Proceedings of the 1994 Winter Simulation Conference
ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila

GENERATING CORRELATED RANDOM VARIATES BASED ON AN ANALOGY BETWEEN CORRELATION AND FORCE

Edward F. Mykytka

Air Force Institute of Technology/ENS
2950 P Street
Wright-Patterson AFB, OH 45433-7765

Chun-Yuan Cheng

Chao Yang Institute of Technology
168, Gi Feng E. Rd.
Wu Feng, Tai Chung County, Taiwan, R.O.C.

ABSTRACT

An approach for generating dependent random variates with specified marginal distributions and measure of correlation is developed based upon an analogy between correlation and force. We first examine the mathematical transformation implied by this analogy and attempt to construct a general algorithm for generating variates that have the desired properties. We then examine the limitations of this approach, noting some potentially surprising distributional consequences associated with linear combinations of random variables. We conclude with the development of an algorithm for generating correlated Uniform(0,1) random variates with (approximately) a specified product-moment correlation.

1 INTRODUCTION

An important issue in simulation concerns the generation of correlated random variates. Most existing methods generally assume that the joint or conditional distributions of the variates can be completely specified. [For some good summaries of methods, see Johnson, Wang, and Ramberg (1984), Devroye (1986), or Johnson (1987)]. In many practical situations, however, such complete information may not be known. Methods particularly designed for such contexts are relatively scarce, with the notable exception of Johnson and Tenenbein (1981). Other relevant works include Plackett (1963), Kimeldorf and Sampson (1975a), and Ali, Mikhail, and Haq (1978).

In this paper, we consider the general problem of generating a pair of simulated values (x,y) from the joint distribution of two random variables X and Y when only the marginal distributions of X and Y and a measure of the correlation between them are known. In particular, we assume that the dependence between the two variates is measured by the product-moment correlation coefficient

$$\rho = \sigma_{XY}/\sigma_X \sigma_Y$$

where σ_{XY} denotes the covariance between X and Y defined by

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$$

and where μ_X, μ_Y, σ_X, and σ_Y denote the means and standard deviations of X and Y, respectively.

If X and Y are independent, then $\rho = 0$ and we can produce the desired point by independently generating x from the marginal distribution of X and y from the marginal distribution of Y. If X and Y are dependent, this method is inadequate since it obviously does not account for the correlation between them. We will account for this dependence by first considering an analogy between correlation and force that was introduced to the authors by Huang (1986) who provides a unique view of the analogy from the perspective of computer-graphics windows.

2 CORRELATION AND FORCE

Correlation is a measure of the degree of dependence between two random variables. Thus, if X and Y are positively correlated, then large (small) values of X tend to be associated with large (small) values of Y, and vice-versa. Although a cause-and-effect relationship may not actually exist between the variables, it is intuitively appealing to envision, however, that the value of one random variable influences (or exerts a force upon) the value of the other.

We thus consider accounting for the dependence between X and Y by first generating x and y as if they were independent and then "repositioning" y to account for the "force" on y exerted by the particular value of x. Although this may seem rather novel, the idea is not new. Indeed, a similar idea applies in the general random variate generation technique known as "conditional sampling" wherein the influence of the particular value of x on y is accounted for by first generating x from its marginal distribution and then generating y from the conditional distribution of Y given x.

Assume now, for clarity of presentation, that X and Y are positively correlated and that values x_0 and y_0 have been obtained independently from the marginal distributions of and Y, respectively. In order to induce the desired correlation, consider exerting a force F, proportional to the distance between x_0 and its mean μ_X, on the point (x_0,y_0) parallel to the y-axis, as shown in Figure 1.

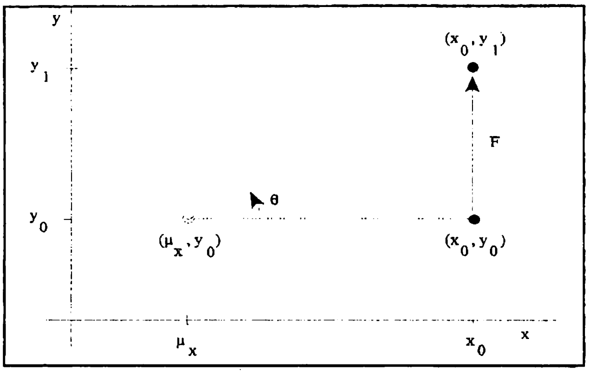

Figure 1 Influence of Force on (x_0, y_0).

The result of this force is to reposition the point (x_0,y_0) to (x_0,y_1) where $y_1 = y_0 + F$. Using the point (μ_X,y_0) as reference, we can relate the direction and magnitude of the force F to the angle θ (shown in Figure 1) via

$$F = \tan(\theta)(x_0 - \mu_X).$$

This repositioning of y_0 is equivalent to transforming the random variable Y into a new random variable Y_1 via

$$Y_1 = F + Y = \tan(\theta)(X - \mu_X) + Y.$$

2.1 Properties of the Transformed Variate Y_1

Since this last equation implicitly defines Y_1 as a linear function of X and Y, it can be immediately established that (i) the mean of Y_1 is the same as that of Y (i.e., $\mu_{Y_1} = \mu_Y$), (ii) its variance is

$$\sigma_{Y_1}^2 = \sigma_Y^2 + \tan^2(\theta)\,\sigma_X^2,$$

and (iii) its correlation with X is

$$\rho_{XY_1} = \tan(\theta)\frac{\sigma_X}{\sigma_{Y_1}} = \frac{\tan(\theta)\sigma_X}{\sqrt{\sigma_Y^2 + \tan^2(\theta)\sigma_X^2}}.$$

If we next select θ so that this correlation is equal to the value of ρ desired, i.e., so that $\rho_{XY_1} = \rho$, we find that

$$\tan(\theta) = \frac{\rho}{\sqrt{1-\rho^2}}\left(\frac{\sigma_Y}{\sigma_X}\right).$$

Choosing θ in this manner thus produces a variate Y_1 with the same mean as Y and the desired correlation with X but, unfortunately *without* the same variance as Y. To overcome this, we transform Y_1 into a new random variate Y_2 via

$$Y_2 = \frac{\sigma_Y}{\sigma_{Y_1}}Y_1 + \left(1 - \frac{\sigma_Y}{\sigma_{Y_1}}\right)\mu_Y.$$

By doing so, we produce a random variate Y_2 that has, by construction, the same mean and variance as Y and also has correlation ρ with X. The overall transformation of Y into Y_2 can be simplified to obtain:

$$Y_2 = \frac{\rho\sigma_Y}{\sigma_X}X + \sqrt{1-\rho^2}\,Y - \frac{\rho\,\sigma_Y\mu_X}{\sigma_x} + \left(1 - \sqrt{1-\rho^2}\right)\mu_Y.$$

2.2 Properties of the Transformed Variate Y_2

Although the preceding transformation produces a variate that has the desired correlation with X and the same mean and variance as Y, Y_2 can not, in general, be expected to have the same marginal distribution as Y. One exception, however, occurs when X and Y are both normally distributed. Then (as a consequence of the reproductive property of the normal) Y_2 will also be normally distributed with the same mean and variance as Y. Indeed, in this case, the approach we have developed leads to a standard method for generating correlated normal random variates.

Unfortunately, such a reproductive property does not apply in general to non-normal distributions. For example, suppose X and Y are two independent, Uniform(0,1) random variables. Then, the marginal probability density /unction (PDF) of Y_2 depends on ρ and has a decidedly non-uniform shape, as evidenced by the examples in Figure 2. These vividly demonstrate that the marginal distribution of Y_2 is *not* Uniform(0,1) nor is its range the [0,1] interval.

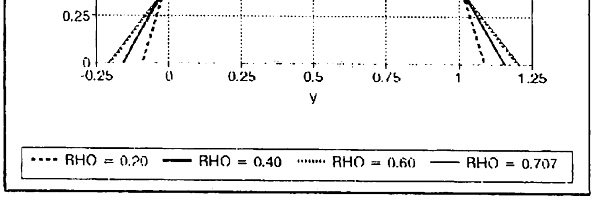

Figure 2 PDF's of Y_2 as a Function of ρ

3 GENERATING CORRELATED UNIFORMS

We have seen that the transformation developed thus far only preserves the mean and variance of Y and its correlation with X but not its marginal distribution. Hence, it does *not* directly yield a viable method for generating random variates with specified marginals and product-moment correlation (unless, of course, we are clever enough to determine the marginal distribution of Y that would produce the desired marginal distribution for Y_2).

We thus retreat from this general objective and consider refining the approach to produce bivariate Uniform(0,1) random variates. Our rationale is that, if we can produce two Uniform(0,1)'s with the desired correlation, then we should, by transforming these Uniform(0,1)'s appropriately, be able to produce variates with other distributions and approximately the same correlation. Such an approach has been used by a number of other authors including Barnett (1980), Devroye (1986), Johnson, Wang and Ramberg (1984) and Kimeldorf and Sampson (1975a, 1975b).

[Whether or not the resulting variates will have *exactly* the same correlation as the bivariate Uniform(0,1)'s will depend, in general, (i) on the type of transformation applied and (ii) the measure of correlation used. For a more rigorous discussion which treats alternate measures of correlation and transformations under which these measures of correlation are invariant, see Cheng (1991).]

Now, in the preceding section, we produced a random variate, Y_2, that has correlation ρ with X but does *not* have the desired marginal distribution. To now produce a variate with a Uniform(0,1) distribution, recall that, if Y is a random variable with cumulative distribution function (CDF) F whose inverse F^{-1} exists, then the transformation $U = F^{-1}(Y)$ yields a random variable U that is Uniform(0,1). Thus, we transform Y_2 into a new variate Y_3 via

$$Y_3 = F_{Y_2}^{-1}(Y_2),$$

where $F_{Y_2}^{-1}(y)$ denotes the inverse CDF of Y_2 (the derivation of this is omitted here).

By construction, then, the transformed variate Y_3 has a Uniform(0,1) distribution but, on the other hand, this variate can not be expected to have correlation ρ with X, as did Y_2. However, it turns out that correlation between X and Y_3 can be expressed as the following function of ρ:

$$\rho_{XY_3} = \begin{cases} \dfrac{\rho}{\sqrt{1-\rho^2}} - \dfrac{3\rho^2}{10(1-\rho^2)}, & 0 \le \rho \le \dfrac{1}{\sqrt{2}}, \\ 1 - \dfrac{1-\rho^2}{2\rho^2} + \dfrac{(1-\rho^2)^{3/2}}{5\rho^3}, & \dfrac{1}{\sqrt{2}} \le \rho \le 1. \end{cases}$$

A plot of this correlation as a function of the desired correlation ρ is provided in Figure 3, which clearly demonstrates that $\rho_{XY_3} \approx \rho$. If one does not consider this approximation to be sufficiently close, then the preceding relation can be used to determine (approximately) the value of ρ which would produce the value of ρ_{XY_3} desired.

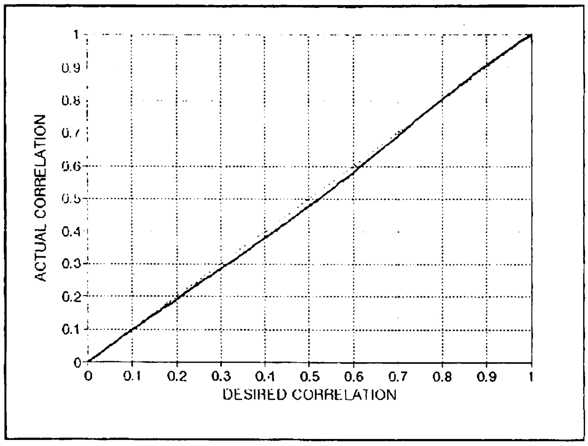

Figure 3 Correlation of X and Y_3 as a Function of ρ

3.1 An Algorithm

The preceding results can be summarized as an algorithm for generating correlated Uniform(0,1) random variates with specified product-moment correlation $\rho^* > 0$. [If negatively correlated variates are desired (i.e., if $\rho^* < 0$), first generate X and Y_3 with correlation $|\rho^*| > 0$ and then return $U_1 = X$ and $U_2 = 1 - Y_3$ as the desired variates.]

Step 0: (Optional) Use the preceding equation to find the value of ρ to yield the desired correlation ρ^*.

Step 1: Set

$$A = \rho^*,$$

$$B = \sqrt{1 - (\rho^*)^2},$$

and

$$C = \frac{1}{2}(1 - \rho^* - \sqrt{1 - (\rho^*)^2})$$

Step 2: Generate two independent Uniform(0,1) random numbers X and Y and let

$$Y_2 = AX + BY + C$$

Step 3: If $0 \le \rho^* \le 1/\sqrt{2}$, then set

$$Y_3 = \begin{cases} \dfrac{(Y_2 - C)^2}{2AB}, & \text{if } C \le Y_2 \le A + C, \\ \dfrac{Y_2 - (A/2) - C}{B}, & \text{if } A + C \le Y_2 \le B + C, \\ 1 - \dfrac{(A + B + C - Y_2)^2}{2AB}, & \text{if } B + C \le Y_2 \le B + C \le A + B + C \end{cases}$$

otherwise, i.e., if $1/\sqrt{2} \le \rho' \le 1$), set

$$Y_3 = \begin{cases} \dfrac{(Y_2 - C)^2}{2AB}, & \text{if } C < Y_2 < B + C, \\ \dfrac{Y_2 - (B/2) - C}{A}, & \text{if } B + C \le Y_2 < A + C, \\ 1 - \dfrac{(A + B + C - Y_2)^2}{2AB}, & \text{if } A + C < Y_2 < B + C \le A + B + C \end{cases}$$

Step 4 Return $U_1 = X$ and $U_2 = Y_3$ as the desired variates.

3.2 A Final Observation

The algorithm we have produced is similar to one developed by Johnson and Tenenbein (1981) who considered using weighted linear combinations of independent random variates to produce correlated random variates. In particular, they proposed generating independent variates U and V from a specified "underlying distribution" and transforming these into a pair of dependent variables via

$$U' = U$$

and

$$V' = cU + (1 - c)V.$$

It can be shown [Cheng (1991)] that our algorithm is equivalent to Johnson and Tenenbein's with a Uniform(0,1) "underlying distribution" and A = c and B = 1 - c.

4 CONCLUDING REMARKS

Our purpose has been to explore the implications of using the correlation/force analogy in the development of procedures for generating correlated random variates. Although this approach does not produce a general method for generating random variables with specified marginals and correlation, it does lead to a viable approach for generating correlated Uniform(0,1) random numbers and helps to establish a more intuitive perspective on the problem.

ACKNOWLEDGEMENTS

The authors would like to gratefully acknowledge the support of the Department of Industrial Engineering at Auburn University for its support during the conduct of the vast majority of this research.

REFERENCES

Ali, M. M., N. N. Mikhail, and M. S. Haq. 1978. A class of bivariate distributions including the bivariate logistic. *Journal of Multivariate Analysis* 8: 405-412.

Barnett, V. 1980. Some bivariate uniform distributions. *Communications in Statistics* A9: 453-461.

Cheng, C. Y. 1991. The generation of random vectors with specified marginal distributions. Ph.D. dissertation, Industrial Engineering Department, Auburn University, Auburn, Alabama.

Devroye, L. 1986. *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.

Huang, J. S. 1986. Development of a risk simulation environment capable of modeling dependencies. Ph.D. dissertation, Industrial Engineering Department, Auburn University, Auburn, Alabama.

Johnson, M. E. 1987. *Multivariate statistical simulation*, New York: John Wiley & Sons.

Johnson, M. E. and A. Tenenbein. 1981. A bivariate distribution family with specified marginals. *Journal of the American Statistical Association* 76: 198-201.

Johnson, M. E., C. Wang, and J. S. Ramberg. 1984. Generation of continuous multivariate distributions for statistical applications. *American Journal of Mathematical and Management Sciences* 4: 225-248.

Kimeldorf, G. and A. Sampson. 1975a. One-parameter families of bivariate distributions with fixed marginals. *Communications in Statistics* 4: 293-301.

Kimeldorf, G. and A. Sampson. 1975b. Uniform representations of bivariate distributions. *Communications in Statistics* 4: 617-627.

Plackett, R. L. 1965. A class of bivariate distributions. *Journal of the American Statistical Association* 60: 51-522.

AUTHOR BIOGRAPHIES

EDWARD F. MYKYTKA is an associate professor of operations research at the Air Force Institute of Technology. He received a Ph.D. degree in Systems Engineering from the University of Arizona. His teaching and research interests lie in applied statistics, simulation, and quality improvement.

CHUN-YUAN CHENG is an associate professor in the Department of Industrial Engineering at the Chao Yang Institute of Technology. She received a Ph.D. in Industrial Engineering from Auburn University. Her research and teaching interests lie in applied statistics, simulation, and computer-integrated manufacturing.