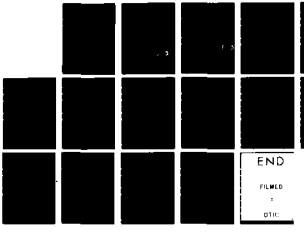
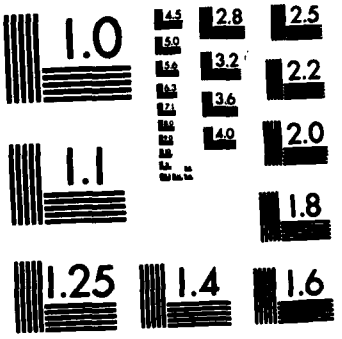
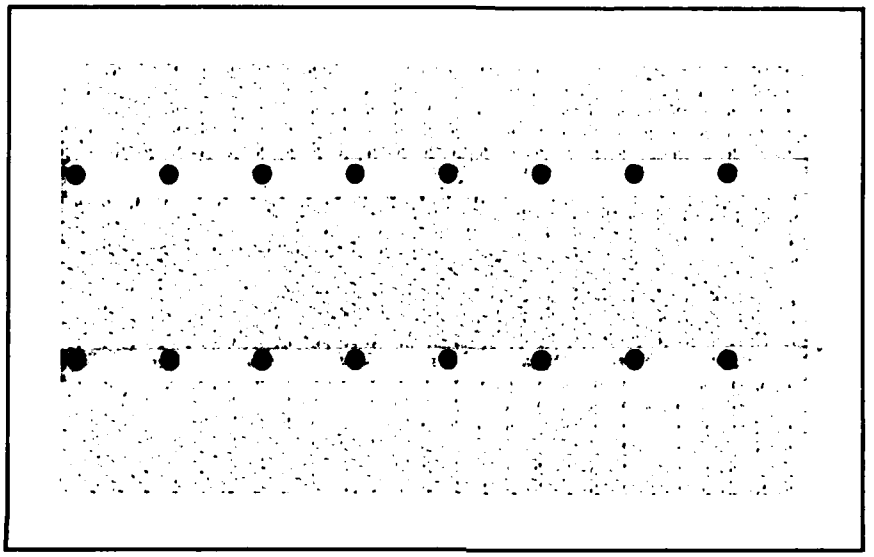
AD-A124 251 VARIANCE REDUCTION IN THE SIMULATION OF STOCHASTIC ACTIVITY NETWORKS(U) NORTH CAROLINA UNIV AT CHAPEL HILL CURRICULUM IN OPERATIONS R.. G S FISHMAN JAN 83 1/1

UNCLASSIFIED UNC/ORSA/TR-82-7 N00014-76-C-0302 F/G 12/1 NL

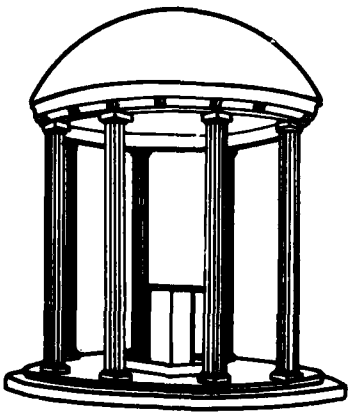




OPERATIONS RESEARCH AND SYSTEMS ANALYSIS



UNIVERSITY OF NORTH CAROLINA
AT CHAPEL HILL



VARIANCE REDUCTION IN THE SIMULATION OF STOCHASTIC ACTIVITY NETWORKS

George S. Fishman

Technical Report No. UNC/ORSA/TR-82/7

January 1983

Curriculum in Operations Research
and Systems Analysis

University of North Carolina at Chapel Hill

This research was supported by the Office of Naval Research under contract N00014-76-C-0302. Reproduction in whole or part is permitted for any purpose of the United States government.

Abstract

This paper describes a Monte Carlo method based on the theory of quasirandom points for estimating the distribution functions and means of network completion time and shortest path time in a stochastic activity network. In particular, the method leads to estimators whose variances converge faster than $1/K$, where K denotes the number of replications collected in the experiment. The paper demonstrates how accuracy diminishes for a given K with increasing dimensionality of the network and shows how a procedure that uses a cutset of the network together with convolution can reduce dimensionality and increase accuracy. Two examples illustrate the benefits of using quasirandom points together with a cutset and then convolution.

Key Words

Monte Carlo method

Networks

Quasirandom points

Variance reduction

Introduction

Activity network analysis is a fairly standard tool in Operations Research. When arc completion times in such a network are deterministic, well-established algorithms exist for finding its network completion time and shortest path time in a computationally efficient manner. For example, see Elmaghraby (1977) and Wagner (1975). However, when arc completion times are stochastic, analysis becomes considerably more difficult, even for relatively small networks. Here the objective often is to compute the distribution functions (d.fs.) and means of network completion time and shortest path time, given the d.fs. of the statistically independent individual arc completion times.

Because severe difficulties exist in deriving analytical solutions to these problems, many analysts have turned to Monte Carlo methods to derive approximate solutions. Van Slyke (1963) proposes the use of importance sampling to gain efficiency in estimating the characteristics of networks. Martin (1965) describes how one can increase statistical efficiency when the arc passage time distributions are polynomials. Burt and Garman (1971) describe how conditional sampling can reduce dimensionality and thus sampling variation. Carrying this idea one step further Sigal, Pritsker and Solberg (1979, 1980) show how one can use a cutset of a network to effect similar reductions. The use of antithetic variates has also been proposed in a number of studies, including Cheng (1980), Sullivan, Hayya and Schaul (1982) and Grant (1982).

Although the benefits of these proposals are well established, they all lead to estimators with variances $O(1/K)$ for K independent replications. The present paper shows how to increase the rate of convergence of this variance using quasirandom points and illustrates the success of the

approach for two networks of N=8 and N=18 arcs. The theory of quasirandom points has its origin in the numerical evaluation of multivariable integrals. The theory concerns the identification of vector sequences that when used in numerical integration lead to approximations whose absolute errors converge to zero faster than if independent random vectors were used. Since one can represent the desired d.fs. and means as multivariable integrals, the appeal of quasirandom points is apparent. Moreover, combining the use of quasirandom points with other approaches such as cutsets and convolution leads to even greater advantage, as we demonstrate.

Section 1 presents the stochastic network problem in detail. Section 2 describes how one can view the problem as one of multivariable numerical integration. Section 3 describes how one would tackle the problem by crude Monte Carlo methods and also describes the benefits of conditional sampling. Section 4 shows the benefits of using cutsets. Section 5 then indicates how cutsets together with convolution can significantly reduce the dimensionality of the problem.

Section 6 presents a short discussion of the principal concepts of multivariable numerical integration using quasirandom points and shows the extent to which known results apply to the problem at hand. Section 7 lays out an experiment whose design is used in Section 8 with two examples to show the benefit of using quasirandom points, especially when combined with convolution and the cutset approach.

1. Definitions

Consider an acyclic directed network with a single source, single sink, N arcs and L paths. Let $X_1,\ldots,X_N$, the passage times for arcs $1,\ldots,N$, be independent random variables where X_i has distribution function (d.f.) F_i on $[0,\infty)$ and inverse distribution function $G_i(u) = \min[x: F_i(x) \geq u, 0 < u \leq 1]$. The completion time of path m is

$$T_m = \sum_{i=1}^{N} a_{im}X_i = \sum_{i \in I_m} X_i \qquad (1)$$

where $a_{im}=1$ if arc i is on path m, $a_{im}=0$ otherwise and I_m denotes the set of arcs on path m.

The principal purpose of this study is to characterize the network completion time

$$T^* = \max(T_1,\ldots,T_L) \qquad (2)$$

and the shortest path time

$$T_* = \min(T_1,\ldots,T_L) \quad . \qquad (3)$$

For T^*, characterizations include:

a. $\text{pr}(T^* \leq t)$ $\qquad 0 \leq t < \infty$

b. $E\,T^*$

c. $\text{pr}(m \text{ is the longest path})$ $\qquad m = 1,\ldots,L$

d. $\text{pr}(T_* \leq t)$ $\qquad 0 \leq t < \infty$

e. $E\,T_*$

f. $\text{pr}(m \text{ is the shortest path})$ $\qquad m = 1,\ldots,L$.

The present study concerns the estimation of a, b, d and e. For expository convenience the main body of the paper concerns the estimation of $\text{pr}(T^* \leq t)$. The Appendix contains details for the estimation of the remaining quantities.

Let $|I_m|$ denote the cardinality of I_m. If

$$N = \sum_{m=1}^{L} |I_m| \tag{4}$$

then no paths share a common arc and T^* and T_* have d.fs.

$$F^*(t) = \prod_{m=1}^{L} F_{I_m}(t) \tag{5}$$

and

$$F_*(t) = 1 - \prod_{m=1}^{L} [1 - F_{I_m}(t)] \qquad 0 \leq t < \infty ,$$

F_{I_m} being the convolution of the $|I_m|$ arcs on path m. As an example, suppose that arc times are exponential with

$$F_i(t) = 1 - e^{-\lambda_i t} \qquad \lambda_i > 0 \qquad i \in I_m . \tag{6}$$

Then one has for distinct λ_i

$$F_{I_m}(t) = 1 - \sum_{i \in I_m} \frac{e^{-\lambda_i t}}{\prod_{\substack{j \in I_m \\ j \neq i}} (1-\lambda_i/\lambda_j)} . \tag{7}$$

Usually the summation in (4) exceeds N in which case the derivations of F^* and F_* are more complicated than in (5). In practice, even when (4) holds, the convolutions for $F_{I_1}, \ldots, F_{I_L}$ may prove too complex to derive analytically.

2. A Solution via Integration

An alternative approach to characterization is through integration; in particular, numerical integration. Observe that one can write

$$F^*(t) = E\, I_{[0,t)}(T^*)$$

$$= \int_0^1 \cdots \int_0^1 h(t;x_1,\ldots,x_n) \prod_{i=1}^{N} dF_i(x_i) \tag{8a}$$

$$= \int_0^1 \cdots \int_0^1 g(t;u_1,\ldots,u_n)\, du_1,\ldots,du_N \tag{8b}$$

where

$$I_{[a,b)}(x) = 1 \quad \text{if } a \le x < b$$
$$= 0 \quad \text{otherwise,}$$

$$h(t;x_1,\ldots,x_N) = I_{[0,t)}\left(\max\left(\sum_{i=1}^{N} a_{i1}x_i,\ldots,\sum_{i=1}^{N} a_{iL}x_i\right)\right) \tag{9a}$$

and

$$g(t;u_1,\ldots,u_N) = h(t;G_1(u_1),\ldots,G_N(u_N)) \ . \tag{9b}$$

Now (8b) involves multiple integration over the N-dimensional unit hypercube $[0,1)\times\ldots\times[0,1)$, hereafter denoted by U_N . It is the expression (9b) on which our analysis is based.

Consider the approximation

$$\bar{g}_K(t) = \frac{1}{K}\sum_{j=1}^{K} g(t;u_{1j},\ldots,u_{Nj}) \tag{10}$$

where $\underset{\sim}{u}_i = \{u_{ij};\ j = 1,2,\ldots\}$ $i=1,\ldots,N$ are infinite sequences of points in U_N chosen in one of several ways. One way is through pure random sampling which we call the <u>crude Monte Carlo method</u>.

3. Crude Monte Carlo Sampling

Let

$$T_j^* = \max\ (T_{j1},\ldots,T_{jL}) \tag{11}$$

where

$$T_{jm} = \sum_{i=1}^{N} a_{im}\, G_i(u_{ij}) \qquad m = 1,\ldots,L \tag{12}$$

and u_{ij} $i = 1,\ldots,N$ $j = 1,\ldots,K$ are i.i.d. from $U[0,1)$. Also define

$$g_j(t) = g(t;u_{1j},\ldots,u_{Nj}) = I_{[0,t)}(T_j^*) \ . \tag{13}$$

Then $\bar{g}_K(t)$ in (10) has expectation $F^*(t)$ and

$$\text{var}\ \bar{g}_K(t) = \frac{F^*(t)\ [1 - F^*(t)]}{K} \ . \tag{14}$$

To improve on the result in (14), Burt and Garman (1971) introduce the use of conditional sampling. Suppose that there exist arcs $i_1,\ldots,i_L$ unique to paths $1,\ldots,L$ respectively. Assume that these arcs are distinct and define

$$S_{jm} = \sum_{\substack{i=1 \\ i\neq i_m}}^{N} a_{im}\, G_i(u_{ij}) \tag{15}$$

and

$$g_j(t) = \prod_{m=1}^{L} F_{i_m}(t - S_{jm}) \quad . \tag{16}$$

Now $\bar{g}_K(t)$ in (10), based on (16), has the correct expectation $F^*(t)$, as can be seen by integrating (16) with respect to the joint p.d.f. of $S_{j_1},\ldots,S_{j_L}$. But $\bar{g}_K(t)$ has smaller variance than (14), as a result of sampling from N-L instead of N arcs. Also note that

$$g_j(t) \leq \prod_{m=1}^{L} F_{i_m}(t) \tag{17}$$

provides an upper bound which also applies to $\bar{g}_K(t)$.

In principle, conditional sampling can be made more effective by working with subsets of arcs unique to each arc. Let

$$J_m = I_m \cap (\cup_{\substack{j=1 \\ j\neq m}}^{L} I_j) \qquad m = 1,\ldots,L \tag{18}$$

so that J_m is the set of arcs unique to path m. Assume $|J_m| > 1$ $m = 1,\ldots,L$ and define

$$S_{jm} = \sum_{i\in I_m - J_m} a_{im} G_i(u_{ij}) \tag{19}$$

and

$$g_j(t) = \prod_{m=1}^{L} F_{J_m}(t - S_{jm}) \tag{20}$$

where F_{J_m} denotes the d.f. of the sum of the passage times on the arcs unique to path m. Since sampling in $N + L - \sum_{m=1}^{L} |J_m|$ dimensions is all that is necessary, the use of (20) in (10) preserves unbiasedness and reduces variance even further.

In essence, the effect of conditional sampling is to replace the coefficient of $1/K$ in (14) by one of smaller magnitude but in no way to change the convergence rate with regard to K as $K \to \infty$. Also, note the upper bound

$$g_j(t) \leq \prod_{m=1}^{L} F_{J_m}(t) \tag{21}$$

which is tighter than the bound in (17).

4. Cutsets

In practice it is conceivable that at least one path does not have a unique arc. That is, $|J_m| = 0$ for at least one path m. Figure 1 illustrates such a case. Although conditional sampling as described by Burt

Insert Fig. 1 about here.

and Garman does not apply here, another more general approach proposed by Sigal, Pritsker and Solberg (1979, 1980) does. Let H denote a cutset of the network. A cutset is a set of arcs that connects a set of nodes W containing the source with a set of nodes $\bar{W}$ containing the sink. Also, assume that each path has only one arc in H. If each arc in H points from W to $\bar{W}$, H is called a uniformly directed cutset (UDC).

Define

$$S_{jm} = \sum_{\substack{i=1 \\ i \notin H}}^{N} a_{im} G_i(u_{ij}) \tag{22}$$

and

$$Y_{ij} = \sup_{m=1,\ldots,L} (a_{jm} S_{jm}) \qquad i \in H \ . \tag{23}$$

Then define

$$g_j(t) = \prod_{i \in H} F_i(t - Y_{ij}) \ . \tag{24}$$

Now $\bar{g}_K(t)$ in (10) based on (24) gives an unbiased estimate of $F^*(t)$

with smaller variance than crude Monte Carlo sampling since one is sampling in only $N - |H|$ dimensions. Since a network may have more than one UDC, using the one with maximal cardinality gives the greatest reduction in dimensionality. However, the determination of this cutset for an arbitrarily large network is not trivial, being an NP complete problem. Nevertheless, among alternative known cutsets for a given network, the one with largest cardinality serves our purpose best. As an example of the use of cutsets note that (e_2, e_3, e_4, e_7) and (e_3, e_5, e_6, e_7) are UDCs with maximal cardinality 4 for the network of Fig. 1, whereas the cutset (e_1, e_2, e_3) with cardinality 3 is not.

As a second example, consider the network in Fig. 2, taken from Battersby (1970), which describes the steps encountered in the partial overhaul of a unit in an oil refinery. Table 1 shows the incidence matrix $\underset{\sim}{a} = || a_{im} ||$

Insert Fig. 2 about here.

---

for arcs (rows) and paths (columns). Observe that arcs $(e_2, e_3, e_5, e_6, e_9, e_{10}, e_{12})$ and arcs $(e_2, e_3, e_5, e_9, e_{10}, e_{12}, e_{15})$ form two UDCs with maximal cardinality 7 . However, the cutset $(e_2, e_3, e_5, e_6 + e_{15}, e_9, e_{10}, e_{12})$, formed by combining arcs e_6 and e_{15} (see Fig. 2) has cardinality 8 and can be used to advantage to reduce dimensionality.

Insert Table 1 about here.

---

5. Further Dimensional Reduction

One can achieve at least one additional reduction in the dimensionality

of sampling by exploiting available knowledge about a network. Let

$$I_m = I_m - I_m \cap H \tag{25}$$

where H denotes a UDC. Then define the subsets

$$J_{m_1,\ldots,m_r} = \cap_{j=1}^{r} I_{m_j} - \sum_{\substack{m=1 \\ m \neq m_1,\ldots,m_r}}^{L} [(\cap_{j=1}^{r} I_{m_j}) \cap I_m] \tag{26}$$

$$1 \leq m_1 < m_2 < \ldots < m_r \leq L; \quad r=1,\ldots,L \ .$$

Here $J_{m_1,\ldots,m_r}$ denotes the subset of arcs, not in H, that are uniquely on paths $m_1,\ldots,m_r$. Suppose there are q such nonempty sets $R_1,\ldots,R_q$.

Let F_{R_i} denote the d.f. of the sum of the $|R_i|$ arc passage times corresponding to the arcs in R_i and let G_{R_i} denote the corresponding inverse d.f. Then define

$$S_{jm} = \sum_{i=1}^{q} (\Pi_{j \in R_i} a_{jm})\, G_{R_i}(U_{ij}) \tag{27}$$

and

$$Y_{ij} = \sup_{m=1,\ldots,L} (S_{jm}) \qquad i = 1,\ldots,|H| \tag{28}$$

and

$$g_j(t) = \Pi_{i \in H} F_i(t - T_{ij}) \ . \tag{29}$$

Now, $\bar{g}_K(t)$ in (10) based on (29) again is an unbiased estimator of $F^*(t)$. However, since $N - |H| + q - \sum_{i=1}^{q} |R_i|$ dimensions arise for sampling per replication, one anticipates that var $\bar{g}_K(t)$ has smaller magnitude than previously described methods produce. Note that this reduced dimensionality comes from the use of UDCs together with convolution.

6. Multivariable Numerical Integration

As (8b) shows, one can view $F^*(t)$ as the result of a multivariable integration over a restricted region in V_N. We now describe how taking this

view of the problem leads to useful results when quasirandom points are used. Our account follows Kuipers and Niederreiter (1974) and Niederreiter (1978). Also see Schmidt (1973).

Suppose that our objective is to evaluate the integral

$$B = \int_0^1 \cdots \int_0^1 f(x_1,\ldots,x_N)\, dx_1 \cdots dx_N \tag{30}$$

by

$$B_K = \frac{1}{K} \sum_{j=1}^{K} f(u_{1j},\ldots,u_{Nj}) \tag{31}$$

and to derive bounds on the error

$$\Delta_K = |B - B_K| \ . \tag{32}$$

Three definitions facilitate our description. Consider the sequences

$\{\underset{\sim}{u}_j = (u_{1j},\ldots,u_{Nj});\quad j = 1,\ldots,K\}$.

Definition 1. $A(R;K)$ = number of points $\underset{\sim}{u}_1,\ldots,\underset{\sim}{u}_K$ that fall in $R \subset V_N$.

Definition 2. $\underset{\sim}{u}_1,\ldots,\underset{\sim}{u}_K$ are uniformly distributed in V_N if

$$\lim_{K\to\infty} \frac{A(R;K)}{K\lambda(R)} = 1$$

for all $R = \{(x_1,\ldots,x_N):\ \alpha_i \le x_i \le \beta_i \quad i = 1,\ldots,N\}$ $R \subset V_N$ and where $\lambda(R)$ is the measure of volume of R in V_N .

Definition 3. The extreme discrepancy associated with a sequence $\underset{\sim}{u}_1,\ldots,\underset{\sim}{u}_K$ is

$$D_K = D_K(\underset{\sim}{u}_1,\ldots,\underset{\sim}{u}_K) = \sup_{R^* \subset V_N} \left|\frac{A(R^*;K)}{K} - \lambda(R^*)\right|$$

where $R^* = \{(x_1,\ldots,x_N):\ 0 \le x_i \le \beta_i \quad i = 1,\ldots,N\}$. Hereafter we assume that $\{\underset{\sim}{u}_j \quad j = 1,\ldots,K\}$ is a uniformly distributed sequence on V_N .

These definitions together with several fundamental theorems from the theory of quasirandom points provide us with a way of characterizing the error (32). In particular,

Theorem 1 (Hlawka 1961). If f is of bounded variation in the sense of Hardy and Krause, then

$$\Delta_K < V(f,N)D_K \tag{33}$$

where $V(f,N)$ is a function of the bounded variation in f in N and lower dimensions.

Theorem 2 (van Aardenne-Ehrenfest 1945). For any infinite sequence $\underset{\sim}{u}_1, \underset{\sim}{u}_2, \ldots$ with $N > 1$

$$\lim_{K\to\infty} KD_K = \infty \ .$$

Theorem 3 (Roth 1954). For any sequence of K points in U_N with $N > 2$

$$D_K > \frac{C_N (\log K)^{(N-1)/2}}{K}$$

where C_N is a function of N only.

Theorem 4 (Roth 1954). For any infinite sequence in U_N with $N > 1$

$$D_K > \frac{C'_N (\log K)^{N/2}}{K}$$

where C'_N is a function of N only.

Theorem 1 shows that one can bound the error (32) by a quantity proportional to the extreme discrepancy D_K, a most convenient result. Theorem 2 implies that one cannot expect the extreme discrepancy to converge as fast as $1/K$. Theorems 3 and 4 provide lower bounds on how fast convergence can occur.

We now describe a uniformly distributed sequence for which upper bounds are known. If $R > 2$ is an integer, then every non-negative integer n has an expansion of the form

$$n = \sum_{i=0}^{m} a_i R^i \qquad a_i \in \{0,1,\ldots,R-1\} \tag{34}$$

$$0 \le i \le m \quad \text{and} \quad m = \lfloor \log_R n \rfloor \ .$$

Corresponding to this expansion one has the radical inverse function

$$\phi_R(n) = \sum_{i=0}^{m} a_i R^{-i-1} \quad . \tag{35}$$

We now define several sequences based on (34):

van der Corput Sequence: $\{\phi_2(j);\ j = 0,1,\ldots,K-1\}$

Hammersley Sequence: $\{\phi_{R_1}(j),\ldots,\phi_{R_{N-1}}(j), \frac{j}{K};\ j = 0,1,\ldots,K-1\}$

where $R_1,\ldots,R_{K-1}$ are pairwise relatively prime.

Halton Sequence: $\{\phi_{R_1}(j),\ldots,\phi_{R_N}(j);\ j=0,1,\ldots,K-1\}$

where $R_1,\ldots,R_N$ are pairwise relatively prime.

The van der Corput sequence is uniformly distributed on V_1 and was the first sequence of this kind proposed for univariate numerical integration. The Hammersley and Halton sequences are uniformly distributed on V_N and as Theorems 5 and 6 show they offer useful upper bounds.

Theorem 5 (Halton 1960). For the Hammersley sequence

$$D_K < \frac{(\log K)^{N-1}}{K} \prod_{i=1}^{N-1} \left(\frac{3R_i-2}{\log R_i}\right) \qquad N > \max_i R_i \ .$$

Theorem 6 (Halton 1960). For the Halton sequence

$$D_K < \frac{(\log K)^{N}}{K} \prod_{i=1}^{N-1} \left(\frac{3R_i-2}{\log R_i}\right) \qquad N > \max_i R_i \ .$$

Several issues of significance need to be mentioned now. Firstly, since the van der Corput, Hammersley and Halton sequences are deterministic the upper bounds can be interpreted as worst case bounds with certainty. Secondly, although $V(f,N)$ in (33) is in theory computable, in the present case f is unknown. Thirdly, the bounds presented here hold for integration the entire N-dimensional hypercube; however, as (8a), (9a) and (9b) show, our problem calls for integration over a restricted region in V_N . Now it is known (Niederreiter 1978, p. 982) that for integration over an arbitrary

region in U_N

$$\Delta_K < C_N''(D_K)^{1/N} \tag{36}$$

where C_N'' is a function of N. This is a considerably weaker bound but since it applies to such a wider set of regions in U_N than (33) does there is reason to believe that the use of the Hammersley and Halton sequences can lead to accelerated convergence.

The fourth issue concerns boundedness. In the case of estimating $F^*(t)$ and $F_*(t)$, this assumption is met. However, it is not met for ET^* and ET_*. This fact together with the integration over a restricted region in U_N makes clear that when we move from theory to practice, some skepticism exists as to how efficient our approach will be. As the examples in Section 8 show, there is little basis for this skepticism.

The fifth issue relates to dimensionality. As Theorems 5 and 6 make clear, it is in one's best interest to reduce dimensionality as much as possible before using quasirandom points. The cutset approach together with convolution are the methods we use to effect this reduction.

The sixth issue of significance concerns the choice between the Hammersley (finite) and Halton (infinite) sequences in practice. If one uses the Hammersley sequence for fixed K and then decides that the accuracy is not acceptable, there is no recourse to continued exploitation of its special structure since $0,1/K,\ldots,(K-1)/K$ are the values in the Nth dimension. By contrast, the Halton sequence enables us to continue exploitation by merely generating additional points $\{\phi_{R_1}(j),\ldots,\phi_{R_N}(j)\}$ $j=k, k+1,\ldots$. The modest degradation in the upper bound that arises when using the Halton rather than

the Hammersley sequence seems worth the price.

7. Experimental Design

This section describes the layout for an experiment designed to determine the extent to which quasirandom points lead to accelerated accuracy when estimating $F^*(t)$, ET^* , $F_*(t)$ and ET_* for the networks in Figs. 1 and 2 . In particular, we introduce a degree of randomness into the experiment in order to compute estimates of the variances of our point estimates. We then study the behavior of these sample variances as K increases.

Consider an experiment consisting of Q statistically independent blocks or macroreplications each of K microreplications. Let $\{U_{im}: i=1,\ldots,N; m=1,\ldots,Q\}$ denote a sequence of i.i.d. random variables from $U[0,1)$ and define

$$S_{im} = (s: \phi_{R_i}(s) = U_{im}) \qquad i=1,\ldots,N \quad m=1,\ldots,Q \ . \tag{37}$$

Then on macroreplication m we use the quasirandom point $\phi_{R_1}(S_{1m} + j-1),\ldots,\phi_{R_N}(S_{Nm} + j-1)$ on microreplication j $j=1,\ldots,K$.

Let $\hat{\theta}_{jm}$ denote the unbiased estimate of a particular quantity θ computed on microreplication j on macroreplication m . Then $\bar{\theta}_1,\ldots,\bar{\theta}_Q$, where

$$\bar{\theta}_m = \frac{1}{K} \sum_{j=1}^{K} \hat{\theta}_{jm} \qquad m=1,\ldots,Q \ , \tag{38}$$

are i.i.d. random variables with sample variance

$$s_K^2 = \frac{1}{Q-1} \sum_{m=1}^{Q} (\bar{\theta}_m - \bar{\theta}_Q)^2 \tag{39}$$

where

$$\bar{\theta}_Q = \frac{1}{Q} \sum_{m=1}^{Q} \bar{\theta}_m \ .$$

It is the behavior of s_K^2 versus K that interests us.

To provide an instructive normalization we also run M statistically independent macroreplications each of K=1 microreplication. Let $\hat{\theta}_1, \ldots, \hat{\theta}_M$ denote the resulting estimates of θ each with sample variance

$$w_Q^2 = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\theta}_m - \tilde{\theta}_Q)^2 \qquad (40)$$

where

$$\tilde{\theta}_m = \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}_m .$$

Then the quantity $w_M^2 / K s_K^2$ should increase as K increases if accelerated convergence is occurring.

We set $M=10^5$. For the quasirandom case we perform runs for $K=2^j$, $j=1,2,\ldots$ and set $Q=2^{16}/K$ for $K<2^9$ and $Q=2^7$ for $K>2^9$. Following this procedure enables us to treat s_K^2 and w_M^2 as highly reliable point estimators of the quantities var $\bar{\theta}_m$ and var $\hat{\theta}_m$ respectively.

8. Examples

We refer to the network in Fig. 1 as the SPS network and to the one in Fig. 2 as the Battersby network. Table 2 lists the relevant cutsets and noncutsets. In both examples the arc passage times are assumed independent

---

Insert Table 2 about here.

---

and exponentially distributed. For the SPS network the means are $\lambda_1=10$, $\lambda_2=15$, $\lambda_3=18$, $\lambda_4=6$, $\lambda_5=7$, $\lambda_6=3$, $\lambda_7=10$ and $\lambda_8=2$. For arc i on microreplication j on macroreplication m the arc passage time is

$$X_i = G_i(\phi_{R_i}(S_{im} + j-1)) = -\lambda_i \ln(1 - \phi_{R_i}(S_{im} + j-1)) . \qquad (41)$$

For the runs using quasirandom points, the cutset $H=(e_2, e_3, e_4, e_7)$ was used reducing dimensionality to $N-|H| = 8-4=4$. An algorithm due to Halton

and Smith (1964) was used to compute the Halton sequences. For the M independent microreplications all N arc times from (41) were used.

Table 3 shows the behavior of w_M^2/Ks_K^2 for $F^*(t)$ estimated at t=15, 30, 40, 50, 70 for ET^*, for $F*(t)$ estimated at t=5, 10, 15, 20, 25 and for ET* . As a point for comparison, Sigal, Pritsker and Solberg (1979)

Insert Table 3 about here.

report variance reductions ranging from 4.18 to 11.00 for the estimation of $F^*(t)$ at t=15, 30, 40, 50 and 70 using the cutset H and independent microreplications.

The second example studies variance reduction for the Battersby network using the means in Table 1. Two separate sets of experiments were run. The first set uses partition 1 in Table 2 with the cutset $H=(e_2, e_3, e_5, e_6, e_9, e_{10}, e_{12})$ with quasirandom points, yielding a dimensionality of $N-|H|=11$. The second set uses partition 2 with the cutset $H=(e_2, e_3, e_5, e_6 + e_{15}, e_9, e_{10}, e_{12})$ yielding a dimensionality of 6 after appropriate convolution within the noncutset. The quantity $F^*(t)$ was estimated for t=90, 110, 120, 130, 140, 150, 160, 180, 190 and $F*(t)$ for t=20, 25, 30, 35, 40, 50, 60, 80, 90 . Table 4 shows the resulting variance reductions w_M^2/Ks_K^2 for $F^*(t)$, ET^* , $F*(t)$ and ET* .

Insert Table 4 about here.

Tables 3 and 4 admit several instructive observations:

1. As K increases variance reduction increases; however, variance reduction is not necessarily monotone in K.
4. For given K , variance reduction tends to decline with increasing dimensionality.
3. For given K variance reduction is greater for $F^*(t)$ than for ET^* and for $F\star(t)$ than for $ET\star$, showing the benefit one obtains from bounded integrands.

Although these observations establish the benefits of using quasirandom points with the cutset approach and convolution, there is a cost issue that also needs to be considered. If one compares the cost of producing one sample point for $F^*(t)$, ET^* , $F\star(t)$ and $ET\star$ using quasirandom points to the cost of producing a sample point using pure random sampling, the ratio never exceeds 2.5 in our experiments. The most costly step turns out to be sampling the summed arc times $x_i + x_j$, in the noncutset of partition 2 in Table 2, from the distribution function

$$F(t) = 1 - \frac{e^{-\lambda_i t}}{1-\lambda_i/\lambda_j} - \frac{e^{-\lambda_j t}}{1-\lambda_j/\lambda_i} \quad .$$

This was done using the Newton-Raphson method.

Since this cost ratio is independent of K , we have strong evidence that for sufficiently large K , the quasirandom point method offers a clear advantage over pure random sampling for networks of arbitrary size. A program called NETWOK is currently under development to implement the proposed method in a computationally efficient manner.

8. References

Battersby, A. (1970). Network Analysis for Planning and Scheduling, John Wiley and Sons, third edition.

Burt, J.M. and M.B. Garman (1971). "Conditional Monte Carlo: A Simulation Technique for Stochastic Network Analysis," Management Science, 18, 207-217.

Cheng, R.C.H. (1981). "The Use of Antithetic Control Variates in Computer Simulations," 1981 Winter Simulation Conference Proceedings, T.I. Oren, C.M. Delfosse and C.M. Schub, eds.

Elmaghraby, S. E. (1977). Activity Networks: Project Planning and Control by Network Models, John Wiley and Sons.

Fishman, G.S. and B.D. Huang. (1980) "Antithetic Variates Revisited," TR 80-4, Curriculum in Operations Research and Systems Analysis, University of North Carolina at Chapel Hill.

Grant, F.H. (1982). "A Note on Efficiency of the Antithetic Variate Method for Simulating Stochastic Networks," University of Evansville, School of Business Administration, Evanston, Indiana. To appear in Management Science.

Halton, J.H. (1960). "On the Efficiency of Certain Quasi-random Sequences of Points in Evaluating Multi-dimensional Integrals," Numerische Mathematik, 2, 84-90.

Halton, J.H. and G.B. Smith (1964). "Algorithm 247: Radical Inverse Quasirandom Point Sequence," Comm. ACM, 12, 701-702.

Hlawka, E. (1961). "Funktionen von beschransker Variation in der Theorie der Gleichverteilung, Ann. Mat. Pura. Appl., 54, 325-333.

Kuipers, L. and H. Niederreiter (1974). Uniform Distributions of Sequences, John Wiley and Sons.

Martin, J.J. (1965). "Distribution of Time Through a Directed Acyclic Network," Operations Research, 13, 46-66.

Niederreiter, H. (1978). "Quasi-Monte Carlo Methods and Pseudo-random Numbers," Bull. Amer. Math. Soc., 84, 957-1041.

Roth, K.F. (1954). "On Irregularities of Distribution," Mathematika, 1, 73-79.

Schmidt, W. M. (1977). Irregularities of Distribution, Tata Institute of Fundamental Research, Bombay.

Sigal, C.C., A.A.B. Pritsker and J.J. Solberg (1979). "The Use of Cutsets in Monte Carlo Analysis of Stochastic Networks," Mathematics and Computers in Simulation, 21, 376-384.

Sigal, C.E., A.A.B. Pritsker and J.J. Solberg, (1980). "The Stochastic Shortest Route Problem," Operations Research, 28, 1122-1129.

Sullivan, R.S., J.C. Hayya and R. Schaul (1982). "Efficiency of the Antithetic Variate Method for Simulating Stochastic Networks," Management Science, 28, 563-572.

van Aardenne-Ehrenfest, T. (1945). "Proof of the Impossibility of a Just Distribution ," Proc. Kon. Ned. Akad. v. Wet., 48, 3-8.

van der Corput, J.C. (1935). "Verteilungsfunktionen", Proc. Kon. Akad. Wet. Amsterdam, 38, 813-821 and 1058-1066.

van Slyke, R.M. (1963). "Monte Carlo Methods and the Pert Problem," Operations Research, 11, 839-860.

Wagner, H. M. (1975). Principles of Operations Research, Prentice-Hall, second edition.

Appendix

Estimation of $F^*(t)$. We derive the estimator based on the cutset H and (22). Let

$$Z_{ij} = \inf_{\substack{m=1,\ldots,L \\ a_{im}=i}} (S_{jm}) \qquad i \in H$$

and define

$$h_j(t) = 1 - \Pi_{i \in H} [1 - F_i(t - Z_{ij})] .$$

Then

$$\bar{h}_K(t) = \frac{1}{K} \sum_{j=1}^{K} h_j(t)$$

gives an unbiased estimator of $F^*(t)$ based on sampling from $N - |H|$ dimensions.

Estimation of ET^* and ET^*. Using (11) one has the unbiased estimator

$$\bar{T}^*_K = \frac{1}{K} \sum_{j=1}^{K} T^*_j$$

for ET^* . Let

$$T^*_j = \min(T_{j1},\ldots,T_{jL}) .$$

Then

$$\bar{T}^*_K = \frac{1}{K} \sum_{j=1}^{K} T^*_j$$

is an unbiased estimator of ET^* .

TABLE 1

Incidence Matrix for Battersby Network

a_{im}

Arc i	λ_i	Path m							
		1	2	3	4	5	6	7	8
e_1	16	1	1						
e_2	16	1							
e_3	8		1						
e_4	6	1	1						
e_5	16			1	1				
e_6	40					1			
e_7	24						1	1	1
e_8	16							1	1
e_9	16								1
e_{10}	24							1	
e_{11}	8							1	1
e_{12}	4						1		
e_{13}	36	1	1	1					
e_{14}	12				1				
e_{15}	8					1			
e_{16}	24						1	1	1
e_{17}	8	1	1	1	1	1	1	1	1
e_{18}	24	1	1	1	1	1	1	1	1

Table 2

Network Analysis

Network	Cutset	Noncutset
Fig. 1	e_2, e_3, e_4, e_7	e_1, e_5, e_6, e_8
Battersby (Fig. 2) Partition 1	e_2, e_3, e_5, e_6, e_9, e_{10}, e_{12}	e_1, e_4, e_7, e_8, e_{11}, e_{13}, e_{14}, e_{15}, e_{16}, e_{17}, e_{18}
Battersby (Fig. 2) Partition 2	e_2, e_3, e_5, $e_6 + e_{15}$, e_9, e_{10}, e_{12}	$e_1 + e_4$, $e_7 + e_{16}$, $e_8 + e_{11}$, e_{13}, e_{14}, $e_{17} + e_{18}$

$e_i + e_j \equiv$ convolution of the d.fs. of arcs i and j.

Table 3

Variance Reduction $w_M^2 / K s_K^2$ for the SPS Network†

$M=10^5$, $N=8$, $L=6$, $|H|=4$

K	F*(t) min	F*(t) max	ET*	F⋆(t) min	F⋆(t) max	ET⋆
2	42	106	7	46	196	8
2^2	53	96	9	44	345	9
2^3	74	101	12	53	773	11
2^4	87	177	17	125	957	19
2^5	111	258	23	149	1,285	24
2^6	146	701	27	313	2,722	36
2^7	248	1,066	51	534	2,984	45
2^8	365	1,409	83	710	3,183	51
2^9	383	2,426	130	1,527	5,323	123
2^{10}	987	4,612	153	2,197	4,512	138
2^{11}	1,000	5,221	266	3,438	8,604	260
2^{12}	1,518	11,341	338	5,391	13,213	303
2^{13}	1,978	16,907	548	17,299	39,357	484
2^{14}	3,289	51,709	1,258	27,943	63,234	838

†min and max for F*(t) denote, respectively, the minimal and maximal variance reduction for t=15, 30, 40, 50, 70 . min and max for F⋆(t) denote, respectively, the minimal and maximal variance reduction for t=5, 10, 15, 20, 25 .

Table 4

Variance Reduction w_M^2 / s_K^2 for the Battersby Network†

$M=10^5$, N=18 , L=8

	Partition 1						Partition 2					
	$F^*(t)$		ET^*	$F_*(t)$		ET_*	$F^*(t)$		ET^*	$F_*(t)$		ET_*
K	min	max		min	max		min	max		min	max	
2	8	11	5	5	8	4	12	18	6	6	11	5
2^2	5	7	3	3	4	2	13	18	6	5	9	4
2^3	4	5	3	2	3	2	19	25	7	7	14	6
2^4	5	3	2	1	2	1	24	46	9	22	26	11
2^5	5	7	4	1	3	2	28	54	16	22	26	12
2^6	11	23	11	5	17	18	38	92	16	26	98	18
2^7	14	32	15	3	14	16	62	153	22	65	156	36
2^8	13	38	14	2	7	10	81	226	19	81	167	44
2^9	15	38	16	1	7	11	100	332	31	223	330	88
2^{10}	17	38	15	2	11	10	203	608	56	297	619	110
2^{11}	72	109	27	19	33	16	473	1,164	104	498	1,178	395
2^{12}	59	176	48	10	59	112	637	1,864	186	1,064	2,743	411
2^{13}	115	250	77	10	78	90						

†min and max for $F^*(t)$ denote, respectively, the minimal and maximal variance reduction for t=90,110,120,130,140,150,160,180,190. min and max for $F_*(t)$ denote, respectively, the minimal and maximal variance reduction for t=20,25,309,35,40,50,60,80,90.

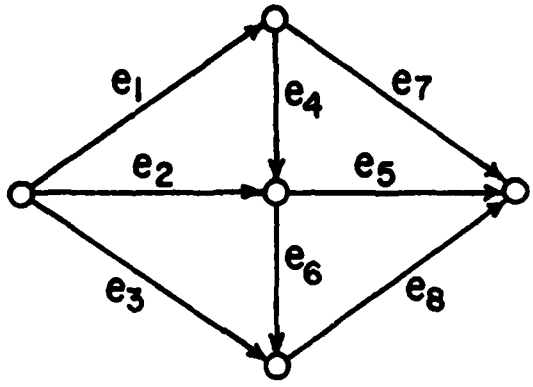


Figure 1. SPS Network

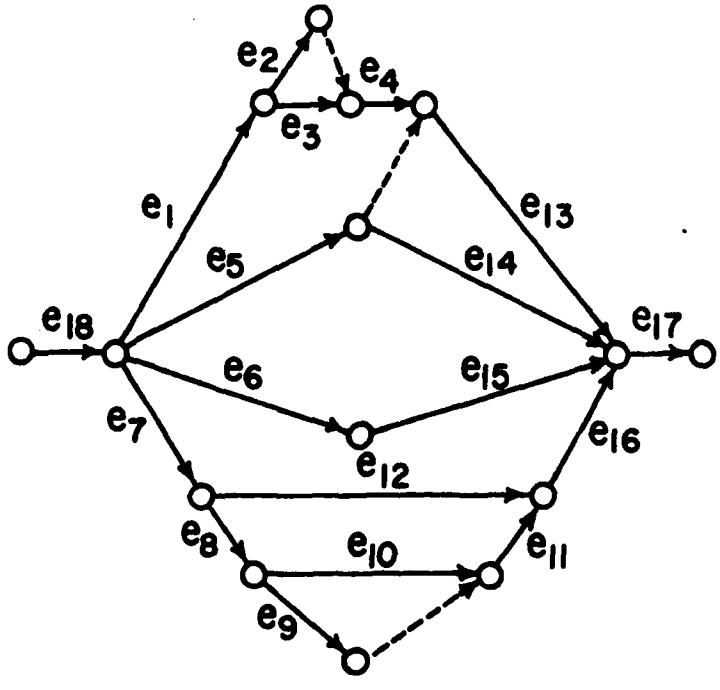


Figure 2. Battersby Network

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: UNC/ORSA/TR-82/7 — 2. GOVT ACCESSION NO.: AD-A124251	3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): VARIANCE REDUCTION IN THE SIMULATION OF STOCHASTIC ACTIVITY NETWORKS	5. TYPE OF REPORT & PERIOD COVERED: TECHNICAL REPORT
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): GEORGE S. FISHMAN	8. CONTRACT OR GRANT NUMBER(s): N00014-76-C-0302
9. PERFORMING ORGANIZATION NAME AND ADDRESS: CURRIC. IN OPERATIONS RESEARCH AND SYSTEMS ANALYSIS SMITH BUILDING UNIV. OF NORTH CAROLINA CHAPEL HILL, NC 27514	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: NAVAL ANALYSIS PROGRAM OFFICE OF NAVAL RESEARCH ARLINGTON, VA 22217	12. REPORT DATE: JANUARY 1983 13. NUMBER OF PAGES: 27
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office):	15. SECURITY CLASS. (of this report): UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

UNCLASSIFIED

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper describes a Monte Carlo method based on the theory of quasirandom points for estimating the distribution functions and means of network completion time and shortest path time in a stochastic activity network. In particular, the method leads to estimators whose variances

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

converge faster than $1/K$, where K denotes the number of replications collected in the experiment. The paper demonstrates how accuracy diminishes for a given K with increasing dimensionality of the network and shows how a procedure that uses a cutset of the network together with convolution can reduce dimensionality and increase accuracy. Two examples illustrate the benefits of using quasirandom points together with a cutset and then convolution.