

Interactive MATLAB-Based Demo Program for Sum of Independent Random Variables

GORDANA JOVANOVIĆ DOLECEK

Institute INAOE, Electronics, E. Erro 1 Tonantzintla, Puebla 72840, Mexico

Received 23 June 2009; accepted 5 July 2010

ABSTRACT: We present an interactive MATLAB-based demo program devoted to teaching undergraduate students the behavior of the sum of independent random variables. The user chooses the type of the variables, and the parameters of these variables. The sum is shown in the time domain followed by the estimation of the probability density of the sum using a histogram. It is subsequently shown that the probability density of the sum of random variables represents the convolution of the density functions of the corresponding variables. The Central Limit Theorem and the progression of the sum towards this limit are illustrated in terms of the number of summands. The software can be used as a complement to theoretical classes or alone as a self-study tool. Student evaluation of the program is also included. © 2010 Wiley Periodicals, Inc. *Comput Appl Eng Educ* 21: 372–385, 2013; View this article online at wileyonlinelibrary.com; DOI 10.1002/cae.20481

Keywords: MATLAB demo program; sum of random variables; probability density; convolution; histogram; Central Limit Theorem

INTRODUCTION

Demo programs that illustrate concepts with visual aids have become one of the more important tools in the field of engineering education [1]. They help students learn faster, remember longer, and identify the fundamental concepts better [2]. Demo programs are especially beneficial in terms of saving time and effort for both teaching and learning [3,4]. Our experience shows that the concepts describing probability and random variables seem somewhat inaccessible and abstract for undergraduate students [5–8]. To this end we present a demo program to aid in understanding the behavior of the sum of independent random variables.

Let us first recall the theoretical underpinning of the sum of independent random variables, starting with two independent random variables. Suppose X_1 and X_2 are random variables with the corresponding densities $f_{X_1(x_1)}$ and $f_{X_2(x_2)}$, and their sum $X = X_1 + X_2$.

We find the probability density (PDF) of the sum $f_X(x)$ by introducing the auxiliary variable Y , as shown in Figure 1.

Suppose the variable Y is equal to X_1 . (We could also substitute X_2 .) Therefore we have

$$\begin{aligned} X &= X_1 + X_2 \\ Y &= X_1 \end{aligned} \quad (1)$$

The joint probability density function of the output variables X and Y is related to the joint probability density of the input variables X_1 and X_2 as [9–11]

$$f_{XY}(x, y) = \frac{f_{X_1 X_2}(x_1, x_2)}{|J(x_1, x_2)|} \quad (2)$$

where J is the Jacobean of the transform (1)

$$J = \begin{vmatrix} \frac{\partial x}{\partial x_1} & \frac{\partial x}{\partial x_2} \\ \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} \end{vmatrix} \quad (3)$$

and according to Equation (1)

$$\begin{aligned} x_1 &= y \\ x_2 &= x - y \end{aligned} \quad (4)$$

where x_1, x_2, x , and y are values of the corresponding variables X_1, X_2, X , and Y , respectively.

From (1) and (3) we have $J = 1$, and finally from (2) we have

$$f_{XY}(x, y) = f_{X_1 X_2}(y, x - y) \quad (5)$$

Correspondence to: G. Jovanovic Dolecek
(gordana.dolecek@gmail.com).
© 2010 Wiley Periodicals, Inc.

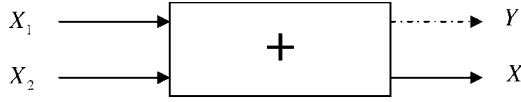


Figure 1 Sum of random variables.

If X_1 and X_2 are independent, it follows

$$f_{X_1 X_2}(z, x - z) = f_{X_1}(z)f_{X_2}(x - z) \quad (6)$$

From (5) and (6) we arrive at [11]

$$f_X(x) = \int_y f_{X_1}(y)f_{X_2}(y - x)dy \quad (7)$$

The right-hand side of Equation (7) is the convolution of the densities f_{X_1} and f_{X_2} . Therefore the probability density function of the sum of the independent variables is the convolution of the corresponding densities. This result can be easily generalized to the sum of m independent variables [9–11].

$$X = \sum_{i=1}^m X_i \quad (8)$$

$$f_X(x) = f_{X_1} * f_{X_2} * \dots * f_{X_m}$$

where $*$ indicates the convolution operator.

The rest of the paper is organized in the following way. Second section describes the motivation followed by a brief description of the content of the demo package. Demonstrations with descriptive details are presented in the third, fourth, fifth, and sixth sections. Last section describes users' evaluation of the programs.

MOTIVATION

Although the field of engineering has changed dramatically in the last years, the study of engineering has changed relatively little [12]. This is particularly true for the teaching of random variables [13]. It is well known that passive learners lose attention quickly in lectures and tutorials and after only 20 min in lecture, students' attention level reduces to one-third of its initial level [12]. "As attention is a prerequisite for learning it is important to capturing and sustaining student's attention in class" [12]. It is believed that the learning can become more efficient when visualization is included in classical lectures, and can thus transform a passive participant in a learning process to an active participant. Moreover, it is sometimes difficult for undergraduate students to understand that the sum of two variables can sometimes be of the same type and sometimes of an entirely different type than the variables that are summed, and that this result depends on the type of the summands. In particular, as the number of terms in the sum increases, the Central Limit Theorem assures that the sum always approaches the same result: the normal random variable [9–11].

Motivation for this work is several-fold. One objective is to raise students' interest through the visualization of the summation of random variables showing the amplitude values of the variables and their sum, as well as the corresponding estimations of density functions. Another objective is to enable students to learn interactively rather than by only passively listening to lectures. Yet another objective is to enhance learning by exposing students to hands-on manipulation of random variables and the corresponding density functions.

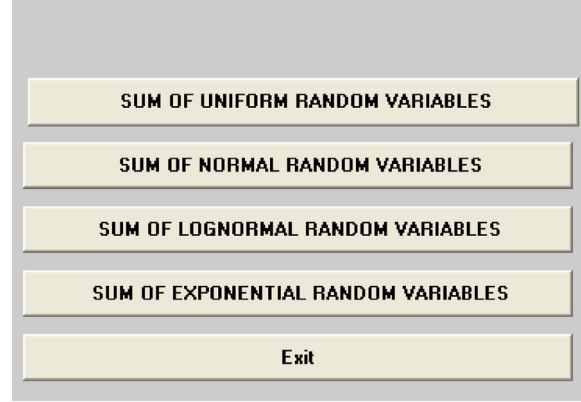


Figure 2 Menu. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

DESCRIPTION OF DEMO PROGRAM

The content of the demo program is given in the Menu shown in Figure 2. The user chooses the type of the variables from the Menu, and the corresponding parameters of the chosen variables. First, the sum of two variables is considered by showing that the probability density function of the resulting random variable is the convolution of the corresponding densities. The sum of m identically distributed random variables is then considered and it is shown how the sum approaches a normal variable, according to the Central Limit Theorem [9,10].

The demo is written in MATLAB. The user need not be a priori proficient in MATLAB or any other programming language, and can thus give all the attention to the presented topic. The demo program can be executed repeatedly with different variables and different parameters.

In the following sections we give short descriptions of the demo program shown in Menu from Figure 2.

SUM OF UNIFORM RANDOM VARIABLES

First, the sum of two uniform random variables X_1 and X_2 in the corresponding intervals $[R_{11}, R_{12}]$, and $[R_{21}, R_{22}]$, respectively, is considered. The user chooses the corresponding values R_{11} , R_{12} , R_{21} , R_{22} . Figure 3 demonstrates the case where $R_{11} = 1$, $R_{12} = 3$, $R_{21} = 4$, and $R_{22} = 5$. Note that the sum of uniform variables is not a uniform variable.

The observation is confirmed in Figure 4 where the probability densities for X_1 and X_2 , as well as for the sum X , are estimated.

The PDF for the uniform random variable X in the range $[A, B]$ is defined as

$$f_X(x) = \begin{cases} \frac{1}{B-A} & \text{for } A \leq x \leq B \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

The convolution of the corresponding uniform densities (9) is shown in Figure 5, thus confirming that the sum of uniform variables is not uniform variable.

Next, we consider the sum of m uniform random variables in the interval $[R_1, R_2]$, where the user chooses the value of m and the values R_1 and R_2 . Consider the zero mean example where $R_1 = -1$ and $R_2 = 1$. The successive convolutions of the

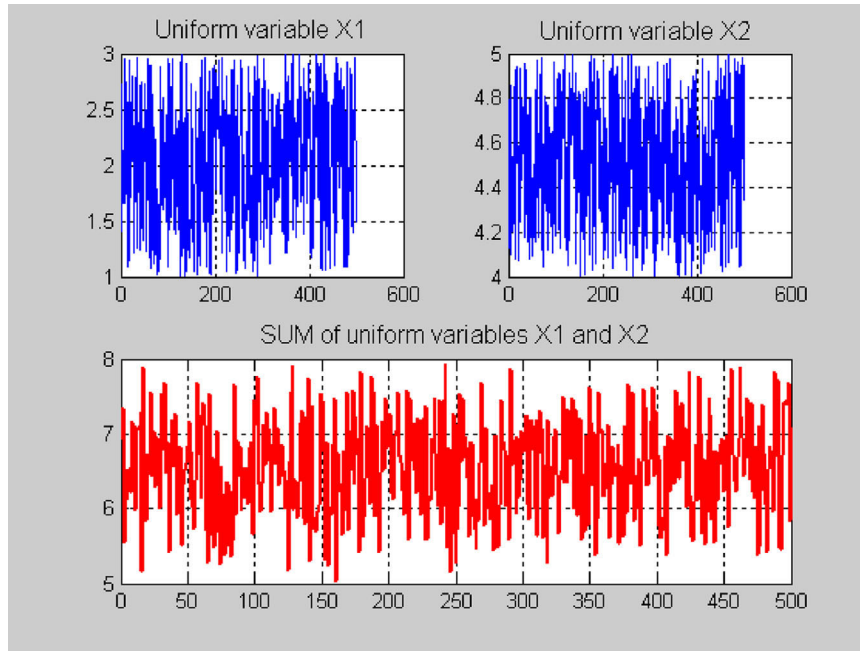


Figure 3 The sum of uniform random variables. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

corresponding densities are shown for different values of m . As an example, Figure 6 shows the densities for $m = 2, 3, 4$, and 5.

Note that the density of the sum of uniform variables approaches the normal density. Figure 7 compares the estimated density of the sum of 10 uniform random variables with the normal density where the mean value and the variance of the normal random variable are equal to the mean value of the sum of 10 uniform

variables, and the variance is equal to the variance of the sum of 10 uniform random variables,

$$\begin{aligned} E\{X\} &= \sum_{m=1}^{10} E\{X_m\} = 0 \\ \text{Var}\{X\} &= 10 \times \frac{2^2}{12} = \frac{10}{3} \end{aligned} \quad (10)$$

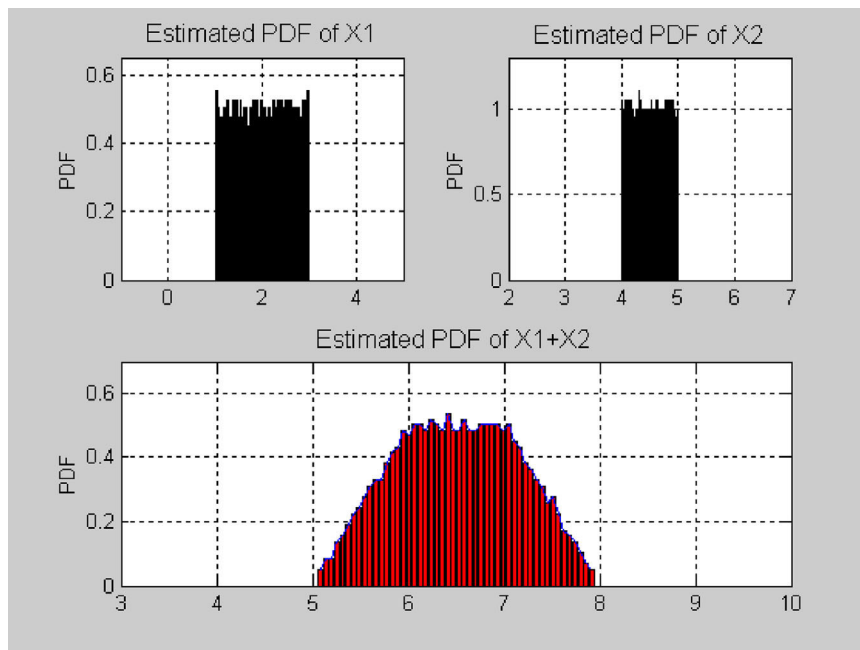


Figure 4 PDF estimates formed by scaled histogram. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

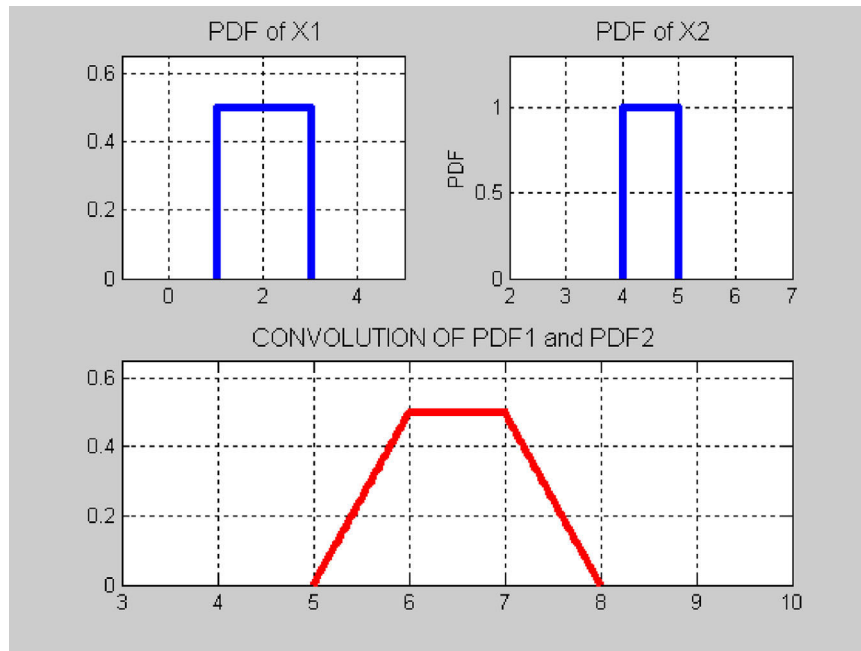


Figure 5 Convolution of densities. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

SUM OF NORMAL RANDOM VARIABLES

In this demo program we demonstrate that the sum of normal random variables is again the normal variable with the mean value and variance equal to the sum of the corresponding mean values m_i and variances σ_i^2 , respectively. First we consider the sum of two

normal random variables as shown in Figure 8. The user chooses the mean value and the variance for each normal random variable. In this example $m_1 = 2$, $m_2 = 4$, $\sigma_1^2 = 4$, and $\sigma_2^2 = 9$.

Note that the sum of the normal variables in Figure 8 is a normal variable.

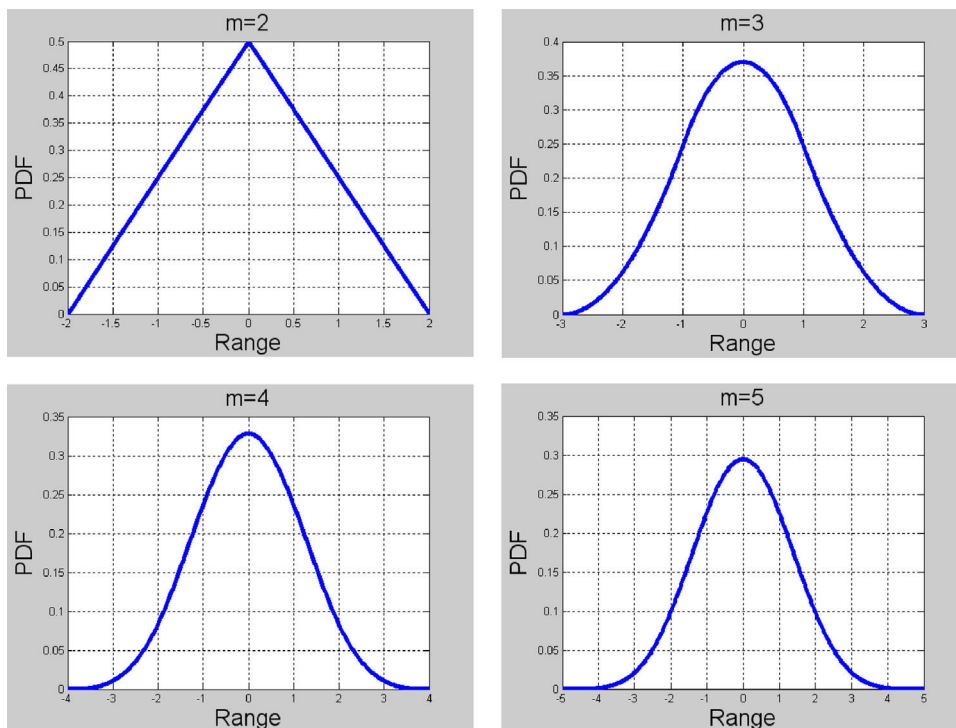


Figure 6 Probability densities for sum of m uniform variables. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

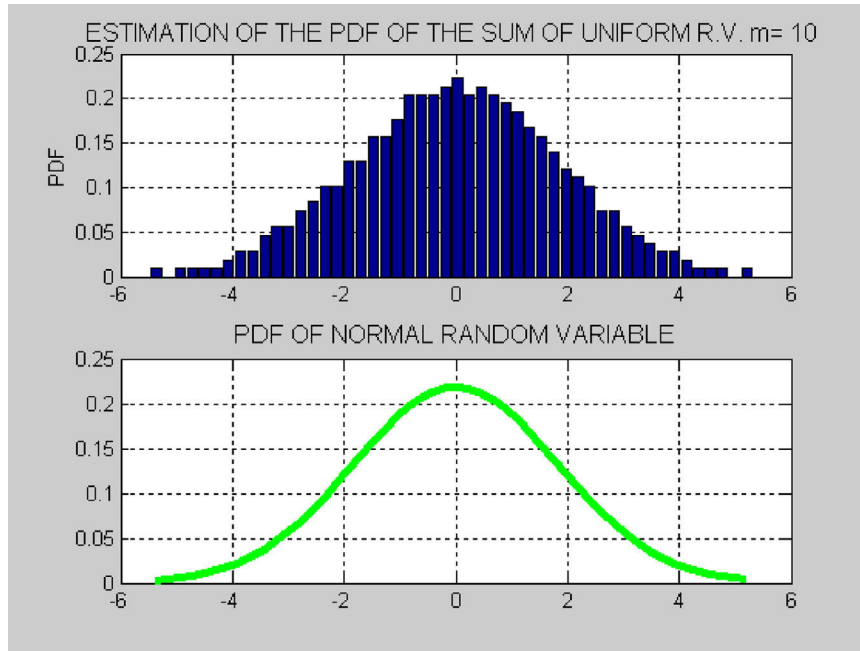


Figure 7 Comparison estimated density with the normal density. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The probability density of the normal variable X with the mean value m and the variance σ^2 is given as

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad -\infty < x < \infty \quad (11)$$

Figure 9 compares the estimation of the density of the sum with the corresponding normal density thus confirming that the

density of the sum of the normal variable is indeed the normal variable. The resulting formula is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-m_1)^2}{2\sigma_1^2}} * \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x-m_2)^2}{2\sigma_2^2}}, \quad (12)$$

$$m = m_1 + m_2; \quad \sigma^2 = \sigma_1^2 + \sigma_2^2$$

where $*$ denotes the convolution.

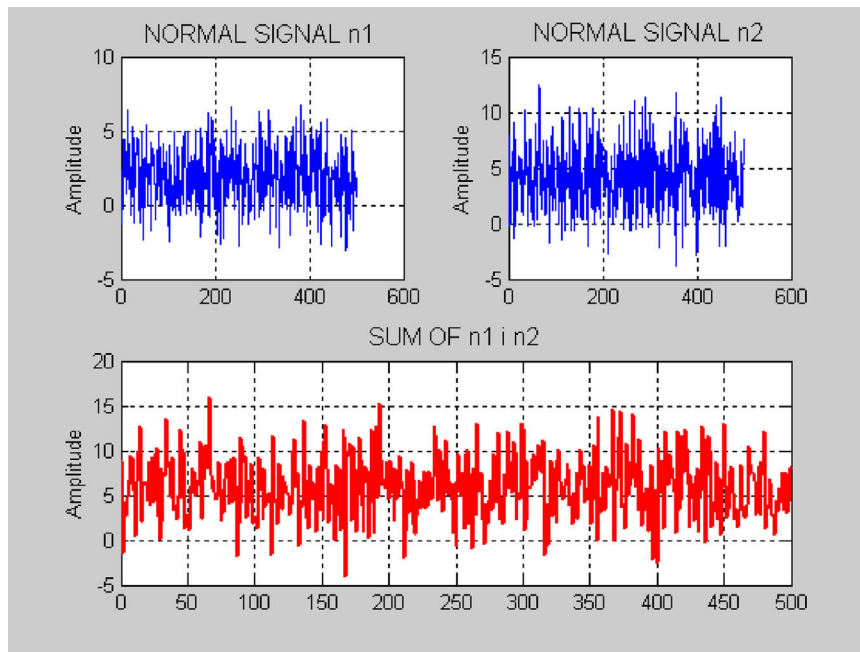


Figure 8 Sum of two normal variables. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

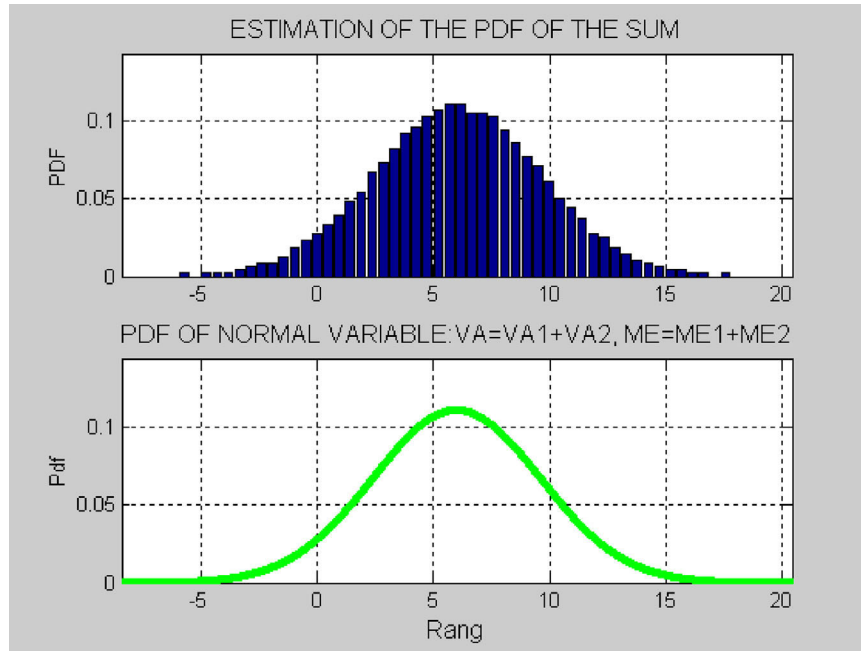


Figure 9 Comparison of the density of the sum with the normal density. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The same observation is demonstrated in Figure 10 where the density of the sum is obtained as the convolution of the corresponding densities.

We next consider the sum of m normal variables. Figure 11 shows the sum of m normal signals with mean value and variance equal to 1 and 4, respectively, for $m = 3, 4, 5$, and 6. It can be noted that the sum of normal variables, irrespective of the number of signals added, is a normal random variable. The same is confirmed in Figure 12 which shows the estimates of the corresponding densities.

SUM OF LOGNORMAL RANDOM VARIABLES

Lognormal random variable Y is obtained by the exponential transform of the normal random variable X with the mean value m_X and the variance σ_X^2

$$Y = e^X \quad (13)$$

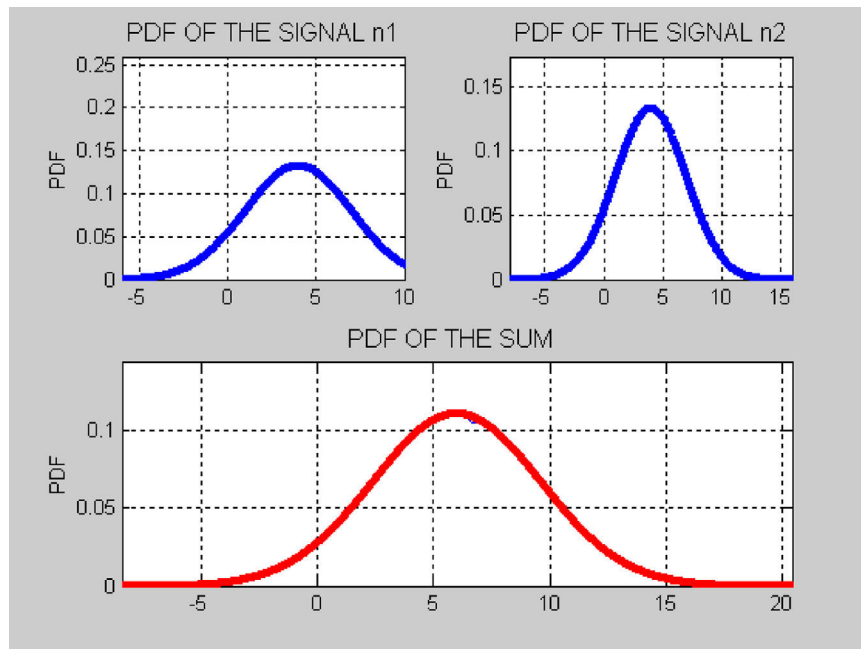


Figure 10 Convolution of the normal densities. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

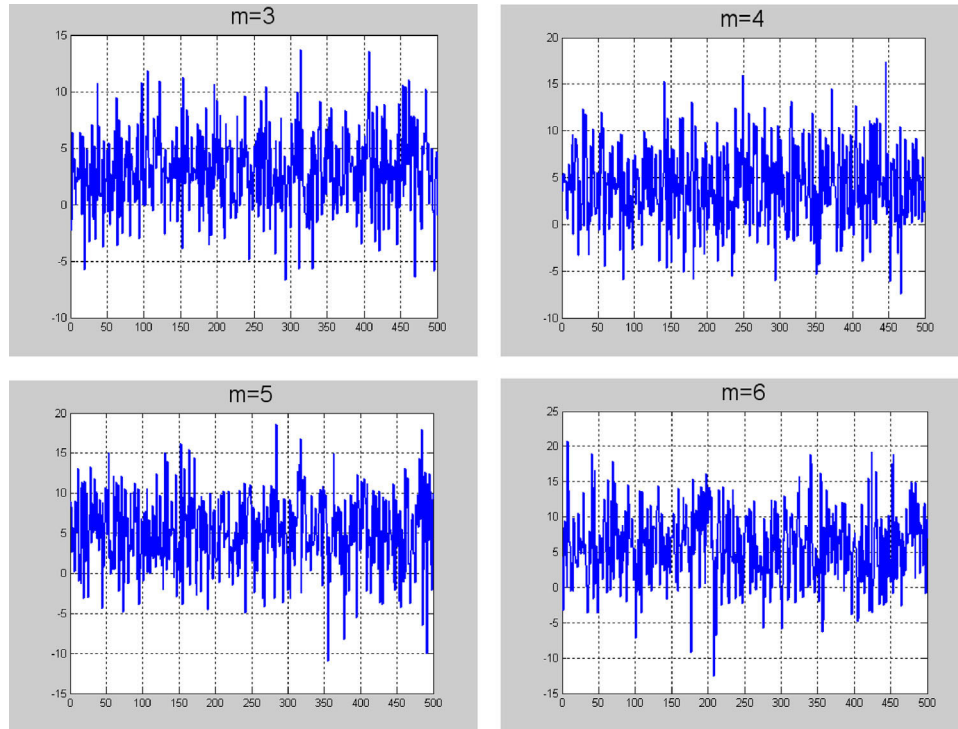


Figure 11 The sum of m normal variables. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The lognormal density is given as

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_X x} e^{-\frac{(\ln x - m_X)^2}{2\sigma_X^2}} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

In this demo is shown that the sum of two independent lognormal random variables is the lognormal random variable. The corresponding lognormal variables are given in Figure 13 for the parameters $m_{X_1} = 2$, $m_{X_2} = 4$, $\sigma_{X_1}^2 = 0.4$, and $\sigma_{X_2}^2 = 0.2$.

The corresponding estimated densities are given in Figure 14.

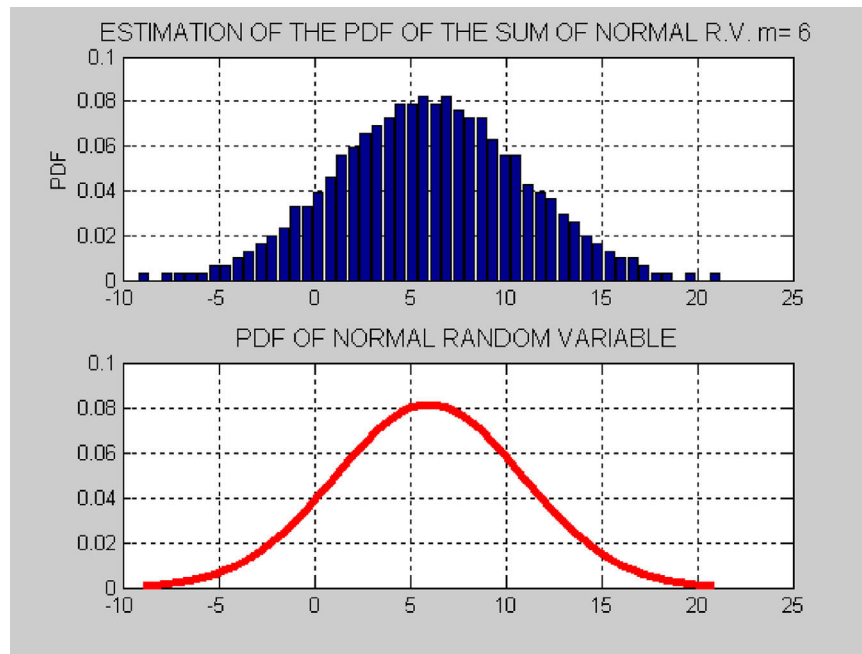


Figure 12 Comparison of the estimated density with that obtained by convolutions of the corresponding normal densities. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

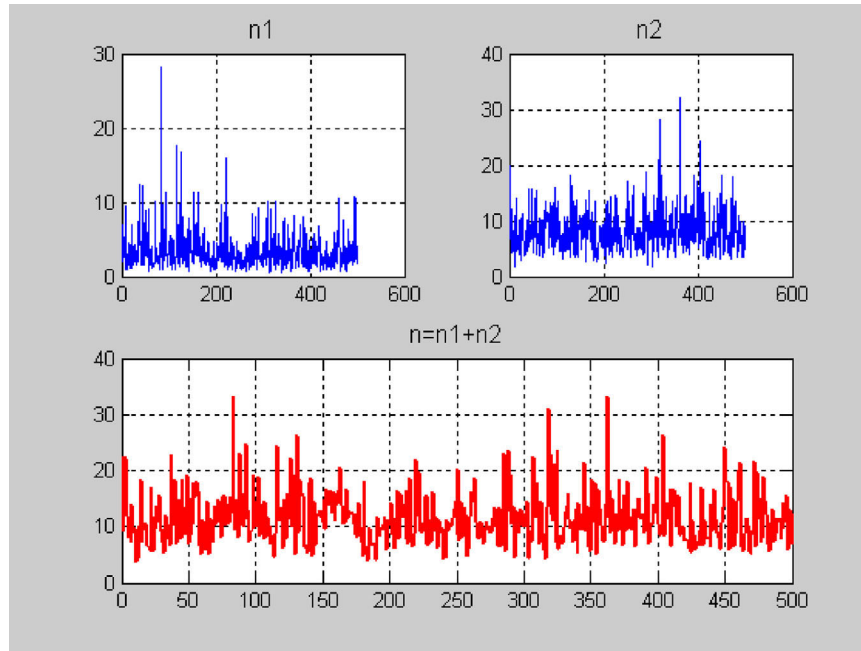


Figure 13 Sum of lognormal variables. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Figure 15 shows the convolution of the corresponding densities.

In the following we consider the sum of lognormal random variables. As an example, Figure 16 shows the sums of lognormal variables for $m=5, 11, 15$, and 20 , where m is the number of variables in the sum.

We can see that the sum of lognormal variables approaches a normal variable, as shown in Figure 17.

SUM OF EXPONENTIAL RANDOM VARIABLES

The exponential probability density function is given as

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases} \quad (15)$$

where λ is the parameter, $\lambda > 0$.

We first consider the sum of two exponential variables. The user chooses two exponential random variables with desired

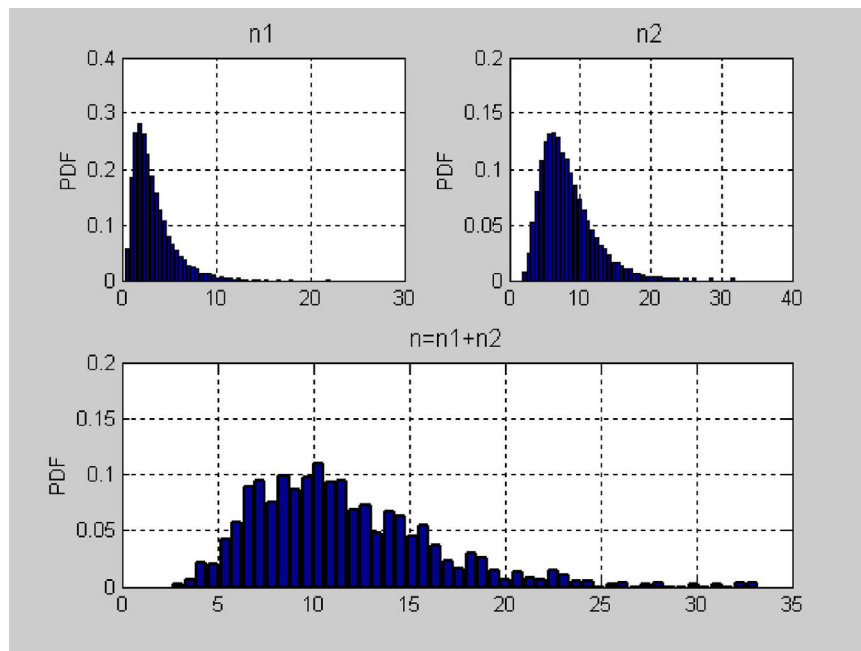


Figure 14 Estimated densities. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

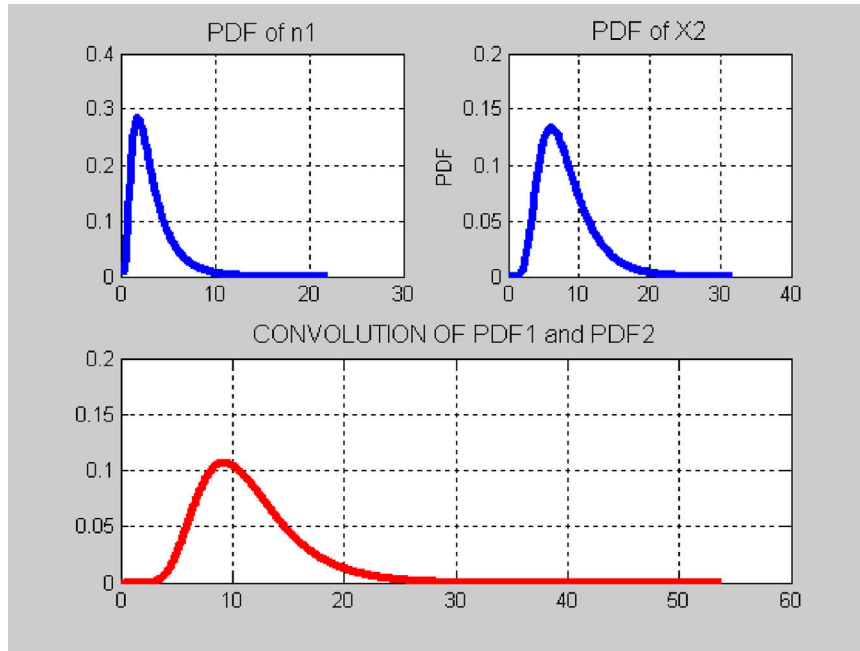


Figure 15 Convolution of the lognormal densities. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

parameters as illustrated in Figures 18 and 19, for $\lambda = 0.6$ and $\lambda = 0.8$, respectively.

The sum of N independent exponential variables with the parameter λ is called the Erlang random variable. The Erlang

density function is given as

$$f_X(x) = \begin{cases} \frac{(\lambda x)^{N-1}}{(N-1)!} \lambda e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

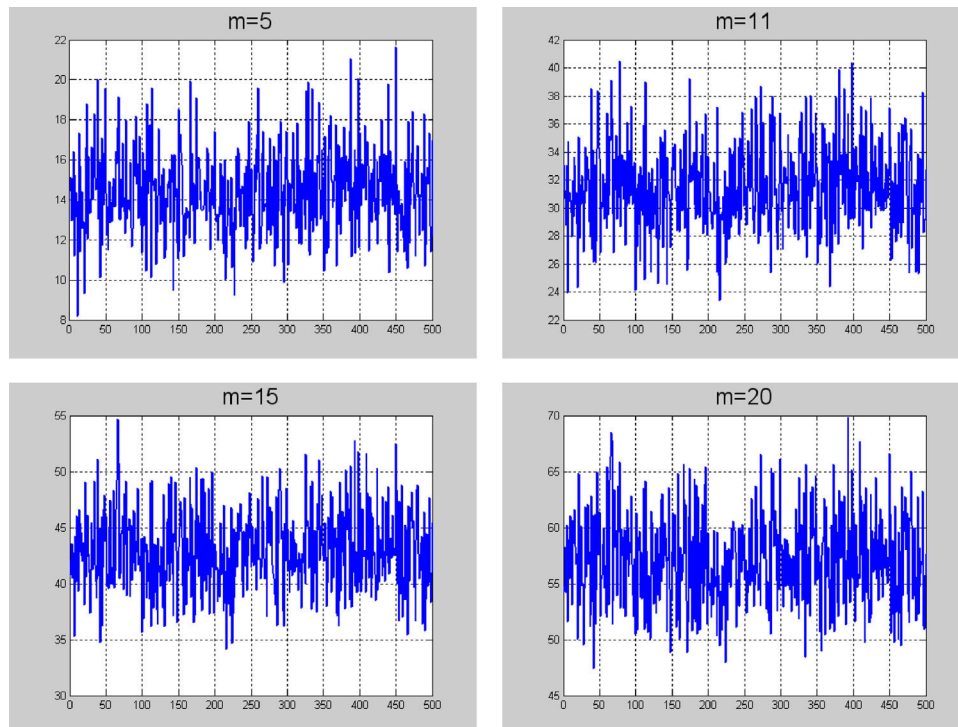


Figure 16 Sum of lognormal variables. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

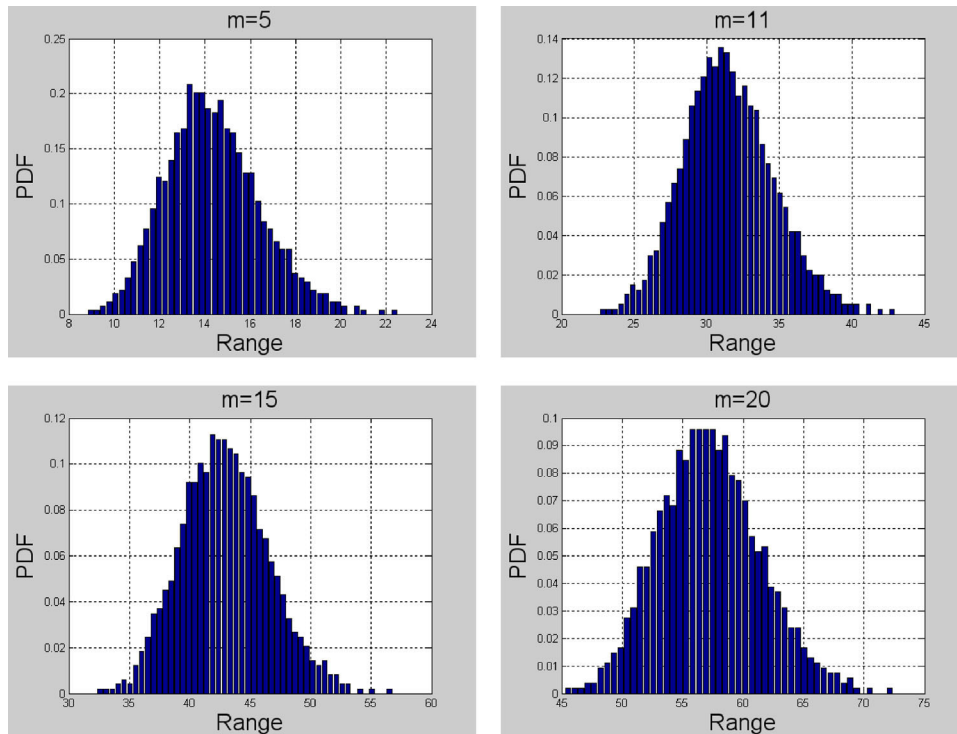


Figure 17 PDF estimations of the sum of lognormal variables. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Figure 20 shows the sum of $N=4$ exponential variables, each with parameter $\lambda=0.9$, that is Erlang random variable with parameters $N=4$, and $\lambda=0.9$. Figure 21 presents the Erlang density along with its estimation. The Erlang density approaches a normal density for large values of the parameter

N , which is in accordance with the Central Limit Theorem. Figure 22 illustrates the Erlang variable for four different values of N : 5, 8, 15, and 20, and confirms that by increasing N , the Erlang random variable is approaching the normal random variable.

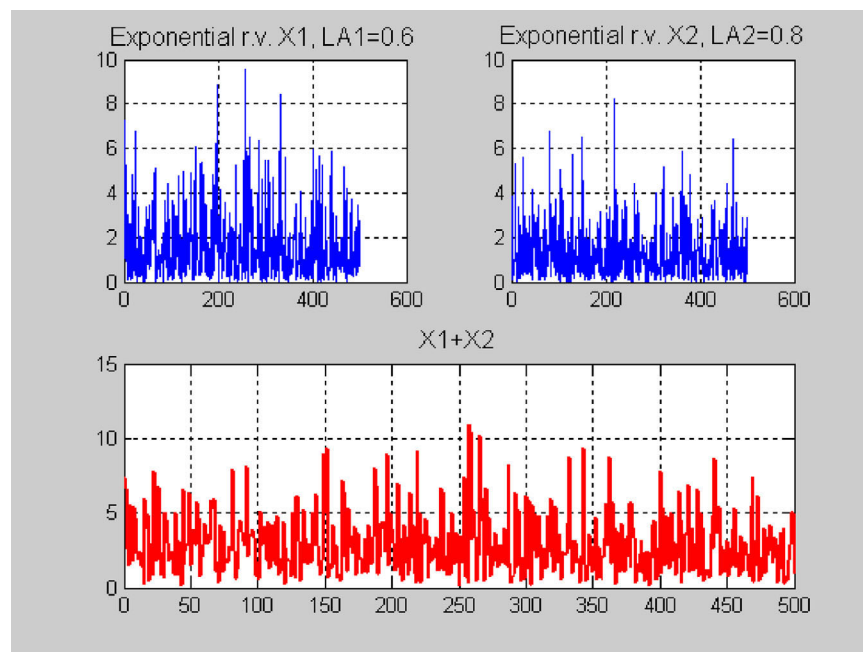


Figure 18 Sum of exponential random variables. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

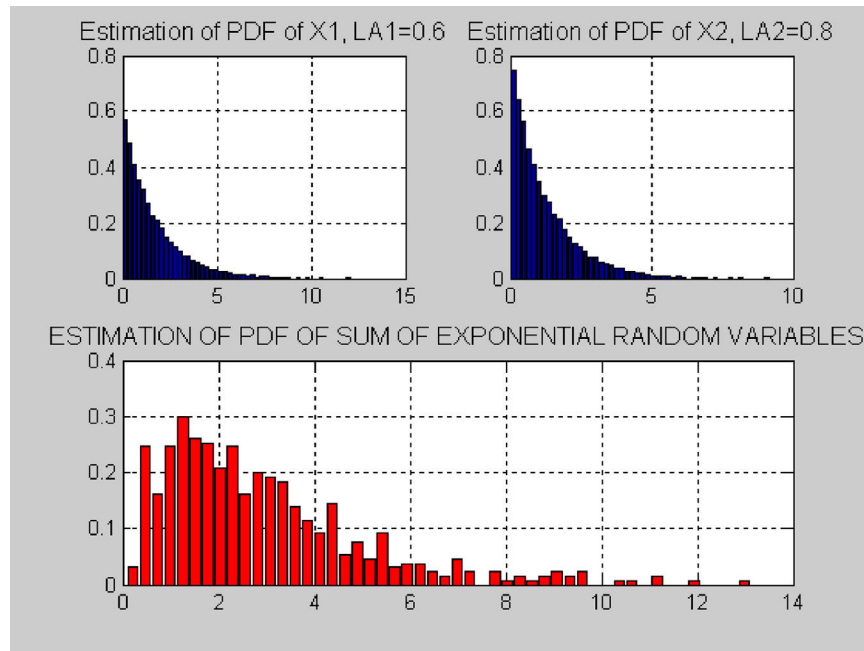


Figure 19 Estimations of densities. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

EVALUATION

Since it is very important to receive information from students about the usefulness of the software in the teaching–learning process, we conducted a survey to evaluate the quality of the demo in the teaching–learning process. The following topics were covered in the survey:

- (A) Usefulness of the software in teaching the sum of random variables.
- (B) Program design aspects.
- (C) Suggestions to improve the demo.

All questions in the forms A and B are rated with marks from 1 to 4 with the latter being the highest mark.

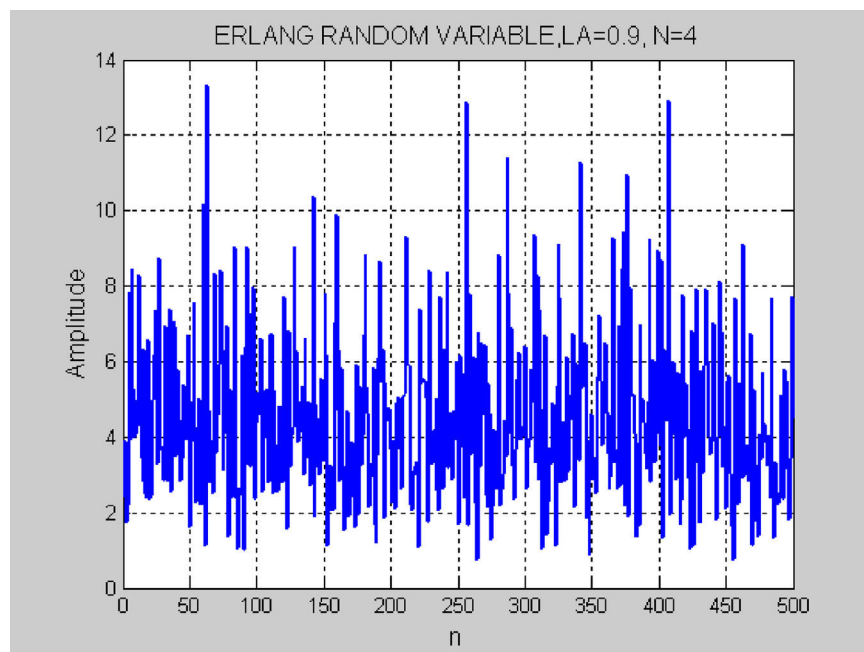


Figure 20 Erlang random variable. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

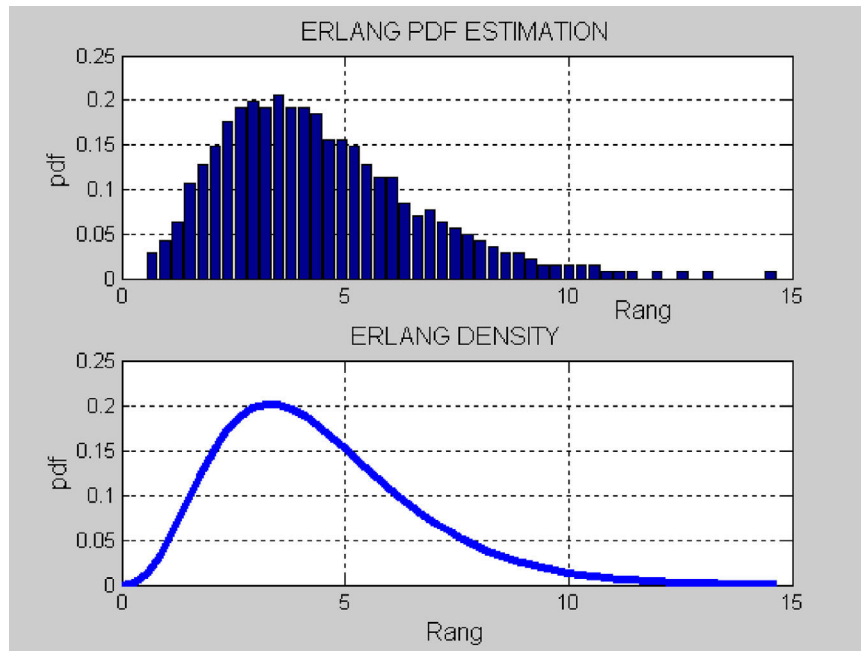


Figure 21 Erlang density. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Evaluation forms:

(A) Usefulness of the software in teaching the sum of random variables.

- (1) Justification for the computer use in teaching random variables. (1 = unjustified; 4 = absolutely justified).
- (2) Contribution to study of the sum of random variables by the demo program usage. (1 = irrelevant; 4 = very effective).

- (3) Clarity of explanations and features of demo. (1 = confusing; 4 = absolutely clear).
- (4) Did this demo help you to better understand the sum of random variables? (1 = NO; 4 = Absolutely YES).
- (5) Did this demo help you to better understand the Central Limit Theorem? (1 = NO; 4 = Absolutely YES).
- (6) Did this demo help you to better understand the sum of uniform random variables? (1 = NO; 4 = Absolutely YES).

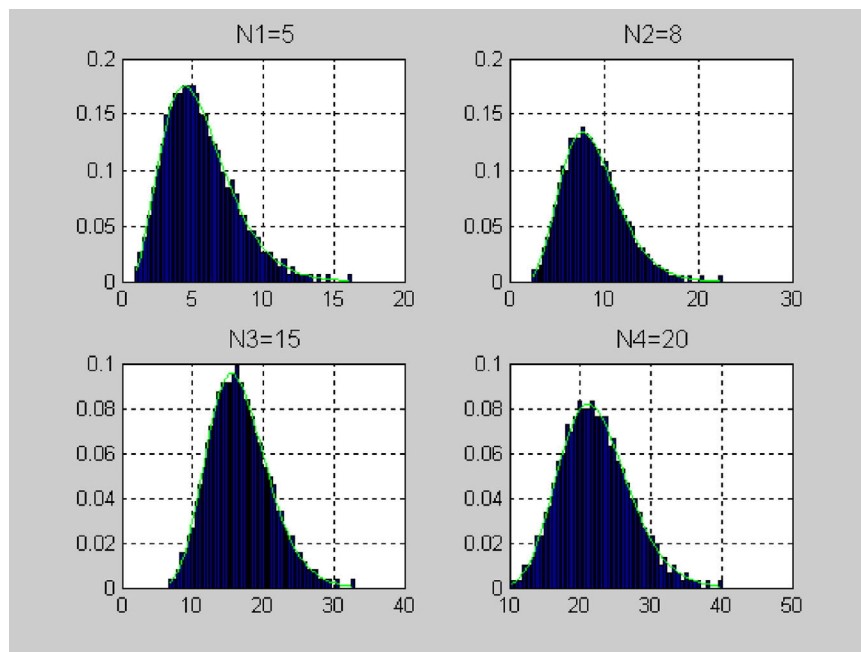


Figure 22 Erlang density for different values of the parameter N . [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

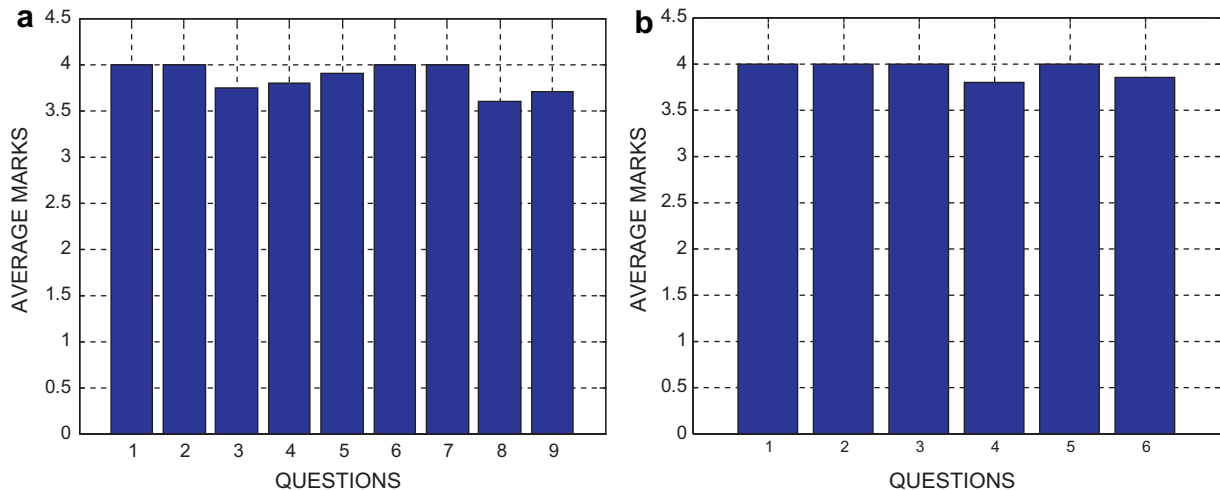


Figure 23 Rating scheme. (a) Usefulness of the software for teaching sum of random variables. (b) Program design aspects. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

- (7) Did this demo help you to better understand the sum of normal random variables? (1 = NO; 4 = Absolutely YES).
- (8) Did this demo help you to better understand the sum of lognormal random variables? (1 = NO; 4 = Absolutely YES).
- (9) Did this demo help you to better understand the sum of exponential random variables? (1 = NO; 4 = Absolutely YES).

(B) Program design aspects

- (1) Special knowledge or programming skills required. (1 = excessive; 4 = none).
- (2) Ease of operation. (1 = complex; 4 = very easy).
- (3) Flexibility and repeatability (possibility to return to previous slide/slides and repeat it/them many times). (1 = unnecessary; 4 = very useful).
- (4) General quality of presentation (figures, resolution, visibility, etc.). (1 = poor; 4 = excellent).
- (5) Ease of operation. (1 = complex; 4 = very easy).
- (6) General quality of presentation (figures, resolution, visibility, etc.). (1 = poor; 4 = excellent).

(C) Suggestions to improve demo

- (1) Please give your suggestions on how to improve efficiency of the software in teaching the sum of random variables.
- (2) Please give your suggestions on how to improve the clarity of presentations and features of the demo.
- (3) Please give your suggestions on how to improve the program design aspects.

The demo program was used as a complementary tool in a basic undergraduate course covering random variables and processes, for electrical engineering majors. The questionnaire took place in the classroom at the end of the course. In total, 65 students participated in the evaluations. Figure 23 presents the result of the evaluation of A and B questionnaires in terms of the average marks for all questions.

The result of the demo evaluation shows that students enjoyed this way of teaching random variables. (The average mark was

3.86.) More specifically, they found this demo to be very useful for better understanding the sum of random variables, Central Limit Theorem, and particularly the sum of uniform, normal, lognormal, and exponential random variables. They especially highly rated the following features of the program: ease of operation, and the fact that no prior programming skills are required to run the program. (The average mark was 3.94.)

Regarding the suggestions to improve the demo, the majority of students proposed to include the analysis of the sum of discrete random variables. They also proposed to improve the readability of formulas and to provide additional explanations in the bottom windows. We plan to include all those recommendations in the next version of the demo.

CONCLUSION

The manuscript's contribution is in the domain of a classroom delivery of learning technology. We presented a demonstration program that permits a student to examine and manipulate time signals, sample densities, and model densities associated with some common random variables. The intent of the program is to invite the students to play with the parameters defining the densities and thereby to obtain a better understanding of the transformations of these time series and density functions formed by simple summation of independent random variables.

REFERENCES

- [1] T. T. Fu, Applications of computer simulation in mechanism teaching, *Comput Appl Eng Educ* 3 (2003), 156–165.
- [2] S. G. Li and Q. Lie, Interactive groundwater (IGW): An innovative digital laboratory for groundwater education and research, *Comput Appl Eng Educ* 4 (2003), 179–202.
- [3] G. Jovanovic-Dolecek, RANDEMO: Educational software for random signal analysis, *Comput Appl Eng Educ* 2 (1997), 93–99.
- [4] G. Jovanovic-Dolecek and V. Champac, CGTDEMO—Educational software for the Central Limit Theorem, *ACM SIGCSE Bull* 2 (2000), 46–48.

- [5] P. Karagiannis, I. Markelis, K. Paparrizos, N. Samaras, and A. Sifaleras, E-learning technologies: Employing MATLAB web server to facilitate the education of mathematical programming, *Int J Math Educ Sci Technol* 37 (2006), 765–782.
- [6] V. Lazaridis, K. Paparrizos, N. Samaras, and A. Sifaleras, Visual LinProg: A web-based educational software for linear programming, *Comput Appl Eng Educ* 15 (2007), 1–14.
- [7] G. Jovanovic Dolecek and A. Fernandez Vazquez, Use of MATLAB in teaching the fundamentals of random variables, *ACM SIGCSE Bull* 40 (2008), 46–51.
- [8] G. Jovanovic Dolecek and F. Harris, On MATLAB demonstrations of narrowband Gaussian noise, *Comput Appl Eng Educ* (published online DOI: 10.1002/cae.20340).
- [9] J. Gubner, *Probability and random processes for electrical and computer engineers*, Cambridge University Press, Cambridge, UK, 2006.
- [10] V. Krishnan, *Probability and random processes*, Wiley-Interscience, Hoboken, NJ, 2006.
- [11] A. Leon-Garcia, *Probability and random processes for electrical engineering*, 3rd edition, Prentice Hall, New Jersey, 2008.
- [12] G. S. Ng, Teaching effectively with visual effects in an image-processing class, *Comput Appl Eng Educ* 2 (1997), 111–114.
- [13] R. Quere, M. Lalande, J. N. Boutin, and C. Valente, An automatic characterization of Gaussian noise source for undergraduate electronics laboratory, *IEEE Trans Educ* 38 (1995), 126–130.

BIOGRAPHY



Gordana Jovanovic Dolecek received a BS degree from the Department of Electrical Engineering, University of Sarajevo, an MSc degree from University of Belgrade, and a PhD degree from the Faculty of Electrical Engineering, University of Sarajevo. She was professor at the Faculty of Electrical Engineering, University of Sarajevo until 1993. In 1995 she joined Institute INAOE, Department for Electronics, Puebla, Mexico, where she works as a full professor. During 2001–2002 and 2006, she was with Department of Electrical & Computer Engineering, University of California, Santa Barbara, as visiting researcher. She was with San Diego State University as visiting researcher on a sabbatical leave during 2008–2009. She is the author of three books, editor of one book, and author of more than 300 papers. Her research interests include digital signal processing, digital communications and educational methods in electrical engineering. She is a senior member of IEEE, a member of Mexican Academy of Sciences, and a member of National Researcher System (SNI) Mexico.