5-1-2006

JMASM22: A Convenient Way Of Generating Normal Random Variables Using Generalized Exponential Distribution

Debasis Kundu
Indian Institute of Technology, Kanpur, India, kundu@iitk.ac.in

Anubhav Manglick
Indian Institute of Technology

JMASM22: A Convenient Way Of Generating Normal Random Variables Using Generalized Exponential Distribution

Debasis Kundu   Anubhav Manglick   Rameshwar D. Gupta
Department of Mathematics and Statistics   Department of Computer Science and Applied Statistics
Indian Institute of Technology   University of New Brunswick

A convenient method to generate normal random variable using a generalized exponential distribution is proposed. The new method is compared with the other existing methods and it is observed that the proposed method is quite competitive with most of the existing methods in terms of the $K - S$ distances and the corresponding p-values.

Key words: Generalized exponential distribution; Kolmogorov-Smirnov distances; random number generator.

Introduction

Generating normal random numbers is an old and very important problem in the statistical literature. Several algorithms are available in the literature to generate normal random numbers like Box-Muller methods, Marsaglia-Bray method, Acceptance-Rejection method, Ahrens-Dieter method, etc. The book of Johnson, Kotz and Balakrishnan (1995) provided an extensive list of references of the different algorithms available today. Among the several methods the most popular ones are

Debasis Kundu is a Professor of Statistics. His research interests include statistical signal processing, reliability analysis, statistical computing and competing risks models. Rameshwar D. Gupta is a Professor and Acting Associate Dean of Graduate Studies. His research interests are statistical inference, multivariate statistical analysis, and statistical analysis of reliability and life-testing models. Anubhav Manglick is a Master's student. His major research interests include generalised exponential models, model discrimination and high frequency financial time series data. He has published in the *Journal of Statistical Planning and Inference* and *Naval Research Logistiscs*.

the Box-Muller transformation method or the improvement suggested by Marsagilia and Bray. Most of the statistical packages like, SAS, IMSL, SPSS, S-Plus, or Numerical Recipes use this method. In this article, a simple and convenient method of generating normal random numbers using generalized exponential distribution is proposed.

Generalized exponential (GE) distribution has been proposed and studied quite extensively recently by Gupta and Kundu (1999; 2001a; 2001b; 2002; 2003a). The readers may be referred to some of the related literature on (GE) distribution by Raqab (2002), Raqab and Ahsanullah (2001), and Zheng (2002). The two-parameter GE distribution has the following distribution function:

$$F_{GE}(x;\alpha,\lambda) = (1 - e^{-\lambda x})^{\alpha}; \ \alpha, \lambda > 0 \qquad (1)$$

for $x > 0$ and 0 otherwise. The corresponding density function is;

$$f_{GE}(x;\alpha,\lambda) = \alpha\lambda(1 - e^{-\lambda x})^{\alpha-1} e^{-\lambda x}; \ \alpha, \lambda > 0, \ (2)$$

for $x > 0$ and 0 otherwise. Here α and λ are the shape and scale parameters respectively. When $\alpha = 1$, it coincides with the exponential distribution. If $\alpha \leq 1$, the density function of a

GE distribution is a strictly decreasing function and for $\alpha > 1$, it has a uni-modal density function. The shape of the density function of the *GE* distribution for different α can be found in Gupta and Kundu (2001a).

In a recent study by Kundu, Gupta and Manglick (2005), it was observed that in certain cases log-normal distribution can be approximated quite well by *GE* distribution and vice versa. In fact, for certain ranges of the shape parameters of the *GE* distributions the distance between the *GE* and log-normal distributions can be very small.

The main idea in this article is to use this particular property of a *GE* distribution to generate log-normal random variables and in turn generate normal random variables. It may be mentioned that the *GE* distribution function is an analytically invertible function, therefore, the generation of *GE* random variables is immediate using uniform random variables.

Methodology

The density function of a log-normal random variable with scale parameter θ and shape parameter σ is denoted as

$$f_{LN}(x;\theta,\sigma) = \frac{1}{\sqrt{2\pi}\,x\sigma}\, e^{-\frac{(\ln x - \ln\theta)}{2\sigma^2}}; \ \theta,\sigma > 0 \quad ,(3)$$

for $x > 0$ and 0 otherwise. If X is a log-normal random variable with scale parameter θ and shape parameter σ, then

$$E(X) = \theta e^{\frac{\sigma^2}{2}}$$

and

$$V(X) = \theta^2 e^{\sigma^2}(e^{\sigma^2} - 1). \tag{4}$$

Note that $\ln X$ is a normal random variable with mean $\ln\theta = \mu$ (say) and variance σ^2.

Similarly, if X is a generalized exponential random variable with the scale parameter λ and shape parameter α, then

$$E(X) = \frac{1}{\lambda}(\psi(\alpha+1) - \psi(1))$$

and

$$V(X) = \frac{1}{\lambda^2}(\psi'(1) - \psi'(\alpha+1)). \tag{5}$$

It was observed by Kundu, Gupta and Manglick (2005) that a generalized exponential distribution can be approximated very well by a log-normal distribution for certain ranges of the shape parameters. The first two moments of the two distribution functions are equated to compute σ and θ from a given α and λ. Without loss of generality, $\lambda = 1$ is taken. For a given $\alpha = \alpha_0$, equating (4) and (5) one obtains

$$\theta e^{\frac{\sigma^2}{2}} = \psi(\alpha_0+1) - \psi(1) = A_0 \tag{6}$$

$$\theta^2 e^{\sigma^2}(e^{\sigma^2} - 1) = \psi'(1) - \psi'(\alpha_0+1) = B_0 \tag{7}$$

Therefore, solving (6) and (7), one obtains

$$\ln\theta_0 = \mu_0 = \ln A_0 - \frac{1}{2}\ln(1 + \frac{B_0}{A_0^2}), \tag{8}$$

$$\sigma_0 = \sqrt{\ln(1 + \frac{B_0}{A_0^2})}. \tag{9}$$

Using (8) and (9), standard normal random variable can be easily generated as follows:

Algorithm

Step 1: Generate U an uniform $(0,1)$ random variable.

Step 2: For a fixed α_0, generate $X = -\ln(1 - U^{\frac{1}{\alpha_0}})$. Note that X is a generalized exponential random variable with shape parameter α_0 and scale parameter 1.

Step 3: Compute $Z = \dfrac{\ln X - \mu_0}{\sigma_0}$. Here

Z is the desired standard normal random variable.

An alternative approximation is also possible. Instead of equating the moments of the two distributions, one can equate the corresponding L-moments also. The L-moments of any distribution are analogous to the conventional moments, but they are based on the quantiles and they can be estimated by the linear combination of order statistics, i.e. by L-statistics (see Hosking, 1990, for details). It is observed by Gupta and Kundu (2003b) in a similar study of approximating gamma distribution by generalized exponential distribution that the L-moments perform better than the ordinary moments.

Let Z be any random variable having finite first moment and suppose $Z_{1:n} \leq .... \leq Z_{n:n}$ be the order statistics of a random sample of size n drawn from the distribution of Z. Then the L-moments are defined as follows:

$$\lambda_r = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E(Z_{r-k:r}); \quad r = 1,2,... \quad (10)$$

The two $L-$moments of a log-normal distribution are

$$\lambda_1 = \theta e^{\frac{\sigma^2}{2}}$$

and

$$\lambda_2 = \theta e^{\frac{\sigma^2}{2}} erf(\frac{\sigma}{2}), \quad (11)$$

where $erf(x) = 2\phi(\sqrt{2}\ x) - 1$ and $\phi(x)$ is the distribution function of the standard normal distribution. Similarly, the two L-moments of a GE random variable are

$$\lambda_1 = \frac{1}{\lambda}(\psi(\alpha+1) - \psi(1))$$

and

$$\lambda_2 = \frac{1}{\lambda^2}(\psi(2\alpha+1) - \psi(\alpha+1)). \quad (12)$$

Therefore, as before equating the first two L-moments for a given $\alpha = \alpha_0$ and for $\lambda = 1$, one obtains

$$\theta e^{\frac{\sigma^2}{2}} = \psi(\alpha_0+1) - \psi(1) = A_0 \quad (13)$$

$$\theta e^{\frac{\sigma^2}{2}} erf(\frac{\sigma}{2}) = \psi(2\alpha_0+1) - \psi(\alpha_0+1) = B_1 \quad (14)$$

Solving (13) and (14), one obtains the solutions of θ and σ as

$$\ln \theta_1 = \mu_1 = \ln A_0 - \frac{\sigma_1^2}{2} \quad (15)$$

$$\sigma_1 = \sqrt{2}\ \phi^{-1}(\frac{1}{2}(1 + \frac{B_1}{A_0})), \quad (16)$$

where ϕ is the cumulative distribution function of standard normal distribution. Therefore, in the proposed algorithm, instead of using $(\mu_0, \sigma_0), (\mu_1, \sigma_1)$ also can be used.

Numerical Comparisons and Discussions

In this section, an attempt is made to determine the value of α_0, so that the distance between the generalized exponential distribution and the corresponding log-normal distribution is minimum. All the computations are performed using Pentium IV processor and the random number generation routines by Press et al. (1993). The distance function between the two distribution functions is considered as the Kolmogorv-Smirnov $(K-S)$ distance only. To be more precise, the $K-S$ distance between the GE is computed, with the shape and scale parameter as σ_0 and 1 respectively, and log-normal distribution with the corresponding shape and scale parameter as $\sigma_0(\sigma_1)$ and $\theta_0(\theta_1)$ respectively. It is believed that the

distance function should not make much difference, any other distance function may be considered also. It is observed that as α_0 increases from 0, the $K-S$ distance first decreases, and then increases. When the moments (L-moments) equations have been used, the minimum $K-S$ distance occurred at $\alpha_0 =12.9(12.8)$. When $\alpha_0 =12.9(12.8)$, then from (8) and (9) ((15) and (16)), the corresponding μ_0 = 1.0820991 (μ_1 = 1.0792510) and σ_0 = 0.3807482 (σ_1 = 0.3820198) was obtained.

To compare the proposed method with the other existing methods, the $K-S$ statistics and the corresponding p-values were mainly used. The method can be described as follows. The standard normal random variables for different sample sizes namely n = 10, 20, 30, 40, 50 and 100 by using Box-Muller (BM) method, Marsaglia-Bray (MB) method, Acceptance-Rejection (AR) method, Ahren-Dieter (AD) method were generated, using moments equations (MM) and using L-moments equations (LM). In each case, the $K-S$ distance and the corresponding p-value between the empirical distribution function and the standard normal distribution function was computed. The process was replicated 10,000 times and the average $K-S$ distances, the average p-values and the corresponding standard deviations were computed. The results are reported in Table 1. In each case the standard deviations are reported within bracket below the average values. From the table values it is quite

clear that, based on the $K-S$ distances and p values the proposed methods work quite well. Also, an effort is made to compute $(Z \le z)$ using the proposed approximation, where Z denotes the standard normal random variable. Note that

$$P(Z \le z) \approx (1 - e^{-e^{z\sigma_0 + \mu_0}})^{12.9}$$

or

$$P(Z \le z) \approx (1 - e^{-e^{z\sigma_1 + \mu_1}})^{12.8}. \tag{17}$$

The results are reported in Table 2. It is clear from Table 2 that using μ_0 and σ_0 the maximum error can be 0.0005, where as using μ_1 and σ_1 the maximum error can be 0.0003. From Table 2, it is clear that L-moments approximations work better than the moments approximations.

Conclusions

A simple and convenient method of generating normal random variables is provided. Even simple scientific calculator can be used to generate normal random number from the uniform generator very quickly. It can be implemented very easily by using a one line program. It is also observed that the standard normal distribution function can be approximated at least up to three decimal places using the simple approximations.

Table 1. The average K-S distances and the corresponding p-values for different methods based on 10,000 replications. The standard deviations are reported within brackets in each case below the average values.

N		BM	MB	AR	AD	MM	LM
10	K-S	0.2587	0.2587	0.2597	0.2591	0.2586	0.2587
		(0.0796)	(0.0796)	(0.0809)	(0.0804)	(0.0794)	(0.0795)
	p	0.5127	0.5128	0.5109	0.5114	0.5135	0.5132
		(0.2938)	(0.2938)	(0.2970)	(0.2955)	(0.2930)	(0.2931)
20	K-S	0.1851	0.1851	0.1871	0.1860	0.1866	0.1867
		(0.0571)	(0.0571)	(0.0575)	(0.0578)	(0.0571)	(0.0572)
	p	0.5178	0.5178	0.5068	0.5135	0.5089	0.5085
		(0.2934)	(0.2934)	(0.2934)	(0.2957)	(0.2927)	(0.2928)
30	K-S	0.1532	0.1532	0.1533	0.1537	0.1524	0.1525
		(0.0467)	(0.0467)	(0.0466)	(0.0477)	(0.0465)	(0.0465)
	p	0.5094	0.5094	0.5086	0.5088	0.5150	0.5145
		(0.2937)	(0.2937)	(0.2923)	(0.2953)	(0.2930)	(0.2930)
40	K-S	0.1331	0.1331	0.1331	0.1335	0.1334	0.1334
		(0.0409)	(0.0488)	(0.0410)	(0.0412)	(0.0410)	(0.0410)
	p	0.5111	0.5111	0.5121	0.5094	0.5097	0.5092
		(0.2923	(0.2923)	(0.2926)	(0.2945)	(0.2927)	(0.2928)
50	K-S	0.1191	0.1191	0.1197	0.1193	0.1199	0.1200
		(0.0370)	(0.0370)	(0.0364)	(0.0368)	(0.0366)	(0.0366)
	p	0.5140	0.5140	0.5071	0.5120	0.5058	0.5053
		(0.2931)	(0.2931)	(0.2924)	(0.2923)	(0.2927)	(0.2927)
100	K-S	0.0852	0.0852	0.0851	0.0854	0.0851	0.0852
		(0.0257)	(0.0257)	(0.0262)	(0.0257)	(0.0259)	(0.0259)
	P	0.5059	0.5059	0.5096	0.5043	0.5082	0.5077
		(0.2914)	(0.2914)	(0.2932)	(0.2895)	(0.2912)	(0.2912)

Table 2. The exact value of $\phi(z)$ and the two approximate values are reported.

Z	L-Moment	Exact	Moment
0.0	0.49984	0.50000	0.50014
0.1	0.53981	0.53983	0.54006
0.2	0.57935	0.57926	0.57955
0.3	0.61808	0.61791	0.61824
0.4	0.65564	0.65541	0.65574
0.5	0.69168	0.69145	0.69174
0.6	0.72594	0.72572	0.72595
0.7	0.75818	0.75800	0.75815
0.8	0.78822	0.78810	0.78814
0.9	0.81593	0.81588	0.81582
1.0	0.84125	0.84127	0.84112
1.1	0.86416	0.86424	0.86400
1.2	0.88469	0.88482	0.88452
1.3	0.90292	0.90308	0.90273
1.4	0.91893	0.91911	0.91875
1.5	0.93288	0.93305	0.93269
1.6	0.94490	0.94505	0.94472
1.7	0.95517	0.95528	0.95500
1.8	0.96385	0.96392	0.96369
1.9	0.97112	0.97114	0.97097
2.0	0.97714	0.97711	0.97701
2.1	0.98209	0.98200	0.98197
2.2	0.98610	0.98597	0.98600
2.3	0.98933	0.98916	0.98924
2.4	0.99189	0.99170	0.99181
2.5	0.99390	0.99370	0.99384
2.6	0.99547	0.99526	0.99542
2.7	0.99667	0.99647	0.99663
2.8	0.99759	0.99739	0.99755
2.9	0.99827	0.99809	0.99825
3.0	0.99878	0.99861	0.99876
3.5	0.99983	0.99976	0.99982
4.0	0.99998	0.99997	0.99998

References

Gupta, R. D. & Kundu, D. (1999). Generalized exponential distributions. *Australian and New Zealand Journal of Statistics, 41*(2), 173-188.

Gupta, R. D. & Kundu, D. (2001). Exponentiated exponential distribution: An alternative to gamma and Weibull distributions. *Biometrical Journal, 43*(1), 117-130.

Gupta, R. D. & Kundu, D. (2001). Generalized exponential distributions: Different methods of estimations. *Journal of Statistical Computations and Simulations, 69*(4), 315-338.

Gupta, R. D. & Kundu, D. (2002). Generalized exponential distributions: Statistical inferences. *Journal of Statistical Theory and Applications, 1*(2), 101-118, 2002.

Gupta, R. D. & Kundu, D. (2003). Discriminating between Weibull and generalized exponential distributions. *Computational Statistics and Data Analysis, 43*, 179-196.

Gupta, R. D. & Kundu, D. (2003). Closeness of gamma and generalized exponential distributions. *Communications in Statistics - Theory and Methods, 32*(4), 705-721.

Hosking, J. R. M. (1990). L-moment: Analysis and estimation of distribution using linear combination of order statistics. *Journal of Royal Statistical Society, 52*(B), 105-124.

Johnson, N., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distribution* (vol. 1). New York, N.Y.: John Willey and Sons.

Kundu, D., Gupta, R. D., & Manglick, A. (2005). Discriminating between log-normal and generalized exponential distribution. *Journal of Statistical Planning and Inference, 127*, 213-227.

Press et al. (1993). *Numerical recipes.* Cambridge: Cambridge University Press.

Raqab, M. Z. (2002). Inference for generalized exponential distribution based on record statistics. *Journal of Statistical Planning and Inference, 104*(2), 339-350.

Raqab, M. Z. & Ahsanullah, M. (2001). Estimation of the location and scale parameters of the generalized exponential distribution based on order statistics. *Journal of Statistical Computations and Simulations, 69*(2), 109-124.

Zheng, G. (2002). On the Fisher information matrix in type-II censored data from the exponentiated exponential family. *Biometrical Journal, 44*(3), 353-357.