UNIVERSITY OF NORTH CAROLINA
Department of Statistics
Chapel Hill, N. C.

ON MONTE CARLO METHODS IN CONGESTION PROBLEMS

II. SIMULATION OF QUEUEING SYSTEMS

by

E. S. Page

February 1963

This research was supported by the Office of Naval Research under contract No. Nonr-855(09) for research in probability and statistics at the University of North Carolina, Chapel Hill, N. C.

Institute of Statistics
Mimeo Series No. 355

ON MONTE CARLO METHODS IN CONGESTION PROBLEMS

II. SIMULATION OF QUEUEING SYSTEMS*

by

E. S. Page**

University of North Carolina

= =

Introduction

Once the often laborious task of model building and checking in a queueing situation has been completed the natural reaction is to proceed immediately to simulation and so to obtain some results as soon afterwards as possible. This note points out that a negligible amount of additional programming of the simulation model can often yield estimates of substantially smaller variance with the same amount of computation or of the same variance with less computing.

The method of antithetic variates (Hammersley and Morton, 1956) was originally devised for the evaluation of definite integrals by Monte Carlo methods. In essence, the method replaced a large number of independent estimates of the integral by a smaller number of linear combinations of suitably chosen correlated estimates. The complexity of many queueing situations prohibits exhaustive analysis into a choice of estimates but some suggestions can be made and so go some way to answering questions asked by Harling (1956) concerning the application of antithetic variates to queueing problems.

Single Server Queue

Consider the simplest single server queue in which customers arrive at a single serving station and are served in order. The interarrival times and the service

*This research was supported by the Office of Naval Research under contract No. Nonr-855(09) for research in probability and statistics at the University of North Carolina, Chapel Hill, N. C. Reproduction in whole or in part is permitted for any purpose of the United States Government.

**Present address: University Computing Laboratory, Newcastle upon Tyne, England.

times are independent random variables with known distribution functions $F(x)$, $G(y)$. A straight forward simulation to estimate, for example, the mean waiting time of a customer would consist of selecting variates $x_1, x_2, \ldots y_1, y_2 \ldots$ from the two distributions and constructing the queueing pattern. An estimate of the variance of the mean waiting time would not be obtained directly from the waiting times of the individual customers because of the correlation between them but from independent trials or from the mean waiting times in sections of the realisation long enough for the correlations to be negligible. In such a simple situation it is clear that a sequence of small x's and large y's i.e. a period of rapid arrivals of customers needing much service, will cause a queue to form. Conversely, a sequence of large x's and small y's will cause a queue in existence at the start of the sequence to diminish. Periods containing such sequences arising in the course of simple random sampling can each give unbiased estimates of the expected waiting time of a customer but the mean of such estimates from a busy and a slack period combined is likely to have much smaller standard error than an estimate from a single period. The aim of the antithetic variate approach is to obtain busy periods corresponding to the slack ones in the simulation and conversely.

Specifically suppose that the simulation is performed by forming variates x_i, y_i from independent unit rectangular variates (i.e. uniform in (0,1)) ξ_i, η_i so that $\xi_i = F(x_i)$, $\eta_i = G(y_i)$. Write $x_i = \phi(\xi_i)$, $y_i = \psi(\eta_i)$. Then an indication whether a queue is tending to disperse or increase is whether $d = x - y$ is positive or negative. In another realisation arrival and service times, x', y' can be generated so that $x_i' = \phi(\xi_i')$, $y_i' = \psi(\eta_i')$ where ξ_i', η_i' are mathematical functions of ξ_i, η_i which also have independent rectangular distributions. We consider choosing ξ_i', η_i' so that $d = x_i - y_i$ and $d' = x_i' - y_i'$ are negatively correlated. Once this choice has been decided the first simulation run proceeds as usual, calling on the uniform variates ξ_i, η_i from the random number generator.

When the run is complete the process is repeated using the <u>same</u> variates ξ_i, η_i and forming from them ξ_i', η_i' before transforming to obtain arrival and service times. Estimates from periods containing corresponding customers in the two runs are then averaged.

Since $d = x - y$, $d' = x' - y'$ we have immediately

$$E(d) = \mu_a - \mu_s = E(d')$$

$$\mathrm{Var}(d) = \sigma_a^2 + \sigma_s^2 = \mathrm{Var}(d')$$

where $\mu_a, \mu_s, \sigma_a^2, \sigma_s^2$ are respectively the mean interarrival and mean service times and the variances of these times.

<u>Transformations of the Variates</u>

Many transformations of ξ, η into ξ', η exist which leave ξ', η' independently and uniformly distributed in (0,1); however, since simplicity is desirable in the simulation we consider only the two cases where (i) $\xi' = \xi'(\xi)$, $\eta' = \eta'(\eta)$, and (ii) $\xi' = \xi'(\eta)$, $\eta' = \eta'(\xi)$.

<u>Case (i).</u>

Since $d' = x' - y' = \phi(\xi') - \psi(\eta')$

$$\mathrm{Corr}\,(d,d') = \frac{E\{\phi(\xi)\phi(\xi')\} + E\{\psi(\eta)\,\psi(\eta')\} - \mu_a^2 - \mu_s^2}{\sigma_a^2 + \sigma_s^2} \qquad (1)$$

Consequently, to obtain a large negative correlation between d, d' we need to reduce the first two terms of the numerator as much as possible. Since $\phi(\xi)$ is monotonic and $\phi(\xi')$ is necessarily some rearrangement of $\phi(\xi)$ it follows (Hardy, et al. 2nd edition, theorem 378) that

$$E\{\phi(\xi)\,\phi(\xi')\} = \int_0^1 \phi(\xi)\,\phi(\xi')\,d\xi$$

is least when $\phi(\xi')$ is monotone in the opposite sense to $\phi(\xi)$. Hence the least (algebraically) correlation is obtained when

$$\xi' = 1 - \xi$$

and (2)

$$\eta' = 1 - \eta \ .$$

To show that this correlation is indeed negative we need to prove that

$$E\left\{\phi(\xi)\ \phi(\xi')\right\} - \mu_a^2 \leq 0 \ ,$$

or since,

$$\mu_a = \int_0^1 \phi(\xi)d\xi = \int_0^1 \phi(1-\xi)d\xi \ ,$$

to prove that

$$\int_0^1 \phi(\xi)\ \phi(1-\xi)d\xi \leq \int_0^1 \phi(\xi)\ d\xi \ . \int \phi(1-\xi)d\xi \ . \qquad (3)$$

More generally, consider

$$I = \int_0^1 \left\{f(x) - \bar{f}\right\} g(x)\ dx = \int_0^1 f(x)g(x)dx - \int_0^1 f(x)dx \int_0^1 g(x)dx$$

where $f(x) \geq 0$ and is continuous and non-decreasing in $(0,1)$, $\bar{f} = \int_0^1 f(x)dx$ and $g(x) \geq 0$. Then there exists an X in $(0,1)$ such that $f(X) = \bar{f}$ and $f(x) \geq \bar{f}$ for $x \geq X$. $f(x) \leq \bar{f}$ for $x \leq X$.

Hence

$$I = \int_0^X + \int_X^1 \left\{f(x) - \bar{f}\right\} g(x)dx \ .$$

If $g(x)$ is monotonic decreasing,

$$I \leq g(X) \int_0^X \left\{f(x) - \bar{f}\right\} dx + g(X) \int_X^1 \left\{f(x) - \bar{f}\right\} dx$$

$$= 0$$

and (3) is established.

Similarly, we obtain the result needed for case (ii) that $I \geq 0$ when $g(x)$ is monotonic increasing.

Case (ii)

When $d' = \phi(\xi') = \psi(\eta')$ and $\xi' = \xi'(\eta)$, $\eta' = \eta'(\xi)$,

$$\text{Corr}\,(d,d') = \frac{2\,\mu_a\mu_s - \int_0^1 \phi(\xi)\,\psi(\eta')d\xi - \int_0^1 \phi(\xi')\,\psi(\eta)\,d\eta}{\sigma_a^2 + \sigma_s^2} \;; \qquad (4)$$

the correlation is least when the last two terms of the numerator are greatest. It follows as in case (i) that $\phi(\xi)$, $\psi(\eta')$ should be monotonic in the same sense and hence that the transformation required is

$$\begin{aligned} \xi' &= \eta \\ \eta' &= \xi \end{aligned} \qquad (5)$$

and that it produces a negative correlation between d, d'.

Example

If arrivals are random and service times are exponential with traffic intensity ρ, the distribution functions can be taken as $F(x) = 1 - e^{-x}$, $G(y) = 1 - e^{-y/\rho}$.

Hence $\phi(\xi) = -\log(1-\xi)$, $\psi(\eta) = -\rho\log(1-\eta)$.

Thus, for the transformation (2), $\xi' = 1 - \xi$, $\eta' = 1 - \eta$, we obtain

$$\text{Corr}\,(d,d') = -0.645\,, \qquad (6)$$

since

$$\int \log x \log(1-x)dx = \sum_1^\infty k^{-1}(k+1)^{-2} = 2 - \frac{\pi^2}{6}\,.$$

In this case (and in some others but not in general) the correlation is independent of the traffic intensity.

For the other transformation (5), $\xi' = \eta$, $\eta' = \xi$

$$\text{corr}\,(d,d') = -2\rho/(1+\rho^2)\,. \qquad (7)$$

Simulation Results

The complexities of congestion systems usually prevent direct consideration of the effects of a proposed antithetic variate sampling procedure on the quantities of interest. Even in the simple example above we have concentrated attention on the differences x-y instead of the waiting times whose mean we wish to estimate. How far these considerations lead to large negative correlations between the new observations and those of the basic simulation can only be found by trial.

The simple single server queue considered above is, of course, only an illustrative example; Monte Carlo methods will only be undertaken when theoretical study is impracticable. However, trials were performed to exemplify the method.

A queue with traffic intensity $\rho = 0.75$ was simulated for four sequences of over 500 customers each and the total waiting times of non-overlapping sequences of 20 customers were recorded. Single runs for traffic intensities $\rho = 0.8, 0.85, 0.9$ were also performed. In each case the direct run (A) using rectangular variates ξ_i, η_i was run, followed by antithetic runs B, C using $1 - \xi_i$, $1 - \eta_i$, and η_i, ξ_i respectively. The efficiency of using one or both antithetic runs relative to the comparable number of replications of the direct run is shown in Table 1; thus the efficiency 2.0 shown for the use of A and B in the first run for $\rho = 0.75$ implies run A would need to be replicated to contain twice as many customers as are considered in runs A and B put together (i.e. four times as many observations as in run A) in order to achieve the same standard error for the estimate of mean waiting time.

Table 1.

Traffic Intensity	0.75	0.75	0.75	0.75	0.80	0.85	0.90
Efficiency: (A, B)	2.0	1.4	1.9	1.7	2.0	1.9	2.1
(A, C)	1.9	1.9	2.2	1.8	1.8	1.7	1.6
(A, B, C)	2.3	1.1	1.7	1.3	2.3	2.4	1.8

Efficiencies of Antithetic Methods

Assessment of the Method

The efficiencies observed in the runs performed always exceeded unity and were usually near two. Naturally these figures will change with the arrival and service time distributions and from model to model and also with the skill with which the different variates are chosen. It may be possible to effect greater efficiency by taking combinations of antithetic variates although in complex situations it may be difficult to select a good combination. As simulation studies of congestion systems are necessarily long in order to represent their behaviour adequately any simple technique which increases the accuracy is welcome. In the queueing situation above the use of this antithetic variate method requires the addition or amendment of only two or three instructions in a computer program and is likely to reduce the computing time substantially.

Summary

When variates in a simulation study of a congestion system are generated from uniform variates ξ_i, η_i, ... by inverting the relevant distribution functions it is suggested that such a direct run be combined with others using the variates $1 - \xi_i$, $1 - \eta_i$, ... or η_i, ξ_i, ... In a simple queueing situation it is shown that such techniques can result in halving the number of observations required for a specified accuracy.

REFERENCES

Hammersley, J. M. and Morton, K. W. (1956). A new Monte Carlo technique: antithetic variates. Proc. Camb. Phil. Soc. 52, 449-475.

Harling, J. (1956). Simulation in Operations Research. Op. Res. 6, 307-319.