*Proceedings of the 1993 Winter Simulation Conference*
*G.W. Evans, M. Mollaghasemi, E.C. Russell, W.E. Biles (eds.)*

# GENERATION OF AUTOCORRELATED RANDOM VARIABLES WITH A SPECIFIED MARGINAL DISTRIBUTION

Wheyming Tina Song
Department of Industrial Engineering
National Tsing-Hua University
Hsinchu, Taiwan, R.O.C.

Li-Ching Hsiao
School of Industrial Engineering
Purdue University
Lafayette, IN 47907, U.S.A.

## ABSTRACT

In this paper we present an easily implemented procedure for the generation of autocorrelated random variables having a specified marginal distribution and a fixed lag-one autocorrelation function.

## 1 INTRODUCTION

Simulation has become a very powerful tool for the analysis of a wide variety of problems. Estimation of a system's performance by means of simulation often requires autocorrelated random variables as input. Some examples of queueing models with correlated arrival and service times are given by Heffes (1973 and 1980), Heffes and Lucantoni (1986), and Lee et al. (1991).

A series of autocorrelated random variables (autocorrelated series) can usually be characterized by two factors: its marginal distribution and its correlation structure. We consider a simple case: generation of a stationary series $\{Y_i\}_{i=1}^n$, which we refer to as a *target series*, with a marginal distribution $F_Y$ and a lag-one autocorrelation $\rho_Y$. Most existing methods for generating autocorrelated series fall into three classes:

(1) Correlation-oriented approach: This approach is exemplified in most of the papers by Lewis and co-workers (Lewis [1980, 1985], Lawrance and Lewis [1981, 1987]). This approach develops a recursive algorithm for $Y_i$ given $Y_{i-1}$. For each marginal distribution $F_Y$, $\rho_Y$ is provided as a parameter to the algorithm. The advantage is that the autocorrelation function is tractable (therefore, we use the name correlation-oriented). The limitation is that a separate algorithm must be devised for different $F_Y$.

(2) Marginal-oriented approach: This approach first transforms a known autocorrelated series, which is referred to as a *reference series*, into its corresponding uniform autocorrelated random variables, then applies an inverse transformation method to generate the target series. The marginal distribution is easily preserved (therefore, we use the name marginal-oriented), while the correlation becomes intractable through two stages of transformation. Lakhan (1981) presented this approach and empirically derived the relationship among the autocorrelations of the uniform, Rayleigh, and exponential distributions corresponding to a given autocorrelation. But Lakhan's results are affected by error in the autocorrelation of his reference series. Schmeiser (1990) also presented this approach to generate a random vector. However, he did not discuss how to get a specified correlation for the random vector. Melamed et al. (1992) presented a special case, called TES methodology, that uses correlated uniform series as a reference series, therefore avoiding the first transformation in the marginal-oriented approach. A heuristic search is used in TES methodology for obtaining the correlation of the target series.

(3) Joint-distribution oriented approach: In this approach, the conditional probability of $Y_{i+1}$ given by $Y_i$ is derived by dividing the bivariate distribution of $Y_i$ and $Y_{i+1}$ by the marginal marginal distribution of $Y_i$. This approach shifts the problem to one of generating bivariate distributions. Johnson and Tenenbein (1981) describe a general scheme for generating continuous bivariate distribution with specified marginals and several dependence measures. Schmeiser and Lal (1982) show how to generate bivariate gamma random vectors with any correlation.

In this paper, we follow the marginal-oriented approach to reach our goal. As mentioned above, autocorrelation is not invariant through the necessary conversions. We give an iterative procedure for determination of a lag-one autocorrelation of a reference series $\{X_n\}$ such that the resulting target series $\{Y_n\}$ exhibits a specific value of $\rho_Y$. Fortran implementation which allows very general choices as to the marginal distribution in the target series is available from the authors.

## 2 PROCEDURE

To generate a stationary series $Y_1, Y_2, \cdots$, with a marginal distribution $F_Y$ and a lag-one autocorrelation $\rho_Y$, we propose an iterative procedure. Each iteration goes through Steps 1-5. Step 0 is a setup for initialization. We denote $k$ as the iteration number, and use $h_Y, l_Y, h_X$, and $l_X$ as intermediate variables. Below is a statement of our algorithm. An explanation follows this statement.

Step 0. Initialization.

$k := 1$, the first iteration,
$\rho_X := \rho_Y$, the initial approximation of $\rho_X$,
$h_Y := 1$,
$l_Y := -1$,
$h_X := 1$,
$l_X := -1$.
(The symbol $:=$ is read "given the value of.")

Step 1. Apply marginal-oriented approach.

(i) Generate a reference series $\{x_i\}_{i=1}^n$, with a lag-one autocorrelation $\rho_X$. e.g. $X_i \sim$ AR(1), i.e., $X_i = \phi X_{i-1} + \epsilon_i$, where $\epsilon_i \sim$ iid Normal(0,$\sigma_\epsilon{}^2 = 1 - \phi^2$ ), and $\phi = \rho_X$. Then $X_i \sim$ Normal(0,1).

(ii) $u_i = F_X(x_i) \equiv \mathrm{P}(X \le x_i)$, $i = 1, 2, \cdots, n$

(iii) $y_i = F_Y{}^{-1}(u_i)$, $i = 1, 2, \cdots, n$

Step 2. Estimate $\mathrm{corr}(Y_i, Y_{i+1})$.

Repeat Step 1.(i)-(iii) $m$ times, each with a size $n$. Each replication generates one estimate of $\mathrm{corr}(Y_i, Y_{i+1})$. We then have $m$ estimates which are denoted as $\hat\rho_Y^{(1)}, \hat\rho_Y^{(2)}, \cdots, \hat\rho_Y^{(m)}$.
The estimator of $corr(y_i, y_{i+1})$ and its standard error at iteration $k$ are defined as

$$\hat\mu(k) \equiv \bar{\hat\rho}_Y \equiv \frac{\sum_{j=1}^m \hat\rho_Y^{(j)}}{m}$$

$$\hat\sigma(k) \equiv \mathrm{s.e.}(\bar{\hat\rho}_Y) = \sqrt{\frac{\sum_{j=1}^m (\hat\rho_Y^{(j)} - \bar{\hat\rho}_Y)^2}{m(m-1)}}$$

If $|\hat\mu(k) - \rho_Y| \le \hat\sigma(k)$, no further iteration is needed and stop.

Step 3. Update intermediate variables.

if $\hat\mu(k) > \rho_Y$,
$h_Y := \hat\mu(k)$,
$h_X := \rho_X$;

if $\hat\mu(k) < \rho_Y$,
$l_Y := \hat\mu(k)$,
$l_X := \rho_X$.

Step 4. Adjust $\rho_X$.

$\rho_X := \rho_X + \triangle$, where

$$\triangle \equiv \begin{cases} \triangle_1 & \text{if } \hat\mu(k) < \rho_Y \quad \text{and if } h_Y = 1 \\ -\triangle_1 & \text{if } \hat\mu(k) > \rho_Y \quad \text{and if } l_Y = -1 \\ \triangle_2 & \text{if } \hat\mu(k) < \rho_Y \quad \text{and if } h_Y < 1 \\ -\triangle_3 & \text{if } \hat\mu(k) > \rho_Y \quad \text{and if } l_Y > -1. \end{cases}$$

The increment or decrement values are defined as

$$\triangle_1 \equiv |\hat\mu(k) - \rho_Y|,$$
$$\triangle_2 \equiv \left|\frac{[\rho_Y - \hat\mu(k)][h_X - \rho_X]}{h_Y - \hat\mu(k)}\right|,$$
$$\triangle_3 \equiv \left|\frac{[\rho_Y - \hat\mu(k)][\rho_X - l_X]}{\hat\mu(k) - l_Y}\right|.$$

Step 5. Update the iteration number.

$k := k + 1$ and go to Step 1.

We start with a selected autocorrelated series as a reference series in Step 1(i). A reference series is not fully determined by the specification of $F_Y$ and $\rho_Y$ for the target series; one may still choose among various autocorrelated series. The first-order autoregressive (AR(1)) process with normal marginal distribution (Box and Jenkins 1976), exponential AR(1) with exponential marginal distribution (Lewis 1980), and Correlated U(0,1) (Melamed 1991) have all been used with success. Inspection of the bivariate joint distribution, shown in Song et al. (1993), motivates selection of AR(1) as a reference series for a broad range of systems. The unusual behavior in the end effect exhibited by Correlated U(0,1), and the truncated characteristic exhibited by exponential AR(1) recommend those series for certain purposes. See Melamed (1991).

We choose AR(1) as a reference series in Step 1(i). As an initial approximation, we choose the lag-one autocorrelation $\rho_X$ equal to the target value $\rho_Y$. Steps 1(i and ii) generate correlated random numbers for use in the inverse transformation in Step 1(iii).

Since Steps 1 (ii and iii) are both nonlinear transformations, the autocorrelation $\rho_Y$ will usually vary from $\rho_X$. Empirical results (Song et al. (1993)) show that the relation between $\rho_Y$ and $\rho_X$ is always monotonically increasing. Thus, if the value of $\rho_X$ used in Step 1(i) does not produce the target $\rho_Y$, we can adjust $\rho_X$ in the same direction as the necessary change in $\rho_Y$.

In Step 2 we estimate the lag-one autocorrelation of $Y$, $\mathrm{corr}(Y_i, Y_{i+1})$. The estimator of $\mathrm{corr}(Y_i, Y_{i+1})$ and its standard error at iteration $k$ are denoted as $\hat{\mu}(k)$ and $\hat{\sigma}(k)$. We take $|\hat{\mu}(k)-\rho_Y| \leq \hat{\sigma}(k)$ as our criterion for accepting $\rho_X$ and terminating the iterative procedure. If $\rho_Y$ fails this criteria, proceed to Step 3.

In Step 3, we update intermediate variables. In Step 4, we adjust $\rho_X$. For the first iteration (i.e. $k = 1$), if $\hat{\mu}(1) > \rho_Y$, we adjust $\rho_X$ by taking $\triangle_1 = |\hat{\mu}(1) - \rho_Y|$ as the decrement. If $\hat{\mu}(1) < \rho_Y$, we take $\triangle_1$ as the increment. In subsequent iterations ($k = 2, 3, \ldots$), we keep taking $\triangle_1 = |\hat{\mu}(k) - \rho_Y|$ as the decrement if $\hat{\mu}(j) > \rho_Y$ for any iteration $j \leq k$. Similarly, we keep taking $\triangle_1$ as the increment if $\hat{\mu}(j) < \rho_Y$ for any iteration $j \leq k$. Otherwise, we use interpolation to adjust $\rho_X$. The corresponding increment and decrement are given by $\triangle_2$ and $\triangle_3$.

In Step 5, we update the iteration number $k$ and proceed by reapplication of Steps 1 through 5.

## 3 EXAMPLES

We apply the procedure of Section 2 to a variety of target series. In all examples the reference and target series length $n = 1000$ and the number of replications $m = 20$. The values of target lag-one correlations $\rho_Y$ are $0.1, 0.2, \ldots, 0.9$. Distributions of target marginals $F_Y$ are Exponential with rate 1, Gamma distribution with shape parameter 7 and scale parameter 1, F distribution with degrees of freedom 7 and 10; see Song et al. (1993) for more examples. For the inverse transformation in Step 1(iii) we use the Fortran code provided by Ding (1987).

In Tables 1 to 3, the first column is the target value $\rho_Y$. The second column gives the number of iterations required to reach a suitable $\rho_X$. The third column is the resulting value of $\rho_X$ (for which $\bar{\hat{\rho}}_Y$ satisfies the termination criterion). Corresponding values of $\bar{\hat{\rho}}_Y$ and s.e.$(\bar{\hat{\rho}}_Y)$ are given for reference in columns 4 and 5, respectively.

The procedure converges rapidly, never requiring more than three iterations in any of the example cases. The computation time per iteration is not more than two and half minutes on 486-based PC computer. Most of the computation time results from computing the inverse transformation in Step 1(iii).

Table 1. Exponential (1)

| target $\rho_Y$ | iterations required | resulting $\rho_X$ | $\bar{\hat{\rho}}_Y$ given $\rho_X$ | s.e.$(\bar{\hat{\rho}}_Y)$ |
|---|---|---|---|---|
| 0.1000 | 3 | 0.1192 | 0.1022 | 0.0076 |
| 0.2000 | 3 | 0.2321 | 0.2006 | 0.0076 |
| 0.3000 | 3 | 0.2405 | 0.3002 | 0.0075 |
| 0.4000 | 3 | 0.4452 | 0.4015 | 0.0070 |
| 0.5000 | 3 | 0.5458 | 0.5024 | 0.0065 |
| 0.6000 | 2 | 0.6501 | 0.6068 | 0.0071 |
| 0.7000 | 2 | 0.7438 | 0.7034 | 0.0062 |
| 0.8000 | 2 | 0.8344 | 0.8017 | 0.0047 |
| 0.9000 | 2 | 0.9201 | 0.8977 | 0.0031 |

Computer time per iteration about 1 min. (486PC,DX-33)

Table 2. Gamma(7,1)

| target $\rho_Y$ | iterations required | resulting $\rho_X$ | $\bar{\hat{\rho}}_Y$ given $\rho_X$ | s.e.$(\bar{\hat{\rho}}_Y)$ |
|---|---|---|---|---|
| 0.1000 | 1 | 0.1000 | 0.0960 | 0.0066 |
| 0.2000 | 1 | 0.2000 | 0.1943 | 0.0068 |
| 0.3000 | 1 | 0.3000 | 0.2934 | 0.0068 |
| 0.4000 | 1 | 0.4000 | 0.3934 | 0.0066 |
| 0.5000 | 1 | 0.5000 | 0.4941 | 0.0061 |
| 0.6000 | 1 | 0.6000 | 0.5959 | 0.0056 |
| 0.7000 | 1 | 0.7000 | 0.6974 | 0.0050 |
| 0.8000 | 1 | 0.8000 | 0.7994 | 0.0043 |
| 0.9000 | 1 | 0.9000 | 0.8989 | 0.0033 |

Computer time per iteration about 1.5 min. (486PC,DX-33)

Table 3. F(7,10)

| target $\rho_Y$ | iterations required | resulting $\rho_X$ | $\bar{\hat{\rho}}_Y$ given $\rho_X$ | s.e.$(\bar{\hat{\rho}}_Y)$ |
|---|---|---|---|---|
| 0.1000 | 2 | 0.1206 | 0.0994 | 0.0084 |
| 0.2000 | 2 | 0.2393 | 0.1977 | 0.0090 |
| 0.3000 | 2 | 0.3538 | 0.2951 | 0.0096 |
| 0.4000 | 2 | 0.4635 | 0.3941 | 0.0101 |
| 0.5000 | 2 | 0.5682 | 0.4947 | 0.0105 |
| 0.6000 | 2 | 0.6674 | 0.5937 | 0.0098 |
| 0.7000 | 2 | 0.7595 | 0.6963 | 0.0072 |
| 0.8000 | 2 | 0.8478 | 0.8032 | 0.0050 |
| 0.9000 | 2 | 0.9275 | 0.9000 | 0.0026 |

Computer time per iteration about 2.5 min. (486PC,DX-33)

For certain types of the target distributions, the result that shows the relationship between $\rho_Y$ and $\rho_X$ is independent of the distribution parameters chosen for the computation. Thus, in Table 1 the exponential (1) was selected, but the reported result is not dependent upon that selection of parameter 1. This is so because the distribution is invariant with linear transformation of the random variable. For the same reason, the mean and variance for Normal distribution, the rate for Rayleigh distribution, the scale parameter for Gamma distribution, and exchanging two parameters for beta distribution do not affect results.

## 4 CONCLUSION

Simulation modeling frequently calls for generating autocorrelated processes. The algorithm presented in this paper can be efficiently applied to any distribution for which the inverse transformation of the cumulative distribution function can be calculated or approximated.

## ACKNOWLEDGMENTS

This work was supported by the National Science Council of Republic of China under Grant No. NSC82-0415-E007-042. The authors are grateful to Bruce Schmeiser, Jung-Hong Chuang, and Mary Johnson for their helpful suggestions.

## REFERENCES

Box, G.E.P. and Jenkins, G.M. 1976. *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

Ding, C.G. 1987. *Computational Tools for Interval Testing.* Ph.D. Dissertation, Department of Statistics, University of Georgia, Athens, Georgia, U.S.A.

Heffes, H. 1973. Analysis of First-Come First-Served Queueing Systems with Peaked Inputs. *Bell System Technical Journal* 52, 1215-1228.

Heffes, H. 1980. A Class of Data Traffic Processes-Covariance Function Characterization and Related Queueing Results. *Bell System Technical Journal* 59, 897-929.

Heffes, H. and Lucantoni, D.M. 1986. A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance. *IEEE Journal on Selected Areas in Communications SAC-4*, 856-868.

Johnson, M.E. and Tenenbein, A. 1981. A Bivariate Distribution Family with Specified Marginals. *Journal of the American Statistical Association* 76, 198-201.

Lee, D.-S., B. Melamed, A. Reibman, and B. Sengupta. 1991. Analysis of a video multiplexer using TES as a modeling methodology. *Proceeding of GLOBCOM'91*, 61-20, Phoenix, Arizona.

Lewis, P.A.W. 1980. Simple Models for Positive -Valued and Discrete-Valued Time series with ARMA Correlation Structure In: *Multivariate Analysis-V* (P.R. Krishnaiah, ed.) North Holland, 151-166.

Lewis, P.A.W. 1985. Some Simple Models for Continuous Variate Time Series. *Water Resources Bulletin* 21, 635-644.

Lawrance, A.J. and Lewis, P.A.W. 1981. A New Autoregressive Time Series Model in Exponential Variables (NEAR(1)). *Adv. Appl. Prob* 13, 826-845.

Lawrance, A.J. and Lewis, P.A.W. 1987. Modeling and Residual Analysis of Nonlinear Autoregressive Time Series in Exponential Variables. *J. of the Royal Statist. Soc.* 47, 165-202.

Lakhan, V.C. 1981. Generating Autocorrelated Pseudo-Random Numbers with Specific Distribution. *J. Stat. Comput.* 12, 303-309.

Melamed, B. 1991. TES: A Class of Methods for Generating Autocorrelated Uniform Variates. *ORSA Journal on Computing* 3, 317-329.

Melamed., B, Goldsman, D., and Hill, J. 1992. The TES Methodology: Modeling Empirical Stationary Time Series. *Proceedings of the 1992 Winter Simulation Conference*, 135-144.

Schmeiser, B.W. 1990. Simulation Experiments. In *Handbook of Operations Research and Management Science:* Vol. 2, (ed. D. Heyman and M. Sobel), New York: North Holland.

Schmeiser, B.W. and R. Lal. 1982. Bivariate Gamma Random Vectors, *Operations Research* 30, 355-374.

Song, W.-M., Chen, H., and Hsiao, L., 1993. Generation of Autocorrelated Random Variables with a Specified Marginal and a Lag-one Correlation. Technical Report 93-6. Department of Industrial Engineering, National Tsing Hua University, Taiwan, R.O.C.

## AUTHOR BIOGRAPHIES

Wheyming Tina Song is an associate professor in the Department of Industrial Engineering at the National Hsing Hua University, Taiwan, R.O.C. She received her undergraduate degree in statistics and master's degree in industrial engineering at Cheng-Kung University in Taiwan. She then received her master's degree in applied mathematics and industrial engineering from the University of Pittsburgh. Her Ph.D. is from the School of Industrial Engineering at Purdue University. Her research interests are statistical aspect of digital-computer stochastic simulation, including input modeling, output analysis, and variance reduction.

Li-Ching Hsiao is a Ph.D candidate in the School of Industrial Engineering at Purdue University, majoring in simulation and applied statistics. She received her undergraduate degree in industrial management at Cheng-Kung University, and she received master's degree in industrial engineering at National Tsing Hua University in Taiwan, R.O.C.