty) = 0$. Figure 3.1 illustrates some of the possible shapes of f for selected values of (a, b) and $\mu = 0$, $\sigma = 1$.

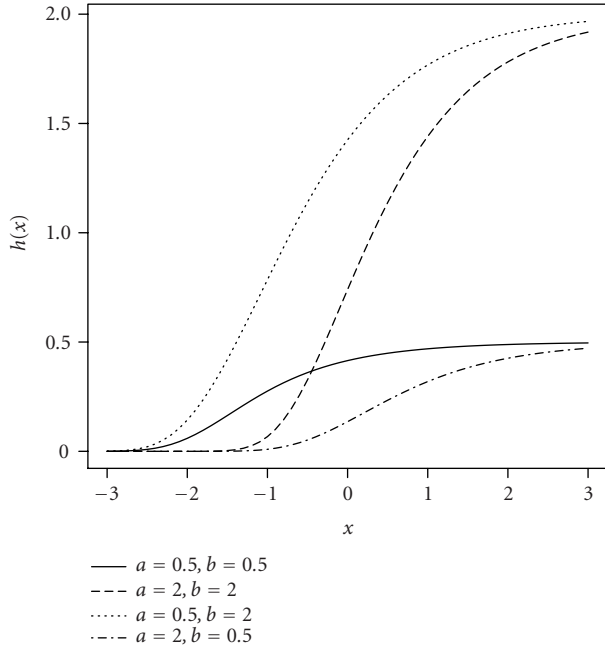


Figure 4.1. Hazard rate function of the beta Gumbel distribution (1.6) for selected values of (a, b) and $\mu = 0, \sigma = 1$.

4. Hazard rate function

The hazard rate function defined by $h(x) = f(x)/\{1 - F(x)\}$ is an important quantity characterizing life phenomena. For the BG distribution, $h(x)$ takes the form

$$h(x) = \frac{u \exp(-au) \{1 - \exp(-u)\}^{b-1}}{\sigma B(a, b) I_{\exp(-u)}(b, a)}, \tag{4.1}$$

where $u = \exp\{- (x - \mu)/\sigma\}$. Calculations using special properties of the incomplete beta function ratio show that $h(x)$ is an increasing function of x . Furthermore, $h(x) \rightarrow 0$ as $x \rightarrow 0$ and $h(x) \rightarrow b$ as $x \rightarrow \infty$. Figure 4.1 illustrates some of the possible shapes of h for selected values of (a, b) and $\mu = 0, \sigma = 1$.

5. Moments

If X has the pdf (1.6), then its n th moment can be written as

$$E(X^n) = \frac{1}{\sigma B(a, b)} \int_{-\infty}^{\infty} x^n \left[1 - \exp \left\{ - \exp \left(- \frac{x - \mu}{\sigma} \right) \right\} \right]^{b-1} \exp \left(- \frac{x - \mu}{\sigma} \right) \times \exp \left\{ - a \exp \left(- \frac{x - \mu}{\sigma} \right) \right\} dx, \tag{5.1}$$

which, on setting $u = \exp\{-(x - \mu)/\sigma\}$, reduces to

$$E(X^n) = \frac{1}{B(a, b)} \int_0^\infty (\mu - \sigma \log u)^n \{1 - \exp(-u)\}^{b-1} \exp(-au) du. \tag{5.2}$$

Using the binomial expansion for $(c + dz)^n$, (5.2) can be rewritten as

$$E(X^n) = \frac{1}{B(a, b)} \sum_{k=0}^n \binom{n}{k} \mu^{n-k} (-\sigma)^k I(k), \tag{5.3}$$

where $I(k)$ denotes the integral

$$I(k) = \int_0^\infty (\log u)^k \{1 - \exp(-u)\}^{b-1} \exp(-au) du. \tag{5.4}$$

Using the representation (1.7), (5.4) can be further expanded as

$$I(k) = \sum_{l=0}^\infty \frac{(-1)^l \Gamma(b) I(k, l)}{l! \Gamma(b - l)}, \tag{5.5}$$

where $I(k, l)$ denotes the integral

$$I(k, l) = \int_0^\infty (\log u)^k \exp\{-(a + l)u\} du. \tag{5.6}$$

Finally, by equation (2.6.21.1) in [7], (5.6) can be calculated as

$$I(k, l) = \left(\frac{\partial}{\partial \alpha}\right)^k [(a + l)^{-\alpha} \Gamma(\alpha)] \Big|_{\alpha=1}. \tag{5.7}$$

By combining (5.3), (5.5), and (5.7), the n th moment of X can be expressed as

$$E(X^n) = \frac{\Gamma(a + b) \Gamma(n + 1) \mu^n}{\Gamma(a)} \sum_{k=0}^n \sum_{l=0}^\infty \frac{(-1)^{k+l} (\sigma/\mu)^k}{k! l! \Gamma(n - k + 1) \Gamma(b - l)} \left(\frac{\partial}{\partial \alpha}\right)^k [(a + l)^{-\alpha} \Gamma(\alpha)] \Big|_{\alpha=1}. \tag{5.8}$$

In particular,

$$\begin{aligned}
 E(X) &= \frac{\Gamma(a+b)}{\Gamma(a)} \sum_{l=0}^{\infty} \frac{(-1)^l \{\mu + C\sigma + \sigma(l+a)\}}{l!(l+a)\Gamma(b-l)}, \\
 E(X^2) &= \frac{\Gamma(a+b)}{\Gamma(a)} \sum_{l=0}^{\infty} (-1)^l \times \frac{1}{\{l!(l+a)\Gamma(b-l)\}} \\
 &\quad \times \{6\mu^2 + (\pi^2 + 6C^2)\sigma^2 + 12C\mu\sigma \\
 &\quad + 12\sigma(\mu + C\sigma)\log(a+l) + 6\sigma^2\log^2(a+l)\}, \\
 E(X^3) &= \frac{\Gamma(a+b)}{2\Gamma(a)} \sum_{l=0}^{\infty} (-1)^l \times \frac{1}{\{l!(l+a)\Gamma(b-l)\}} \\
 &\quad \times \{(\pi^2 + 6C^2)\mu\sigma^2 + 6C\sigma\mu^2 + 2\mu^3 - (4\eta(3) + \pi^2C + 2C^3)\sigma^3 \\
 &\quad + (12C\sigma^2\mu + 6\sigma\mu^2 - \sigma^3\pi^2 - 6\sigma^3C^2)\log(l+a) \\
 &\quad + 6\sigma^2(\mu - \sigma C)\log^2(l+a) - 2\sigma^3\log^3(l+a)\}, \\
 E(X^4) &= \frac{\Gamma(a+b)}{20\Gamma(a)} \sum_{l=0}^{\infty} (-1)^l \times \frac{1}{\{l!(l+a)\Gamma(b-l)\}} \\
 &\quad \times \{-40\sigma^3(\pi^2C + 2C^3 + 4\eta(3))\mu + 20\sigma^2(\pi^2 + 6C^2)\mu^2 + 80C\sigma\mu^3 + 20\mu^4 \\
 &\quad + (3\pi^4 + 20C^4 + 160C\eta(3) + 20\pi^2C^2)\sigma^4 \\
 &\quad + 40\sigma\{(\pi^2C + 4\eta(3) + 2C^3)\sigma^3 - (\pi^2 + 6C^2)\sigma^2\mu + 6C\sigma\mu^2 + 2\mu^3\}\log(l+a) \\
 &\quad + 20\sigma^2(\pi^2\sigma^2 + 6C^2\sigma^2 - 12C\sigma\mu + 6\mu^2)\log^2(l+a) - 80\sigma^3\mu\log^3(l+a) \\
 &\quad + 20\sigma^4(1 + 4C)\log^4(l+a)\},
 \end{aligned} \tag{5.9}$$

where C denotes Euler’s constant. The variance, skewness, and kurtosis measures can now be calculated using the relations

$$\begin{aligned}
 \text{Var}(X) &= E(X^2) - E^2(X), \\
 \text{Skewness}(X) &= \frac{E(X^3) - 3E(X)E(X^2) + 2E^3(X)}{\text{Var}^{3/2}(X)}, \\
 \text{Kurtosis}(X) &= \frac{E(X^4) - 4E(X)E(X^3) + 6E(X^2)E^2(X) - 3E^4(X)}{\text{Var}^2(X)}.
 \end{aligned} \tag{5.10}$$

The skewness and kurtosis measures are controlled mainly by the parameters a and b and [Figure 5.1](#) illustrates their variation.

It is evident that (1.5) is much more flexible than the Gumbel distribution.

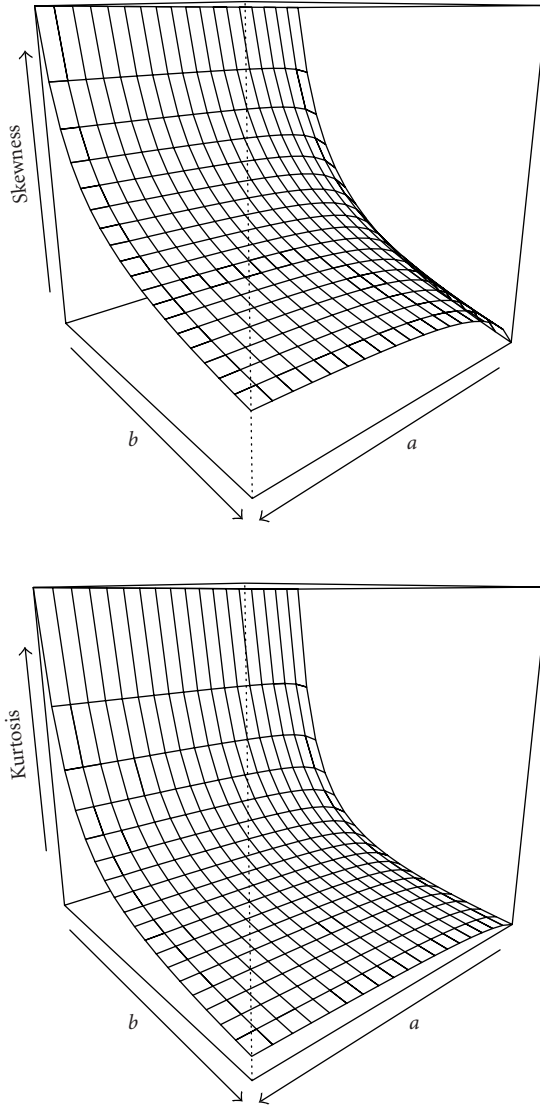


Figure 5.1. Skewness and kurtosis measures versus $a = 1, 1.5, \dots, 10$ and $b = 1, 1.5, \dots, 10$ for the beta Gumbel distribution.

6. Asymptotics

If X_1, \dots, X_n is a random sample from (1.6) and if $\bar{X} = (X_1 + \dots + X_n)/n$ denotes the sample mean, then, by the usual central limit theorem, $\sqrt{n}(\bar{X} - E(X))/\sqrt{\text{Var}(X)}$ approaches the standard normal distribution as $n \rightarrow \infty$. Sometimes one would be interested in the asymptotics of the extreme values $M_n = \max(X_1, \dots, X_n)$ and $m_n = \min(X_1, \dots, X_n)$. Note

from (1.6) that

$$1 - F(t) \sim \frac{1}{bB(a, b)} \exp\left(-b \frac{t - \mu}{\sigma}\right) \tag{6.1}$$

as $t \rightarrow \infty$ and that

$$F(t) \sim \frac{1}{aB(a, b)} \exp\left\{-a \exp\left(-\frac{t - \mu}{\sigma}\right)\right\} \tag{6.2}$$

as $t \rightarrow -\infty$. Thus, it follows that

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1 - F(t + \sigma x/b)}{1 - F(t)} &= \exp(-x), \\ \lim_{t \rightarrow -\infty} \frac{F(t + (\sigma/a)x \exp((t - \mu)/\sigma))}{F(t)} &= \exp(x). \end{aligned} \tag{6.3}$$

Hence, it follows from Theorem 1.6.2 in [4] that there must be norming constants $a_n > 0$, $b_n, c_n > 0$, and d_n such that

$$\begin{aligned} \Pr\{a_n(M_n - b_n) \leq x\} &\rightarrow \exp\{-\exp(-x)\}, \\ \Pr\{c_n(m_n - d_n) \leq x\} &\rightarrow 1 - \exp\{-\exp(x)\} \end{aligned} \tag{6.4}$$

as $n \rightarrow \infty$. The form of the norming constants can also be determined. For instance, using Corollary 1.6.3 in [4], one can see that $a_n = b/\sigma$ and that b_n satisfies $1 - F(b_n) \sim 1/n$ as $n \rightarrow \infty$. Using the fact (6.1), one can see that

$$b_n = \mu - \frac{\sigma}{b} \log\left\{\frac{bB(a, b)}{n}\right\} \tag{6.5}$$

satisfies $1 - F(b_n) \sim 1/n$. The constants c_n and d_n can be determined by using the same corollary.

7. Estimation

We consider estimation by the method of maximum likelihood. The log-likelihood for a random sample x_1, \dots, x_n from (1.6) is

$$\begin{aligned} \log L(\mu, \sigma, a, b) &= -n \log B(a, b) - n \log \sigma + (b - 1) \sum_{i=1}^n \log \left[1 - \exp\left\{-\exp\left(-\frac{x_i - \mu}{\sigma}\right)\right\} \right] \\ &\quad - \sum_{i=1}^n \frac{x_i - \mu}{\sigma} - a \sum_{i=1}^n \exp\left(-\frac{x_i - \mu}{\sigma}\right). \end{aligned} \tag{7.1}$$

The first-order derivatives of (7.1) with respect to the four parameters are

$$\begin{aligned}
 \frac{\partial \log L}{\partial \mu} &= \frac{n}{\sigma} - \frac{1}{\sigma} \sum_{i=1}^n \exp\left(-\frac{x_i - \mu}{\sigma}\right) \\
 &\quad + \frac{b-1}{\sigma} \sum_{i=1}^n \frac{\exp(-(x_i - \mu)/\sigma) \exp\{-\exp(-(x_i - \mu)/\sigma)\}}{1 - \exp\{-\exp(-(x_i - \mu)/\sigma)\}}, \\
 \frac{\partial \log L}{\partial \sigma} &= -\frac{n}{\sigma} + \sum_{i=1}^n \frac{x_i - \mu}{\sigma^2} \left\{1 - a \exp\left(-\frac{x_i - \mu}{\sigma}\right)\right\} \\
 &\quad + \frac{b-1}{\sigma^2} \sum_{i=1}^n \frac{(x_i - \mu) \exp(-(x_i - \mu)/\sigma) \exp\{-\exp(-(x_i - \mu)/\sigma)\}}{1 - \exp\{-\exp(-(x_i - \mu)/\sigma)\}}, \\
 \frac{\partial \log L}{\partial a} &= n\Psi(a+b) - n\Psi(a) - \sum_{i=1}^n \exp\left(-\frac{x_i - \mu}{\sigma}\right), \\
 \frac{\partial \log L}{\partial b} &= n\Psi(a+b) - n\Psi(b) + \sum_{i=1}^n \log \left[1 - \exp\left\{-\exp\left(-\frac{x_i - \mu}{\sigma}\right)\right\}\right],
 \end{aligned} \tag{7.2}$$

where $\Psi(x) = d \log \Gamma(x) / dx$ is the digamma function. Setting these expressions to zero and solving them simultaneously yields the maximum-likelihood estimates of the four parameters.

References

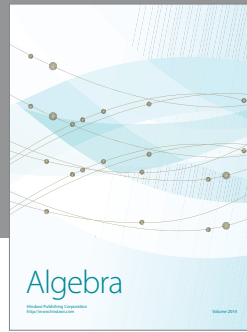
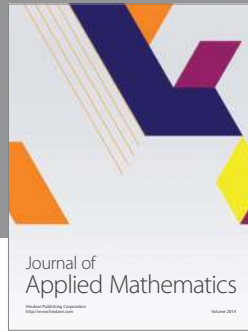
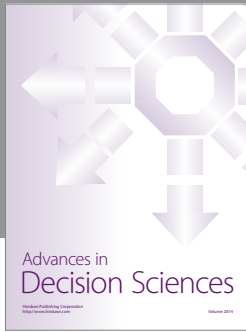
- [1] N. Eugene, C. Lee, and F. Famoye, *Beta-normal distribution and its applications*, Comm. Statist. Theory Methods **31** (2002), no. 4, 497–512.
- [2] A. K. Gupta and S. Nadarajah, *On the moments of the beta normal distribution*, Comm. Statist. Theory Methods **33** (2004), no. 1, 1–13.
- [3] S. Kotz and S. Nadarajah, *Extreme Value Distributions. Theory and Applications*, Imperial College Press, London, 2000.
- [4] M. R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*, Springer Series in Statistics, Springer-Verlag, New York, 1983.
- [5] Y. W. Lee, *Statistical Theory of Communication*, John Wiley & Sons, New York, 1960.
- [6] D. Middleton, *An Introduction to Statistical Communication Theory*, McGraw-Hill, New York, 1960.
- [7] A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series. Vol. 3*, Gordon and Breach Science Publishers, New York, 1986.

Saralees Nadarajah: Department of Mathematics, University of South Florida, Tampa, FL 33620, USA

E-mail address: snadaraj@chuma1.cas.usf.edu

Samuel Kotz: Department of Engineering Management and Systems Engineering, The George Washington University, Washington, DC 20052, USA

E-mail address: kotz@seas.gwu.edu



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

