Research Article

An Estimate of the Probability Density Function of the Sum of a Random Number N of Independent Random Variables

Angelo Gifuni, Antonio Sorrentino, Giuseppe Ferrara, and Maurizio Migliaccio

Dipartimento di Ingegneria, Università degli Studi di Napoli "Parthenope", Centro Direzionale, Isola C4, 80143 Napoli, Italy

Correspondence should be addressed to Angelo Gifuni; angelo.gifuni@uniparthenope.it

Received 30 October 2014; Revised 1 March 2015; Accepted 3 March 2015

Academic Editor: Liudong Xing

Copyright © 2015 Angelo Gifuni et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A new estimate of the probability density function (PDF) of the sum of a random number of independent and identically distributed (IID) random variables X_i ($i = 1, 2, \dots, N$) is shown. The sum PDF is represented as a sum of normal PDFs weighted according to the N PDF. The analytical model is verified by numerical simulations. The comparison is made by the Chi-Square Goodness-of-Fit test.

1. Introduction

The probability density function (PDF) of the sum of a random number of independent random variables is important for many applications in the scientific and technical area [1]. Such a problem is not at all straightforward and has a theoretical solution only in some cases [2–5]. Further, in some cases the theoretical solution is not engineering viable (see, e.g., [2–6] and the references therein, especially in [4, 5] where the rates of convergence in various metrics are studied as well).

The purpose of this paper is to find a viable and good estimate of the PDF of a random variable (RV) Y which is the sum of a random number N of independent and identically distributed (IID) real RVs X_i ($i = 1, 2, \dots, N$). The new model has also the benefit of providing an intuitive support to physical interpretation. In simple terms, the sum PDF is represented as a sum of normal PDFs weighted according to the N PDF. Two particular results are achieved. They are the conditions under which the PDF of the sum approximates a normal PDF again and those under which it approximates the envelope of the PDF of N , except at the extremes when they are high.

The theory can be applied all the time where a random number of IID RVs has to be summed. For instance, the total energy stored in the volume of reverberation chambers (RCs) can be statistically achieved by the sum of a random number N of IID RVs representing the energy stored in as

many appropriate subvolumes N . By following this rationale, the PDF of the total stored energy in an RC can be achieved as well as the PDF of the relevant quality factor Q . A similar rationale is used to achieve a generalized stochastic field model for RCs [7].

Another important applicative case is experienced in coherent narrow-band sensor when the speckle arises and an appropriate model cannot be achieved to best mitigate it or exploit it as a geophysical signal [8].

2. The Model

The RV Y is expressed as follows:

$$Y = X_1 + X_2 + \dots + X_N, \quad (1)$$

where the real IID RVs X_i are assumed with finite variance. N is an integer and positive RV independent of the X_i ; the values of N are denoted with n_j , ($j = \min, \min+1, \min+2, \dots, \min+m$), where $n_{\min}$ and $n_{\min+m} = n_{\max}$ are the minimum and maximum value of N , respectively. The value $n_{\min}$ cannot be too small since the central limit theorem (CLT) has to be applied for each N value. $n_{\min} \geq 30$ can be certainly considered an adequate value [9].

The N PDF $f(n)$ can be written as follows:

$$f(n) = \sum_{j=\min}^{j=\min+m} P(n_j) \delta(n - n_j), \quad (2)$$

where $P(n_j)$ is the discrete probability relevant to n_j and $\delta(\cdot)$ is the delta function.

The mean μ_Y and variance σ_Y^2 of Y are as follows [2]:

$$\mu_Y = N_0 \mu_X, \quad (3)$$

$$\sigma_Y^2 = N_0 \sigma_X^2 + \sigma_N^2 \mu_X^2, \quad (4)$$

where N_0 and σ_N are the mean and standard deviation of N , respectively; μ_X and σ_X are the mean and standard deviation of X_i , respectively.

The random sum Y represents m mutually exclusive random sums, which are selected by the N values; they are

$$\begin{aligned} Y_{n_{\min}} &= \sum_{i=1}^{n_{\min}} X_i; \\ Y_{n_{\min+1}} &= \sum_{i=1}^{n_{\min+1}} X_i, \\ Y_{n_{\min+2}} &= \sum_{i=1}^{n_{\min+2}} X_i, \\ &\vdots \\ Y_{n_{\min+m}} &= \sum_{i=1}^{n_{\min+m}} X_i. \end{aligned} \quad (5)$$

The set of the values of Y , which is the certain event, is given by the sum of the mutually exclusive events, which are the subsets formed by the values of the m RVs Y_{n_j} ; that is, it can be written as follows:

$$Cy = Cy_{n_{\min}} + Cy_{n_{\min+1}} + Cy_{n_{\min+2}} + \cdots + Cy_{n_{\min+m}}, \quad (6)$$

where Cy is the set of the values of Y and Cy_{n_j} , $n_j = n_{\min}, \dots, n_{\min+m}$, are the subsets of the values of the corresponding Y_{n_j} . Therefore, it results in

$$\begin{aligned} \int_{-\infty}^{+\infty} f(y) dy &= \int_{-\infty}^{+\infty} P(n_{\min}) f(y_{n_{\min}}) dy_{n_{\min}} \\ &+ \int_{-\infty}^{+\infty} P(n_{\min+1}) f(y_{n_{\min+1}}) dy_{n_{\min+1}} + \cdots \\ &+ \int_{-\infty}^{+\infty} P(n_{\min+m}) f(y_{n_{\min+m}}) dy_{n_{\min+m}} = 1, \\ \int_{-\infty}^{+\infty} f(y_{n_j}) dy_{n_j} &= 1. \end{aligned} \quad (7)$$

The PDF of the random sum Y is given by the sum of m PDFs, $f(y_{n_j})$, which are weighted by the relevant probability $P(n_j)$. That is, it can be written as follows:

$$\begin{aligned} f(y) &= P(n_{\min}) f(y_{n_{\min}}) + P(n_{\min+1}) f(y_{n_{\min+1}}) \\ &+ P(n_{\min+2}) f(y_{n_{\min+2}}) + \cdots \\ &+ P(n_{\min+m}) f(y_{n_{\min+m}}). \end{aligned} \quad (8)$$

Considering the CLT, the PDFs $f(y_{n_j})$ can be assumed normal [9, 10]. Hence [9],

$$f(y_{n_j}) = \frac{1}{\sqrt{2\pi} \sqrt{n_j \sigma_X^2}} e^{-(y_{n_j} - n_j \mu_X)^2 / 2n_j \sigma_X^2}. \quad (9)$$

By considering (8) and (9), the estimate of $f(y)$ can be written as follows:

$$\begin{aligned} f(y) &= \sum_{j=\min}^{\min+m} P(n_j) f(y_{n_j}) \\ &= \sum_{j=\min}^{\min+m} P(n_j) \frac{1}{\sqrt{2\pi} \sqrt{n_j \sigma_X^2}} e^{-(y_{n_j} - n_j \mu_X)^2 / 2n_j \sigma_X^2}. \end{aligned} \quad (10)$$

It can be noted that if $\mu_X = 0$, then the Y PDF is the sum of m normal PDFs which are centered at zero with increasing variances $n_j \sigma_X^2$ and weighted by the relevant probability $P(n_j)$.

We stress that (9) and (10) are written under the condition of identically distributed RVs X_i ; it is precautionary with respect to the validity of the CLT [9–13]. However, even though the extension of (9) is practicable when the RVs X_i are not identically distributed, by considering the further conditions on the RVs X_i for the possible convergences different from the normal distribution [9–13], the corresponding extension of (10) requires more stringent conditions as specified below (see the end of Section 4). Note that when the PDF of the sum of the RVs X_i is exactly expressible for any constant value n_j apart from the PDFs of X_i , which could also be different, then (10) can be exactly written and no validation procedure is necessary. The sums of independent exponentially distributed RVs with different scale parameters as well [14–17] (and not only them) are an example where the X_i are not identically distributed and (9) and (10) can also be exactly achieved.

3. Model Applicability

Equation (10) can be easily implemented by a simple algorithm, which can be solved by a calculator to achieve the $f(y)$ values; they are denoted here as analytical solution values.

The error on a single distribution $f(y_{n_j})$ depends on n_j and on the form of the density $f(x)$ [9, 18–21]; it is maximized for $n_j = n_{\min}$; however, under the assumed hypothesis that the random variables X_i are IID, the value $n_j = 30$ is adequate for most applications [9]. The total error depends on $n_{\min}$ and $n_{\max}$, and it can be estimated by considering the error on each distribution $f(y_{n_j})$. The approximation in (10) improves as $n_{\min}$ increases. However, the analytical determination of the total error is too arduous. Here, in order to find the conditions such that the overall approximation is adequate for applications, we compare the histogram achieved by (10) and that achieved by numerical simulation. This last one is

achieved starting from the generation of the RVs having the same PDFs of X_i and N , which are known, and then by summing the X_i according to the generated values of N . The histogram originated from the numerical simulation is used as a null hypothesis for the comparison. The comparison is accomplished by Chi-Square Goodness-of-Fit test to the significance of 0.05. The analytical solution, which includes the implementation of (10), the numerical simulation, and Chi-Square Goodness-of-Fit test are accomplished by the software LabVIEW of National Instruments. Clearly, the width of the bins and the relevant centers along the x -axis is the same both for the analytical histogram and for the numerically simulated one.

Even though the results of the tests shown here are achieved by using as a null hypothesis the PDF of the simulated data, it has been verified that the outcome of the test is the same when one chooses as a null hypothesis the achieved model (10). Actually, since (10) is an analytical model, it allows us to calculate probability and samples for each bin with no statistical fluctuations, so that it could be used as a null hypothesis for the comparison. This inversion of rationale is again consistent with the theory of the Chi-Square Goodness-of-Fit test. In fact, it corresponds to consider the achieved model as an exact model so that it can originate the theoretical frequencies, which are compared with the frequencies achieved by numerical simulations; these last ones have the typical statistical fluctuations. In order to calculate the number of degrees of freedom for Chi-Square test, it is specified that N_0 , σ_N^2 , μ_X , and σ_X^2 are theoretical parameters as the PDF of N is known as well as the mean and variance of X_i . When (10) is considered exact, its mean and variance are promptly achieved; they turn out to be equal to (3) and (4), respectively.

It is specified that the comparisons are made so that the bins whose samples are less than four are discarded. It is also specified that the sample number of Y , which is denoted with k , is equal to 10000 and the bin number is 60.

In some cases, in order to avoid N values close to zero (too much low) or simply to test different values of $n_{\min}$ and $n_{\max}$ (namely, different N_0 values) with the same shape of the N PDF, the mean of the N PDF is shifted by adding a fixed value to it. In these cases, the relevant PDF is shown with the same nomenclature except for the addition of the adjective “shifted” as we have a shift of the mean value only.

The comparison is made by using various pairs of PDFs for N and X_i with different means and variances. In particular, two meaningful PDFs are tested, which are the uniform distribution and the one chi squared with two freedom degrees; this last one is shifted along the x -axis (shifted exponential distribution). However, both for N and X_i , PDFs as a triangular distribution, a chi distribution with two degrees of freedom (Rayleigh distribution), and a chi square distribution with six degrees of freedom are tested as well. To save space, only the results relative to PDFs as the uniform, (un)shifted exponential and (un)shifted Rayleigh are shown here.

All the tests made are accepted at the significance level of 0.05; some results are shown below, Figures 1–19.

FIGURE 1: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDF of N is formed by three delta of Dirac, with $N_1 = 485$, $N_2 = 500$, and $N_3 = 515$; $P(N_1) = P(N_2) = P(N_3) = 0.3(3)$; the PDF of X_i is a chi square distribution with two degrees of freedom. The null hypothesis is the normal PDF with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 500$; $\sigma_N = 12.3$; $\mu_X = 20000$; $\sigma_X = 20000$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $T_1 = 0.3$; $T_2 = 0.3$; $R_3 = 1$.

4. Discussion on the Behaviour of the Y PDF and Two of Its Useful Approximations

In this section, we consider the behavior of the Y PDF by some parameters defined in terms of mean and variance of N and X_i . In other words, we write the general conditions in order that the Y PDF approximates the normal PDF and the N one; they include implicitly the conditions why the Y PDF progressively changes from a normal PDF to the N one. The conditions, why the weighted normal PDFs approximating the Y PDF are separated, are also examined.

To start from (4), by also taking a cue from (10) and (24) in [2], and by considering that $N_0\sigma_X^2 \gg \sigma_N^2\mu_X^2$ implies $\sigma_N^2\mu_X^2/N_0\sigma_X^2 \ll 1$, it can be inferred that if

$$R_1 = \frac{\sigma_N^2}{N_0} \leq T_1, \quad (11)$$

$$R_2 = \frac{\sigma_N^2 \mu_X^2}{N_0 \sigma_X^2} \leq T_2 < 1, \quad (12)$$

where T_1 and T_2 are adequate threshold values depending on the PDFs of the RVs N and X_i , then we can again assume $f(y)$ approximately normal. Note that $\sigma_N^2/N_0 = N_0 cv_N$, where cv_N is the variation coefficient of N . Normally, the thresholds T_1 and T_2 are both less than one; however, if $\mu_X = 0$ then T_1 can turn out to be considerably greater than one. By setting

FIGURE 2: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDF of N is formed by three delta of Dirac, with $N_1 = 460$, $N_2 = 500$, and $N_3 = 540$; $P(N_1) = P(N_2) = P(N_3) = 0.3(3)$; the PDF of X_i is a chi square distribution with two degrees of freedom. The null hypothesis is the PDF of the simulated data with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 500$; $\sigma_N = 32.7$; $\mu_X = 20000$; $\sigma_X = 20000$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $R_1 = 2.1$; $R_2 = 2.1$; $R_3 = 1$.

$R_3 = \mu_X^2/\sigma_X^2$, it is noted that $R_2 = R_1 R_3$; namely, as (11)-(12) show, the thresholds T_1 and T_2 are not independent in sense that if one decreases, the other can increase and vice versa; clearly, it can occur within adequate limits depending on the PDFs of N and X_i . It can be noted that (11)-(12) are similar to (10) and (24) in [2]; actually, they are as a whole less stringent than (10) and (24) in [2]. However, (11)-(12) are very useful for the practical applications, where it is sufficient to know an approximate solution and not just that exact or asymptotically exact.

Similarly, by considering that $\sigma_N^2 \mu_X^2 \gg N_0 \sigma_X^2$ implies $R_2 = \sigma_N^2 \mu_X^2 / N_0 \sigma_X^2 \gg 1$, it can be inferred that if

$$R_1 = \frac{\sigma_N^2}{N_0} > 1, \quad (13)$$

$$1 \ll T_3 \leq R_2 = \frac{\sigma_N^2 \mu_X^2}{N_0 \sigma_X^2} \leq T_4, \quad (14)$$

where T_3 and T_4 are adequate threshold values depending on the PDFs of N and X_i , then $f(y)$ approximates the envelope of the N PDF, as it will be cleared below.

The thresholds T_3 and T_4 depend on the R_3 ; as this last one increases, T_3 decreases and T_4 increases.

For both inferences (11)-(12) and (13)-(14), it is necessary that the next weighted normal PDFs forming the Y PDF markedly interfere. In order to find the conditions under

FIGURE 3: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDF of N is formed by three delta of Dirac, with $N_1 = 450$, $N_2 = 500$, and $N_3 = 550$; $P(N_1) = P(N_2) = P(N_3) = 0.3(3)$; the PDF of X_i is a chi square distribution with two degrees of freedom. The null hypothesis is the PDF of the simulated data with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 500$; $\sigma_N = 40.8$; $\mu_X = 20000$; $\sigma_X = 20000$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $R_1 = 3.3$; $R_2 = 3.3$; $R_3 = 1$.

which the next weighted normal PDFs markedly interfere, the next unweighted normal PDFs can be considered. Mean and standard deviation of each normal component can be assumed equal to $\mu_{y_{n_j}} = n_j \mu_X$ and $\sigma_{y_{n_j}}^2 = n_j \sigma_X^2$, respectively; it can be written as follows:

$$\mu_{Y_{n_{j+1}}} - \mu_{Y_{n_j}} = (n_{j+1} - n_j) \mu_X = \mu_X. \quad (15)$$

Therefore, it has to be $\sigma_{Y_{n_j}} = \sqrt{n_j} \sigma_X \gg \mu_X$ (the double symbol $\gg$ means “the same order of magnitude and greater than”); that is

$$R_3 = \frac{\mu_X^2}{\sigma_X^2} \gg n_{\min}. \quad (16)$$

Indeed, if (16) is satisfied for $n_{\min}$ it is satisfied for any N belonging to its distribution. For the inference (11)-(12), in (16) the symbol meaning “greater than” is prevalent, whereas, for the inference (13)-(14), in (16) the symbol meaning “the same order of magnitude” is prevalent.

Note that condition (16) is included in condition (12). Note also that for a given pair of PDFs of N and X_i , by fixing a certain value of R_1 , as R_3 increases, R_2 increases as well; so, gradually T_4 is exceeded and condition (16) is no longer satisfied. Under these conditions, the weighted normal PDFs approximating the Y PDF are partially separated; this

FIGURE 4: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDF of N is formed by three delta of Dirac, with $N_1 = 440$, $N_2 = 500$, and $N_3 = 560$; $P(N_1) = P(N_2) = P(N_3) = 0.3(3)$; the PDF of X_i is a chi square distribution with two degrees of freedom. The null hypothesis is the PDF of the simulated data with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 500$; $\sigma_N = 49$; $\mu_X = 20000$; $\sigma_X = 20000$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $R_1 = 4.8$; $R_2 = 4.8$; $R_3 = 1$.

separation increases with the increasing of R_3 with respect to $n_{\min}$, so that (10) always approximates the Y PDF, but this last one does not approximate the envelope of the N PDF again. We stress that the limitation T_4 is due to condition (16).

Clearly, (15) and (16) can be written for the cases where the difference between two next N values is greater than 1; that is, it can be written as follows:

$$\mu_{Y_{n_{j+I}}} - \mu_{Y_{n_j}} = (n_{j+I} - n_j) \mu_X = I \mu_X, \quad (17)$$

where I is an integer given by the difference between n_{j+1} and n_j .

Therefore, it has to be $\sigma_{Y_{n_j}} = \sqrt{n_j} \sigma_X \cong I \mu_X$; that is,

$$R_3 I^2 = I^2 \frac{\mu_X^2}{\sigma_X^2} < \cong n_{\min}. \quad (18)$$

Shortly, the conditions in order that the Y PDF approximates a normal PDF again or the envelope of the N PDF are (11)-(12) and (13), (14), and (16), respectively. Below, both the inferences will be verified by (10) and numerical simulations; the values of the thresholds for which $f(y)$ approximates a normal PDF and those for which $f(y)$ approximates the envelope of the N PDF will also be shown for some examples.

FIGURE 5: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDF of N is formed by three delta of Dirac, with $N_1 = 350$, $N_2 = 500$, and $N_3 = 650$; $P(N_1) = P(N_2) = P(N_3) = 0.3(3)$; the PDF of X_i is a chi square distribution with two degrees of freedom. The null hypothesis is the PDF of the simulated data with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 500$; $\sigma_N = 122.5$; $\mu_X = 20000$; $\sigma_X = 20000$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $R_1 = 30.0$; $R_2 = 30.0$; $R_3 = 1$.

We recall that if the N PDF is normal then $f(y)$ is approximately normal apart from the conditions given here [2].

5. Experimental Results

In order to verify the model achieved for the Y PDF, some simulation results are shown; various pairs of PDFs are used for N and X_i with different means and variances. In particular, two PDFs with very different kurtosis index are tested, so as to make the tests very meaningful. Such PDFs are the uniform and chi squared with two freedom degrees shifted along the x -axis (shifted exponential distribution). The uniform distribution is platykurtic whereas the exponential one is leptokurtic.

A PDF formed by three delta of Dirac only ($m = 3$; $j = 1, 2, 3$) is first used for N ; see Figures 1–6. To each delta is coupled the relative probability. The X_i PDF is chi square distribution with two degrees of freedom so that R_3 is a constant equal to 1. Since the forms of histograms and relevant PDFs are the same, the marker points representing the height of the histogram bins are interpolated by a simple line in all the figures shown here.

When the normal approximation is tested, it is specified that the null hypothesis is the theoretical normal with mean and variance given by (3) and (4), respectively. Similarly,

FIGURE 6: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDF of N is formed by three delta of Dirac, with $N_1 = 350$, $N_2 = 500$, and $N_3 = 650$; $P(N_1) = P(N_3) = 0.2$, $P(N_2) = 0.6$; the PDF of X_i is a chi square distribution with two degrees of freedom. The null hypothesis is the PDF of the simulated data with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 500$; $\sigma_N = 94.9$; $\mu_X = 20000$; $\sigma_X = 20000$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $R_1 = 18.0$; $R_2 = 18.0$; $R_3 = 1$.

when the envelope of the N PDF is tested, it is specified that the null hypothesis is the N PDF, which is considered continuous, with mean and variance given by (3) and (4) again. However, the captions of all the figures are exhaustive and the null hypothesis is specified.

In Figures 1–6, the behavior of the Y PDF with the increasing of the ratio R_1 for the same probability (n_j) can be noted. In particular, when (11) and (12) are satisfied the Y PDF approximates the normal one; see Figure 1; then as the ratios R_1 and R_2 increase, the Y PDF converges to the envelope of the PDF of N except at its extremes when they are high; see Figures 2 and 3; finally, the Y PDF becomes a sequence of separated approximate normal PDFs; see Figures 5 and 6. In Figure 4, the normal PDFs forming the Y PDF are only partially separated. Figure 1 shows that if the thresholds T_1 and T_2 are limited to 0.3 and the PDFs of N and X_i are as specified in the caption of the same figures, then the Y PDF can be considered approximately normal again.

In Figures 5 and 6, it can be noted that each weighted normal PDF forming the PDF of the sum lowers and widens with the increasing of the relative n_j , as expected. In this connection, it is worth to recall that the variation coefficient of each weighted normal component tends to zero as the relative n_j tends to infinite. In Figure 6, note that the probabilities $P(n_j)$ are different and that in Figures 5 and 6, the weighted normal components are separated as the

condition (18) is not satisfied. Clearly, it is why the N PDF is not formed by consecutive deltas of Dirac but the next deltas of Dirac are far. In particular, $n_3 - n_2 = n_2 - n_1 = I = 150$ in Figures 5 and 6.

It is specified that different PDFs for X_i do not change the essence of the results, except the thresholds T_1 and T_2 for which $f(y)$ can be considered approximately normal again and those for which $f(y)$ approximates the envelope of the N PDF, with mean and variance given by (3) and (4). As an example, in Figures 7(a), 7(b), and 7(c), the normal approximation is tested where the PDFs of N and X_i are both uniform distributions; their parameters are shown in the captions of the relative figures. In Figures 8–10, the PDFs of N and X_i are both uniform distributions again; the results show that $f(y)$ approximates gradually the envelope of the N PDF. In particular, in Figure 9, where $R_1 = 10$, it is shown that $T_3 \cong 30$ and $T_4 \cong 80$. In Figure 11, the case where (16) is not satisfied so that the single normal PDFs are separated is shown; in this case, in order to resolve well the PDF graphs, the bin number is 1000 and k is 100000.

Figures 12–15 show the results where the PDF of N is a shifted chi square distribution with two degrees of freedom and the one of X_i is a uniform distribution. It can be noted again that $f(y)$ converges to the envelope of the N PDF as R_1 and R_2 increase. In Figure 15, the thresholds T_3 and T_4 for $R_1 = 71$ are shown.

Similarly, Figures 16–18 show the results where the PDFs of N and X_i are a shifted chi distribution with two degrees of freedom and a chi square distribution with two degrees of freedom, respectively. In Figure 19, the PDF of X_i is a uniform distribution. In Figure 19, the thresholds T_3 and T_4 for $R_1 = 54.79$ are also shown.

It is understood that (13), (14), and (16) or (18) are satisfied; it is important to note that with the increasing of R_2 the Y PDF converges to the envelope of the PDF of N except at the extremes when they are high. In fact, the extremes of $f(y)$ are always tapered and the tapering at the lower extremity is steeper than the one at the upper extremity, as the theoretical model (10) shows. It is shown in Figures 9–10 where the envelope of N PDF is uniform. Similarly, when the envelope of the N PDF has an extreme high as that of a shifted exponential PDF, where the first extreme is high and the second tends to zero, then for values sufficiently high of R_2 , the Y PDF assumes the same form of the envelope of the PDF of N , except for a starting steep rise to which is associated a small amount of probability only. However, for high values of R_2 , the probabilities associated with the tapers are very low in all cases apart from the N PDF, so that the Chi-Square test is as a rule accepted at the significance level of 0.05, even if the null hypothesis is with no tapers; it is shown in Figures 10, 15, and 19.

Hence, the Y PDF changes from a normal PDF to the envelope of the N PDF, by assuming various forms between the two, and finally, it gradually becomes a sequence of separated approximate normal PDFs according to the ratios R_1 , R_2 , and R_3 . Note that the envelope of the normal PDFs forming the Y PDF has the same form of the envelope of the N PDF also when they are totally separated as they are weighted according to the probabilities $P(n_j)$.

FIGURE 7: (a) Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDFs of N and X_i are both uniform distributions. $n_{\min}$ and $n_{\max}$ are 9307 and 10693, respectively. The null hypothesis is the normal PDF with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 10000$; $\sigma_N = 400$; $\mu_X = 0$; $\sigma_X = 57.73$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $T_1 = 15.96$; $T_2 = 0$; $R_3 = 0$. (b) Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDFs of N and X_i are both uniform distributions. $n_{\min}$ and $n_{\max}$ are 9307 and 10693, respectively. The null hypothesis is the normal PDF with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 10000$; $\sigma_N = 400$; $\mu_X = 8$; $\sigma_X = 57.73$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $T_1 = 15.9$; $T_2 = 0.3$; $R_3 = 0.019$. (c) Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDFs of N and X_i are both uniform distributions. $n_{\min}$ and $n_{\max}$ are 9948 and 10052, respectively. The null hypothesis is the normal PDF with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 10000$; $\sigma_N = 30$; $\mu_X = 100$; $\sigma_X = 57.73$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $T_1 = 0.09$; $T_2 = 0.27$; $R_3 = 3$.

FIGURE 8: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDFs of N and X_i are both uniform distributions. $n_{\min}$ and $n_{\max}$ are 948 and 1052, respectively. The null hypothesis is the PDF of the simulated data with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 1000$; $\sigma_N = 30$; $\mu_X = 10$; $\sigma_X = 57.7$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $R_1 = 0.9$; $R_2 = 2.7$; $R_3 = 3$.

FIGURE 9: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDFs of N and X_i are both uniform distributions. $n_{\min}$ and $n_{\max}$ are 38 and 142, respectively. The null hypothesis is the PDF of the simulated data with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 90$; $\sigma_N = 30$; $\mu_X = 100$; $\sigma_X = 57.73$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $R_1 = 10$; $R_2 = 30$; $R_3 = 3$; $R_1 = 10 \Rightarrow T_3 \approx 30$ and $T_4 \approx 80$.

FIGURE 10: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDFs of N and X_i are both uniform distributions. $n_{\min}$ and $n_{\max}$ are 2292 and 57703, respectively. The null hypothesis is a uniform PDF with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 30000$; $\sigma_N = 16000$; $\mu_X = 100$; $\sigma_X = 57.73$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $R_1 = 8533$; $R_2 = 25600$; $R_3 = 3$.

FIGURE 11: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDFs of N and X_i are both uniform distributions. $n_{\min}$ and $n_{\max}$ are 97 and 103, respectively. The null hypothesis is the PDF of the simulated data with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 100$; $\sigma_N = 2$; $\mu_X = 100$; $\sigma_X = 0.577$. In this case, the sample number k is 100000 and the number of bins is 1000. The chi squared test is accepted at the significance level of 0.05. $R_1 = 0.04$; $R_2 = 1200$; $R_3 = 30000$.

FIGURE 12: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDF of N is a chi square distribution with two degrees of freedom, which is shifted by 10000 unit, and the one of X_i is a uniform distribution. $n_{\min}$ and $n_{\max}$ are 10000 and 10164, respectively. The null hypothesis is the normal PDF with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 10018$; $\sigma_N = 18$; $\mu_X = 100$; $\sigma_X = 57.73$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $T_1 = 0.03$; $T_2 = 0.09$; $R_3 = 3$.

FIGURE 14: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDF of N is a chi square distribution with two degrees of freedom, which is shifted by 50 unit, and the one of X_i is a uniform distribution. $n_{\min}$ and $n_{\max}$ are 50 and 845, respectively. The null hypothesis is the PDF of the simulated data with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 122$; $\sigma_N = 72$; $\mu_X = 100$; $\sigma_X = 57.73$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $R_1 = 42.49$; $R_2 = 127.47$; $R_3 = 3$.

FIGURE 13: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDF of N is a chi square distribution with two degrees of freedom, which is shifted by 1000 unit, and the one of X_i is a uniform distribution. $n_{\min}$ and $n_{\max}$ are 1000 and 1645, respectively. The null hypothesis is the PDF of the simulated data with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 1072$; $\sigma_N = 72$; $\mu_X = 100$; $\sigma_X = 57.73$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $R_1 = 4.83$; $R_2 = 14.50$; $R_3 = 3$.

FIGURE 15: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDF of N is a chi square distribution with two degrees of freedom, which is shifted by one unit, and the one of X_i is a uniform distribution. $n_{\min}$ and $n_{\max}$ are 1 and 711, respectively. The null hypothesis is chi square with two degrees of freedom with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 73$; $\sigma_N = 72$; $\mu_X = 100$; $\sigma_X = 57.73$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $R_1 = 71.0$; $R_2 = 213.0$; $R_3 = 3$; $R_1 = 71 \Rightarrow T_3 \approx 450$ and $T_4 \approx 32000$.

FIGURE 16: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDF of N is a chi distribution with two degrees of freedom, which is shifted by 400 unit, and the one of X_i is a uniform distribution. $n_{\min}$ and $n_{\max}$ are 400 and 441, respectively. The null hypothesis is the normal PDF with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 412.53$; $\sigma_N = 6.55$; $\mu_X = 100$; $\sigma_X = 57.73$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $T_1 = 0.1$; $T_2 = 0.3$; $R_3 = 3$.

FIGURE 17: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDF of N is a chi distribution with two degrees of freedom, which is shifted by 10 unit, and the one of X_i is a uniform distribution. $n_{\min}$ and $n_{\max}$ are 10 and 56, respectively. The null hypothesis is the PDF of the simulated data with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 22.53$; $\sigma_N = 6.55$; $\mu_X = 100$; $\sigma_X = 57.73$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $R_1 = 1.9$; $R_2 = 5.71$; $R_3 = 3$.

FIGURE 18: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDF of N is a chi distribution with two degrees of freedom, which is shifted by one unit, and the one of X_i is a uniform distribution. $n_{\min}$ and $n_{\max}$ are 1 and 43, respectively. The null hypothesis is the PDF of the simulated data with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 13.53$; $\sigma_N = 6.55$; $\mu_X = 100$; $\sigma_X = 57.73$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $R_1 = 3.17$; $R_2 = 9.51$; $R_3 = 3$.

In order to give results ready for the applications, the next progress of this work is to map the areas of a Cartesian plane (R_1, R_2) to show where the Y PDF is approximately normal for the pairs of PDFs of N and X_i most common and useful for applications. Similarly, the areas of the Cartesian plane (R_1, R_2) where the Y PDF converges to the envelope of the N PDF should be mapped. In this last one case, R_1 can be taken as a parameter so that the areas can be mapped as a function of T_3 and T_4 .

In principle, by considering the PDF of N as a continuous function, the analytical model for the PDF of Y could also be written as an integral; that is, it could be written as follows:

$$f(y) \cong \int_{n_{\min}}^{n_{\max}} \frac{1}{\sqrt{2\pi} \sqrt{n\sigma_X^2}} e^{-(y-n\mu_X)^2/2n\sigma_X^2} f(n) dn. \quad (19)$$

Clearly, the continuity of N affects the quality of the approximation on $f(y)$. However, the analytical solution of the integral in (19) is enough difficult also in the case in which $f(n)$ is uniform; therefore, the attempt to achieve an analytical solution of (19) is not very feasible.

The above developments could be extended to the case where the RVs X_i are not identically distributed. In this case, by also considering the above on the further conditions

FIGURE 19: Y histogram: cross markers (black color), numerical simulation; circle markers (red color), analytical solution. The PDF of N is a chi distribution with two degrees of freedom and the one of X_i is a chi square distribution with two degrees of freedom. $n_{\min}$ and $n_{\max}$ are 2 and 776, respectively. The null hypothesis is chi with two degrees of freedom with mean and variance given by (3) and (4), respectively. The means and standard deviations of N and X_i are $\mu_N = 200.5$; $\sigma_N = 104.8$; $\mu_X = 20000$; $\sigma_X = 20000$. The sample number k is 10000 and the number of bins is 60. The chi squared test is accepted at the significance level of 0.05. $R_1 = 54.79$; $R_2 = 54.79$; $R_3 = 1$; $R_1 = 54.79 \Rightarrow T_3 \approx 54.79$ and $T_4 \approx 20000$.

for the convergence to the normal distribution [9–13], the estimates of $f(y_{n_j})$ and $f(y)$ can be written as follows:

$$f(y_{n_j}) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{Y_{n_j}}^2}} e^{-\frac{(y_{n_j} - \mu_{Y_{n_j}})^2}{2\sigma_{Y_{n_j}}^2}},$$

$$f(y) = \sum_{j=\min}^m P(y_j) f(y_{n_j}) \quad (20)$$

$$= \sum_{j=\min}^m P(n_j) \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{Y_{n_j}}^2}} e^{-\frac{(y_{n_j} - \mu_{Y_{n_j}})^2}{2\sigma_{Y_{n_j}}^2}},$$

where $\mu_{Y_{n_j}} = \mu_{X_1} + \mu_{X_2} + \dots + \mu_{X_{n_j}}$ and $\sigma_{Y_{n_j}}^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 + \dots + \sigma_{X_{n_j}}^2$ are mean and variance of the sum of X_i ($i = 1, 2, \dots, n_j$). Practically, (20) imply that $n_{\min}$ be very high so that $\mu_{Y_{n_j}}$ and $\sigma_{Y_{n_j}}^2$ have stable values; otherwise, the situation is much more difficult as it is necessary to know the means and variances $\mu_{Y_{n_j}}$ and $\sigma_{Y_{n_j}}^2$, as well as the PDFs of Y_{n_j} , so that (20) can be exactly written.

6. Conclusions

A model for the estimate of the PDF of the sum of a random number N of IID distributed random variables is shown. The

CLT is repeatedly applied, so that the PDF of the sum is represented by a sum of weighted normal PDFs; therefore, it is also an intuitive model and as a rule, it allows us to make out quickly the form of the PDF of Y . The validity of the model is verified by Chi-Square Goodness-of-Fit test by comparing the histogram correspondent with that achieved by numerical simulations starting from the generation of the RVs having the same PDFs of X_i and N , which are known, and then by summing the X_i according to the generated values of N .

Two particular results are achieved; they are the conditions under which the sum PDF can be approximated again by a normal PDF and those under which it approximates the envelope of the PDF of N with mean and variance given by (3) and (4), respectively.

The model is usable for practical applications, where it is useful and sufficient to know an approximate solution. In order to give results ready for the applications, the next progress of this work is to map the areas of a Cartesian plan to show the areas where the Y PDF is approximately normal and those where it converges to the envelope of the N PDF for the pairs of PDF of N and X_i interesting for applications.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported in part by the POR Campania FESR 2007/2013 O.O. 2.2—*Contract of Regional Program for the Development of the Production Chains in Campania*. The authors are grateful to Luigi De Luca (GGM Lab at the Università degli Studi di Napoli Parthenope) for the useful help in the paper editing.

References

- [1] B. V. Gnedenko and V. Y. Korolev, *Random Summation, Limit Theorems and Applications*, CRC Press, Boca Raton, Fla, USA, 1996.
- [2] H. Robbins, “The asymptotic distribution of the sum of a random number of random variables,” *Bulletin of the American Mathematical Society*, vol. 54, pp. 1151–1161, 1948.
- [3] T. L. Hung and T. T. Thanh, “Some results on asymptotic behaviors of random sums of independent identically distributed random variables,” *Korean Mathematical Society. Communications*, vol. 25, no. 1, pp. 119–128, 2010.
- [4] J. K. Sunklodas, “Some estimates of normal approximation for the distribution of a sum of a random number of independent random variables,” *Lithuanian Mathematical Journal*, vol. 52, no. 3, pp. 326–333, 2012.
- [5] J. K. Sunklodas, “On the normal approximation of a sum of a random number of independent random variables,” *Lithuanian Mathematical Journal*, vol. 52, no. 4, pp. 435–443, 2012.
- [6] L. B. Klebanov, A. V. Kakosyan, S. T. Rachev, and G. Temnov, “On a class of distributions stable under random summation,” *Journal of Applied Probability*, vol. 49, no. 2, pp. 303–318, 2012.

- [7] P. Corona, G. Ferrara, and M. Migliaccio, "Generalized stochastic field model for reverberating chambers," *IEEE Transactions on Electromagnetic Compatibility*, vol. 46, no. 4, pp. 655–660, 2004.
- [8] M. Migliaccio, G. Ferrara, A. Gambardella, F. Nunziata, and A. Sorrentino, "A physically consistent speckle model for marine SLC SAR images," *IEEE Journal of Oceanic Engineering*, vol. 32, no. 4, pp. 839–847, 2007.
- [9] A. Papoulis, *Probability, Random Variables and Stochastic Process*, McGraw-Hill, New York, NY, USA, 1991.
- [10] H. Fischer, *A History of the Central Limit Theorem*, Springer, New York, NY, USA, 2010.
- [11] B. V. Gnedenko, *Theory of Probability*, Mir, Moscow, Russia, 1978.
- [12] C. M. Grinstead and J. L. Snell, *Introduction to Probability*, 1997, http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.pdf.
- [13] J. K. Sunklodas, "Some estimates of the normal approximation for independent non-identically distributed random variables," *Lithuanian Mathematical Journal*, vol. 51, no. 1, pp. 66–74, 2011.
- [14] M. Balazs, "Sum of independent exponentials," <http://www.maths.bris.ac.uk/~mb13434/sumexp.pdf>.
- [15] M. Bibinger, "Notes on the sum and maximum of independent exponentially distributed random variables with different scale parameters," <http://arxiv.org/abs/1307.3945>.
- [16] M. Akkouchi, "On the convolution of exponential distributions," *Journal of the Chungcheong Mathematical Society*, vol. 21, 2008.
- [17] A. M. Mathai, "Storage capacity of a dam with gamma type inputs," *Annals of the Institute of Statistical Mathematics*, vol. 34, no. 3, pp. 591–597, 1982.
- [18] A. C. Berry, "The accuracy of the Gaussian approximation to the sum of independent variates," *Transactions of the American Mathematical Society*, vol. 49, pp. 122–136, 1941.
- [19] R. N. Bhattacharya and R. R. Rao, *Normal Approximation and Asymptotic Expansions*, SIAM (Society for Industrial and Applied Mathematics), Philadelphia, Pa, USA, 2010.
- [20] I. S. Tyurin, "On the accuracy of the Gaussian approximation," *Doklady Mathematics*, vol. 80, pp. 840–843, 2009.
- [21] B. L. Goldstein, "Bounds on the constant in the mean central limit theorem," *The Annals of Probability*, vol. 38, no. 4, pp. 1672–1689, 2010.