COMPUTER GENERATION OF RANDOM VECTORS WITH GUARANTEE OF THE STATISTICAL PARAMETERS USING THE MONTE-CARLO METHOD

Dimitar Tyanev

Abstract: *The objects of comment in this paper are the deficiencies of computer generated standard normal statistical samples yielding to the* ***N(0,I)*** *law and the impossibility of the generators to guarantee the statistical parameters assigned in advance. An approach for appropriation of the statistical samples of a priori known mean vector and covariance matrix in absolute conformity with the* $\boldsymbol{N^{(k)}(v,C)}$ *law is suggested.*

Categories and Subject Descriptors: ; G.3 [**Probability and Statistics**]: Random Number Generation; I.5.1 [**Pattern Recognition**]:Models.

General Terms: Theory, Statistical Sequences, Experimentation.

Additional Key Words and Phrases: random-number generation; pattern recognition.

1. Computer Generated Samples – Deficiencies

The use of statistic estimation for evaluating the values of a certain integer, also known as the Monte-Carlo Method, is becoming more and more topical, due to the universal abilities of its computer applications [4, 8, 12, 16].

The attention here is to be focused on the problems associated with the solution of the following problem: a statistical study is being carried out on a specially synthesised [18] distributing function [15], with the purpose of determining its probable possibilities to recognise patterns with precisely defined statistical features. In order to implement the respective survey, we need to dispose of statistical samples yielding to a definite distribution law. The survey is fulfilled over computer generated samples. However, the actual condition with this approach is such that there are no guarantees for the full satisfaction of the required statistical parameters of the distribution law. For instance, the numeric sequences that have to subordinate to a uniform law, are in fact, non-uniformly distributed, and, what is more, they are usually pseudorandom, due to their periodicity [2, 8, 12]. On the other hand, if they are uniform [17], they are not random. The multidimensional space environment involves the requirement for independence of the one-dimensional generators [5]. The practical situations, however, are characterised by the presence of strong correlation.

After studying many methods and programs for generation of standard normally distributed samples from random k-dimensional vectors [4, 6, 8, 11, 14, etc.], the following major conclusions have been ascertained:

1). The samples are characterised with displacement and their mean vector is non-zero;

2). The samples have hyperellipsoidal space dispersion, instead the expected hyperspherical dispersion;

3). Their own eigenvector basis is oriented randomly and in general position, instead of coinciding with the co-ordinate system of axes on the set of one-dimensional and independent generators, as expected to be;

4) The samples have dense covariance matrices instead of the predictable identity matrix.

The above-stated conclusions are supported by experimental results, represented in Table 1. The formulation of the experiment, which results in these particular examples, consists of the following: using the library subroutine RNNOA in FORTRAN-90 [11], a lot of statistical samples of m in number standard and normally distributed k-dimensional random vectors are generated. Here is a short quotation from the library Programmer's Handbook, regarding the above-mentioned subroutine.

"Generate pseudorandom numbers from a standard normal distribution using an acceptance/rejection method.

Usage

CALL RNNOA (NR, R)

Arguments

NR — Number of random numbers to generate. (Input)

R — Vector of length NR containing the random standard normal deviates. (Output)

Algorithm

Routine RNNOA generates pseudorandom numbers from a standard normal (Gaussian) distribution using an acceptance/rejection technique due to Kinderman and Ramage (1976). In this method, the normal density is represented as a mixture of densities over which a variety of acceptance/rejection methods due to Marsaglia (1964), Marsaglia and Bray (1964), and Marsaglia et al. (1964) are applied. This method is faster than the inverse CDF technique used in RNNOR to generate standard normal deviates."

The volumes m, for the purpose of which the numeric values have been produced, are picked out of the interval from 200 to 5000, which has been considered to be representative enough. The numeric values refer to the Euclidean vector norm of the mean vector μ and to the eigenvalues λ_i of the covariance matrix of each sample. The results in Table 1 are related to the samples, generated in 7-dimensional space. The mean vector characterises the distribution displacement compared to the initiation of the co-ordinate system, and its eigenvalues – its deviation from the hyperspherical shape.

Table 1

	m=200	m=500	m=1500	m=3000	m=5000
$\lvert\mu\rvert$	0.21848	0.15392	0.05995	0.03903	0.02322
λ_1	0.85786	0.90645	1.00208	0.98623	1.02536
λ_2	0.95876	1.03759	0.98528	1.03707	1.00983
λ_3	0.94530	0.96545	1.03142	0.99591	1.00133
λ_4	0.91762	1.07776	0.96478	0.97215	1.01239
λ_5	0.97808	0.94318	0.95656	0.95317	0.98087
λ_6	0.95815	1.06364	1.05786	1.03465	0.99783
λ_7	1.07937	0.95622	0.97793	0.99684	0.99186

From the above-displayed results, one could easily estimate that the deviations of eigenvalues, as related to 1, reach up to 14% (see Table 1 – λ_1, for m=200). For the largest volume – m=5000, the deviation decreases to 2,5%. The situation with the sample displacement is similar. The magnitude of mean vector, which is expected to be zero, reaches to $|\mu|=0.21848$ (see Table 1 – for m=200). Considering that in this case the radius of the statistical dispersion in the space is less than $3.\sigma=3$, one could estimate that the sample displacement is more than 7%. For the largest volume – m=5000 the displacement decreases under 1%.

We find such deviations of the indicated parameters of the required values to be rather significant.

The shown deviation parameters relate to the particular generator and the particular samples, that is, other values, higher and lower, could be produced with other samples, derived from other generators. This fact of random behaviour and impossibility a priori to be guaranteed the statistical parameters of particularly generated samples is inadmissible in cases where the pursue is for a statistical estimation of a given value and when the conditions for the conduct of the estimation have been assigned a priori. These are the grounds for the search of a way to fulfil these conditions with generation of experimental statistical data.

2. Formulation of the Problem

The conclusions stated so far unambiguously lead to formulation of the task for a “correction” of the generated statistical sample in terms of its constant ability to generate the required statistical structure. The formal treatment reads: the given matrix **X**(m,k) whose rows are the generated (m in number) k-dimensional random and normally distributed in space vectors, i.e., the matrix **X** represents a generated sample. The actual parameters of this sample could be computed and let they are: $\mu(k)$ –

mean vector and **K**(k,k) – covariance matrix. Thus the distribution, which represents the matrix **X**, satisfies the normal law **N**(μ,**K**). At the same time, the samples are required to have the parameters of the **N**(ν,**C**) law. It is, therefore, clearly realised that two formal problems are distinguished:

a) the problem of statistical centre shift from point μ to point ν;

b) the problem of the statistical structure conversion defined by the covariance matrix shown by **K** to the required one – shown by the matrix **C**.

3. Solution

The problems, formulated above, are not immediately associated, which facilitates the solution. The first problem is easy to solve – the needed translation of the statistical centre is achieved by means of the differentiating vector **r**.

The second problem requires that the statistical sample **X** converts into a new one – **Y**, which possesses the necessary statistical parameters. The solution is to be found in the form of a linear operator, with converting matrix **S**. Using this approach, every vector $\mathbf{y}_i$ is to be derived as follows:

$$\mathbf{y}_i = \mathbf{S}^t.\mathbf{x}_i + \mathbf{r} \ , \quad i = \overline{1,m} \ . \tag{1}$$

The most probable values of the statistical parameters of the sample **Y** are estimated like this [9]:
- for the mean vector:

$$\underset{i}{\mathrm{E}}\{\mathbf{y}_i\} = \nu = \underset{i}{\mathrm{E}}\{\mathbf{S}^T.\mathbf{x}_i + \mathbf{r}\} = \mathbf{r} + \underset{i}{\mathrm{E}}\{\mathbf{S}^T.\mathbf{x}_i\} = \mathbf{r} + \mathbf{S}^T.\underset{i}{\mathrm{E}}\{\mathbf{x}_i\} = \mathbf{r} + \mathbf{S}^T.\mu \ . \tag{2}$$

By (2) we find the differentiating vector **r**:

$$\mathbf{r} = \nu - \mathbf{S}^T.\mu \ ; \tag{3}$$

- for the elements of the covariance matrix:

$$\begin{aligned}
&\underset{i}{\mathrm{E}}\{(\mathbf{y}_i - \nu).(\mathbf{y}_i - \nu)^T\} = \mathbf{C} = \\
&\quad = \underset{i}{\mathrm{E}}\{(\mathbf{S}^T.\mathbf{x}_i + \mathbf{r} - \nu)(\mathbf{S}^T.\mathbf{x}_i + \mathbf{r} - \nu)^T\} = \\
&\quad = \underset{i}{\mathrm{E}}\{[\mathbf{S}^T.\mathbf{x}_i + \mathbf{r} - (\mathbf{r} + \mathbf{S}^T.\mu)].[\mathbf{S}^T.\mathbf{x}_i + \mathbf{r} - (\mathbf{r} + \mathbf{S}^T.\mu)]^T\} = \\
&= \underset{i}{\mathrm{E}}\{[\mathbf{S}^T.(\mathbf{x}_i - \mu)].[\mathbf{S}^T.(\mathbf{x}_i - \mu)]^T\} = \mathbf{S}^T.\underset{i}{\mathrm{E}}\{(\mathbf{x}_i - \mu).(\mathbf{x}_i - \mu)^T\}.\mathbf{S} = \mathbf{S}^T.\mathbf{K}.\mathbf{S} \ .
\end{aligned} \tag{4}$$

The last formula shows how the converting matrix **S** of the linear operator (1) conjoins the actual covariance matrix **K** with the needed **C**. According to [1] the sufficient condition for the existence of the matrix **S** is that the matrices **K** and **C** need to be commutative. Here this condition is fulfilled as the covariance matrices are symmetrical.

The construction of the linear operator (1) is based on the following theorem:

Theorem: Let **K** and **C** be symmetrical positively determined matrices and let their orthogonal decompositions be given by:

$$\mathbf{K} = \mathbf{V}.\Lambda.\mathbf{V}^T \tag{5}$$

and

$$\mathbf{C} = \Phi.\mathbf{D}.\Phi^T \ , \tag{6}$$

where Λ and **D** are diagonal matrices, containing their relevant eigenvalues, and **V** and Φ are the orthonormal matrices of their eigenvectors, corresponding to them, then the matrix of linear conversion **S** is given as:

$$\mathbf{S} = \mathbf{V}.\Lambda^{-1/2}.\mathbf{D}^{1/2}.\Phi^T \tag{7}$$

Proof: We substitute directly (7) in (4) to obtain:

$$\left(\mathbf{V}.\Lambda^{-1/2}.\mathbf{D}^{1/2}.\Phi^{T}\right)^{T}.\mathbf{K}.\left(\mathbf{V}.\Lambda^{-1/2}.\mathbf{D}^{1/2}.\Phi^{T}\right) = \Phi.\mathbf{D}^{1/2}.\Lambda^{-1/2}.\left(\mathbf{V}^{T}.\mathbf{K}.\mathbf{V}\right).\Lambda^{-1/2}.\mathbf{D}^{1/2}.\Phi^{T}.$$

According to equation (5), the bracketed expression in the last formula equals to Λ, which results in:

$$\Phi.\mathbf{D}^{1/2}.\Lambda^{-1/2}.\Lambda.\Lambda^{-1/2}.\mathbf{D}^{1/2}.\Phi^{T} = \Phi.\mathbf{D}.\Phi^{T} = \mathbf{C}. \qquad \textbf{(8)}$$

The above-mentioned expression shows that the matrix **S** converts the matrix **K** into **C**, as required by equation (4). This completes the proof.

In case the covariance matrix **K** is positively semidetermined, with rank n (n<k), then it has (k-n) in number eigenvalues, equal to zero. Therefore, the diagonal matrix Λ in decomposition (5), is given as the following block matrix:

$$\Lambda = \text{diag}(\lambda_1,\lambda_2,...,\lambda_n,\lambda_{n+1} = 0,...,\lambda_k = 0) = \begin{bmatrix} \Lambda_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \qquad \textbf{(9)}$$

Such a matrix has a signature number q=k-n and inertia matrix [13] $\mathbf{I}(\Lambda) = \text{diag}(\mathbf{I}_n;\mathbf{0}_q)$. Under these actual conditions the multidimensional statistical sample **X** has no dispersion in the direction of those eigenvectors, to which zero eigenvalues correspond. This fact makes the forming of the matrix (7) impossible, as it includes the reciprocal square root $\Lambda^{-1/2}$.

In this case, in order to guarantee the conversion (1), we will nullify the signature number of the matrix Λ, while accepting the identity matrix to be its inertia matrix and forming the following block matrix:

$$\tilde{\Lambda} = \begin{bmatrix} \Lambda_n & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{bmatrix}. \qquad \textbf{(10)}$$

Such a matrix has a full rank and therefore the matrix $\left(\tilde{\Lambda}\right)^{-1/2}$ exists. Thus, under the conditions of the above theorem, the conversion (1) could be achieved by means of the following converting matrix:

$$\tilde{\mathbf{S}} = \mathbf{V}.\left(\tilde{\Lambda}\right)^{-1/2}.\mathbf{D}^{1/2}.\Phi^{T}. \qquad \textbf{(11)}$$

By substituting (11) in (4), we obtain:

$$\left(\mathbf{V}.\left(\tilde{\Lambda}\right)^{-1/2}.\mathbf{D}^{1/2}.\Phi^{T}\right)^{T}.\mathbf{K}.\left(\mathbf{V}.\left(\tilde{\Lambda}\right)^{-1/2}.\mathbf{D}^{1/2}.\Phi^{T}\right) =$$

$$= \Phi.\mathbf{D}^{1/2}.\left(\left(\tilde{\Lambda}\right)^{-1/2}.\Lambda.\left(\tilde{\Lambda}\right)^{-1/2}\right).\mathbf{D}^{1/2}.\Phi^{T} = \Phi.\mathbf{D}^{1/2}.\mathbf{I}(\Lambda).\mathbf{D}^{1/2}.\Phi^{T} =$$

$$= \Phi.\begin{bmatrix} \mathbf{D}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.\Phi^{T}. \qquad \textbf{(12)}$$

In the last expression, the product of the three diagonal matrices has appeared in the form of (9), which shows that the statistical sample **Y** will obtain the required parameters, but only in the n-dimensional space, determined by the inertia matrix $\mathbf{I}(\Lambda)$. In this situation, the statistical sample **Y** would not be able to obtain the structure assigned through any matrix **C**. The latter will be achieved partly (only in the n-dimensional space). In the remaining part, (q-dimensional complementing subspace) the needed structure would not be possible to achieve – there it would be ignored, which is in full conformity with Sylvester's Inertia Theorem [13], according to which the connection (4) is possible, if both matrices **K** and **C** have the same inertia.

4. Conclusion

The ensuing results, that experiments completely suit, confirm the expectations. For example, when the required as a statistical sample covariance matrix **C**(3, 3) is assigned:

$$\mathbf{C} = \begin{bmatrix} 7 & -3 & 0 \\ -3 & 7 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

then without fulfilment of the correction algorithm under the already mentioned conditions and means, a random statistical sample of 1000 3-dimensional vectors is generated, whose covariance matrix has elements, substantially distinctive from the required ones:

$$\mathbf{C}' = \begin{bmatrix} 0.6637366\mathrm{E}+01 & -0.2957354\mathrm{E}+01 & -0.1840812\mathrm{E}-01 \\ -0.2957354\mathrm{E}+01 & 0.6627336\mathrm{E}+01 & -0.1899240\mathrm{E}+00 \\ -0.1840812\mathrm{E}-01 & -0.1899240\mathrm{E}+00 & 0.7171562\mathrm{E}+00 \end{bmatrix}.$$

At the same time, if the generated sample is processed by the means of the correction algorithm, we get a statistical sample, possessing the following covariance matrix:

$$\mathbf{C} = \begin{bmatrix} 0.7000002\mathrm{E}+01 & -0.3000002\mathrm{E}+01 & 0.7675064\mathrm{E}-07 \\ -0.3000002\mathrm{E}+01 & 0.7000001\mathrm{E}+01 & -0.1636595\mathrm{E}-06 \\ 0.7675064\mathrm{E}-07 & -0.1636595\mathrm{E}-06 & 0.1000000\mathrm{E}+01 \end{bmatrix}.$$

It could be assumed that the correspondence is accomplished as it depends solely on the accuracy of the computer calculations.

In conclusion, we could deduce that the method herein presented is able to convert any normally distributed statistical structure into another randomly chosen one, but under the conditions of the above theorem. This common statement also holds good for the particular case with positively semidetermined covariance matrix **C**. In this case, the matrix **D** in decomposition (6), will appear in the form of (9) and the matrix **C** will yield to the properties expressed in this connection. However, since in (7), as well as in (11), its positive square root $\mathbf{D}^{1/2}$ participates, this fact does not obstruct the execution of the conversion (1). It is, yet, important to apprehend, that in this particular case, the initially given k-dimensional sample **X**, after the conversion into the space **Y**, will lose the dispersion in the direction of those eigenvectors, to which correspond zero eigenvalues of the matrix **C**. It could be therefore affirmed, that in this case the conversion (1) changes also the dimension of the eigenspace (k⇒n) – a property that could successfully be used for the solution of some problems connected with pattern recognition.

5. In addition to Deficiencies

Although assisted by the suggested conversion method, any statistical structure could alter as required, it keeps storing some of its deficiencies. Such defects that could not be eliminated by this conversion, we could call non-eliminateable, and we use different means for trouble shooting. A typical instance of such defect in the statistical structure represents the random shift – a least probable realisation, disobeying the (3.σ) rule, when generating the random normally distributed vectors. In the multidimensional space such realisation represents a point, standing non-typically (indiscreet) and obviously out of the main group of points, as shown on the figure 1, right bellow.

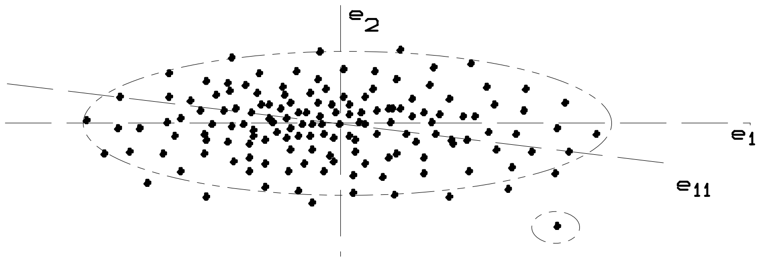

Figure 1: Influence of a randomly shifted point to the main directions

As a consequence of the presence of such a randomly shifted point in the statistical sample, the main directions of the maximum dispersion will be determined with deviation towards their most probable position, which would occur if this point did not exist - and we will have $\mathbf{e}_{11}$ instead of $\mathbf{e}_1$. The presence of randomly shifted points in the samples is determined by the properties of uniformly distributed generators, as well as the properties of the normally distributed generators. Therefore the opposition against these deficiencies is most likely to lead to increase in their quality.

References:

[1]. Bellman R., *Introduction to matrix analysis,* McGraw-Hill Book company, Inc., 1960.

[2]. Deley D.W., *Computer Generated Random Numbers, SYNOPSIS,* 30 Apr. 1996, http:// WWW.VIRTUALSCHOOL.EDU/MON/CRIPTO/RANDOMNUMBERMATH,

[3]. Jambu M., *Classification automatique pour l'analyse des donnes,* Dunod, Paris, 1978.

[4]. Forsythe G., Malcolm M., Moler C., *Computer Methods for Mathematical Computations,* Prentice-Hall, Inc., 1977.

[5]. Fukunaga K., *Introduction to Statistical Pattern Rcognition,* Academic Press, 1972.

[6]. Haas A., *The multiple prime random number generator,* ACM Transaction on Mathematical Software, vol.13, 4, XII.1987, p. 368-381.

[7]. Horn R.A., Johnson C.R., *Matrix Analysis,* Cambridge, University Press, 1986.

[8]. Knuth D.E., *The Art of Computer Programming. Seminumerical Algorithms, vol. 2,* 2nd ed., 1981, Addison Wesley, Reading, MA.

[9]. Lloyd E., Ledermann W., *Handbook of applicable mathematics, vol. 6: Statistics*, John Wiley&Sons Ltd., 1984.

[10]. Manasiev L., Konstantiniva P., Djerassi E., *Using sequential URAND random generator in distributed random generation,* Proceedings of 11-th International Conference "Systems for automation of engineering and research and DECUS NUG Seminar'97, St. Konstantin resort - Varna, Bulgaria, 20-21 Sept. 1997, p. 86-90.

[11]. Microsoft FORTRAN Power Station 4.0.

[12]. Niederreiter H., *Random Number Generation and Quasi-Monte Carlo Methods,* SIAM, Philadelphia, PA, 1992.

[13]. Parlett B.N., *The Symmetric Eigenvalue Problem,* Prentice-Hall, 1980.

[14]. Rabiner L.R., Gold B., *Theory and application of digital signal processing,* Prentice-Hall, New Jersey, 1975.

[15]. Tyanev D.S., *Recognition Rule for Normally Distributed Vectors after Secondary Orthogonal Transformations.* Proceedings of 11-th International Conference "Systems for automation of engineering and research and DECUS NUG Seminar'97, St. Konstantin resort - Varna, Bulgaria, 20-21 Sept. 1997, p. 117-121.

[16]. Sobol I. M., *The Monte-Carlo Method*, Nauka, Moscow, 1985. (in Russian)

[17]. Sobol I. M. *Points, Uniformly Distributed in a Multidimentional Cube*, Znanie, Moscow, 1985. (in Russian)

[18]. Tyanev D.S., *Secondary ortogonal transformations in pattern recognition*. Proceedings of the Maritime Scientific Forum for Communication Systems and Automation - Varna, Bulgaria, 1996, Vol.2, p. 30-37. (in Bulgarian)