A fast method to seek the mean and variance of the summation of lognormal variables

Un método rápido para buscar la media y varianza de la suma de variables lognormales

J. G. Torres [iD][1], R. Bustamante [iD][2], C. E. Caicedo [iD][3]

[1]Departamento de Ingeniería de Software, Universidad Manuela Beltrán. Avenida Circunvalar # 60-00. C. P. 254046. Bogotá, Colombia.
[2]Departamento de Ingeniería Eléctrica y Electrónica, Universidad de los Andes. Carrera 1 # 18A-12. C. P. 4976. Bogotá, Colombia.
[2]School of information studies, Syracuse University. 900 South Crouse. P. C. 13244. New York, United States.

ABSTRACT: The sum of *lognormal* variables has been a topic of interest in several fields of research such as engineering, biology and finance, among others. For example, in the field of telecommunications, the aggregate interference of radio frequency signals is modeled as a sum of *lognormal* variables. To date, there is no closed expression for the probability distribution function (PDF) of this sum. Several authors have proposed approximations for this PDF, with which they calculate the mean and variance. However, each method has limitations in its range of parameters for mean, variance and number of random variables to be added. In other cases, long approximations as power series are used, which makes the analytical treatment impractical and reduces the computational performance of numerical operations. This paper shows an alternative method for calculating the mean and variance of the sum of *lognormal* random variables from a computational performance approach. Our method has been evaluated extensively by Monte Carlo simulations. As a result, this method is computationally efficient and yields a low approximation error computation for a wide range of mean values, variances and number of random variables.

RESUMEN: La suma de las variables *lognormal* ha sido un tema de interés en varios campos de investigación como ingeniería, biología y finanzas entre otros. Por ejemplo, en el campo de las telecomunicaciones la interferencia agregada de señales de radiofrecuencia se modela como una suma de variables *lognormal*. A la fecha, no existe una expresión cerrada para la probabilidad función de distribución (PDF) de esta suma. Varios autores han propuesto aproximaciones para esta PDF con el fin de calcular el valor medio y varianza. Pero cada método tiene limitaciones en su rango de parámetros para la media, la varianza y el número de variables que se suman. En otros casos, existen aproximaciones con series de potencias muy largas, lo que hace que el análisis y tratamiento analítico sea impráctico, además, reduce el rendimiento computacional de operaciones numéricas. En este artículo, mostramos un nuevo método para calcular la media y varianza de la suma de variables aleatorias de tipo *lognormal* y desarrollamos todo el análisis desde un enfoque de eficiencia computacional. El desarrollo ha sido evaluado extensamente por simulaciones de Monte Carlo. Como resultado, este método es computacionalmente eficiente y produce un cálculo de error de aproximación bajo para una amplia gama de valores medios, varianzas y número de variables aleatorias. Lo cual lo hace útil al momento de hacer simulaciones con este tipo de variables.

1. Introduction

The normal distribution is the most common probability distribution function (PDF) used in any scientific field if we want to characterize a random variable (RV). This occurs because many natural phenomena are modeled according to the central limit theorem. Sometimes, random variable

values are pretty small, and it is necessary to expand those values through an exponential function, which is also useful for analytic treatments. Due to this, the *lognormal* random variable arises. It means that if we characterize a RV as $X \sim$ Normal $[m, s]$ with mean m and standard deviation s, and then we calculate $x = 10^X$, we will yield $x \sim lognormal [\mu, \sigma]$ with mean μ and standard deviation σ, where: $\mu = \kappa \cdot m$, $\sigma = \kappa \cdot s$ and $\kappa = \frac{ln(10)}{10}$.

Several works shown in [1, 2] have suggested that the sum of *lognormal* RVs can be approximated by a *lognormal* distribution. It means that for:

$$x_i \sim lognormal [\mu_i, \sigma_i] \tag{1}$$

If we have N RVs

$$y = \sum_{i=1}^{N} x_i \tag{2}$$

Then:

$$y \sim lognormal [\tilde{\mu}, \tilde{\sigma}] \tag{3}$$

Now, through two examples, we will show whether or not in a telecommunications scenario known as *characterization of the aggregate interference with N interferents*, it is possible to approximate the sum of *lognormal* variables with a *lognormal* PDF.

Explanation of Monte Carlo simulations 1 and 2 We'll consider a scenario where $N = 5$ is the number of random variables to sum. 10^4 samples are generated for each of the five RVs $X_i, i = [1, .., 5]$ using a Normal distribution function with parameters randomly selected within the range of an interference signal where $m = [-80, -60]$, and $s = [4, 12]$. The parameters selected for the mean value and variance were $m = \{-63, -61, -78, -61, -67\}$, $s = \{4, 5, 7, 9, 9\}$ respectively. Then the following operation is performed: $y = 10^{\frac{X_1}{10}} + 10^{\frac{X_2}{10}} + 10^{\frac{X_3}{10}} + 10^{\frac{X_4}{10}} + 10^{\frac{X_5}{10}}$ and finally we calculate $Y = 10 \cdot log_{10}(y)$.

Since it has been said that y follows a *lognormal* PDF, then Y should have a Normal PDF. Experimentally, the mean value $\tilde{\mu}$, variance $\tilde{\sigma}^2$ and standard deviation $\tilde{\sigma}$ are obtained using the *mean, var* and *std* commands of the software Matlab. With these parameters, we generate a normal analytical PDF to compare it against the histogram of the simulation.

Simulation 2 used the following parameters $m = \{-74 - 60 - 76 - 69 - 71\}$, $s = \{6, 6, 5, 7, 7\}$. As a result of the simulations, Figure 1 shows that for $N = 5$, sometimes this approximation could work, but even with a small set of variables, this approximation fails.

In this way, we show two examples where there is not a straightforward method to find an approximated PDF that fits in all cases to the real PDF of the sum of *lognormal* variables. This wide-used approximation must be carefully handled because of the statement mentioned above; although, it is frequently used in telecommunications [1]. Our work will focus only on an efficient method to find the mean value and the variance of the sum of *lognormal* RV within a wide range of parameters.

The rest of this paper is organized as follows. In section II we give a background and show previous works about the sum of *lognormal* variables, we review some of the approximation methods and assess their accuracy. In section III, we propose a new method for computing the sum of *lognormal* random variables and provide equations to calculate the mean value and variance of the sum. We validate our method in section IV and finally, in section V the conclusions are given.

2. Background and previous work

Some papers such as [1, 3, 4] show a summary of several works that address the sum of *lognormal* variables. From them, we can say that the main problem with this kind of RV is due to the fact that it has neither a characteristic function nor a moment-generating function. Some works address this issue by approximating the characteristic function and obtaining either the PDF or the cumulative distribution function (CDF). However, the curve for the experimental CDF or PDF does not completely match with the analytical function [3], and their results are bounded by asymptotic tail approximations for specific scenarios. For example, in finance [5], there are simple approximations when σ tends to 1.

The errors in the CDF or PDF approximations also depend on the number of RV in the sum. On the other hand, most of the analytical approximations must be solved using numerical computation, Monte Carlo simulations, recursive algorithms or other computational techniques. So, this means that all traditional methods are inefficient and impractical for problems where the *lognormal* parameters change dynamically.

A paper from Wilkinson in 1934 is perhaps one of the first works to approximate the sum of *lognormal* variables with a PDF that is also *lognormal*. Although this work was not published, it was first referenced in [6]. The work of [7] approximates the sum of *lognormal* variables to a *lognormal* PDF, but it is not always correct and is only useful when $\sigma < 4$ for each RV. This approach is also known as the Fenton-Wilkinson method, since both use the same mechanism to approximate the first two moments of the sum.
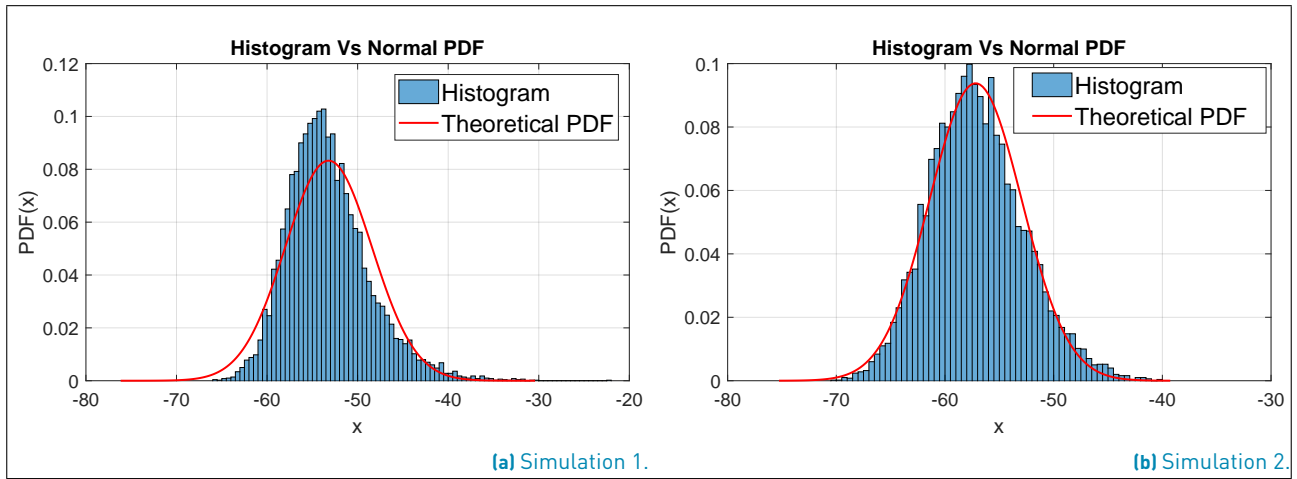
(a) Simulation 1.

(b) Simulation 2.

Figure 1 Comparison of the histogram of Y against an analytical Normal PDF with the parameters $\tilde{\mu}, \tilde{\sigma}$

In [8], the sum is approximated with a *lognormal* PDF by means of a recursive process for the calculation of the moments, but the expressions are not in closed form. In addition, it uses approximations by power series with 40 terms, which increases the computational cost. This method is useful when both the mean value and the variance of the sum are required. Then in [9] the mean value and the variance are approximated in a similar way to the work shown in [8], with the difference that the approximation is developed by a Taylor series around 0. This method is used in the area of finance. Work of [10] is based on [8], and it assumes correlated variables. A review of these approximation methods is shown in [11]. In this work, a $minmax$ optimization approach is proposed to adjust the parameters of the *lognormal* PDF. This work shows that these approximations do not provide a good fit in the tails of the PDF.

A later work in [12], proposed a closed solution for the CDF, but their most relevant contribution is that they demonstrate through several results that according to the parameters used, it is not correct always to approximate the sum of *lognormal* variables with a *lognormal* PDF due to the error on the tails of the PDF function.

The work in [13] develops a closed expression to adjust the lower tail of a *lognormal* PDF. In [14], authors approximate the moment-generating function with a *Gauss Hermite* series for each variable and in [15] the authors propose an approximation of the PDF with a *Log-shifted-gamma* function.

[16] offers a review of related works with some focus on the approximation of the CDF for the sum of independent and correlated variables based on the characteristic function given in [17] (quadrature rules). As a result, [16] also uses

the *Log-shifted-gamma* approximation, but it is only useful for the calculation of the CDF of the sum of 6 variables or less. Additionally, the authors propose a computational method using the Epsilon algorithm to calculate a CDF when 40 or more variables are added. In [18] the $\alpha - \mu$, distribution is used for the PDF, and the adjustment of parameters is done by means of a least squares regression approach. Finally, [5] approximates the PDF with a quasi-normal distribution, but this method is useful only when $\mu, \sigma \leq 1$.

2.1 Review of the approximation methods

The Fenton-Wilkinson method

One of the most widely used works for the computation of the sum of *lognormal* random variables is the Fenton-Wilkinson method [7] due to its good results and the simplicity of its formulas. This method states the following: a random variable X_i follows a Normal distribution with mean m_i standard deviation s_i. $X_i \sim$ Normal $[m_i, s_i]$, if $x_i = 10^{\frac{X_i}{10}}$; then: $x_i \sim lognormal[\mu_i, \sigma_i]$. Where: $\mu_i = \kappa \cdot m_i, \sigma_i = \kappa \cdot s_i$ and $\kappa = \frac{ln(10)}{10}$. The 1st $(n = 1)$ and 2nd $(n = 2)$ moments of x_i are given as:

$$\mathbb{E}[x_i^n] = e^{n \cdot u_i + \frac{(n\sigma_i)^2}{2}} \tag{4}$$

$$\mathbb{V}\text{ar}[x_i] = e^{2u_i + \sigma_i^2}\left(e^{\sigma_i^2} - 1\right) \tag{5}$$

If we have:

$$y = \sum_{i=1}^{N} x_i \tag{6}$$

Assuming that $y \sim lognormal[\tilde{\mu}, \tilde{\sigma}]$. then:

$$\mathbb{E}\left[y\right] = \sum_{i=1}^{N} \mathbb{E}\left[x_i\right] \tag{7}$$

$$\mathbb{V}\text{ar}\left[y\right] = \sum_{i=1}^{N} \mathbb{V}\text{ar}\left[x_i\right] \tag{8}$$

Finally:

$$\tilde{\sigma} = \kappa^2 ln\left(\frac{\mathbb{V}\text{ar}\left[y\right]}{\mathbb{E}\left[y\right]^2} + 1\right) \tag{9}$$

$$\tilde{\mu} = \kappa \cdot ln\left(\frac{\mathbb{E}\left[y\right]}{\sqrt{\frac{\mathbb{V}\text{ar}[y]}{\mathbb{E}[y]} + 1}}\right) \tag{10}$$

Now, if we do:

$$Z = 10 \cdot log_{10}\left(y\right) \tag{11}$$

Then:

$$Z \sim \text{Normal}\left[\tilde{\mu}, \tilde{\sigma}\right] \tag{12}$$

Now, we assess whether this method works correctly for the calculation of the mean value and variance through a Monte Carlo simulation.

Explanation of Monte Carlo simulation 3 used to analyze the Fenton-Wilkinson method As an example, we will use three RV's (N=3) with 10^7 samples for each. $X_i, i = [1, 2, 3]$ follows a Normal PDF. We calculate $x_i = 10^{\frac{X_i}{10}}$ to generate *lognormal* RV's and $y = \sum_{i=1}^{N} x_i$. Then, we get the experimental $\tilde{\mu}, \tilde{\sigma}$. Analytically Equations 7 and 8 are used to get $\mathbb{E}\left[y\right]$ and $\mathbb{V}\text{ar}\left[y\right]$. After that, we use Equations 9 and 10 to find the theoretical $\tilde{\mu}$ and $\tilde{\sigma}$. We compare both experimental and analytical results with different parameters m_i, s_i. Those results are shown in Table 1.

Note that experimental and analytical results begin to fail when $s_i > 4$, even for only 3 RVs.

The Schwartz-Yeh method

This method was proposed by [8] in 1982, which calculates the mean value $\mathbb{E}\left[Z\right]$ and variance $\mathbb{V}\text{ar}\left[Z\right]$ within a set of N *lognormal* random variables $x_i \sim lognormal\left[\mu_i, \sigma_i\right]$, $i \in \{1, 2, ..., N\}$. When $i = 1$, we have Equation 13:

$$Z_1 = X_1 \tag{13}$$

With: $\mathbb{E}\left[Z_1\right] = m_1$; $\mathbb{V}\text{ar}\left[Z_1\right] = s_1$.
And for $i > 1$:

$$Z_i = Z_{i-1} + 10log_{10}\left(1 + 10^{\frac{W_i}{10}}\right), \forall i \in \{2, ..., N\} \tag{14}$$

With Equation 15 as :

$$W_i = X_i - Z_{i-1}, \forall i \in \{2, ..., N\} \tag{15}$$

Now, we focus on calculating the mean value.

$$\mathbb{E}\left[Z_i\right] = \mathbb{E}\left[Z_{i-1}\right] + \mathbb{E}\left[10 \cdot log_{10}\left(1 + 10^{\frac{W_i}{10}}\right)\right] \tag{16}$$

We will use the term:

$$\xi\left(W_i\right) = 10 \cdot log_{10}\left(1 + 10^{\frac{W_i}{10}}\right) \tag{17}$$

Rewritten:

$$Z_i = Z_{i-1} + \xi\left(W_i\right) \tag{18}$$

$$\mathbb{E}\left[Z_i\right] = \mathbb{E}\left[Z_{i-1}\right] + \mathbb{E}\left[\xi\left(W_i\right)\right] \tag{19}$$

The term $\xi\left(W_i\right)$ is originally expanded through a power series with 40 terms. After that, we calculate $\mathbb{E}\left[\xi\right]$ and $\mathbb{V}\text{ar}\left[\xi\right]$. Equation 14 shows that this method is recursive, and we have to expand this term $\xi\left(W_i\right)$ in each i iteration which reduces the computational performance.

Observations: $W_i \sim \text{Normal}\left[u_i, s_i\right]$, $u_i = m_i - \mathbb{E}\left[Z_{i-1}\right]$, $s_i = \sqrt{s_i^2 + \mathbb{V}\text{ar}\left[Z_{i-1}\right]}$

3. Proposed method

In order to improve the computational performance and the analytical treatment of the method proposed by Schwartz-Yeh, we need to find a function that approximates $\xi\left(\cdot\right)$ (Equation 17) using a fewer number of terms. It also must be easily integrable. Therefore, the proposed solution in our work for $\xi\left(\cdot\right)$ is hereafter Equation 20 with the parameters: a_n, b_n, l_n, where n only takes the value 1 or 2 to indicate the 1st and 2nd polynomial degree, respectively. Figure 2 shows the comparison between Equation 20 against Equation 17 which evidences such an approach.

$$\xi\left(W_i\right)^n \cong \begin{cases} e^{n \cdot \frac{a_n + W_i}{b_n}} & -\infty < W_i < l_n \\ W_i^n & W_i \geq l_n \end{cases} \tag{20}$$

Parameters a_n and b_n were found through a nonlinear regression. The parameter l_n is the closest intersection point of $\xi\left(\cdot\right)$ with the approximation given by Equation 20 (see Figure 2). To find this point, we define l_n as:

$$\min_{l_n \in \mathbb{R}} \Delta(l_n) \tag{21}$$

Where $\Delta(l_n)$ is given as a Riemann integral.

$$\Delta(l_n) = \int_{-\infty}^{\infty} \xi\left(W_i\right)^n \cdot dW_i - \left(\int_{-\infty}^{l_n} e^{\frac{a_n + W_i}{b_n}} \cdot dW_i + \right.$$

Table 1 Comparison of Monte Carlo simulations and analytical results using the Fenton-Wilkinson method. **Expe=Experimental**

{M} {S}	E $[y]$	Expe. mean$[y]$	Var $[y]$	Expe. Var$[y]$	$\widetilde{\mu}$	Expe. mean$[Z]$	$\widetilde{\sigma}$	Expe. Var$[Z]$
{M_1} {S_1}	12.69×10^{-6}	12.69×10^{-6}	9.93×10^{-6}	9.92×10^{-6}	-49.46	-49.41	7.78	7.02
{M_2} {S_2}	11.11×10^{-6}	11.11×10^{-6}	5.40×10^{-6}	5.40×10^{-6}	-49.47	-49.46	3.34	3.20
{M_3} {S_3}	11.11×10^{-5}	11.11×10^{-5}	5.40×10^{-5}	5.40×10^{-5}	-39.94	-39.94	3.92	3.89
{M_4} {S_4}	45.48×10^{-4}	45.46×10^{-4}	20.68×10^{-2}	14.23×10^{-2}	-39.99	-39.06	143.95	116.96
{M_5} {S_5}	46.05×10^{-6}	45.99×10^{-6}	4.28×10^{-6}	1.31×10^{-6}	-59.89	-57.09	143.54	87.07

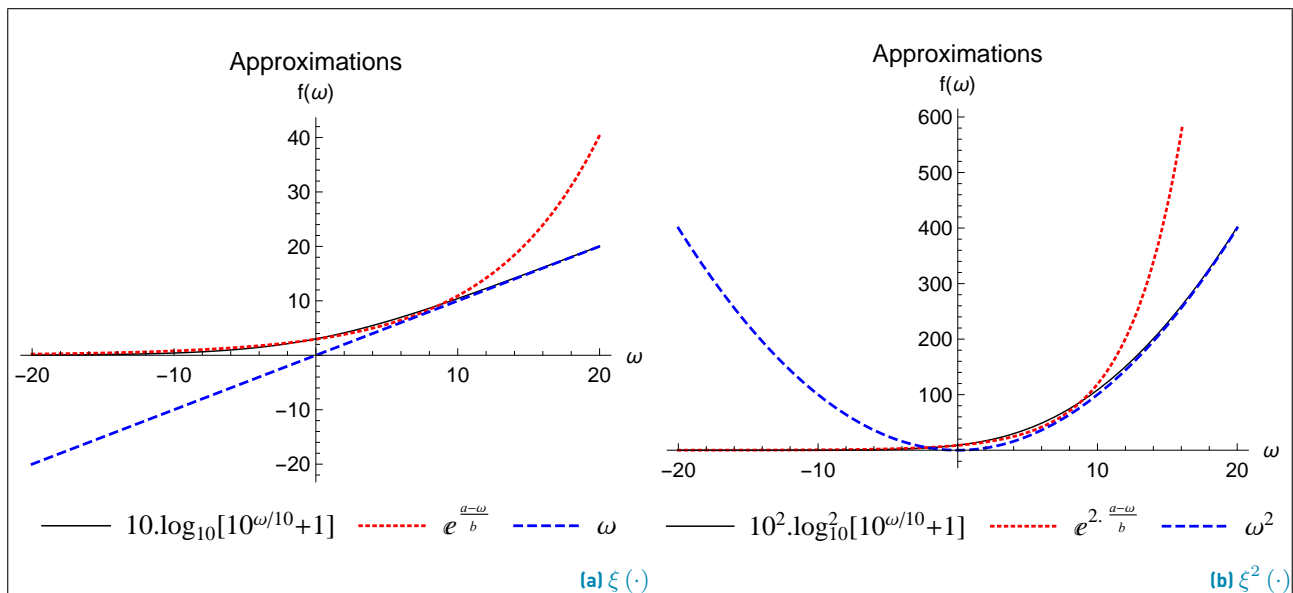| |
|---|
| $M_1=\{m_1, m_2, m_3\}=\{-70, -60, -50\}$, $S_1 = \{s_1, s_2, s_3\} = \{1, 2, 3\}$ |
| $M_2=\{m_1, m_2, m_3\}=\{-70, -60, -50\}$, $S_2 = \{s_1, s_2, s_3\} = \{1, 2, 2\}$ |
| $M_3 = \{m_1, m_2, m_3\} = \{-70, -60, -40\}$, $S_3 = \{s_1, s_2, s_3\} = \{1, 2, 2\}$ |
| $M_4 = \{m_1, m_2, m_3\} = \{-70, -60, -80\}$, $S_4 = \{s_1, s_2, s_3\} = \{4, 8, 12\}$ |
| $M_5 = \{m_1, m_2, m_3\} = \{-70, -60, -80\}$, $S_5 = \{s_1, s_2, s_3\} = \{8, 12, 4\}$ |

Figure 2 Comparison between $\xi(\omega)$ and its approximations, using the parameters in Table 2

$$\left. \int_{l_n}^{\infty} W_i^n \cdot dW_i \right) \tag{22}$$

Equation 22 seeks to approximate the areas of both functions Equation 17 and Equation 20. The evaluation of the minimization has been done through numerical integration, and the exact parameters used with Equation 20 are listed in Table 2.

Next, we proceed to calculate the moments of the function $\xi(\cdot)$. We have for any $i > 1$ that:

$$W_i \sim \mathsf{N}[u_i, s_i] \tag{23}$$

Then:

$$f_{W_i}(W_i) = \frac{e^{-\frac{1}{2}\left(\frac{W_i - u_i}{s_i}\right)^2}}{\sqrt{2\pi}s_i} \tag{24}$$

The first moment is:

$$\mathbb{E}[\xi(W_i)] = \int_{-\infty}^{\infty} \xi(W_i) \cdot f_{W_i}(W_i) \cdot dW_i =$$

$$\int_{-\infty}^{l_1} e^{\frac{a_1 + W_i}{b_1}} \cdot f_{W_i}(W_i) \cdot dW_i + \int_{l_1}^{\infty} W_i \cdot f_{W_i}(W_i) \cdot dW_i \tag{25}$$

$$= \frac{u_i}{2}\left[\text{erf}\left(\frac{u_i - l_1}{\sqrt{2}s_i}\right) + 1\right] + \frac{s_i \cdot e^{-\frac{(u_i - l_1)^2}{2s_i^2}}}{\sqrt{2\pi}} +$$

$$\frac{1}{2}e^{\frac{2b_1(a_1 + u_i) + s_i^2}{2b_1^2}}\text{erfc}\left[\frac{b_1(u_i - l_1) + s_i^2}{\sqrt{2}b_1 \cdot s_i}\right] \tag{26}$$

n	a	b	l	$\Delta(l)$ (see equation 22)
1	7.78279	1/0.136807	10.8040	1.2197×10^{-4}
2	7.67784	1/0.13826	11.1620	5.6627×10^{-4}

And the second moment is:

$$\mathbb{E}\left[\xi^2\left(W_i\right)\right] = \int_{-\infty}^{\infty} \xi^2\left(W_i\right) \cdot f_{W_i}\left(W_i\right) \cdot dW_i \quad (27)$$

$$= \int_{-\infty}^{l_2} e^{2 \cdot \frac{a_2 + W_i}{b_2}} \cdot f_{W_i}\left(W_i\right) \cdot dW_i +$$

$$\int_{l_2}^{\infty} W_i^2 \cdot f_{W_i}\left(W_i\right) \cdot dW_i \quad (28)$$

$$= \frac{1}{2} e^{\frac{2\left(b_2(a_2 + u_i) + s_i^2\right)}{b_2^2}} \operatorname{erfc}\left[\frac{b_2(u_i - l_2) + 2s_i^2}{\sqrt{2}.b_2.s_i}\right] +$$

$$\frac{1}{2}\left(s_i^2 + u_i^2\right)\left(\operatorname{erf}\left[\frac{u_i - l_2}{\sqrt{2}s_i}\right] + 1\right) + \frac{s_i(l_2 + u_i)e^{-\frac{(u_i - l_2)^2}{2s_i^2}}}{\sqrt{2\pi}} \quad (29)$$

We can use Equations 26 and 29 to calculate the variance:

$$\mathbb{V}\mathrm{ar}[\xi(\cdot)] = \mathbb{E}\left[\xi^2\left(\cdot\right)\right] - \mathbb{E}\left[\xi\left(\cdot\right)\right]^2 \quad (30)$$

Hereinafter, we must calculate the mean value and variance in each i iteration as follows.

Calculating the mean value $\mathbb{E}\left[Z_i\right]$ We have Equation 19 as:

$$\mathbb{E}\left[Z_i\right] = \mathbb{E}\left[Z_{i-1}\right] + \mathbb{E}\left[\xi\left(W_i\right)\right]$$

Calculating the variance $\mathbb{V}\mathrm{ar}\left[Z_i\right]$ Due to

$$\mathbb{V}\mathrm{ar}\left[Z_i\right] \neq \mathbb{V}\mathrm{ar}\left[Z_{i-1}\right] + \mathbb{V}\mathrm{ar}\left[\xi\left(W_i\right)\right] \quad (31)$$

because W_i and Z_{i-1} are correlated, we need to seek an alternative way to give a solution. Then, calculating first $\mathbb{E}\left[Z_i^2\right]$.

$$\mathbb{E}\left[Z_i^2\right] = \mathbb{E}\left[\left\{Z_{i-1} + \xi\left(W_i\right)\right\}^2\right] \quad (32)$$

$$= \mathbb{E}\left[Z_{i-1}^2\right] + 2\mathbb{E}\left[Z_{i-1} \cdot \xi\left(W_i\right)\right] + \mathbb{E}\left[\xi\left(W_i\right)^2\right] \quad (33)$$

Then

$$\mathbb{V}\mathrm{ar}\left[Z_{i-1}\right] + \mathbb{E}\left[Z_{i-1}\right]^2 + 2\mathbb{E}\left[Z_{i-1} \cdot \xi\left(W_i\right)\right] + \mathbb{E}\left[\xi\left(W_i\right)^2\right] \quad (34)$$

The term $2\mathbb{E}\left[Z_{i-1} \cdot \xi\left(W_i\right)\right]$ must be treated carefully because both variables are not independent. Based on [8], we have that:

$$2\mathbb{E}\left[Z_{i-1} \cdot \xi\left(W_i\right)\right] = \mathcal{A} - \mathcal{B} \quad (35)$$

Where:

$$\mathcal{A} = 2 \cdot Z_{i-1}\mathbb{E}\left[\xi\left(W_i\right)\right] \quad (36)$$

$$\mathcal{B} = 2\rho^2 \mathbb{E}\left[(W_i - m_{w_i})\xi\left(W_i\right)\right] \quad (37)$$

$$m_{w_i} = \mathbb{E}\left[\xi\left(W_i\right)\right] \quad (38)$$

$$\rho = -\frac{s_i}{s_{w_i}} \quad (39)$$

The term $\mathbb{E}\left[(W_i - m_{w_i})\xi\left(W_i\right)\right]$ is calculated using our approximation proposed in Equation 20.

$$\mathbb{E}\left[(W_i - m_{w_i})\xi\left(W_i\right)\right] = \quad (40)$$

$$\int_{-\infty}^{l_1} (W_i - m_{w_i}) e^{\frac{a_1 + W_i}{b_1}} \cdot f_{W_i}\left(W_i\right) \cdot dW_i +$$

$$\int_{l_1}^{\infty} (W_i - m_{w_i}) W_i \cdot f_{W_i}\left(W_i\right) \cdot dW_i \quad (41)$$

So:

$$\mathbb{E}\left[(W_i - m_{w_i})\xi\left(W_i\right)\right] = \quad (42)$$

$$\frac{1}{2}s^2 \left[\frac{\mathsf{A} \cdot \operatorname{erf}\left(\sqrt{\mathsf{B}}\right)}{|\mathsf{A}|} + \left(\frac{l_1\sqrt{\frac{2}{\pi}}}{s \cdot e^{\mathsf{B}}} + 1\right)\right] +$$

$$\frac{1}{2}\left(s \cdot e^{\frac{a_1}{b_1} + \mathsf{D}}\right)\left[\frac{s}{b_1} - \frac{s\,|\mathsf{C}|\,\operatorname{erf}\left(\frac{|\mathsf{C}|}{\sqrt{2}b_1.s}\right)}{b_1\mathsf{C}} - \sqrt{\frac{2}{\pi}}e^{\frac{l_1}{b_1} - \mathsf{B} - \mathsf{D}}\right] \quad (43)$$

Where:
$\mathsf{A} = u_{i_2} - l_1$
$\mathsf{B} = \frac{\mathsf{A}^2}{2s_i^2}$
$\mathsf{C} = s_i^2 + b_1 \cdot \mathsf{A}$
$\mathsf{D} = \frac{s_i^2 + 2b_1 \cdot u_i}{2b_1^2}$
And finally, we have:

$$\mathbb{V}\mathrm{ar}\left[Z_i\right] = \mathbb{E}\left[Z_i^2\right] - \mathbb{E}\left[Z_i\right]^2 \quad (44)$$

$$= \mathbb{V}\mathrm{ar}\left[Z_{i-1}\right] + \mathbb{E}\left[Z_{i-1}\right]^2 +$$

$$2\mathbb{E}\left[Z_{i-1} \cdot \xi\left(W_i\right)\right] + \mathbb{E}\left[\xi\left(W_i\right)^2\right] -$$

$$\left\{ \mathbb{E}\left[Z_{i-1}\right]^2 + 2\mathbb{E}\left[Z_{i-1}\right]\mathbb{E}\left[\xi\left(W_i\right)\right] + \mathbb{E}\left[\xi\left(W_i\right)\right]^2 \right\} \quad (45)$$

$$\mathbb{V}\text{ar}\left[Z_i\right] = \mathbb{V}\text{ar}\left[Z_{i-1}\right] + \mathbb{V}\text{ar}\left[\xi\left(W_i\right)\right] +$$

$$2\mathbb{E}\left[Z_{i-1}\cdot\xi\left(W_i\right)\right] - 2\mathbb{E}\left[Z_{i-1}\right]\mathbb{E}\left[\xi\left(W_i\right)\right] \quad (46)$$

Now with $\mathbb{E}\left[\xi\left(W_i\right)\right]$ using Equation 26, $\mathbb{V}\text{ar}\left[\xi\left(W_i\right)\right]$ using Equation 30 and $2\mathbb{E}\left[Z_{i-1}\cdot\xi\left(W_i\right)\right]$ using Equation 35, we have all equations to be used together with Equation 46, which can also be rewritten as:

$$\mathbb{V}\text{ar}\left[Z_i\right] = \mathbb{V}\text{ar}\left[Z_{i-1}\right] + \mathbb{V}\text{ar}\left[\xi\left(W_i\right)\right] +$$

$$2\cdot\text{Cov}\left[Z_{i-1}\cdot\xi\left(W_i\right)\right] \quad (47)$$

Where: **Cov** means covariance. Finally in this section, we have found all the formulations to evaluate $\mathbb{E}\left[\xi\left(W_i\right)\right]$, $\mathbb{V}\text{ar}\left[\xi(W_i)\right]$, $\mathbb{E}\left[Z_i\right]$, and $\mathbb{V}\text{ar}\left[Z_i\right]$ through the Equations 26, 30, 3, and 46, respectively.

4. Validation and analysis

We are going to validate our formulas by comparing them against the results of the Monte Carlo simulation. For this aim, we sum different numbers of RV's with a wide range of parameters. The difference between the theoretical and experimental results are shown in Tables 3, 4, and 5. In this way, we can show that the approximation for $\xi\left(\cdot\right)$ let us calculate the mean value and variance for Z with a low error, even when the range of RVs are $N = 20$, and the parameters are $m \in [-80, 80]$ and $s \in [6, 12]$.

Explanation of Monte Carlo simulations 4,5 and 6 used to analyze Equations for $\mathbb{E}\left[Z\right]$ and $\mathbb{V}\text{ar}\left[Z\right]$ We want to assess Equations 3 and 46 through simulations. Simulation number 4 analyzes the sum of only two *lognormal* RVs, $N = 2$. Then, we generate 10^7 samples for X_i following a Normal PDF with parameters m_i, s_i and finally, we calculate Equation 49. From this result, we get the experimental mean value and variance. We repeat this procedure 4 times with different parameters m_i and s_i. Simulation 5 and 6 follow the previous procedure, using $N = 10$ and $N = 20$ RVs, respectively.

$$Z_i = 10\cdot log_{10}\left(10^{\frac{X_{i-1}}{10}}\left[1 + 10^{\frac{X_i - X_{i-1}}{10}}\right]\right) \quad (48)$$

$$\forall i = [1, 2, .., N] \quad (49)$$

The 3rd result in Table 4 and the 4th result in Table 5 show the maximum variance difference between the theoretical and the experimental. On the other hand, this method is based on the work of Schwartz-Yeh, but with a lower number of terms. To validate the performance of the method, two metrics are taken into account: *error of the mean value* and *error of the variance*.

- The perceptual error of the mean value: It is the difference between the experimental (by means of Monte Carlo simulation) and the theoretical result. This error will be percentual using as a reference the experimental result. The way to calculate this value is:

$$[\%]\,\text{mean error} = \frac{100\cdot\|\text{Expe.}\mathbb{E}\left[Z\right] - \text{Theoretical}\mathbb{E}\left[Z\right]\|}{\text{Expe.}\mathbb{E}\left[Z\right]} \quad (50)$$

- Percentual error of the variance: It is the difference between the experimental and the theoretical result of the variance. This error will be percentual using as a reference the experimental result, and the way to calculate this value is:

$$[\%]\,\text{var error} = \frac{100\cdot\|\text{Expe.}\mathbb{V}\text{ar}\left[Z\right] - \text{Theorical}\mathbb{V}\text{ar}\left[Z\right]\|}{\text{Expe.}\mathbb{V}\text{ar}\left[Z\right]} \quad (51)$$

The proposed method has no restrictions with respect to m values and s values of each RV. However, the numerical evaluation must take into account the quantization error. That is why we must look for methods to reduce the error with respect to the analytical model proposed. Our initial approach was aimed at characterizing the sum of *lognormal* variables in a telecommunications scenario (where it is usual to take values for $m = [-80, -60]$); however, in our assessment, a greater range for the mean value was used $m = [-80, 80]$.

When we evaluate $10^{\frac{X_i}{10}}$, very small or large numerical values can be generated which leads to errors due to rounding. To reduce the rounding error, it is important that when we calculate Equation 15, the order of the variables is chosen in such a way as to guarantee that $W_i \geq 0$. Another important aspect is to analyze how the value of $\tilde{\mu}, \tilde{\sigma}$ changes in each iteration. The order of a set of N *lognormal* RVs does not affect the result of $\tilde{\mu}, \tilde{\sigma}$ analytically. But in the numerical evaluation, it is observed that for the same set of random variables, the result can change according to its order. The results of simulation 6 show that the error of the mean value is usually below 0.5 %, whereas that the error of $\tilde{\sigma}$ can change drastically from 20 to 0.1 % just with an adequate ordering of random variables. Through experimental processes, it was found that $\tilde{\sigma}$ is more sensitive to the values of m_i instead of s_i.

Hence, the ordering of random variables is made simply with respect to the value of m_i using the command *sort* of Matlab. To validate which ordering is appropriate, the following experiment was carried out.

Explanation of simulation 7 used to determine the effect of ordering For different orderings, we are going to calculate the error metrics of our method using as a reference Monte Carlo simulation. In this case, we use 10 Normal RV's $X_i, i = \{1, ..., 10\}$ with parameters

Table 3 Assessment with $N = 2$, samples for each RV $=10^6$

$\{m_1, m_2\}$	$\{s_1, s_2\}$	Theo. $\mathbb{E}[Z]$	Expe. $\mathbb{E}[Z]$	Mean Error	Theo. $\mathbb{V}\mathrm{ar}[Z]$	Expe $\mathbb{V}\mathrm{ar}[Z]$	Var Error
51, 65	5, 12	66.44	66.47	0.03	104.41	103.38	1,03
-71, 10	5, 8	10.00	10.02	0.02	64	64.04	0.04
27, 24	9, 10	31.84	31.84	0.00	57.16	56.44	0.72
-33, 80	11, 7	80	79.97	0.03	49	48.83	0.17

Theo.=theoretical, Expe.=experimental, Mean error=$\| $ Theo. $\mathbb{E}[Z] - $ Expe. $\mathbb{E}[Z] \|$,

Var. error=$\| $ Theo. $\mathbb{V}\mathrm{ar}[Z] - $ Expe. $\mathbb{V}\mathrm{ar}[Z] \|$

Table 4 Assessments with $N = 10$, samples for each RV $=10^6$

$\{m_1, m_{10}\}$	$\{s_1, s_{10}\}$	Theo. $\mathbb{E}[Z]$	Expe. $\mathbb{E}[Z]$	Mean Error	Theo. $\mathbb{V}\mathrm{ar}[Z]$	Expe $\mathbb{V}\mathrm{ar}[Z]$	Var Error
M_1	S_1	71.75	72.19	0.44	45.79	47.41	1.62
M_2	S_2	82.79	82.64	0.15	63.96	65.54	1.58
M_3	S_3	73.13	73.61	0.48	94.62	101.71	7.09
M_4	S_4	63.02	62.82	0.20	24.89	24.94	0.05

$M_1 = \{-50, 68, -43, -19, 6, 64, 20, -25, 20, -20\}$, $S_1 = \{9, 9, 11, 7, 4, 7, 4, 10, 11, 10\}$
$M_2 = \{70, -56, 35, -37, -48, 6, 80, -65, 47, 58\}$, $S_2 = \{9, 10, 10, 7, 9, 6, 10, 11, 7, 10\}$
$M_3 = \{56, 72, -30, -11, -10, 23, 7, -72, -19, 3\}$, $S_3 = \{7, 11, 9, 8, 6, 4, 11, 8, 9, 11\}$
$M_4 = \{-27, -47, -38, -37, 63, 15, 31, -76, -19, 37\}$, $S_4 = \{5, 11, 8, 10, 5, 6, 5, 4, 4, 4\}$

Theo.=theoretical, Expe.=experimental, Mean error=$\| $ Theo. $\mathbb{E}[Z] - $ Expe. $\mathbb{E}[Z] \|$,

Var. error=$\| $ Theo. $\mathbb{V}\mathrm{ar}[Z] - $ Expe. $\mathbb{V}\mathrm{ar}[Z] \|$

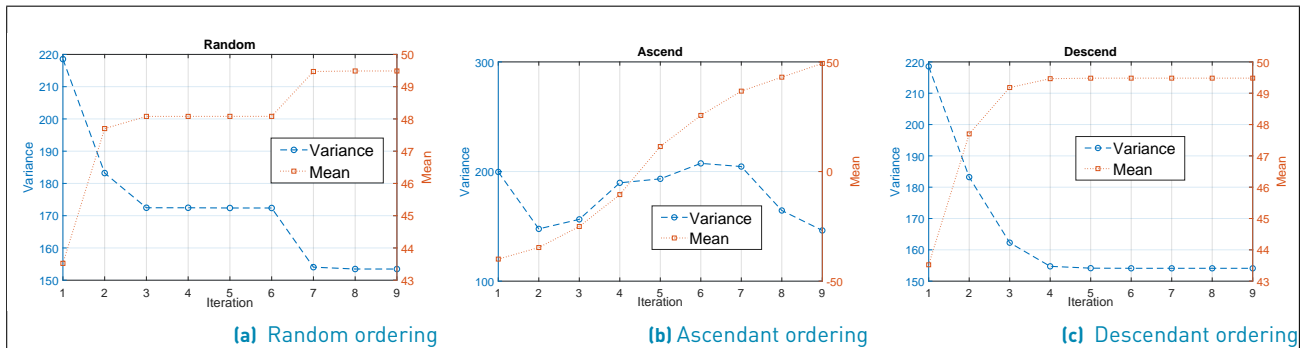**(a)** Random ordering   **(b)** Ascendant ordering   **(c)** Descendant ordering

Figure 3 Tracking the evolution of $\tilde{\mu}, \tilde{\sigma}^2$ for different ordering
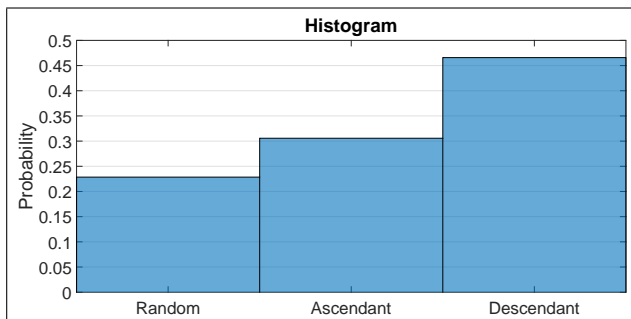
Figure 4 Histogram: Sorting method to get the smallest error of $\tilde{\sigma}^2$

following operation is performed: $y = \sum_{i=1}^{10} 10^{\frac{X_i}{10}}$; and we calculate $Y = 10 \cdot log_{10}(y)$.

For each set of m and s, we sort *lognormal* RVs according to m in three different ways: random, ascendant and descendant. We calculate the theoretical mean value and variance of Y in each order, and these results are compared against the experimental results given by Monte Carlo .

Finally, we choose the ordering with the smallest error. This procedure is done 10^6 times to know which ordering gives the smallest average error for the variance. To understand this phenomenon even better, as an example, we plot the evolution of $\tilde{\mu}$ and $\tilde{\sigma}^2$ in each iteration for each kind of ordering (Figure 3a-3c). The overall result is shown in Figure 4 as a histogram, where we can see that

randomly selected for the mean value and standard deviation from $m = [-80, 80]$, $s = [4, 20]$ respectively. For each RV 10^6 samples are generated. After that, the

Table 5 Assessment with $N = 20$, samples for each RV $=10^6$

$\{m_1, m_{20}\}$	$\{s_1, s_{20}\}$	**Theo. $\mathbb{E}[Z]$**	**Expe. $\mathbb{E}[Z]$**	**Mean Error**	**Theo. $\mathbb{V}\text{ar}[Z]$**	**Expe $\mathbb{V}\text{ar}[Z]$**	**Var Error**
M_1	S_1	84.55	84.78	0.23	28.96	32.53	3.57
M_2	S_2	80.11	80.25	0.14	34.83	32.32	2.51
M_3	S_3	77.95	77.90	0.05	102.78	103.45	0.67
M_4	S_4	85.42	85.15	0.27	65.11	69.10	3.99

$M_1 = \{54, 3, 25, -17, -59, 28, 55, 37, 78, 77, 70, 56, -18, 57, -74, 60, 8, 51, -37, -22\}$, $S_1 = \{5, 5, 6, 4, 10, 4, 5, 5, 8, 5, 12, 7, 7, 8, 9, 10, 7, 10, 5, 6\}$
$M_2 = \{70, 31, -16, -21, -6, 23, 0, -66, 76, -19, -32, 40, 62, 6, -66, 4, -66, -12, -39, -37\}$, $S_2 = \{12, 6, 9, 10, 9, 7, 8, 12, 5, 4, 4, 10, 9, 8, 12, 9, 5, 5, 8, 11\}$
$M_3 = \{1, -62, 19, 33, 76, -1, -57, 32, 52, 21, -8, -48, 9, -66, -58, 33, 55, 18, 37, 37\}$, $S_3 = \{7, 12, 10, 11, 12, 4, 7, 8, 12, 4, 8, 7, 7, 12, 6, 12, 12, 8, 8, 6\}$
$M_4 = \{-39, 43, 8, 28, 74, -9, 3, -5, -44, 68, -70, -37, 80, -7, 17, 19, 42, -68, -20, -64\}$, $S_4 = \{6, 11, 11, 4, 12, 4, 7, 5, 9, 7, 9, 4, 11, 7, 5, 9, 4, 5, 6, 11\}$

Theo.=theoretical, Expe.=experimental, Mean error$=\|$ Theo. $\mathbb{E}[Z]$ $-$ Expe. $\mathbb{E}[Z]$ $\|$,

Var. error$=\|$ Theo. $\mathbb{V}\text{ar}[Z]$ $-$ Expe. $\mathbb{V}\text{ar}[Z]$ $\|$

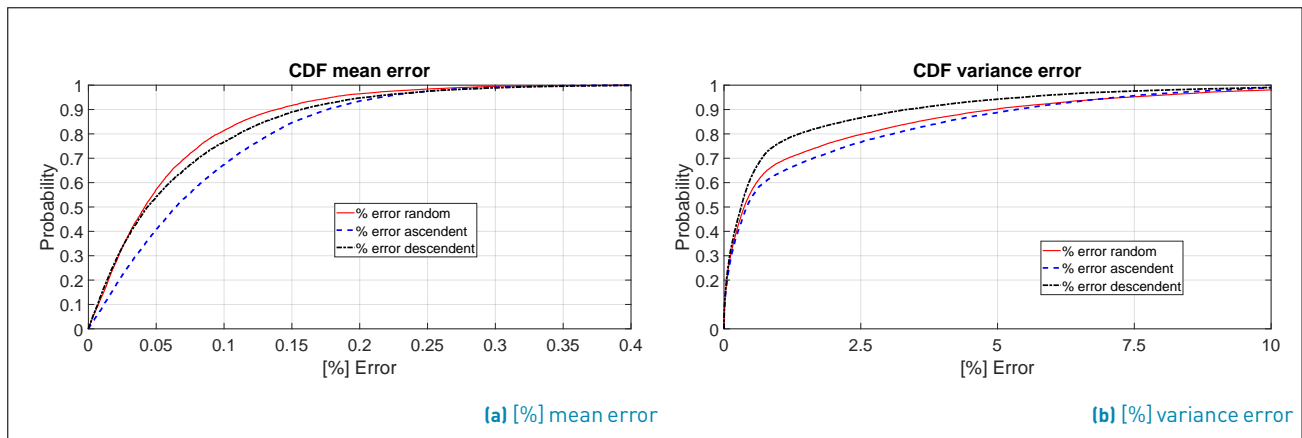**(a)** [%] mean error   **(b)** [%] variance error

Figure 5 Comparison of the error of the variance for different orderings

the smallest error occurs when the order is descending (with a probability of 47%). Note that the smaller error is reached with a descending order because the values converge asymptotically in each iteration. Meanwhile, in random and ascending ordering, abrupt changes of these metrics in each iteration are observed. (see Figure 3a-3c)

In order to give conclusions, we characterize statistically the behavior of the error metrics through their CDF. Figure 5 shows the CDF of the error for both metrics: *the error of the mean value* and *the error of the variance* for each ordering. It is noted that the error in the mean value $\tilde{\mu}$ is below of 0.2% (with a probability of 90% of the time) in all orderings, while the error in $\tilde{\sigma}$ is below 3% (with a probability of 90% of the time) for descending order.

5. Conclusions

In this paper, we have proposed an efficient method based on the Schwartz-Yeh's work for calculating the mean value

and variance of the sum of *lognormal* variables. We use Equation 17 as a recursive process in conjunction with our Equations 3 and 46 to find the mean value and variance. When compared to other methods in the literature, the benefits of our method are: 1) Computational efficiency because Equation 17 was expanded in 40 terms in previous works, now we use only two terms. 2) Low error and useful in many fields because it can be used with a wide range of parameters. Take into account that the order of the terms of the sum does matter due to the numerical evaluation (the rounding errors). Therefore, we want to highlight that the best results are obtained when (a) the variables are processed in descending order with respect to their mean, and (b) $W_i \geq 0$ is guaranteed.

The validation of the proposed method has considered ranges for random variables with parameters $m \in [-80, 80]$ and $s \in [6, 12]$, and the descending ordering showed that the error for $\tilde{\mu}$ is below 0.2% and the error for $\tilde{\sigma}$ is below 3%, both cases with a probability of 90% of

the time.

6. Declaration of competing interest

We declare that we have no significant competing interests, including financial or non-financial, professional, or personal interests interfering with the full and objective presentation of the work described in this manuscript.

7. Funding

8. Author contributions

Juan Guillermo Torres: As a PhD student, developed the set of formulas and the new proposed method Roberto Bustamante: guided this work based on telecommunication field.

9. Data availability statement

The used instrument was based on computer simulations, All data were generated through Monte Carlo Simulation.

References

[1] S. Nadarajah, "A Review of Results on Sums of Random Variables," *Acta Applicandae Mathematicae*, vol. 103, no. 2, pp. 131–140, sep 2008.

[2] G. L. Stuber, *Principles of mobile communication*. Springer, 2011.

[3] E. Limpert, W. A. Stahel, and M. Abbt, "Log-normal Distributions across the Sciences: Keys and Clues," *BioScience*, vol. 51, no. 5, p. 341, 2001.

[4] D. Dufresne, "Sums of lognormals," *Society of actuaries, University of Melbourne*, 2009.

[5] A. Messica, "A simple low-computation-intensity model for approximating the distribution function of a sum of non-identical lognormals for financial applications," in *AIP Conference Proceedings*, vol. 1773, 2016.

[6] N. A. Marlow, "A Normal Limit Theorem for Power Sums of Independent Random Variables," *Bell System Technical Journal*, vol. 46, no. 9, pp. 2081–2089, nov 1967.

[7] L. Fenton, "The Sum of Log-Normal Probability Distributions in Scatter Transmission Systems," *IEEE Transactions on Communications*, vol. 8, no. 1, pp. 57–67, 1960.

[8] S. C. Schwartz and Y. S. Yeh, "On the Distribution Function and Moments of Power Sums With Log-Normal Components," *Bell System Technical Journal*, vol. 61, no. 7, pp. 1441–1462, sep 1982.

[9] E. L. C. K. Shimizu, *Lognormal Distributions: Theory and Application*. New York: Dekker, 1988.

[10] A. Safak, "Statistical analysis of the power sum of multiple correlated log-normal components," *IEEE Transactions on Vehicular Technology*, vol. 42, no. 1, pp. 58–61, 1993.

[11] N. Beaulieu and Q. Xie, "An Optimal Lognormal Approximation to Lognormal Sum Distributions," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 2, pp. 479–489, mar 2004.

[12] N. Beaulieu and F. Rajwani, "Highly Accurate Simple Closed-Form Approximations to Lognormal Sum Distributions and Densities," *IEEE Communications Letters*, vol. 8, no. 12, pp. 709–711, dec 2004.

[13] S. S. Szyszkowicz and H. Yanikomeroglu, "On the Tails of the Distribution of the Sum of Lognormals," in *2007 IEEE International Conference on Communications*. IEEE, jun 2007, pp. 5324–5329.

[14] N. Mehta, J. Wu, A. Molisch, and J. Zhang, "Approximating a Sum of Random Variables with a Lognormal," *IEEE Transactions on Wireless Communications*, vol. 6, no. 7, pp. 2690–2699, jul 2007.

[15] C. Lam and T. Le-Ngoc, "Log-shifted gamma approximation to lognormal sum distributions," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 4 II, 2007.

[16] A. S. H. Mahmoud and A. H. Rashed, "Efficient Computation of Distribution Function for Sum of Large Number of Lognormal Random Variables," *Arabian Journal for Science and Engineering*, vol. 39, no. 5, pp. 3953–3961, may 2014.

[17] J. Gubner, "A New Formula for Lognormal Characteristic Functions," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 5, pp. 1668–1671, sep 2006.

[18] B. Wang, G. Cui, W. Yi, L. Kong, and X. Yang, "Approximation to independent lognormal sum with α-μ Distribution and the application," *Signal Processing*, vol. 111, 2015.