Nuclear Instruments and Methods in Physics Research A 690 (2012) 75–78

Contents lists available at SciVerse ScienceDirect

Nuclear Instruments and Methods in Physics Research A

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

Correlated random sampling for multivariate normal and log-normal distributions

Gašper Žerovnik*, Andrej Trkov, Ivan A. Kodeli

Jožef Stefan Institute, Jamova cesta 39, SI-1000 Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 20 February 2012
Accepted 24 June 2012
Available online 29 June 2012

Keywords:
Random correlated sampling
Multivariate probability distributions
Normal distribution
Log-normal distribution
Resonance parameters

ABSTRACT

A method for correlated random sampling is presented. Representative samples for multivariate normal or log-normal distribution can be produced. Furthermore, any combination of normally and log-normally distributed correlated variables may be sampled to any requested accuracy. Possible applications of the method include sampling of resonance parameters which are used for reactor calculations.

1. Introduction

Along with the development of fast computer systems, accurate but numerically relatively inefficient Monte Carlo methods are being more frequently used in neutron transport calculations [1]. One step further is the so-called Total Monte Carlo method, described in Ref. [2], which couples random sampling of nuclear data (taking into account their uncertainties and correlations) with Monte Carlo particle simulations. Nuclear data are usually given in the form of expected values and covariance matrices. The corresponding distributions are multivariate normal (for location parameters) or log-normal (for scale or inherently positive parameters). According to Ref. [3], normal distribution can be used with satisfactory accuracy for most applications if relative uncertainties of the inherently positive parameters are small ($\lesssim 0.3$). However, this is not always the case.

Random sampling of single or independent arbitrarily distributed parameters is straightforward. When several parameters are correlated, the correlated sampling problem can be reduced to the problem of single parameter sampling by simply diagonalizing the covariance matrix. However, non-Gaussian distributions become distorted by the diagonalization which is a linear operation. This is a consequence of the fact that in general distribution shape is not conserved when performing linear combinations. Since following the Central limit theorem [4] Gaussian is the only distribution which is conserved when convoluted, non-Gaussian distributions are non-linear in this aspect.

Following Ref. [5], inherently positive parameters with known (estimate of the) expected value and uncertainty are best described with log-normal distribution. However, the diagonalization method experiences problems when dealing with several inter-correlated log-normally distributed parameters with large relative uncertainties. The diagonalization method is limited by the fact that only normal distribution is preserved when performing linear transformations thus producing distorted original parameter distributions and negative values. In the first approximation, the multivariate log-normal distribution can be taken into account by applying the multivariate normal distribution in the logarithmic parameter space, however at the expense of producing significant biases in mean values and standard deviations of parameters with large uncertainties [6]. Alternatively, unphysical negative parameter values may be simply set to 0, thereby again producing biases in mean values and standard deviations of the sampled parameters.

The final solution to the problem of multivariate log-normal random sampling is presented in this paper. It is based on a different approach, arbitrarily called the 'Correlated sampling method', according to which the parameters are calculated from independently sampled variables exactly taking into account the entire distributions of the parameters. The dependence of parameters on the sampled variables is a function of the parameter distributions. Exact analytical expressions for multivariate normal and log-normal distributions have been derived. Also, employing a simple numerical trick, the method can be extended to arbitrary combinations of coupled normally and log-normally distributed parameters, which can be sampled to any requested precision. Such combination of distributions appears for example when dealing with resonance parameters [7]: the resonance energies

* Corresponding author. Tel.: +386 15885326; fax: +386 15885454.
E-mail addresses: gasper.zerovnik@ijs.si (G. Žerovnik), andrej.trkov@ijs.si (A. Trkov), ivan.kodeli@ijs.si (I.A. Kodeli).

0168-9002/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.nima.2012.06.036

are normally while the inherently positive resonance widths are log-normally distributed; for example as described in Ref. [8].

Since multivariate log-normal distribution (we believe the reader should be well aware of applicability of multivariate normal distribution) is used in many other applications, from radiation protection [9] to numerical weather prediction [10] and astrophysics [11], from hazard modelling [12] and pattern recognition [13] to financial mathematics [14], the applicability of the method described in this paper reaches beyond nuclear engineering.

2. Correlated sampling method

The method is based on correlated sampling of parameters, which in principle allows the use of arbitrary probability distribution functions for parameters. A useful property of the method is that it exactly preserves the entire parameter distributions and correlations of the sampled parameters. The main limitation of the method is the need to solve a system of n^2 equations, where n is the number of inter-correlated parameters, before applying the parameter sampling. In this work, only normal and log-normal distributions are considered. Luckily, for these two distributions the system of equations is quadratic and may be written in a form of a matrix equation with one solution obtainable by performing the matrix square root operation.

The basic idea of the method is very simple. We would like to produce samples of n parameters distributed according to their distribution functions by sampling n independent (e.g. uniformly or normally distributed) random variables ξ_i. A sample would then be produced by calculating a known function $\vec{F}$ of the random variables, which would depend on the parameter distributions and correlations. Formally, the m-th sample $\vec{x}^{(m)}$ would be

$$\vec{x}^{(m)} = \vec{F}(\vec{\xi}^{(m)}). \tag{1}$$

In practice, one has to derive $\vec{F}$ for every distribution or family of distributions, which is in general far from trivial. The function $\vec{F}$ is non-linear except for normal distribution of $\vec{x}$.

2.1. Normal distribution

Due to its linearity, normal distribution is probably the most simple example. It is well-known that any linear combination of normally distributed variables is again a normally distributed variable. If in $\vec{\xi}^{(m)}$ all components are normally distributed variables, the samples could be produced by

$$\vec{x}^{(m)} = A \cdot \vec{\xi}^{(m)} + \vec{\mu} \tag{2}$$

where matrix A is defined such that mean values, standard deviations and correlations between the components of $\vec{x}$ are preserved. Without any loss of generality, all $\vec{\xi}$ can have zero means and unit standard deviations, i.e. can be distributed according to the so-called standard normal distribution. Consequently, $\vec{\mu}$ represents the vector of expected values of $\vec{x}$:

$$\vec{\mu} = \langle \vec{x} \rangle. \tag{3}$$

Furthermore, matrix A has to satisfy

$$\begin{aligned} V_{ij} &= \lim_{M\to\infty} \frac{1}{M-1} \sum_{m=1}^{M} (x_i^{(m)} - \mu_i)(x_j^{(m)} - \mu_j) \\ &= \lim_{M\to\infty} \frac{1}{M-1} \sum_{m=1}^{M} \sum_{k,l=1}^{n} A_{ik}\xi_k^{(m)} A_{jl}\xi_l^{(m)} \\ &= \sum_{k,l=1}^{n} A_{ik}A_{jl} \lim_{M\to\infty} \frac{1}{M-1} \sum_{m=1}^{M} \xi_k^{(m)}\xi_l^{(m)} = \sum_{k,l=1}^{n} A_{ik}A_{jl}\delta_{kl} \\ &= \sum_{k=1}^{n} A_{ik}A_{jk} \end{aligned} \tag{4}$$

where V_{ij} are the absolute covariances of the parameters x_i and x_j. Equality

$$\lim_{M\to\infty} \frac{1}{M-1} \sum_{m=1}^{M} \xi_k^{(m)}\xi_l^{(m)} = \delta_{kl} \tag{5}$$

holds since the left hand side equals variance ($k=l$) or covariance ($k \neq l$) of independent variables by definition assuming samples $\xi_k^{(m)}$ are representative for the standard normal distribution.

Eq. (4) defines n^2 quadratic equations for n^2 unknowns (the elements of the matrix A). When system equation (4) is solved, unlimited number of normally distributed parameter set $\vec{x}$ samples can be produced, taking into account all correlations between individual parameters. Probably the easiest way to find a solution A (any solution is sufficient for our purposes since though they all produce different random samples $\vec{x}^{(m)}$, the latter all obey the same distribution) is to first write Eq. (4) in matrix form:

$$V_{ij} = \sum_{k=1}^{n} A_{ik}A_{jk} = \sum_{k=1}^{n} A_{ik}(A^T)_{kj} = (AA^T)_{ij} \Rightarrow V = AA^T. \tag{6}$$

Since the covariance matrix V is real symmetric (because, of course, parameter i is correlated to parameter j as much as j is correlated to i), it is diagonalizable with non-negative eigenvalues [15, p. 288]:

$$V = QDQ^T \tag{7}$$

where D is a diagonal matrix of the eigenvalues, and Q is an orthogonal matrix of the eigenvectors of V. Then, because $Q^T = Q^{-1}$ and $D^T = D$,

$$A = QD^{1/2}Q^T \tag{8}$$

is a solution to Eq. (6). Here, we take advantage of the fact that the square root of a diagonal matrix is simply a matrix with square roots of diagonal elements. Eq. (9) also clarifies the need for matrix V diagonalisability.

To summarize, matrix equation (6) may have several solutions, one of them being the square root of matrix V:

$$A = V^{1/2}. \tag{9}$$

With explicitly given A and $\vec{\mu}$, arbitrary random multivariate normally distributed samples can be produced by employing Eq. (2).

2.2. Log-normal distribution

For normal distribution, the correlated sampling method is equivalent to the diagonalization method. However, the correlated sampling method can be extended to consistently deal with the multivariate log-normal distribution.

Natural logarithms of log-normally distributed variables are by definition normally distributed. In the following discussion, this property will be used several times.

Let us start with independent random variables $\xi_i^{(m)}$, $i = 1,\dots,n$, this time log-normally distributed. Then variables $\ln \xi_i^{(m)}$ are normally distributed and so are all of their linear combinations:

$$\ln x_i^{(m)} = \sum_{j=1}^{n} A_{ij} \ln \xi_j^{(m)} + \mu_i \tag{10}$$