Probabilistic Engineering Mechanics 25 (2010) 315–322

Contents lists available at ScienceDirect

Probabilistic Engineering Mechanics

journal homepage: www.elsevier.com/locate/probengmech

A simple algorithm for generating spectrally colored, non-Gaussian signals

J.M. Nichols [a,*], C.C. Olson [b], J.V. Michalowicz [c], F. Bucholtz [a]

[a] Naval Research Laboratory, 4555 Overlook Ave. Washington, DC 20375, United States

[b] National Research Council, Postdoctoral Research Associated, Naval Research Laboratory, United States

[c] Global Strategies Inc., Crofton MD, 21114, United States

ARTICLE INFO

Article history:
Received 30 March 2009
Received in revised form
19 January 2010
Accepted 26 January 2010
Available online 2 February 2010

Keywords:
Surrogate data
Colored noise
Non-Gaussian
Random process

ABSTRACT

This work describes a simple method for generating signals conforming to a stationary random process for which the practitioner specifies both the power spectral density function and the marginal probability density function. The general approach is to first create a Gaussian random process with the appropriate spectral density and then apply a memoryless nonlinear transformation to achieve the desired marginal density. The transformation is not specified *a priori* but rather is simulated via an iterative "shuffling" procedure. The method is very simple to implement and yields results that are comparable to some of the more complicated methods.

Published by Elsevier Ltd

1. Introduction

Random vibration problems often require simulating the response of a system to inputs that are often modeled as stationary random processes. Examples might include the response of a ship hull to wave loading, the vibrational response of an airfoil to gust loading, or even a building response to wind excitation. In these cases we cannot write down a set of equations that will exactly describe the loading a structure will be subject to, rather it is more appropriate to use probabilistic models and think of what a *typical* loading history might look like. Probabilistic models for sequences of observations (signals) are referred to as random process models. In order to simulate the response of a structure to a particular random process one first has to generate a realization of that random process. This work presents an approach for generating the output of a random process with a prescribed probability density function (PDF) and power spectral density function (PSD).

Here we define a stationary random process to be one that results in a sequence of observations $x(t_1), x(t_2), \ldots, x(t_N)$ for which the joint probability distribution $p(x(t_1), x(t_2), \ldots, x(t_N))$ is invariant to temporal shifts i.e. $p(x(t_1), x(t_2), \ldots, x(t_N)) = p(x(t_1+\tau), x(t_2+\tau), \ldots, x(t_N+\tau))$. Typically, however, when we talk about a random process we do not specify the entire joint distribution. Rather we often describe a random process in terms of (1) the marginal distribution associated with each observation, $p(x) \equiv p(x(t_1)) = p(x(t_2)) = \cdots p(x(t_N))$, and (2) the auto-covariance function $R_{xx}(\tau) = E\left[\tilde{x}(t)\tilde{x}(t+\tau)\right]$ where $\tilde{x}(t) = x(t) - E[x(t)]$. These two quantities are perhaps the most commonly used descriptors of a random process. It is also useful to note that for a stationary random process the auto-covariance function and power spectral density (PSD) function, to be denoted as $S_{xx}(\omega)$, are related by Fourier transform via the Wiener–Khintchine relation (assuming the Fourier transform of $R_{xx}(\tau)$ exists, a sufficient condition being that $\int_{-\infty}^{\infty} |R_{xx}(\tau)|\mathrm{d}\tau < \infty$). Thus specifying the auto-covariance function is the same as specifying the PSD. In engineering applications it is more common for a random process to be described by $S_{xx}(\omega)$.

This work will consider the problem of generating a sequence of observations with a specific $p(x)$ and $S_{xx}(\omega)$ only. Other quantities, e.g. $R_{xxx}(\tau_1, \tau_2) = E[\tilde{x}(t)\tilde{x}(t+\tau_1)\tilde{x}(t+\tau_2)]$ (equivalently, the bispectrum) are typically not specified when describing random processes and are often implicitly assumed to be negligible. This may or may not be a good assumption. A number of researchers have detected the presence of significant third-order correlations in a variety of random processes. For example, Richardson and Hodgkiss found clear evidence of higher-order correlations in underwater acoustic data. Kim and Powers [1] detected the presence of third-order correlations in plasma density fluctuations [1] as did Hajj et al. [2] in fluid flow. Gurley et al. found evidence of higher-order correlations in wave height data and, in fact, proposed an approach for generating time-series consistent with such correlations [3]. The question of whether or not $p(x)$ and $S_{xx}(\omega)$ are sufficient descriptors of a random process is not discussed here.

The goal of this work is to provide the practitioner with a simple, but accurate means of generating a sequence of observations

* Corresponding author. Tel.: +1 202 404 5433; fax: +1 202 767 5792.
E-mail address: jonathan.nichols@nrl.navy.mil (J.M. Nichols).

0266-8920/$ – see front matter. Published by Elsevier Ltd
doi:10.1016/j.probengmech.2010.01.005

with specified $p(x)$ and $S_{xx}(\omega)$. This is accomplished by first generating a stationary sequence of random variables conforming to $p(x)$ and then iteratively "shuffling" these values to provide the correct spectral coloring (i.e., match $S_{xx}(\omega)$). In the event that samples conforming to $p(x)$ are not readily available, we also provide a simple method for "rejection sampling" from the desired distribution. The end result is a very simple approach that works even in situations where certain other approaches (to be discussed in Section 2) have difficulty. Code for the proposed algorithms is also provided.

2. Review of previous work and some basic concepts

The problem of how to generate observations $x(n), n = 0 \cdots N-1$ with a given PDF and auto-covariance (or PSD) has been tackled by a number of researchers. While the details vary considerably, all approaches tend to follow the same general prescription. This procedure is illustrated schematically in Fig. 1. First, a spectrally white, Gaussian distributed sequence is linearly filtered in order to produce a signal with the proper auto-covariance (equivalently PSD). This step makes use of the fact that if the input to a linear filter is Gaussian distributed, so too will be the output. The job of the filter is to impose the correct spectral coloring on the output signal. The data are then subject to a zero memory, nonlinear (ZMNL) transformation in order to produce a signal with the appropriate non-Gaussian PDF. The requirement of a memoryless transformation ensures that the spectral properties of the signal generated in the first step will not be significantly altered. Gujar and Kavanagh [4] based their linear filtering operation on the desired (target) PSD function (as will be done in this work). The nonlinear transformation they used was obtained by expanding the filtered signals' autocorrelation function as a power series using Hermite polynomials (this is possible because of the Gaussian PDF of the input [5]). This same general scheme was applied by Liu and Munson [6] but with a much more detailed analysis of the method.

Following this approach the ZMNL function, denoted $g(\cdot)$, is given by

$$g(\cdot) = F^{-1}F_y(\cdot)$$

where F_y is the Gaussian cumulative distribution function (CDF) and F is the desired (target) CDF. The ZMNL function $g(\cdot)$ is expanded in terms of Hermite polynomials, which enables the autocorrelation of the ZMNL output to be written as a power series of the autocorrelation of $y(n)$. One then solves for the autocorrelation associated with $y(n)$ which makes the output of the ZMNL best approximate the desired autocorrelation, and then designs the linear filter to approximate this autocorrelation arbitrarily closely. The main problem with this method is that F is sometimes not invertible analytically (e.g. the Gamma and Beta CDFs), and finding F^{-1} numerically reduces the simplicity and accuracy of the method. In response to this, Filho and Yacoub [7] proposed an approach that used a combination of the Hermite polynomials (for generating the proper autocorrelation) followed by a rank-reordering step for transforming the PDF to the desired non-Gaussian target. This rank-reordering or "shuffling" procedure attempts to approximate the influence of $g(\cdot)$ and has the distinct advantage of not requiring an analytical expression for the static, nonlinear transformation. A similar approach to approximating this function was tried earlier by Hunter and Kearney [8] and dubbed "Stochastic Minimization".

In the context of structural dynamics there are numerous papers on the topic of colored noise generation. A recent review of many of these techniques has been provided by Bocchini and Deodatis [9]. The foundations for generating a signal with a marginal Gaussian distribution but with a prescribed power spectral density function were given by Shinozuka and Jan [10]. This approach simply involves converting the PSD to the associated Fourier amplitudes, scaling, and applying the real inverse Fourier transform (the linear filtering step in Fig. 1. By the central limit theorem the resulting signal will be normally distributed. It is the second step, finding the appropriate transformation for altering the PDF that is the main source of difficulty. Yamazaki and Shinozuka [11] proposed what appears to be the core algorithm for matching both spectral properties *and* the marginal distribution. The method iteratively adjusts the original Gaussian distributed signal (generated using the method of [10]) to have the correct PDF and then corrects for the changes this operation caused to the PSD. This same general scheme has been improved upon with works by Deodatis and Micaletti [12], Shi and Deodatis [13], and Bocchini and Deodatis [9]. Again, the goal is to approximate the influence of the unknown $g(\cdot)$ without adversely influencing the PSD. The work of Grigoriu [14] considers a similar strategy and uses the term "translation process" to describe the ZMNL function.

We should also point out that solutions to the colored noise generation problem bear close resemblance to the generation of so-called "surrogate sets", frequently used in the physics community for determining whether or not an observed signal is consistent with a linear random process [15]. The goal of most surrogate algorithms is to produce a new (surrogate) signal with the same PSD and marginal PDF as the original data. Surrogate algorithms are therefore excellent candidates for our problem.

In this work the linear filter is created using a well-known approach using the Fourier amplitudes associated with the target power spectral density. The static nonlinearity is approximated using a rank-reordering procedure — similar in spirit to that proposed in [7], but modified in accordance with the aforementioned surrogate generation methods. The iterative portion of our algorithm is, in fact, identical to that put forward by Schreiber [16]. The algorithm does require the practitioner to draw samples from the desired PDF. For complicated PDFs this may not be accomplished by standard methods. For this reason we also briefly discuss the concept of rejection sampling and provide a simple algorithm for generating iid random sequences for any given probability distribution.

3. Iterative Fourier transform method

Assume that we wish to create a sequence of observations $x(n)\ \ n = 0, \ldots, N-1$ with probability density function $p(x)$ and a two-sided power spectral density $S_{xx}(\omega)$. The first step is to sample the desired PSD function at N discrete frequencies $\omega_k = (k - N/2)\Delta_\omega, k = 0, \ldots, N-1$ giving $S_{xx}(\omega_k)$. The frequency bin width is dictated by the desired temporal sampling interval Δ_t. The two are related via $\Delta_\omega = 2\pi/N\Delta_t$ and, as usual, Δ_t should be chosen in accordance with the Nyquist criterion for the maximum resolvable frequency. We also define the discrete Fourier transform

$$X(k) = \mathrm{FT}(x(n)) \equiv \sum_{n=0}^{N-1} x(n)\mathrm{e}^{-\mathrm{i}2\pi kn/N} \tag{1}$$

and inverse Fourier transform

$$x(n) = \mathrm{FT}^{-1}(X(k)) \equiv \frac{1}{N}\sum_{k=0}^{N-1} X(k)\mathrm{e}^{\mathrm{i}2\pi kn/N}. \tag{2}$$

Before proceeding, it should be mentioned that $p(x)$ and $S_{xx}(\omega)$ cannot be specified independently as they are linked through the signal mean $\bar{x}$ and variance σ_x^2. For the signal mean we have the relationship

$$\frac{X(0)}{N} = \frac{1}{N}\sum_{n=0}^{N-1} x(n) = \bar{x} \tag{3}$$