

Simulating Non-Realizable Processes

José Cândido Silveira Santos Filho, *Student Member, IEEE*, Michel Daoud Yacoub, and Gustavo Fraidenraich, *Student Member, IEEE*

Abstract— Consider an output process defined as the sum of M ergodic identically distributed input random processes. Moreover, assume that some output statistics of interest can be found as explicit functions of M with no constraints for M real. Then, extending the investigation of those statistics for non-integer values of M may be desired, mainly if the output process is in effect an idealized approximate model of a realistic random phenomenon. Within this context and for simulation purposes, this work presents a generic, efficient method for generating a sample sequence matching the output statistics for a given non-integer M . To the best of the authors' knowledge, no solutions to the problem addressed here exist yet.

Index Terms— Simulation, random processes, statistics for communications.

I. INTRODUCTION

IN several areas of knowledge, such as engineering, physics, mathematics, and economics, some processes frequently arise as a sum of M ergodic identically distributed input random processes. These output processes are usually investigated in terms of certain statistical functions of interest, such as the probability density function (PDF) and the autocorrelation function (ACF). Sometimes, these metrics can be written as explicit functions of M and impose no mathematical constraints for M real, although, by construction, M is discrete (integer). Under these circumstances, it may be desired to extend the use of the output statistical functions for non-integer values of M . In effect, the output process is often an idealized approximate model for a realistic stochastic phenomenon. Hence, allowing for M non-integer in the resulting statistics is an algebraic artifice which reduces the limitations of the model in describing the true process.

Various examples of the above statistical scenario are reported in the literature. For instance, in wireless communications, the fading phenomena affecting the transmitted signal are commonly modeled as the summation of multipath signal clusters, with non-integer numbers of clusters having been found in practice [1], [2] (and the references therein). As the theoretical counterpart of these practical verifications, some fading models incorporate the number of clusters as a generalized real-valued distribution parameter [3]. Similarly, in the study of high-frequency ultrasonic backscattering from in vivo human skin, the signal received by an ultrasound transducer can be modeled as a phasor sum of the returns from several scatterers [4]. This model is presented in [4] as an approximate description for the realistic scattering occurring

The authors are with the Department of Communications, School of Electrical and Computer Engineering, University of Campinas (DECOM-FEEC-UNICAMP), PO Box 6101, 13083-852 Campinas, SP, Brazil (Tel.: +55 (19) 3788 5106, Fax: +55 (19) 3289 1395, E-mail:[candido,michel,gf]@decom.fee.unicamp.br).

continuously throughout the surface and not at a discrete number of scatterers.

As illustrated above, there is often a need for using the output statistical expressions to establish the description of a *generalized* output process with non-integer M . Of course, such an output process is *non-realizable* via the original output definition as the summation of M input processes, which is meaningless for M non-integer. Indeed, the generalized output process lacks a realization model, although certain statistical functions of interest are specified. Thus, simulating such a *non-realizable process*, i.e. generating a sample sequence whose statistics match the output statistical functions for a given non-integer M , poses a challenging task.

In this work, we present a generic method to simulate the above-defined class of non-realizable processes. To this end, a novel simulation paradigm is introduced. The optimal solution is envisaged, but it presents a prohibitive computational cost. Alternatively, a simple, effective suboptimal solution is proposed. The corresponding results proved very satisfactory in all of the cases investigated by the authors, as shall be seen from an application example. To the best of the authors' knowledge, no solutions to the problem addressed here have been published in the open literature.

II. THE PROBLEM

In the following, a discrete-time simulation problem is formulated and solved. Notwithstanding, because the simulation of continuous-time processes is also attained in a discrete-time fashion, the solution derived here applies to both types of processes (discrete- and continuous-time).

Consider M ergodic identically distributed discrete-time input random processes $X_i[n]$, $i = 1, \dots, M$, and an output process $Y[n]$ defined as

$$Y[n] = \sum_{i=1}^M X_i[n] \quad (1)$$

Let $x_i[n]$ and $y[n]$ denote sample sequences of $X_i[n]$ and $Y[n]$, respectively. Assume that, by knowing the joint statistics of the input processes, certain output statistics of interest can be found as explicit functions of M . For instance, let the PDF $f_Y(y; M)$ of $Y[n]$ and its ACF¹ $R_Y(k; M) \triangleq E\{Y[n+k]Y[n]\}$ be the referred output statistics ($E\{\cdot\}$ denotes expectation). In addition, assume that $f_Y(y; M)$ and $R_Y(k; M)$ impose no mathematical constraints for M real.

As previously discussed, for non-integer values of M , it may be desired to use $f_Y(y; M)$ and $R_Y(k; M)$ to establish the statistics of a generalized output process. Such a

¹The ACF has been defined here according to [5].

generalization enhances the output model if, for instance, it constitutes an idealized mathematical description which in effect only approximates the behavior of a realistic random phenomenon. Nevertheless, for M non-integer, the original output formulation (1) is meaningless. In other words, the generalized output process introduced here lacks a realization model, although certain statistical functions of interest are specified. For this reason, such a process shall be called *non-realizable*.

Within this context and for simulation purposes, consider the following problem:

How to simulate the above-defined class of non-realizable processes, i.e. how to generate an N -sample output sequence $y[n]$ matching $f_Y(y; M)$ and $R_Y(k; M)$ for a given non-integer M ?

III. THE APPROACH

In this section, we introduce a novel simulation paradigm in order to solve the above problem. First, note that, for M integer, the output sample sequence can be obtained directly via (1), i.e., the M input sequences $x_i[n]$ can be generated and then summed to yield $y[n]$. This is a standard simulation approach, which uses the realization model of the process to generate simultaneously the correct samples in the correct order. However, for M non-integer, (1) makes no sense and a realization model no more exists. Accordingly, the standard approach does not apply. To circumvent this impasse, we shall decompose the simulation task into two subtasks:

- 1) To obtain N appropriate samples in order to compose the sequence
- 2) To arrange (order) those samples appropriately

This decomposition establishes a novel simulation approach, which shall be used next in order to solve the problem stated in section II.

IV. THE OPTIMAL SOLUTION

Implementing the proposed approach requires solving the two subtasks above. In the following, we address each of them.

A. The First Subtask (Static Behavior)

Note that the first subtask concerns the histogram (static behavior) of $y[n]$, which, from ergodicity, must approach $f_Y(y; M)$ for N sufficiently large. Thus, for all practical purposes (large N), drawing N samples according to $f_Y(y; M)$ provides the appropriate elements to compose $y[n]$. This can be accomplished by well-established methods of generating random variables with a specified PDF, such as the percentile transformation method or the rejection method [5].

B. The Second Subtask (Dynamic Behavior)

The second subtask is rather complex. It concerns the dynamics of the process, characterized here by the ACF. Note that rearranging the N samples obtained in the first subtask does not alter the realized PDF but the realized ACF. Unfortunately, to the best of the authors' knowledge, there are no analytical tools to find the appropriate arrangement

in a direct manner. Indeed, this is a quite new problem which has not been previously addressed in the literature. Accordingly, in order to find which arrangement provides the best fit to $R_Y(k; M)$, all the $N!$ possible arrangements should be investigated and compared under some deviation measure. This exhaustive search provides the optimal solution to the problem but, for practical purposes (large N), it would require a prohibitive computational effort. For instance, $N = 10^4$ leads to more than 2.8×10^{35659} sample arrangements. For $N = 10^5$, this number exceeds $2.8 \times 10^{456573}!$

V. A SIMPLE SUBOPTIMAL SOLUTION

In view of the prohibitive computational cost of the optimal solution, we present in this section an alternative, simple suboptimal method for solving the second subtask.

First, note that, given a set of N samples, a certain arrangement of them can be specified by a N -length vector whose i th element gives the statistical rank² of the i th element in the arrangement. In the following, this vector is called *statistical rank vector*. For instance, with $N = 4$, the statistical rank vector $V = [3, 2, 4, 1]$ (rank in ascending order), applied on the sample set $S = \{5.2, 7.3, 0.5, 3.7\}$, leads to the arrangement $A = [5.2, 3.7, 7.3, 0.5]$. Let us introduce the notation $A = \mathcal{R}(S; V)$, denoting A as the result of a rank operator $\mathcal{R}(\cdot; \cdot)$ applied on S given the rank information of V .

From the above discussion, the second subtask can be thought of as finding a suitable N -length statistical rank vector which establishes the appropriate arrangement for the N samples generated in the first subtask. In principle, as said before, an exhaustive search over all the $N!$ possible statistical rank vectors would be required to determine the optimal one.

Alternatively, as a suboptimal solution, we propose to arrange the N output samples according to the statistical rank vector $v_i[n]$ of a N -sample sequence $x_i[n]$ of any one of the input processes. This proposal finds an heuristical basis on the symmetry (*identically distributed* inputs) and linearity (*summation* of inputs) of the original output model (1). We reinforce that, although the proposed suboptimal solution has no exact analytical basis, it provides very good results, as shall be seen from an application example. Moreover, the complexity of this alternative approach is virtually negligible as compared to that of the optimal solution given in the last section. The cost now is to order a sample set according to a rank vector, which is indeed very simple!

In short, the following suboptimal method is proposed in order to generate an N -sample output sequence $y[n]$ matching $f_Y(y; M)$ and $R_Y(k; M)$ for a given non-integer M :

- Draw N samples of a random variable following $f_Y(y; M)$, obtaining a sample set $\tilde{y}[n]$
- Generate an N -sample sequence $x_i[n]$ for any one of the input processes
- Arrange the samples of $\tilde{y}[n]$ according to the statistical rank vector $v_i[n]$ of $x_i[n]$, obtaining $y[n] = \mathcal{R}(\tilde{y}[n]; v_i[n])$

²The statistical rank is the ordinal number of a value in a list arranged in a specified order.

It is worthy noting that, although in section II the ACF has been adopted as the dynamic metric of interest, the simulation method derived here clearly applies to any other high-order statistical metric or even to a set of high-order statistics.

VI. AN APPLICATION EXAMPLE

In this section, in order to illustrate the use and the efficiency of the simulation method proposed in section V, an application example is presented. Consider M independent zero-mean white Gaussian processes each of which filtered through a second-order lowpass digital Butterworth filter with cutoff frequency 0.1 and transfer function $H(z)$ [5]

$$H(z) = \frac{0.0201 + 0.0402z^{-1} + 0.0201z^{-2}}{1.0000 - 1.5610z^{-1} + 0.6414z^{-2}} \quad (2)$$

Let $U_i[n]$, $i = 1, 2, \dots, M$, be the outputs of the filters. Define $X_i[n] = U_i^2[n]$ as the input processes in (1). The resultant system output $Y[n]$ can be shown to be gamma distributed

$$p_Y(y; M) = \frac{y^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} \exp(-y/\beta) \quad (3)$$

where $\alpha = M/2$ and $\beta = 2E\{X_i^2\}$ are the shape and scale parameters, respectively. The ACF $R_Y(k; M)$ of $Y[n]$ can be obtained as

$$R_Y(k; M) = MR_{X_i}(k) + M(M-1)R_{U_i}^2(0) \quad (4)$$

where $R_{U_i}(k)$ and $R_{X_i}(k)$ are the autocorrelation functions of $U_i[n]$ and $X_i[n]$, respectively. Since the random variables $U_i[n+k]$ and $U_i[n]$ are jointly normal [5], $R_{X_i}(k)$ is found as [5]

$$R_{X_i}(k) = R_{U_i}^2(0) + 2R_{U_i}^2(k) \quad (5)$$

Replacing (5) into (4) yields

$$R_Y(k; M) = M(2R_{U_i}^2(k) + MR_{U_i}^2(0)) \quad (6)$$

where $R_{U_i}(k)$ is given by [5]

$$R_{U_i}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 e^{jk\omega} d\omega \quad (7)$$

with $H(\cdot)$ as in (2). (Evaluating (7) requires numerical integration.) Note that output statistics (3) and (6) are explicit functions of M and no mathematical constraints exist for allowing for non-integer M .

Now, following the steps summarized at the end of section V, $N = 10^6$ samples according to (3) have been generated for some arbitrarily chosen non-integer values of M , obtaining the sample sets $\tilde{y}[n]$. An N -sample sequence $x_i[n]$ of any one of the input processes has been also generated, by filtering a Gaussian sequence through $H(z)$. Finally, the statistical rank vector $v_i[n]$ of $x_i[n]$ has been calculated and then used to order the initial output sample set $\tilde{y}[n]$, yielding $y[n] = \mathcal{R}(\tilde{y}[n]; v_i[n])$.

Fig. 1 shows the autocorrelation of $y[n]$ against the desired exact autocorrelation function given by (6). (The discrete points have been joined for visualization purposes.) The differences are seen to be minimal. As M increases, the fitting becomes slightly poorer for low k . A substantial number of other application examples (not shown here) have been investigated by the authors of this work. The corresponding results proved indeed very satisfactory.

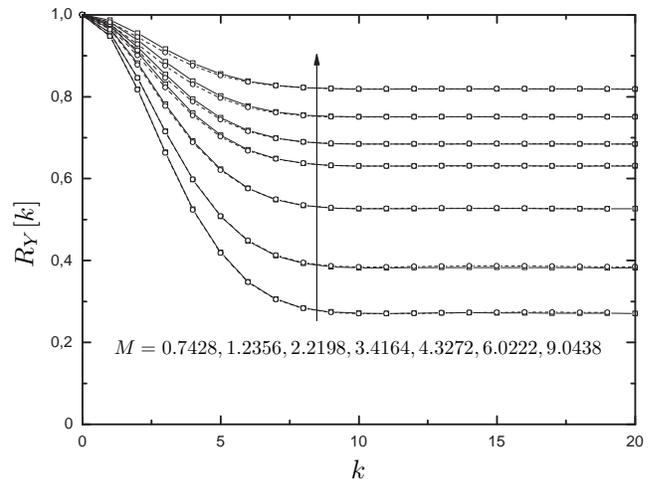


Fig. 1. Autocorrelation function of $Y[n]$ (exact: solid; simulated: dashed).

VII. CONCLUSIONS

This work addressed a class of output processes originally defined as the sum of M ergodic identically distributed input random processes. In particular, certain output statistics of interest were assumed to be found as explicit functions of M with no constraints for M real. Then, allowing for M non-integer in the resulting statistics is mainly important if the output process under study is an idealized approximate description for a realistic stochastic phenomenon. Within this context and for simulation purposes, this work presented a generic framework to generate a sample sequence whose realized statistics match the output statistical functions for a given non-integer M . To this end, a novel simulation paradigm was introduced. The optimal solution was envisaged, but it presents a prohibitive computational cost. Alternatively, a simple, effective suboptimal solution was then proposed. The corresponding results proved very satisfactory in all of the cases investigated by the authors. To the best of the authors' knowledge, no solutions—optimal or suboptimal—to the problem addressed here have been published in the literature.

REFERENCES

- [1] H. Asplund, A. F. Molisch, M. Steinbauer, and N. B. Mehta, "Clustering of scatterers in mobile radio channels—evaluation and modeling in the COST259 directional channel model," in *Proc. IEEE International Conference on Communications*, New York, 2002.
- [2] W. R. Braun and U. Dersch, "A physical mobile radio channel model," *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, pp. 472–482, May 1991.
- [3] M. Nakagami, "The m -distribution - a general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*, W. C. Hoffman, Ed. Oxford, England: Pergamon, 1960.
- [4] B. I. Raju and M. A. Srinivasan, "Statistics of envelope of high-frequency ultrasonic backscatter from human skin in vivo," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 49, no. 7, pp. 871–882, July 2002.
- [5] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York: McGraw-Hill, 2002.



José Cândido Silveira Santos Filho (S'04) received the B.S. and M.S. degrees in electrical engineering from the University of Campinas, Brazil, in 2001 and 2003, respectively, where he is currently working toward the Ph.D. degree. His research interests include modeling, analysis and simulation of diversity fading channels.



Michel Daoud Yacoub was born in Brazil in 1955. He received the B.S.E.E. and the M.Sc. degrees from the School of Electrical and Computer Engineering of the State University of Campinas, UNICAMP, Brazil, in 1978 and 1983, respectively, and the Ph.D. degree from the University of Essex, U.K., in 1988. From 1978 to 1985, he worked as a Research Specialist at the Research and Development Center of Telebrás, Brazil, in the development of the Tropico digital exchange family. He joined the School of Electrical and Computer Engineering, UNICAMP,

in 1989, where he is presently a Full Professor. He consults for several operating companies and industries in the wireless communications area. He is the author of *Foundations of Mobile Radio Engineering* (Boca Raton, FL: CRC, 1993), *Wireless Technology: Protocols, Standards, and Techniques* (Boca Raton, FL: CRC, 2001), and the co-author of *Telecommunications: Principles and Trends* (São Paulo, Brasil: Erica, 1997, in Portuguese). He holds two patents. His general research interests include wireless communications.



Gustavo Fraidenraich received the B.S.E.E. degree from the Federal University of Pernambuco, UFPE, Brazil, in 1997. He received his M.Sc. and Ph.D. degrees from the State University of Campinas, UNICAMP, Brazil, in 2002 and 2006, respectively. He is currently as a Postdoctoral Fellow in the Star Lab Group at Stanford University, USA. His research interests include Fading Channels, MIMO systems, and Wireless Communications in general. Dr. Fraidenraich is also an IEEE Member.