Probability in the Engineering and Informational Sciences, **20**, 2006, 251–256. Printed in the U.S.A.

MAKING SIMULATIONS MORE EFFICIENT WHEN ANALYZING POISSON ARRIVAL SYSTEMS AND MEANS OF MONOTONE FUNCTIONS

Sheldon M. Ross
Epstein Department of Industrial and Systems Engineering
University of Southern California
Los Angeles, CA 90089-0193
smross@usc.edu

For a system in which arrivals occur according to a Poisson process, we give a new approach for using simulation to estimate the expected value of a random variable that is independent of the arrival process after some specified time t. We also give a new approach for using simulation to estimate the expected value of an increasing function of independent uniform random variables. Stratified sampling is a key technique in both cases.

1. INTRODUCTION AND SUMMARY

In Section 2 we consider a model in which arrivals occur according to a Poisson process, and we then present an efficient way of using stratified sampling to estimate the expected value of a random variable whose mean depends on the arrival process only through arrivals up to some specified time t. In Section 3 we show how to use stratified sampling to efficiently estimate the expected value of a nondecreasing function of random numbers.

2. SYSTEMS HAVING POISSON ARRIVALS

Consider a system in which arrivals occur according to a Poisson process and suppose that we are interested in using simulation to compute $E[D]$, where the value of D depends on the arrival process only through those arrivals before time t. For

 0269-9648/06 $16.00

https://doi.org/10.1017/S0269964806060153 Published online by Cambridge University Press

instance, D might be the sum of the delays of all arrivals by time t in a parallel multiserver queuing system. We suggest the following approach to using simulation to estimate $E[D]$. First, with $N(t)$ equal to the number of arrivals by time t, note that

$$E[D] = \sum_{j=0}^{m} E[D|N(t)=j]\,\frac{e^{-\lambda t}(\lambda t)^j}{j!} + E[D|N(t)>m]\left(1-\sum_{j=0}^{m}\frac{e^{-\lambda t}(\lambda t)^j}{j!}\right). \quad \textbf{(1)}$$

Each run of our suggested simulation procedure generates an independent estimate of $E[D]$. At each stage of a run, we will have a set S whose elements are arranged in increasing value and which represents the set of arrival times. For simplicity, we will present our approach under the assumption that $E[D|N(t)=0]$ can be easily computed and also that D can be determined by knowing the arrival times along with the service time of each arrival. A run is as follows:

1. Let $N=1$. Generate a random number U_1 and let $S=\{tU_1\}$.
2. Suppose $N(t)=1$, with the arrival occurring at time tU_1. Generate the service time of this arrival and compute the resulting value of D. Call this value D_1.
3. Let $N=N+1$.
4. Generate a random number U_N and add tU_N in its appropriate place to the set S so that the elements is S are in increasing order.
5. Suppose $N(t)=N$, with S specifying the N arrival times; generate the service time of the arrival at time tU_N, and using the previously generated service times of the other arrivals, compute the resulting value of D. Call this value D_N.
6. If $N<m$, return to Step 3. If $N=m$, use the inverse transform method to generate the value of $N(t)$ conditional on it exceeding m. If the generated value is $m+k$, generate k additional random numbers, multiply each by t, and add these k numbers to the set S. Generate the service times of these k arrivals and, using the previously generated service times, compute D. Call this value $D_{>m}$.

With $D_0 = E[D|N(t)=0]$, the estimate from this run is

$$EST = \sum_{j=0}^{m}\frac{D_j e^{-\lambda t}(\lambda t)^j}{j!} + D_{>m}\left(1-\sum_{j=0}^{m}\frac{e^{-\lambda t}(\lambda t)^j}{j!}\right).$$

Using the fact that the set of unordered arrival times, given that $N(t)=j$, is distributed as a set of j independent uniform $(0,t)$ random variables, it follows that the preceding is an unbiased estimator of $E[D]$. Generating multiple runs and taking the average value of the resulting estimates yields the final simulation estimator.

https://doi.org/10.1017/S0269964806060153 Published online by Cambridge University Press

We now show that EST has a smaller variance than does the raw simulation estimator D.

THEOREM 1:

$$\mathrm{Var}(EST) \leq \mathrm{Var}(D).$$

PROOF: The quantity D can be simulated as follows:

1. Generate the value of N', a random variable whose distribution is the same as that of $N(t)$ condition to exceed m. That is,

$$P\{N' = k\} = \frac{(\lambda t)^k/k!}{\sum_{k=m+1}^{\infty} (\lambda t)^k/k!}, \qquad k > m.$$

2. Generate the values of $A_1, \ldots, A_{N'}$, independent uniform $(0, t)$ random variables.
3. Generate the values of $S_1, \ldots, S_{N'}$, independent service time random variables.
4. Generate the value of $N(t)$, a Poisson random variable with mean λt.
5. If $N(t) \leq m$, use the arrival times $A_1, \ldots, A_{N(t)}$ along with their service times $S_1, \ldots, S_{N(t)}$ to compute the value of D.
6. If $N(t) > m$, use the arrival times $A_1, \ldots, A_{N'}$ along with their service times $S_1, \ldots, S_{N'}$ to compute the value of D.

It is now easy to check, by conditioning on $N(t)$, that

$$EST = E[D \mid N', A_1, \ldots, A_{N'}, S_1, \ldots, S_{N'}].$$

The result now follows from the conditional variance formula. ■

Remarks:

1. The use of Eq. (1) is the use of stratified sampling (see [3]), which has been previously suggested when analyzing Poisson arrival systems (see, for instance, [1, p. 228]). However, previous suggestions were to first estimate $E[D \mid N(t) = 1]$ by a sequence of independent runs; then to estimate $E[D \mid N(t) = 2]$ by a new sequence of independent runs (that are also independent of the runs used to compute $E[D \mid N(t) = 1]$), and so on. This differs from our idea of estimating all of the quantities $E[D \mid N(t) = j], j = 1, \ldots, m$, and $E[D \mid N(t) > m]$ in a single run, sequentially making use of the set of generated data values used to estimate $E[D \mid N(t) = j]$ to speed the estimation of $E[D \mid N(t) = j + 1]$, and so on.
2. It should be noted that the variance of our estimator $\sum_{j=0}^{m} D_j e^{-\lambda t}(\lambda t)^j/j! + D_{>m}(1 - \sum_{j=0}^{m} e^{-\lambda t}(\lambda t)^j/j!)$ is, because of the positive correlations introduced by reusing the same data, larger than it would be if the D_j were inde-

https://doi.org/10.1017/S0269964806060153 Published online by Cambridge University Press

pendent estimators. However, we believe that the increased speed of the simulation will usually more than make up for this increased variance.

3. When computing D_{j+1}, we can make use of quantities used in computing D_j. For instance, suppose $D_{i,j}$ was the delay of arrival i when $N(t) = j$. Then if the new arrival time tU_{j+1} is the kth smallest of the new set S, then $D_{i,j+1} = D_{i,j}$ for $i < k$.
4. Other variance reduction ideas can be used in conjunction with our approach. For instance, we can improve the estimator by using a linear combination of the service times as a control variable.

It remains to determine an appropriate value of m. A reasonable approach might be to choose m to make $E[D|N(t) > m](1 - \sum_{j=0}^{m} e^{-\lambda t}(\lambda t)^j/j!)$ sufficiently small. Because $\text{Var}(N(t)) = \lambda t$, a reasonable choice would be of the form $m = \lambda t + k\sqrt{\lambda t}$ for some positive integer k. One can often bound $E[D|N(t) > m]$ and use this bound to determine the appropriate value of m. For instance, suppose D is the sum of the delays of all arrivals by time t in a single server system with mean service time 1. Then because this quantity will be maximized when all arrivals come simultaneously, we see that

$$E[D|N(t)] \leq \sum_{i=1}^{N(t)-1} i.$$

Because the conditional distribution of $N(t)$ given that it exceeds m will, when m is at least five standard deviations greater than $E[N(t)]$, put most of its weight near $m + 1$, we see from the preceding equation that one can reasonably assume that

$$E[D|N(t) > m] \leq (m + 1)^2/2.$$

Using that for a standard normal random variable Z (see [2]),

$$P(Z > x) \leq \left(1 - \frac{1}{x^2} + \frac{3}{x^4}\right)\frac{e^{-x^2/2}}{c\sqrt{2\pi}}, \qquad x > 0,$$

we see, upon using the normal approximation to the Poisson, that for $m = \lambda t + k\sqrt{\lambda t}$, we can reasonably assume that

$$E[D|N(t) > m]P\{N(t) > m\} \leq (m + 1)^2 \frac{e^{-k^2/2}}{2k\sqrt{2\pi}}.$$

For instance, with $\lambda t = 10^3$ and $k = 6$, the preceding upper bound is about 0.0008.

3. COMPUTING THE EXPECTED VALUE OF AN INCREASING FUNCTION

Suppose that we want to use simulation to compute $E[g(U_1, \ldots, U_n)]$, where $U_1, \ldots, U_n$ are independent uniform(0,1) random variables and g is nondecreasing in each of its coordinates. Because of the monotonicity of g, $\prod_{i=1}^{n} U_i$ will often be a good

https://doi.org/10.1017/S0269964806060153 Published online by Cambridge University Press

predictor of g in the sense that $E[\text{Var}(g(U_1,\ldots,U_n)|\prod_{i=1}^n U_i)]$ will be relatively small. Because of this, we suggest generating $U_1,\ldots,U_n$ by first generating the value of $\prod_{i=1}^n U_i$ and then generating $U_1,\ldots,U_n$ conditional on the generated value of the product. In generating the value of $\prod_{i=1}^n U_i$ on different simulation runs, we suggest using stratified sampling.

To implement the preceding idea, we need to first show how to generate $\prod_{i=1}^n U_i$ and how to generate $U_1,\ldots,U_n$ conditional on $\prod_{i=1}^n U_i$. Note the following:

(a) $-\ln(U_1\cdots U_n)$ is a gamma$(n,1)$ random variable.
(b) $-\ln(U_1\cdots U_j), j = 1,\ldots,n$, can be regarded as the first j event times of a Poisson process with rate 1.
(c) Given that the nth event of a Poisson process occurs at time t, the first $n-1$ event times are distributed as the ordered values of a set of $n-1$ uniform$(0,t)$ random variables.

Now the generation can proceed as follows:

1. Generate the value of T, a gamma random variable with parameters $(n,1)$. (T will equal $-\ln(U_1\cdots U_n)$.)
2. Generate $n-1$ random numbers, $V_1,\ldots,V_{n-1}$, and order them to obtain $V_{(1)} < \cdots < V_{(n-1)}$. (So, $TV_{(j)} = -\ln(U_1\cdots U_j)$.)
3. Let $V_{(0)} = 0$ and $V_{(n)} = 1$, and set

$$U_j = e^{-T[V_{(j)}-V_{(j-1)}]}, \qquad j = 1,\ldots,n.$$

Let G_n denote the distribution function of a gamma random variable with parameters $(n,1)$. Suppose that we are planning to do m simulation runs. Then on the kth simulation run, a random number U should be generated and T should be taken to equal $G_n^{-1}((U+k-1)/m)$ (i.e., we use stratified sampling when generating the successive values of T). We should then follow Steps 2 and 3 of the preceding and then for the values of U_i obtained in Step 3, compute $g(U_1,\ldots,U_n)$. The average of the values of $g(U_1,\ldots,U_n)$ obtained in the m runs is then taken as the estimator of $E[g(U_1,\ldots,U_n)]$.

Remarks:

1. It follows from the fact that a gamma random variable with parameters $(n,1)$ has the same distribution as does $\frac{1}{2}\chi^2_{2n}$, where χ^2_{2n} is a chi-squared random variable with $2n$ degrees of freedom, that

$$G_n^{-1}(x) = \frac{1}{2} F^{-1}_{\chi^2_{2n}}(x),$$

where $F_{\chi^2_{2n}}$ is the distribution function of a chi-squared random variable with $2n$ degrees of freedom. Approximations for the inverse of the chi-square distribution function are readily available in the literature.

https://doi.org/10.1017/S0269964806060153 Published online by Cambridge University Press

2. We suggest using $G_n^{-1}((k-0.5)/m)$ as the value of T on the kth simulation run; that is, rather than generating the value of a uniform random variable on $((k-1)/m, k/m)$ to determine T, just use the value $(k-0.5)/m$.
3. Another way to do the stratification is to choose r values $0 = a_0 < a_1 < \cdots < a_r < a_{r+1} = \infty$. Determine $p_i = P\{a_{i-1} < T < a_i\}$, $i = 1, \ldots, r+1$, and then perform mp_i of the m planned simulation runs with T simulated conditional on its being in the interval (a_{i-1}, a_i). This conditional value can be simulated by using the rejection procedure based on an exponential random variable that is conditioned to lie in the same interval. Indeed, even better is to do a small preliminary simulation to estimate the quantities $\text{Var}(g(U_1, \ldots, U_n) | a_{i-1} < T < a_i)$. If $s_i^2, i = 1, \ldots, r+1$, are the estimates, then do a total of $m(p_i s_i^2 / \sum_j p_j s_j^2)$ of the runs conditional on $a_{i-1} < T < a_i$, either by using the inverse transform or the rejection method. (If the inverse transform is used, then you should stratify within the subintervals.) The final estimate is $\sum_{j=1}^{r+1} p_j \bar{X}_j$, where $\bar{X}_j$ is the average of the simulation runs done conditional on $a_{i-1} < T < a_i$.
4. If $g(u_1, \ldots, u_n)$ is only increasing in k of its variables, say in $u_1, \ldots, u_k$, then we can generate $U_1, \ldots, U_n$ on successive runs by using stratified sampling on $\prod_{i=1}^{k} U_i$, generating $U_1, \ldots, U_k$ conditional on the value of this product and generating $U_i, i > k$, as independent uniforms.
5. If $g(u_1, \ldots, u_n)$ is increasing in some of its variables and decreasing in the others, then we can still utilize the preceding idea. For if g is, say, increasing in its first r variables and decreasing in its final $n-r$, then

$$E[g(U_1, \ldots, U_n)] = E[g(U_1, \ldots, U_r, 1-U_{r+1}, \ldots, 1-U_n)],$$

which shows that we can just work with the increasing function $h(u_1 \ldots, u_n) = g(u_1, \ldots, u_r, 1-u_{r+1}, \ldots, 1-u_n)$.

References

1. Glasserman, P. (2004). *Monte Carlo methods in financial engineering*. New York: Springer.
2. Ross, S.M. (2002). *Probability models for computer science*. New York: Academic Press.
3. Ross, S.M. (2002). *Simulation*. New York: Academic Press.