7 Variance Reduction Techniques

In a simulation study, we are interested in one or more performance measures for some stochastic model. For example, we want to determine the long-run average waiting time, $E(W)$, of customers in a $G/G/1$ queue. To estimate $E(W)$ we can do a number of, say K, independent runs, the i-th one yielding the output random variable W_i with $E(W_i) = E(W)$. After K runs have been performed, an estimator of $E(W)$ is given by $\bar{W} = \sum_{i=1}^{K} W_i/K$. However, if we were able to obtain a different unbiased estimator of $E(W)$ having a smaller variance than $\bar{W}$, we would obtain an estimator with a smaller confidence interval. In this section we will present a number of different methods that one can use to reduce the variance of the estimator $\bar{W}$. We will successively describe the following techniques:

1. Common Random Numbers
2. Antithetic Variables
3. Control Variates
4. Conditioning

The first method is typically used in a simulation study in which we want to compare performance measures of two different systems. All other methods are also useful in the case that we want to simulate a performance measure of only a single system.

7.1 Common Random Numbers

The idea of the method of common random numbers is that if we compare two different systems with some random components it is in general better to evaluate both systems with the same realizations of the random components. Key idea in the method is that if X and Y are two random variables, then

$$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\mathrm{Cov}(X, Y).$$

Hence, in the case that X and Y are *positively* correlated, i.e. $\mathrm{Cov}(X, Y) > 0$, the variance of $X - Y$ will be smaller than in the case that X and Y are independent. In general, the use of common random numbers leads to positive correlation of the outcomes of a simulation of two systems. As a consequence, it is better to use common random numbers instead of independent random numbers when we compare two different systems.

Let us illustrate the method using the following scheduling example. Suppose that a finite number of N jobs has to be processed on two identical machines. The processing times of the jobs are random variables with some common distribution function F. We want to compare the completion time of the last job, $C_{\max}$, under two different policies. In the LPTF policy, we always choose the remaining job with the longest processing time first, in the SPTF policy we choose the remaining job with the shortest processing time first.

In Table 5 and Table 6 we compare for $N = 10$ and $F(x) = 1 - e^{-x}$ the estimators and confidence intervals for $E(C_{\max}^{\mathrm{SPTF}} - C_{\max}^{\mathrm{LPTF}})$ when we do not, resp. do, use common random numbers.

We conclude that in this example the use of common random numbers reduces the standard deviation of the estimator and hence also the length of the confidence interval with a factor 5.

# of runs	mean	st. dev.	95% conf. int.
1000	0.8138	2.5745	[0.645, 0.973]
10000	0.8293	2.4976	[0.780, 0.878]
100000	0.8487	2.4990	[0.833, 0.864]
1000000	0.8398	2.4951	[0.835, 0.845]

Table 5: Estimation of $E(C_{\max}^{\text{SPTF}} - C_{\max}^{\text{LPTF}})$ without using common random numbers

# of runs	mean	st. dev.	95% conf. int.
1000	0.8559	0.5416	[0.822, 0.889]
10000	0.8415	0.5230	[0.831, 0.852]
100000	0.8394	0.5164	[0.836, 0.843]
1000000	0.8391	0.5168	[0.838, 0.840]

Table 6: Estimation of $E(C_{\max}^{\text{SPTF}} - C_{\max}^{\text{LPTF}})$ using common random numbers

When we want to use common random numbers, the problem of synchronization can arise: How can we achieve that the same random numbers are used for the generation of the same random variables in the two systems?

In the previous example, this synchronization problem did not arise. However, to illustrate this problem, consider the following situation. In a $G/G/1$ queueing system the server can work at two different speeds, v_1 and v_2. Aim of the simulation is to obtain an estimator for the difference of the waiting times in the two situations. We want to use the same realizations of the interarrival times and the sizes of the service requests in both systems (the service time is then given by the sizes of the service request divided by the speed of the server). If we use the program of the discrete event simulation of Section 3 of the $G/G/1$ queue, then we get the synchronization problem because the order in which departure and arrival events take place depends on the speed of the server. Hence, also the order in which interarrival times and sizes of service requests are generated depend on the speed of the server.

The synchronization problem can be solved by one of the following two approaches:

1. Use separate random number streams for the different sequences of random variables needed in the simulation.

2. Assure that the random variables are generated in exactly the same order in the two systems.

For the example of the $G/G/1$ queue, the first approach can be realized by using a separate random number stream for the interarrival times and for the service requests. The second approach can be realized by generating the service request of a customer already at the arrival instant of the customer.

7.2 Antithetic Variables

The method of antithetic variables makes use of the fact that if U is uniformly distributed on $(0, 1)$ then so is $1 - U$ and furthermore U and $1 - U$ are negatively correlated. The key idea is that, if

W_1 and W_2 are the outcomes of two successive simulation runs, then

$$\text{Var}\left(\frac{W_1 + W_2}{2}\right) = \frac{1}{4}\text{Var}(\text{W}_1) + \frac{1}{4}\text{Var}(W_2) + \frac{1}{2}\text{Cov}(W_1, W_2)].$$

Hence, in the case that W_1 and W_2 are *negatively* correlated the variance of $(W_1 + W_2)/2$ will be smaller than in the case that W_1 and W_2 are independent.

The question remains how we can achieve that the outcome of two successive simulation runs will be negatively correlated. From the fact that U and $1 - U$ are negatively correlated, we may expect that, if we use the random variables $U_1, \ldots, U_m$ to compute W_1, the outcome of the first simulation run, and after that $1 - U_1, \ldots, 1 - U_m$ to compute W_2, the outcome of the second simulation run, then also W_1 and W_2 are negatively correlated. Intuition here is that, e.g., in the simulation of the $G/G/1$ queue large realizations of the U_i's corresponding to large service times lead to large waiting times in the first run. Using the antithetic variables, this gives small realizations of the $1-U_i$'s corresponding to small service times and hence leading to small waiting times in the second run.

We illustrate the method of antithetic variables using the scheduling example of the previous subsection.

# of runs	mean	st. dev.	95% conf. int.
1000	5.0457	1.6201	[4.945, 5.146]
10000	5.0400	1.6020	[5.009, 5.071]
100000	5.0487	1.5997	[5.039, 5.059]
1000000	5.0559	1.5980	[5.053, 5.059]

Table 7: Estimation of $E(C_{\max}^{\text{LPTF}})$ without using antithetic variables

# of pairs	mean	st. dev.	95% conf. int.
500	5.0711	0.7216	[5.008, 5.134]
5000	5.0497	0.6916	[5.030, 5.069]
50000	5.0546	0.6858	[5.049, 5.061]
500000	5.0546	0.6844	[5.053, 5.056]

Table 8: Estimation of $E(C_{\max}^{\text{LPTF}})$ using antithetic variables

In Table 7 and Table 8 we compare, again for $N = 10$ and $F(x) = 1 - e^{-x}$, the estimators and confidence intervals for $E(C_{\max}^{\text{LPTF}})$ when we do not, resp. do, use antithetic variables. So, for example, we compare the results for 1000 independent runs with the results for 1000 runs consisting of 500 pairs of 2 runs where the second run of each pair uses antithetic variables. We conclude that in this example the use of antithetic variables reduces the length of the confidence interval with a factor 1.5.

Finally, remark that, like in the method of common random numbers, the synchronization problem can arise. Furthermore, it should be noted that the method is easier to implement if all random variables are generated using the inversion transform technique (only one uniform random number is needed for the realization of one random variable) than if we use, e.g., the rejection method to generate random variables (a random number of uniform random numbers are needed for the realization of one random number).

7.3 Control Variates

The method of control variates is based on the following idea. Suppose that we want to estimate some unknown performance measure $E(X)$ by doing K independent simulation runs, the i-th one yielding the output random variable X_i with $E(X_i) = E(X)$. An unbiased estimator for $E(X)$ is given by $\bar{X} = (\sum_{i=1}^{K} X_i)/K$. However, assume that at the same time we are able to simulate a related output variable Y_i, with $E(Y_i) = E(Y)$, where $E(Y)$ is known. If we denote by $\bar{Y} = (\sum_{i=1}^{K} Y_i)/K$, then, for any constant c, the quantity $\bar{X} + c(\bar{Y} - E(Y))$ is also an unbiased estimator of $E(X)$. Furthermore, from the formula

$$\text{Var}(\bar{X} + c(\bar{Y} - E(Y))) = \text{Var}(\bar{X}) + c^2\text{Var}(\bar{Y}) + 2c\text{Cov}(\bar{X}, \bar{Y}).$$

it is easy to deduce that $\text{Var}(\bar{X} + c(\bar{Y} - E(Y)))$ is minimized if we take $c = c^*$, where

$$c^* = -\frac{\text{Cov}(\bar{X}, \bar{Y})}{\text{Var}(\bar{Y})}.$$

Unfortunately, the quantities $\text{Cov}(\bar{X}, \bar{Y})$ and $\text{Var}(\bar{Y})$ are usually not known beforehand and must be estimated from the simulated data. The quantity $\bar{Y}$ is called a *control variate* for the estimator $\bar{X}$. To see why the method works, suppose that $\bar{X}$ and $\bar{Y}$ are positively correlated. If a simulation results in a large value of $\bar{Y}$ (i.e. $\bar{Y}$ larger than its mean $E(Y)$) then probably also $\bar{X}$ is larger than its mean $E(X)$ and so we correct for this by lowering the value of the estimator $\bar{X}$. A similar argument holds when $\bar{X}$ and $\bar{Y}$ are negatively correlated.

In the simulation of the production line of section 2, a natural control variate would be the long-term average production rate for the line with zero buffers.

7.4 Conditioning

The method of conditioning is based on the following two formulas. If X and Y are two arbitrary random variables, then $E(X) = E(E(X|Y))$ and $\text{Var}(X) = E(\text{Var}(X|Y)) + \text{Var}(E(X|Y)) \geq \text{Var}(E(X|Y))$. Hence, we conclude that the random variable $E(X|Y)$ has the same mean as and a smaller variance than the random variable X.

How can we use these results to reduce the variance in a simulation? Let $E(X)$ be the performance measure that we want to estimate. If Y is a random variable such that $E(X|Y = y)$ is known, then the above formulas tell us that we can better simulate Y and use $E(X|Y)$ than that we directly simulate X.

The method is illustrated using the example of an $M_\lambda/M_\mu/1/N$ queueing model in which customers who find upon arrival N other customers in the system are lost. The performance measure that we want to simulate is $E(X)$, the expected number of lost customers at some fixed time t. A direct simulation would consist of K simulation runs until time t. Denoting by X_i the number of lost customers in run i, then $\bar{X} = (\sum_{i=1}^{K} X_i)/K$ is an unbiased estimator of $E(X)$. However, we can reduce the variance of the estimator in the following way. Let Y_i be the total amount of time in the interval $(0, t)$ that there are N customers in the system in the i-th simulation run. Since customers arrive according to a Poisson process with rate λ, it follows that $E(X|Y_i) = \lambda Y_i$. Hence an improved estimator would be $\lambda\bar{Y}$, where $\bar{Y} = (\sum_{i=1}^{K} Y_i)/K$.

In Table 9 and Table 10 we compare, for $\lambda = 0.5$, $\mu = 1$, $N = 3$ and $t = 1000$, the estimators and confidence intervals for $E(X)$ when we do not, resp. do, use conditioning. We conclude that

# of runs	mean	st. dev.	95% conf. int.
10	33.10	10.06	[26.86, 39.33]
100	34.09	9.21	[32.28, 35.90]
1000	33.60	8.88	[33.05, 34.15]

Table 9: Estimation of $E(X)$ without using conditioning

# of runs	mean	st. dev.	95% conf. int.
10	33.60	7.16	[29.17, 38.05]
100	33.82	6.91	[32.46, 35.18]
1000	33.36	6.57	[32.95, 33.76]

Table 10: Estimation of $E(X)$ using conditioning

in this example the use of conditioning reduces the standard deviation of the estimator and hence also the length of the confidence interval with a factor 1.3.