Proceedings of the 1983
Winter Simulation Conference
S. Roberts, J. Banks, B. Schmeiser (eds.)

MONOTONIC GENERATION OF POSITIVE RANDOM VARIABLES

LOUIS R. MOORE
SCHOOL OF BUSINESS ADMINISTRATION
CARROLL HALL 012A
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27514

This paper presents methods to invert efficiently the distribution function of an arbitrary positive random variable. Such methods preserve the monotonic relationship between the random variable and the uniform deviate which generates it. This relationship is necessary to reduce the variance of an estimator in simulation experiments. For discrete distributions, an indexed search method is employed utilizing variable spaced cutpoints. For continuous distributions, a piecewise continuous increasing quadratic spline is fit to prespecified values of the inverse distribution function. The index number of the piece is then generated by the indexed search method. The spacing of the cutpoints for the search is chosen to minimize the expected number of comparisons required per variate generated.

1. INTRODUCTION

Suppose that a positive finite random variable X has distribution function, $F(t)=P[X\leq t]$ for all real t. The inverse distribution function may be defined for $0\leq u<1$ as

$$G(U)=\inf\{x\geq 0: F(x)\geq u\} .$$

A straightforward way to generate X is to sample a uniform (0,1) deviate U and to calculate $G(U)$. This method, called the inverse distribution function method, is desirable and perhaps costly.

This desirability stems from the observation of Hoeffding (1940) that the antithetic variates $G(U)$ and $G(1-U)$ achieve the most negative correlation possible. See also Whitt (1976). Fishman and Huang (1980) and Fishman (1983) use an extension of antithetic variates called rotation sampling to significantly reduce the variation in the sample mean generaged by a simulation experiment. The cost of such a method is high for arbitrary distributions since a search of the distribution function is generally required. Even when the distribution function is assumed to have a certain form, say the exponential distribution, the inversion operation may require expensive operations, such as logarithms. Computationally efficient inversion techniques applicable to arbitrary distribution functions would allow the benefits of the antithetic variate techniques to be enjoyed more fully by the users of simulation.

2. DISCRETE DISTRIBUTIONS

Without loss of generality we will assume that X takes on only positive integral values. Then the inverse distribution function becomes for $0\leq u<1$.

$$G(u)=\min\{i: F(i)\geq u\} .$$

For arbitrary distributions a sequential search takes j comparisons to identify $X=j$ and thus EX comparisons on average. As the architecture of computers developes the number of comparisons will be the governing criteria in evaluating the complexity of calculations. This is especially true for the inversion of F if the values of $\{F(i)\}$ are tabled.

Chen and Asau (1974) first described an indexed search procedure to calculate G. Fishman and Moore (1981) analyzed such a procedure and demonstrated its efficiency. Their procedure defined the cutpoints

$$I_{\ell}=G((\ell-1)/M)$$

for $\ell=1,2,\ldots,M$ which are tabled during an initial setup. Thus, if $L[M U]$, where $[x]$ denotes the integer part of x, then only the integers between I_L and I_{L+1} need to be searched to identify X. The expected number of comparisons necessary to generate a single deviate is $EX-\bar{I}+1$, where $\bar{I}$ is the average of the set of cutpoints. An example is given which demonstrates

CH1953-9/83/0000-0189 $01.00

a significant increase in efficiency over the sequential search.

Instead of basing the cutpoints on a uniform partition of the unit interval, variable spacing may be employed as well. Ahrens and Kohrt (1981) vary the spacing in the interval $((M-1)/M,1)$ to handle long tailed distributions. In general, we wish to determine a set $U^*=\{u_i;\ i=0,\ldots,M\}$ in the unit interval with $u_0=1-u_M=0$ and cutpoints $\bar{I}_\ell=G(u_{\ell-1})$ maximizing $\bar{I}=\sum_{i=1}^{M}(u_i-u_{i-1})I_i$. Such a cutset will minimize the average number of comparisons required to generate a single deviate.

This problem is a form of the M-level quantizer which was shown to have a solution by Trushkin (1982) in the case that F has a log-concave density. One may find an optimal quantizer using the method of Lloyd (1982). For computational convenience the determination that $u_j\le U\le u_{j+1}$ should not require either expensive computations or a sequential search, a property the optimal quantizer may not possess. In practice the use of $u_j=jC/(MC-M+j)$ for $j=0,1,\ldots,M$ allows one to calculate easily $L[MU(C-1)/(C-U)]$ from the uniform deviate U and to start the sequential search at I_L . This choice of cutpoints is nearly optimal for concave distribution functions having long upper tails when

$C=0.5+\sqrt{0.16+G(0.9)/G(0.1)}$. As C increases the spacing of U^* becomes more uniform.

3. CONTINUOUS DISTRIBUTION

In this section an approximant, $\hat{G}$, to G will be developed for a continuous distribution function F . The approximant will interpolate G at prespecified points

$$K=\{(y_i=G(x_i),s_i):x_{i+1}\ x_i,\ i=0,1,\ldots,n\}$$

called knots. The quantizer problem may be solved to determine an optimal set of knots if desired. A piecewise increasing quadratic spline will be employed for $\hat{G}$. Passow and Roulier (1977) found that such a function may not exist on a fixed K but does exist on a superset $\bar{K}$ of K where the added knots lie between the knots in K .

The approach taken here is to solve for non-negative $\{d_i\}$ in the linear program

$$\max \sum_{i=1}^{M}(y_i-y_{i-1})\min(d_{i-1},d_i)$$

$$\text{s.t.}\quad d_i+d_{i-1}\le s_i \quad \text{for all } i$$

where $s_i=2(x_i-x_{i-1})/(y_i-y_{i-1})$. If for any i $d_i+d_{i-1}<s_i$ a new knot is added to K between (y_{i-1},x_{i-1}) and (y_i,x_i) and a value $\bar{d}_i$ is calculated. The resulting $\bar{K}$ yields the desired piecewise quadratic on $\bar{K}$ with minimum non-negative slopes $\{d_i\}$. That is, $\hat{G}$ has as flat a density as possible.

The use of $\hat{G}$ is a two-step process. First, the piece index J is generated from a uniform deviate U using the procedure discussed in Section 2. Then the quadratic function on that piece is evaluated at U to obtain $\hat{G}$. For an exact evaluation of G, the second step may be followed by a localized search at $\hat{G}(U)$ if such is deemed necessary.

REFERENCES

Ahrens, J.H. and K.D. Kohrt (1981). "Computer Methods for Efficient Sampling from Largely Arbitrary Statistical Distributions", Computing, 26, 19-31.

Chen, Hui-Chuan, and Yoshinori Asau (1974). "On Generating Random Variates from an Empirical Distribution," AIIE Transactions, 6, 163-166.

Fishman, George S. and Baosheng D. Huang (1980). "Antithetic Variates Revisited", Comm. ACM, to appear.

Fishman, George S. (1983). "Accelerated Accuracy in the Simulation of Markov Chains", Operations Research, Vol. 31, No. 3, 466-487.

Fishman, George S. and Louis R. Moore (1983). "Sampling from a Discrete Distribution While Preserving Monotonicity", Technical Report No. UNC/ORSA/TR-81/7, Curriculum in Operations Research and Systems Analysis, University of North Carolina at Chapel Hill.

Hoeffding, Wassily (1940). "Masstabinvariante Korrelationstheorie," Schriften des Mathematischen Instituts and des Instituts fur Angewandte Mathematik der Universitat Berlin, 5, 197-233.

Lloyd, S.P. "Least Squares Quantization in PCM", IEEE Trans. Inform. Theory, Vol. IT-28, 129-137.

McAllister and J.A. Roulier (1978), "Interpolation by Convex Quadratic Splines", Math. of Computation, Vol. 32, No. 144, 1154-1162.

Passow, E. and J.A. Roulier (1977). "Monotone and Convex Spline Interpolation", SIAM J. Numer. Anal., Vol. 14, No. 5, 904-909.

Trushkin, A.V. (1982). "Sufficient Conditions for Uniqueness of a Locally Optimal Quantizer for a Wide Class of Error Weighting Functions", IEEE Trans. Inform. Theory, Vol. IT-28, 187-198.

Whitt, Ward (1976). "Bivariate Distributions with Given Marginals", Ann. Stat., 4, 1280-1289.