**Far East Journal of Theoretical Statistics**


Published Online: December 2014
Available online at http://pphmj.com/journals/fjts.htm
Volume 49, Number 1, 2014, Pages 31-40

# DISTRIBUTION OF CORRELATED LOGNORMAL RANDOM SUM

**Aceng Komarudin Mutaqin**

Department of Statistics
Faculty of Mathematics and Natural Sciences
Universitas Islam Bandung
Ranggamalela 1 Bandung 40116
Indonesia
e-mail: aceng.k.mutaqin@gmail.com

## Abstract

Correlated lognormal random sum (CLRS) is random sum, $Z_N = X_1 + X_2 + \cdots + X_N$, $N = 0, 1, 2, \ldots$, when $X_1, X_2, \ldots, X_N$ are correlated and the distribution of $X_1, X_2, \ldots, X_N$ given $N = l$ is multivariate lognormal. This paper discusses the approximation distribution for CLRS. The approximation percentile, mean and variance of CLRS are also discussed. Monte Carlo simulation is conducted to investigate the approximations. The simulation results show that the approximation distribution, percentile, mean and variance of CLRS are valid and good approximations.

## 1. Introduction

The sum of independent random variables also called the *random sum*, $Z_N = X_1 + X_2 + \cdots + X_N$, $N = 0, 1, 2, \ldots$, where $Z_N = 0$, has been

Received: March 4, 2014; Accepted: May 9, 2014

2010 Mathematics Subject Classification: 60G50, 60G99, 62H10.

Keywords and phrases: correlated random sum, multivariate lognormal distribution, moment generating function, Gauss-Hermite expansion, Monte Carlo simulation.

discussed by many authors (for examples, see Taylor and Karlin [12], and Klugman et al. [6]). It is assumed that $X_1, X_2, ..., X_N$ are independent random variables, and $N$ is a discrete random variable and independent of $X_1, X_2, ..., X_N$. In the real applications, $X_1, X_2, ..., X_N$ can be correlated, such as in insurance, finance, sports, cellular mobile systems, Internet traffic, and communications. Some authors have discussed the distributions of the sum of correlated random variables and the estimation of its parameters (for example, see Kaas et al. [4], Mehta et al. [9], Chatelain et al. [2], and Gao et al. [3]). Unfortunately, they assumed that $N$ be fixed not a discrete random variable. Whereas, in some applications, for examples in insurance and sports, $N$ can be a discrete random variable.

In this paper, phrase correlated random sum (CRS) will be used for the sum of correlated random variables in case where $N$ be a discrete random variable and independent of $X_1, X_2, \ldots.$ The formulas to calculate mean and variance of CRS have been derived by Mutaqin et al. [10]. They set some assumptions on the moments of $N$ and $X$. The formulas have been numerically verified by comparing between the empirical moments and theoretical moments of CRS via Monte Carlo simulation.

Mutaqin et al. [11] have proposed three approaches to estimate the mean and variance of CRS. The first approach is based on the empirical moments of CRS. The second approach is based on the empirical moments of $X_j$'s. The third approach is based on the weighted mean of the sum of empirical moments of $X_j$'s. Monte Carlo simulation is used to compare the performance of the approaches.

This paper derives the distribution, percentile, mean, and variance of CRS when $X_1, X_2, ..., X_N$ are correlated lognormal random variables. Therefore, we can call that $Z_N$ as a *random sum* of correlated lognormal random variables or *correlated lognormal random sum* (*CLRS*). It is assumed that conditional distribution of $X_1, X_2, ..., X_N$ given $N = l$ is multivariate lognormal. The multivariate lognormal distribution has widespread application, such as in insurance, social sciences, economics, engineering, environment, and physical sciences. We use research results of Mehta et al. [9] to derive the distribution of CLRS. The Monte Carlo simulation will be conducted to validate the proposed distribution, mean, variance and percentile.

The rest of the paper is organized as follows. Section 2 presents the CLRS. The distribution and its properties of CLRS are given in Section 3. Section 4 provides the numerical results. The conclusions follow in the last section.

## 2. The CLRS

Let correlated random sum

$$Z_N = X_1 + X_2 + \cdots + X_N, \quad N = 0, 1, 2, \ldots,$$

where $Z_0 = 0$, $X_1, X_2, \ldots, X_N$ are correlated random variables, $N$ be a discrete random variable with the probability mass function, $p_N(n) = P(N = n)$ for $n = 0, 1, 2, \ldots,$ and independent of $X_1, X_2, \ldots$. CLRS is constructed by assuming that the distribution of $X_1, X_2, \ldots, X_N$ given $N = l$ is multivariate lognormal, with probability density function

$$f(x_1, x_2, \ldots, x_l)$$

$$= \frac{1}{(2\pi)^{l/2} |\Lambda|^{1/2} (x_1 x_2 \cdots x_l)} \exp\left\{-\frac{1}{2}(\ln \boldsymbol{x} - \boldsymbol{\theta})^T \Lambda^{-1} (\ln \boldsymbol{x} - \boldsymbol{\theta})\right\},$$

$$x_i > 0, \quad i = 1, 2, \ldots, l,$$

where

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_l)^T, \qquad \ln \boldsymbol{x} = (\ln x_1, \ln x_2, \ldots, \ln x_l)^T,$$

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_l)^T,$$

and $\Lambda = (\lambda_{i,j})$. More details about multivariate lognormal distribution can be found in Kotz et al. [7, pp. 27-28, 219-230], and Kleiber and Kotz [5, pp. 124-126].

## 3. Distribution of CLRS and Its Properties

Mehta et al. [9] showed that the distribution of $Z_N \mid N$ can be approximated by the univariate lognormal distribution. Let $\mu_{Z_l}$ and $\sigma^2_{Z_l}$ denote the parameters of the distribution of $Z_l$ (or $Z_N \mid N = l$), then those parameters can be estimated numerically using the following two equations:

$$\hat{\Psi}_{Z_l}(s_i;\, \mu_{Z_l},\, \sigma^2_{Z_l}) = \hat{\Psi}^{(c)}_{\sum_{j=1}^{l} X_j}(s_i;\, \theta,\, \mathbf{C}); \quad i = 1,\, 2,$$

where

$$\hat{\Psi}_{Z_l}(s_i;\, \mu_{Z_l},\, \sigma^2_{Z_l}) \approx \sum_{k=1}^{M} \frac{w_k}{\sqrt{\pi}} \exp[-s_i \exp(\sqrt{2}\sigma_{Z_l} a_k + \mu_{Z_l})], \tag{1}$$

and

$$\hat{\Psi}^{(c)}_{\sum_{j=1}^{l} X_j}(s_i;\, \theta,\, \mathbf{C})$$

$$\approx \sum_{k_1=1}^{M} \cdots \sum_{k_l=1}^{M} \left[\prod_{t=1}^{l} \frac{w_{k_t}}{\sqrt{\pi}}\right] \exp\left[-s_i \sum_{t=1}^{l}\left[\exp\left(\sqrt{2}\sum_{j=1}^{l} c_{tj} a_{k_j} + \theta_t\right)\right]\right]. \tag{2}$$

The equations (1) and (2) are the approximation moment generating function using Gauss-Hermite expansion for real $s_i$ of $Z_l$ and $\sum_{j=1}^{l} X_j$, respectively, $M$ is the Hermite integration order. The value of $M = 12$ was found to be accurate for the correlated case (Mehta et al. [9]). The weights, $w_k$, and abscissas, $a_k$, for $M$ up to 20 are tabulated in Abramowitz and Stegun [1, Tbl. 25.10, p. 924]. The $c_{tj}$ is the $(t,\ j)$th element of matrix **C**, where **C** be the square root of the matrix $\Lambda$, i.e., $\Lambda = \mathbf{C}\mathbf{C}^T$.

According to the results of Mehta et al. [9], it is easy to show that the approximation distribution function of CLRS, $Z_N$, is given by

$$\tilde{F}(z) = \begin{cases} 0; & z < 0, \\ p_N(0); & z = 0, \\ p_N(0) + \sum_{l=1}^{\infty} p_N(l)\Phi\left(\dfrac{\ln z - \mu_{Z_l}}{\sigma_{Z_l}}\right); & z > 0, \end{cases} \quad (3)$$

where $\Phi(t)$ denote the standard normal distribution function evaluated at $t$. The approximation probability density function of CLRS, $Z_N$, is given by

$$\tilde{f}(z) = \begin{cases} p_N(0); & z = 0, \\ \sum_{l=1}^{\infty} \dfrac{p_N(l)}{z\sigma_{Z_l}\sqrt{2\pi}} \exp\left[-\dfrac{1}{2}\left(\dfrac{\ln z - \mu_{Z_l}}{\sigma_{Z_l}}\right)^2\right]; & z > 0. \end{cases} \quad (4)$$

If $N$ is constant, and $N > 0$, then the resulting approximation distribution of $Z_N$ is univariate lognormal distribution. The approximation expectation of CLRS, $Z_N$, is given by

$$\begin{aligned} \tilde{E} &= E[Z_N] = \int_0^{\infty} z f_{Z_N}(z)dz \\ &\approx 0 p_N(0) + \int_0^{\infty} z \sum_{l=1}^{\infty} \frac{p_N(l)}{z\sigma_{Z_l}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln z - \mu_{Z_l}}{\sigma_{Z_l}}\right)^2\right]dz \\ &= \sum_{l=1}^{\infty} p_N(l) \int_0^{\infty} \frac{z}{z\sigma_{Z_l}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln z - \mu_{Z_l}}{\sigma_{Z_l}}\right)^2\right]dz \\ &= \sum_{l=1}^{\infty} p_N(l) \exp\left(\mu_{Z_l} + \frac{\sigma_{Z_l}^2}{2}\right). \end{aligned} \quad (5)$$

The approximation variance of CLRS, $Z_N$, is given by

$$\begin{aligned}
\tilde{V} &= Var[Z_N] = E[Z_N^2] - (E[Z_N])^2 \\
&= \int_0^\infty z^2 \sum_{l=1}^\infty \frac{p_N(l)}{z\sigma_{Z_l}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log z - \mu_{Z_l}}{\sigma_{Z_l}}\right)^2\right] dz - \tilde{E}^2 \\
&= \sum_{l=1}^\infty p_N(l) \int_0^\infty \frac{z^2}{z\sigma_{Z_l}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log z - \mu_{Z_l}}{\sigma_{Z_l}}\right)^2\right] dz - \tilde{E}^2 \\
&= \sum_{l=1}^\infty p_N(l) \exp(2\mu_{Z_l} + 2\sigma_{Z_l}^2) - \tilde{E}^2. \qquad (6)
\end{aligned}$$

Let $\tilde{z}_p$ denotes the approximation 100*p*th percentile of $Z_N$. Then

$$F_{Z_N}(\tilde{z}_p) = \Pr(Z_N \le \tilde{z}_p) = p.$$

Therefore $\tilde{z}_p = 0$ for $p \le p_N(0)$, or $\tilde{z}_p$ is the unique solution of the following equation:

$$\sum_{l=1}^\infty p_N(l)\Phi\left(\frac{\ln \tilde{z}_p - \mu_{Z_l}}{\sigma_{Z_l}}\right) = p - p_N(0), \qquad (7)$$

for $p > p_N(0)$.

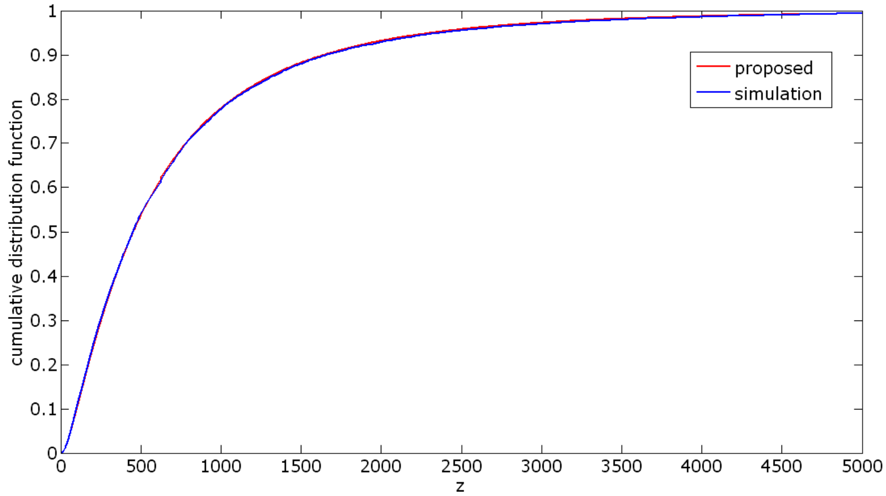


**Figure 1.** Comparison between the proposed distribution and simulation for $N \sim DU(1, 5)$.

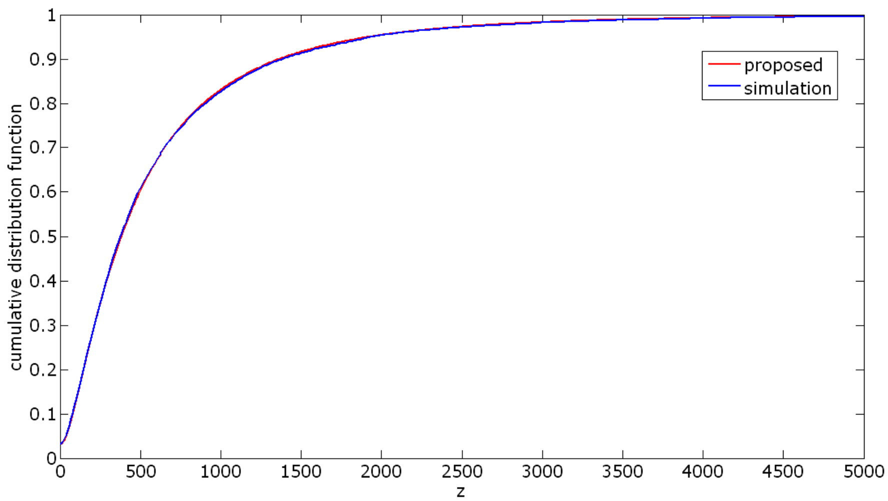


**Figure 2.** Comparison between the proposed distribution and simulation for $N \sim bin(3, 0.5)$.

## 4. Numerical Results

This section contains numerical results to validate the distribution of CLRS and its properties are discussed in Section 3. Monte Carlo simulation is conducted 10.000 times to simulate CLRS, where the distribution of $X_1, X_2, \ldots, X_N$ given $N = l$ is multivariate lognormal, with parameters $\theta_i = 5$, $\lambda_{i,i} = 1$, and $\lambda_{i,j} = 0, 62$, for $i \neq j = 1, 2, \ldots$. Empirical distribution from simulated CLRS can give an appropriate approximation to exact distribution of CLRS. Figure 1 and Figure 2 show comparison between the proposed approximation cumulative distribution function and empirical distribution from simulated CLRS, where the distributions of *N* are discrete uniform, $DU(1, 5)$ (Law and Kelton [8]), and binomial, $bin(5, 0.5)$ (Law and Kelton [8]), respectively. Comparisons between the proposed approximation 100*p*th percentile and simulation are displayed in Table 1 and Table 2. Based on Figures 1-2, and Tables 1-2, we can conclude that the proposed distribution of CLRS is not too different from the exact distribution of CLRS.

**Table 1.** Comparison between the proposed approximation 100*p*th percentile and simulation for distribution of *N* is discrete uniform

| $p$ | $N \sim DU(1, 3)$ | | $N \sim DU(1, 5)$ | |
|---|---|---|---|---|
| | Proposed | Simulation | Proposed | Simulation |
| 0.10 | 72.62 | 72.80 | 100.98 | 98.25 |
| 0.25 | 144.09 | 143.00 | 211.29 | 207.01 |
| 0.50 | 299.03 | 293.33 | 451.28 | 449.00 |
| 0.75 | 598.01 | 596.52 | 907.38 | 917.67 |
| 0.90 | 1087.86 | 1102.73 | 1642.22 | 1656.67 |

**Table 2.** Comparison between the proposed approximation 100*p*th percentile and simulation for distribution of *N* is binomial

| $p$ | $N \sim bin(3; 0.5)$ | | $N \sim bin(5; 0.5)$ | |
|---|---|---|---|---|
| | Proposed | Simulation | Proposed | Simulation |
| 0.10 | 0.00 | 0.00 | 80.84 | 78.71 |
| 0.25 | 79.64 | 80.81 | 180.14 | 179.19 |
| 0.50 | 208.74 | 208.06 | 380.20 | 374.91 |
| 0.75 | 454.55 | 458.10 | 755.43 | 763.66 |
| 0.90 | 863.85 | 875.79 | 1363.09 | 1384.38 |

Tables 3-4 display comparison between the proposed approximation values and its exact values for mean and variance of CLRS. It can be seen that the proposed approximation values are not too different from its exact values. These proofs show that the formulas of the proposed approximation mean and variance of CLRS are valid.

**Table 3.** Comparison between the proposed approximation and simulation for mean and variance of CLRS when the distribution of $N$ is discrete uniform

| Moment | $N \sim DU(1, 3)$ | | $N \sim DU(1, 5)$ | |
|---|---|---|---|---|
| | Exact | Approximation | Exact | Approximation |
| Mean | 8.9634 | 8.9618 | 13.4451 | 13.4427 |
| Variance | 128.4210 | 127.8852 | 281.7252 | 280.7758 |

**Table 4.** Comparison between the proposed approximation and simulation for mean and variance of CLRS when the distribution of $N$ is binomial

| Moment | $N \sim bin(3; 0.5)$ | | $N \sim bin(5; 0.5)$ | |
|---|---|---|---|---|
| | Exact | Approximation | Exact | Approximation |
| Mean | 6.7225 | 6.7216 | 11.2042 | 11.2021 |
| Variance | 92.7111 | 92.3718 | 197.6486 | 196.8718 |

## 5. Conclusions

We have proposed a simple method to approximate the distribution of CLRS. The resulting distribution is mixture of lognormal distribution, when $N > 0$. The result of Mehta et al. [9] is special case of the resulting distribution when $N$ is constant. We also have derived percentile, mean and variance of CLRS. The numerical results showed that the distribution, percentile, mean and variance of CLRS are valid and good approximations.

## References

[1] M. Abramowitz and I. Stegun, Handbook of Mathematical Functions with Formulas: Graphs, and Mathematical Tables, Dover, 1972.

[2] F. Chatelain, P. Borgnat, J.-Y. Tourneret and P. Abry, Parameter estimation for sums of correlated gamma random variables application to anomaly detection in Internet traffic, Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP-08, Las Vegas, 2008, pp. 3489-3492.

[3] X. Gao, H. Xu and D. Ye, Asymptotic behavior of tail density for sum of correlated lognormal variables, Inter. J. Math. Math. Sci. 2009 (2009), Article ID 630857, 28 pp.

[4] R. Kaas, J. Dhaene and M. J. Goovaerts, Upper and lower bounds for sums of random variables, Insurance: Mathematics and Economics 27 (2000), 151-168.

[5] C. Kleiber and S. Kotz, Statistical Size Distributions in Economics and Actuarial Sciences, John Wiley & Sons, New York, 2003.

[6] S. A. Klugman, H. H. Panjer and G. E. Willmot, Loss Models: From Data to Decision, John Wiley & Sons, New York, 2004.

[7] S. Kotz, N. Balakrishnan and N. L. Johnson, Continuous multivariate distributions, Volume 1: Models and Applications, John Wiley & Sons, New York, 2000.

[8] A. M. Law and W. D. Kelton, Simulation Modeling and Analysis, 3rd ed., McGraw-Hill, New York, 2000.

[9] N. B. Mehta, J. Wu, A. F. Molisch and J. Zhang, Approximating a sum of random variables with a lognormal, IEEE Trans. Wireless Comm. 6 (2007), 2690-2699.

[10] A. K. Mutaqin, D. R. Tampubolon, S. Darwis and R. A. Djojosugito, The moments of sum of correlated random variables, Far East J. Theo. Stat. 31(1) (2010), 49-57.

[11] A. K. Mutaqin, D. R. Tampubolon, S. Darwis and R. A. Djojosugito, Estimating mean and variance of sum of correlated random variables using empirical moments, Adv. Appl. Stat. 16(1) (2010), 65-81.

[12] H. M. Taylor and S. Karlin, An Introduction to Stochastic Modeling, Academic Press, New York, 1994.