An Alternative Method for Generating Almost Infinite Sequence of Gaussian Variables

Nyah C. Temaneh, F. A. Phiri, E. Ruhunga

Abstract—Most of the well known methods for generating Gaussian variables require at least one standard uniform distributed value, for each Gaussian variable generated. The length of the random number generator therefore, limits the number of independent Gaussian distributed variables that can be generated meanwhile the statistical solution of complex systems requires a large number of random numbers for their statistical analysis. We propose an alternative simple method of generating almost infinite number of Gaussian distributed variables using a limited number of standard uniform distributed random numbers.

Keywords—Gaussian variable, statistical analysis, simulation of Communication Network, Random numbers.

I. INTRODUCTION

The development and design of a simulation tool for the planning and optimization of radio communication systems is a multi-dimensional process due to the large number of different design requirements and systems parameters and the degree of randomness involved. Such planning tools considerably simplify and increase the effectiveness of the designing, planning and optimization of radio communication systems. Amongst the tasks involved are the simulation of the electromagnetic environment and the estimation of its effect on radio communication devices by mathematical modelling methods. Monte Carlo simulation has become one of the most important tools in the field of science and engineering [1] for the statistical analysis of large and complex systems. The Monte-Carlo simulation provides the most accurate predictions in network planning and optimization tools through the utilization of Monte Carlo algorithms, thus providing reliable and credible results [1] –

4. The use of a sum of large number of independent uniform distributed random variables (this sum according to the law of large numbers approximates to a normal distribution).

Most of the well known methods for generating the value of a Gaussian variable described above are different in procedure but have in common the fact that each Gaussian value generated, requires at least one sample from the standard uniform distribution. The length [6] (the number of random samples that can be generated by a random number generator without repetition) of the random number generator therefore, limits the number of independent Gaussian distributed variables that can be generated. At the same time the statistical solution of complex problems e.g. the planning and optimization of a radio communication system like a mobile cellular communication network [2, 4] require a large number of random numbers for their statistical analysis. Therefore the use of the existing methods of generating a normally distributed variable is limited.

In this paper, we propose an alternative simple method of generating two sets of almost infinite number of Gaussian distributed variables $x_k$ and $y_k$ on the bases of the complex...
exponential $\exp(j2\pi k)$; where, $k = 1, 2, \ldots$ and using only a limited number of standard uniform distributed random numbers.

In section II, the proposed algorithm for generating the Gaussian distributed variables and the investigation of the important properties necessary to establish the Gaussian nature of the generated data are presented. The next sections will be dedicated to the investigations of the properties of normality as mentioned in this section. In section III, we investigate the correlation between the two sets of data $x_k$ and $y_k$ and its dependence on the sample size $N$ where we found that the two sets of data are uncorrelated. In section IV, we investigate the behaviour of the two sets of data as the sample size $N$ is increased towards infinity. By constructing the histograms of the data and calculating the mean value, the standard deviation for each set of data and the co-variation matrices were obtained for different sample sizes. In section V, we conclude on the results obtained.

II. PROPOSED ALGORITHM FOR GENERATING GAUSSIAN DISTRIBUTED VARIABLES

Any complex number has a “real” ($x$) and “imaginary” ($y$) part, and is expressed as $Z = x + jy$. The imaginary part is the square root of a negative number, which is really a non-existent number. Complex numbers can be used to define a two dimensional variable. The representation of a complex number requires two orthogonal axes. It can also be visualized as a point on the complex number plane, or as a vector originating at the origin and terminating at the point. The complex exponential $\exp(j2\pi k)$ can be expressed in the form $\exp(j2\pi k) = \cos(2\pi k) + j\sin(2\pi k)$, where, $k = 1, 2, \ldots$.

A sequence of complex random variables $Z_k$, can be formed using the complex exponential $\exp(j2\pi k)$ and a sequence of numbers $X_i$ ($i = 1, 2, 3, \ldots, n$) which are independent and standard uniformly distributed as

$$Z_k = \sqrt{\frac{2}{n}} \sum_{i=1}^{n} \exp(j2\pi kX_i)$$

$$= \sqrt{\frac{2}{n}} \sum_{i=1}^{n} [\cos(2\pi kX_i) + j\sin(2\pi kX_i)]$$

(1)

Therefore, it is possible to assign a value for $n$ (example, between 10 and 50) and then generate a sequence of standard uniform distributed random numbers $X_i$ then, performing calculations according to the algorithm (Fig. 1) based on expression (1), we obtain two random sequences $x_k$ and $y_k$. In Fig.1, the random number generator (RNG) provides the standard uniform distributed variable $X$.

In order to establish that the random sequences $x_k$ and $y_k$ are each normally distributed, we have to investigate the following important properties:
1. That the sequences of random variables $x_k, y_k$ are uncorrelated.

2. That as the length of each of the sequences $x_k$ and $y_k$ turns to infinity, the random variables are independent and each asymptotically normal (central limit theorem).

3. That the mathematical expectations of each of the random sequences $x_k$ and $y_k$ equals zero and the co-variation matrix equals unity.

The next sections will be dedicated to the investigations of the above properties for the normality of $x_k$ and $y_k$.

III. CORRELATION BETWEEN RANDOM SEQUENCES $x_k$ AND $y_k$

The correlation coefficient ($r$) is a statistical calculation that is used to examine the relationship between two sets of data, in this case the sequences $x_k$ and $y_k$. The value of the correlation coefficient will always be between 1 and -1 and it tells us about the strength and the nature of the relationship.

A correlation coefficient that has a value of exactly 1 or -1 would be a perfectly straight line when the data sets are plotted on a graph. Values of $r$ that are close to 0 tell us there is no relationship between the sets of data $x_k$ and $y_k$.

Positive correlation coefficients tell us that when one variable is increased, the other variable will increase as well. Negative correlation coefficients tell us that as one of the variables is increased, the other variable will decrease.

The correlation coefficient $r$ will be calculated according to the expression [7]:

$$ r = \frac{\sum_{k=1}^{N} x_k y_k - \frac{1}{N} \left( \sum_{k=1}^{N} x_k \right) \left( \sum_{k=1}^{N} y_k \right)}{\sqrt{\left( \sum_{k=1}^{N} x_k^2 - \frac{1}{N} \left( \sum_{k=1}^{N} x_k \right)^2 \right) \times \left( \sum_{k=1}^{N} y_k^2 - \frac{1}{N} \left( \sum_{k=1}^{N} y_k \right)^2 \right)}} $$

With the value of $n$ taken as 40, the sequences of data $x_k$ and $y_k$ were each generated according to expression (1) above for values of $k = 1, 2, ..., N$, where $N$ is the length (sample or data size) of each of the sequences and was taken to be between 10000 and 100000 samples. The correlation coefficient $r$ as a function of generated data size $N$ was obtained and graphically (Fig. 2).

IV. THE BEHAVIOUR OF THE RANDOM SEQUENCES $x_k$ AND $y_k$ AS $N \rightarrow \infty$

The random sequences $x_k$ and $y_k$ were generated as before according to the proposed algorithm based on expression (1) for values of data size $N$ between 10000 and 50000 samples. The histogram plot for both generated sequences $x_k$ and $y_k$ for the different values of generated data size $N$ were obtained (Fig. 3).

For each data size, the mean values $(\bar{x}_k, \bar{y}_k)$, standard deviations $(\sigma_{x_k}, \sigma_{y_k})$ and the co-variation matrix $(\text{cov}_x, \text{cov}_y)$ for each sequence were obtained (Table 1-3) and graphically represented (Fig. 4 -6) respectively.
V. DISCUSSION AND ANALYSIS OF RESULTS

From the histogram (Fig. 3), it can be seen that as the sample size $N$ increases, the random sequences $x_k$ and $y_k$ generated by the proposed algorithm approximate to a normal distribution. The results in Table 1, Table 2 and Table 3 show that as the sample size $N$ increases, the mean value of the sequences $x_k$ and $y_k$ is approximately zero, the standard deviation and covariance matrix each approximate to one respectively. Therefore, the algorithm (Fig. 1) based on the expression (1) can be used to generate independent random sequences $x_k$ and $y_k$ each having an approximate standard normal distribution (mean of zero and standard deviation of one) as $k$ increases.

The length of each sequence $N$ is defined as $N = 10^p - 1$, where $p$ is the number of decimal digits of the random number $X$. If $k > 10^p - 1$, the values of $Z_k$ ($x_k$ and $y_k$) starts to repeat i.e. $Z_{k+N} = Z_k \Rightarrow x_{k+N} = x_N$ and $y_{k+N} = y_N$.

Increasing the value of $p$ by one, the value of $N$ increases by $\frac{10^p \cdot 10}{10^p} \approx 10$ times. Therefore the length of the random sequence can be increased considerably by increasing the number of decimal digits of the random number $X$.
VI. CONCLUSION

The implementation of the proposed simple method of generating almost infinite sequence of normally distributed variables and the analysis of the results showed that indeed an almost infinite sequence of random numbers distributed normally can be achieved using only a limited number of standard uniform distributed random numbers.

This method will be of interest to those who for example use simulation for the solution of complex system involving a high degree of randomness like in cellular mobile communication networks.

REFERENCES


[5] ERC Report 68,“Monte-Carlo Simulation Methodology for the use in sharing and compatibility studies between different radio services or systems” Naples, February 2000


Nyah C. Temaneh (M’2009) was born in Cameroon. He received the Master of Science (MSc.) in Engineering from the Faculty of Radio communication, Radio broadcasting and Television at the Ukrainian State Telecommunications Academy – Odessa in 1995 and a Ph.D. degree in Telecommunications from the Yaroslavl State University – Russia in 2007. Presently he is Senior lecturer at the University of Namibia. His research interest is on Telecommunication’s network performance simulation models, for network management, planning, optimisation and analysis including electromagnetic compatibility analysis.