Available online at www.sciencedirect.com

ScienceDirect

European Journal of Operational Research 192 (2009) 468–478

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/ejor

Stochastics and Statistics

A matching algorithm for generation of statistically dependent random variables with arbitrary marginals

Nesa Ilich *

University of Calgary, 7128-5 Street NW, Calgary, Alberta, Canada T2K 1C8

Received 8 March 2006; accepted 14 September 2007
Available online 25 September 2007

Abstract

Simulation has gained acceptance in the operations research community as a viable method for analyzing complex problems. While random generation of variables with various marginal distributions has been studied at length, developing ability to preserve a given degree of statistical dependence among them has been lagging behind. This paper includes a short summary of the previous work and a description of the proposed algorithm for efficient re-arranging of generated random variables such that a desired product moment correlation matrix is induced. The proposed approach is different from similar algorithms that induce a desired rank-order correlation among random variables. The algorithm is demonstrated using three numerical examples, one of which also includes a comparison with @RISK commercial package. Its main features are simplicity, ease of implementation and the ability to handle either theoretical or empirical distribution functions.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Simulation; Regression; Stochastic processes; Statistical dependence; Correlation

1. Introduction

The need to generate statistically dependent random variables arises in various fields where simulation has proven to be a useful tool, such as finance, production, natural resources management or scheduling. This paper offers a new approach for generating stochastic variables with desired correlation structure and arbitrary marginal distributions aimed at preserving the product moment correlations.

The basic premise for generating dependencies among random variables is to formulate the process of generation of new dependent variables as linear combinations of independent random variables. The notion of "independent random variables" here denotes de facto random variables that were already generated in previous steps of the generation process. This approach is known to work well with normal distribution. However, Iman and Conover (1982) identified a difficulty associated with the case of stratified samples where the intent is to preserve the desired bounds on the generated variables. The bounds may be violated since the normalized random component in a regressive process must be unbounded to preserve the desired correlation structure and normal distribution. For example, in some regressive processes variables that are not allowed to be negative may actually become negative after addition of normalized random terms. Also, many processes cannot be adequately represented with a normal distribution, while the use of linear combinations in the generating scheme is only guaranteed to preserve normal distribution. Hence, much of the previous research was based on an attempt to find a universal transformation function from a normal distribution to an arbitrary distribution such that both the distribution properties and desired correlations are preserved. The anticipation was that such a transformation would allow a mathematical transition from a set of correlated variables with normal distribution to a set of correlated variables with arbitrary distribution. The following section contains a survey of published work in this regard.

* Tel.: +1 403 7304480; fax: +1 403 2744031.
E-mail address: nilich@optimal-solutions-ltd.com

0377-2217/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2007.09.024

Early efforts by Mardia (1970) were restricted to finding transformations of bivariate random variables with normal distribution into other distributions. Johnson and Ramberg (1977) also considered marginal distributions as functional transformations of normal distributions. They first transformed the original normally distributed vector to a correlated multivariate normal vector, and then converted it from normal into desired marginal distributions. The problem was that the statistics of the transformed vector (i.e. means, variances and correlation) could not be easily controlled and often deviated from the desired targets. A mathematical treatment for some types of distributions (e.g. lognormal) was developed, however this approach lacked generality since it was not applicable to all marginal distributions. It has been recognized that correlated multivariate random vectors and marginal distributions from the same family of distributions have been covered in the literature (Devorye, 1986; Johnson, 1987). However, correlated random variables with distributions that do not originate from the same family have been given much less attention.

The work of Iman and Conover (1982) provided an algorithm which is used today by two commercial simulation software vendors that provide general purpose simulation models, although with a disclaimer that they are only capable of matching the rank correlations between the generated random variables. A valuable aspect of this approach is that the marginal distributions of the original random vectors remain intact, the algorithm merely provides for a key to re-ordering of the elements of the original vectors. In that sense, this was the first truly "distribution free" algorithm since it guaranteed that the original marginal distributions would not change. In this approach the target correlation matrix contains rank correlations, not the Person correlations. Although rank correlation is most frequently used as a measure of statistical dependence, in certain simulation studies a desired goal is to matching the product moment correlations. In such cases the use of the available commercial packages can only be made under the assumption that matching rank correlations was a sufficiently close approximation to matching the Pearson product moment correlations. This assumption may lead to errors of unacceptable magnitude.

Cario and Nelson (1996, 1997) designed the NORTA ("Normal to Anything") method, which has generated significant interest in the research community, although to this date it has found no application in a general commercial simulation software, but is rather restricted to specific applications related to finance. This algorithm begins with generation of a random vector with multivariate normal distribution, which is then transformed to a random vector with desired marginal distributions and correlation matrix. The authors developed a numerical procedure which determines the correlation structure of the initial normal vector such that the correlation structure of the resulting transformed vector with desired marginal distributions is maintained. However, as documented by Li and Hammond (1975) as well as Lurie and Goldberg (1998), some attempts to generate random vectors with arbitrary marginal distributions and with arbitrary feasible correlation have failed. Ghosh and Henderson suggested an adjustment to the method, and a different adjustment was also suggested by Clemen and Reilly (1999). The recent variants of this approach are the QUARTA method ("Quasi-Random to Anything") from Henderson et al., 2000) and VARTA ("Vector Auto-Regressive to Anything") from Biller and Nelson (2003) which relies on the approximation of input vector with Johnson type distributions. The common feature to all variants is an iterative numerical procedure for matching the correlation structure of the initial random vector until the desired correlation structure of the resulting marginal vector is achieved. To help prevent the iterative procedure from failing, Ghosh and Henderson (2002) resorted to the use of semidefinite programming (SDP). Much of their recent efforts were related to developing a procedure that would determine if the target correlation matrix was feasible for a set of given random variables with arbitrary distribution functions. They introduced the notion of 'NORTA defective correlation matrices' if they are feasible and yet cannot be matched using the NORTA method, and conducted numerical experiments in which the failures of the NORTA method were related to increase in dimensionality of the generated random vector (Ghosh and Henderson, 2003). They noted that the probability of failure of the NORTA method is over 80% for random vectors with dimensions that are above 10. Additionally, the recent use of SDP that they proposed has significantly slowed down the execution, reporting for example 10 min run times for simulating correlated random vectors of dimension 10 (Ghosh and Henderson, 2003).

In their work, Ghosh and Henderson (2003) repeatedly state that "for two-dimensional random vectors, the NORTA method can match any feasible correlation matrix. This follows immediately from the characterizations in Whitt (1975)." The idea in this paper is based on extending this concepts for vectors which are multi-dimensional, using a systematic approach of re-arranging the elements of vectors $X_k, k = 2, \ldots, n$ based on the use of multiple regression fit as a measure of statistical dependence. Hence, the focus of this paper is a method of re-arranging the elements of each generated random variable in order to induce a desired statistical dependence. In addition to execution speed and the ease of implementation, additional advantages of the proposed method are

(a) The method preserves the Pearson correlations instead of the rank correlations. This may sometimes be preferable to preserving the rank correlation.
(b) The method can be used to induce desired statistical dependence among random variables which are derived from empirical distributions. Recent advances in kernel distribution functions (Silverman, 1986; Scott, 1992) have gained momentum among researchers since they offer more flexibility for statis-

Download English Version:

https://daneshyari.com/en/article/481184

Download Persian Version:

https://daneshyari.com/article/481184

Daneshyari.com