Solutions to Problems in Chapter 11 of
Simulation Modeling and Analysis, 5th ed., 2015, McGraw-Hill, New York
by Averill M. Law

11.1. Assuming a terminating-simulation environment in which we simply replicate for statistical analysis, we can look at this in three ways, any of which leads to the same conclusion.

First, suppose we want to achieve a prespecified variance r in the result averaged over replications. Making n_0 replications via straightforward simulation, the variance of the result is V_0/n_0, while making n_1 replications with the VRT leads to a final variance of V_1/n_1. We want sample sizes n_0 and n_1 that are large enough so that $V_0/n_0 = r = V_1/n_1$, leading to $n_0 = V_0/r$ and $n_1 = V_1/r$ (assume that r divides exactly into both V_0 and V_1). The cost, then, of achieving the desired precision using straightforward simulation is $C_0n_0 = C_0V_0/r$, and would be $C_1n_1 = C_1V_1/r$ if we instead use the VRT. For the VRT to make economic sense, it must cost less to get us to the desired precision, i.e., it must be good enough so that $C_1V_1/r < C_0V_0/r$, i.e., $C_1V_1 < C_0V_0$. Rewriting this condition as $V_1/V_0 < C_0/C_1$, the interpretation is that the ratio of the VRT variance V_1 to the straightforward variance V_0 must be less than the ratio of the straightforward cost C_0 per replication to the per-replication cost C_1 when using the VRT. In percent-change terms, the condition is also equivalent to

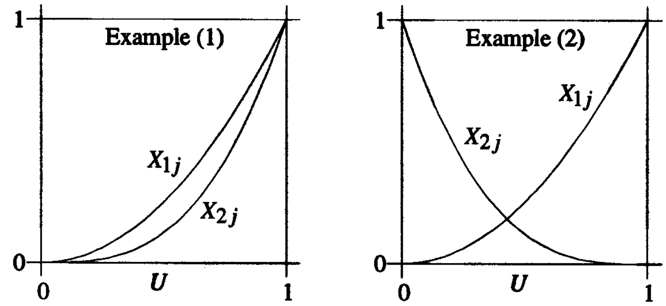
$$100\left(1-\frac{V_1}{V_0}\right) > 100\left(1-\frac{C_0}{C_1}\right)$$

The left-hand side is the percent reduction in variance if we *do* use the VRT, which must exceed the right-hand side, being the percent reduction in cost if we *don't* use the VRT. Note that neither the condition for the VRT to be worthwhile nor its interpretations depend on the amount r of precision required.

Alternatively, suppose we are given a budget constraint b for the simulation, and want to get the lowest-variance final result we can. If we do straightforward simulation, we could afford $n_0 = b/C_0$ replications, and if we use the VRT we could make $n_1 = b/C_1$ replications (assume that b is exactly divisible by both C_0 and C_1). The variance of the result under straightforward simulation would thus be $V_0/n_0 = V_0C_0/b$, and using the VRT our result would have variance $V_1/n_1 = V_1C_1/b$. For the VRT to pay off, then, it is required that $V_1C_1/b < V_0C_0/b$, i.e., $V_1C_1 < V_0C_0$, which is the same condition (with the same interpretations) as before. Note that this condition does not depend on the amount b of the budget.

Finally, we can quantify the overall effectiveness of a VRT, taking into account the amount of the variance reduction and any difference in cost at the same time. Define the *efficiency ratio* of the VRT (relative to straightforward simulation) to be $(C_0V_0)/(C_1V_1)$. To motivate the name, suppose that there is no added cost to using the VRT (i.e., $C_1 = C_0$), and that the VRT cuts the variance in half (i.e., $V_1 = V_0/2$). Then it seems reasonable to claim that using the VRT doubles our efficiency, and indeed the efficiency ratio is $(C_0V_0)/(C_0V_0/2) = 2$. Thus, for the VRT to pay off the efficiency ratio must exceed 1; this is equivalent to our original condition $C_1V_1 < C_0V_0$.

11.2. (*a*)



(*b*) First note that for any positive integer k, $E(U^k) = \int_0^1 u^k du = [u^{k+1}/(k+1)]_{u=0}^{u=1} = 1/(k+1)$ since $U \sim$ U(0, 1), and recall that for any two random variables A and B, Cov(A, B) = $E(AB) - E(A)E(B)$.

Example (1): Using the above formula directly, we get $E(X_{1j}) = E(U^2) = 1/3$, $E(X_{2j}) = E(U^3) = 1/4$, and $E(X_{1j}X_{2j}) = E(U^5) = 1/6$, so $\mathrm{Cov}(X_{1j}, X_{2j}) = 1/6 - (1/3)(1/4) = 1/12 \approx 0.0833$. Since this covariance is positive, CRN should lead to a variance reduction; this is consistent with the graph for this example in (*a*), showing that the two curves are monotone in the same direction.

Example (2): $E(X_{1j}) = E(U^2) = 1/3$, and since $1 - U \sim$ U(0, 1) as well, $E(X_{2j}) = E[(1 - U)^3] = E(U^3) = 1/4$. Now $E(X_{1j}X_{2j}) = E[U^2(1-U)^3] = E(-U^5 + 3U^4 - 3U^3 + U^2) = -E(U^5) + 3E(U^4) - 3E(U^3) + E(U^2) = -1/6 + 3(1/5) - 3(1/4) + 1/3 = 1/60$, so $\mathrm{Cov}(X_{1j}, X_{2j}) = 1/60 - (1/3)(1/4) = -1/15 \approx -0.0667$. Since this covariance is negative, CRN will backfire; this might have been anticipated from looking at the graph for this example in (*a*), where the curves are monotone in *opposite* directions.

(*c*) Recall that for any two random variables A and B, $\mathrm{Var}(A - B) = \mathrm{Var}(A) + \mathrm{Var}(B) - 2\mathrm{Cov}(A, B)$, and that $\mathrm{Var}(A) = E(A^2) - [E(A)]^2$.

Example (1): $\mathrm{Var}(X_{1j}) = \mathrm{Var}(U^2) = E(U^4) - [E(U^2)]^2 = 1/5 - (1/3)^2 = 4/45$, and $\mathrm{Var}(X_{2j}) = \mathrm{Var}(U^3) = E(U^6) - [E(U^3)]^2 = 1/7 - (1/4)^2 = 9/112$. Thus, $\mathrm{Var}(X_{1j} - X_{2j})$ would be $4/45 + 9/112 = 853/5040 \approx 0.1692$ if we did independent sampling, and would be $4/45 + 9/112 - 2(1/12) = 13/5040 \approx 0.0026$ using CRN. So we would expect a variance reduction of some 98 percent from CRN; intuitively, this strong variance reduction is due to the fact that the relationship between X_{1j} and X_{2j} is nearly linear, as can be seen from the plot in (*a*).

Example (2): $\mathrm{Var}(X_{1j}) = \mathrm{Var}(U^2) = 4/45$ as before, and since $1 - U \sim$ U(0, 1), $\mathrm{Var}(X_{2j}) = \mathrm{Var}[(1 - U)^3] = \mathrm{Var}(U^3) = 9/112$, also as before. Thus, $\mathrm{Var}(X_{1j} - X_{2j})$ would be $4/45 + 9/112 = 853/5040 \approx 0.1692$ if we did independent sampling, but would be $4/45 + 9/112 - 2(-1/15) = 1525/5040 \approx 0.3026$ using CRN. Thus, CRN *increases* the variance by some 79 percent due to the negative covariance, so does backfire.

(*d*) For each example we sampled $n = 1000$ X_{1j}'s and $n = 1000$ X_{2j}'s using both independent sampling and CRN; we used stream 1 of the random-number generator in App. 7A for the X_{1j}'s and the CRN-based X_{2j}'s, and stream 2 for the independently sampled X_{2j}'s. We estimated $\mathrm{Cov}(X_{1j}, X_{2j})$ by

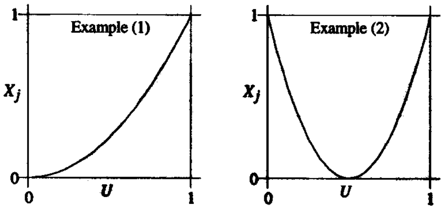
$$\widehat{\mathrm{Cov}}\left(X_{1j}, X_{2j}\right) = \frac{\sum_{j=1}^{n}[X_{1j} - \bar{X}_1(n)][X_{2j} - \bar{X}_2(n)]}{n-1}$$

and used the variance estimator $S^2(n)$ from Eq. (4.4) applied to the values $Z_j = X_{1j} - X_{2j}$. We got

	Example (1)		Example (2)	
	Independent	CRN	Independent	CRN
$\widehat{\mathrm{Cov}}(X_{1j}, X_{2j})$	0.0008	0.0859	0.0022	−0.0667
$S^2(n)$	0.1707	0.0025	0.1640	0.3037

which are all very close to their analytically computed counterparts in (*b*) and (*c*).

11.3.(*a*)



(*b*) First note that for any positive integer k, $E(U^k) = \int_0^1 u^k du = [u^{k+1} / (k+1)]_{u=0}^{u=1} = 1/(k+1)$ since $U \sim$ U(0, 1), and recall that for any two random variables A and B, Cov(A, B) = $E(AB) - E(A)E(B)$.

Example (1): $E(X_j^{(1)}) = E(X_j^{(2)}) = E(U^2) = 1/3$, and under AV $E(X_j^{(1)} X_j^{(2)}) = E[U^2(1-U)^2] = E(U^2 - 2U^3 + U^4) = E(U^2) - 2E(U^3) + E(U^4) = 1/3 - 2/7 + 1/5 = 1/30$. Thus, $\text{Cov}(X_j^{(1)}, X_j^{(2)}) = 1/30 - (1/3)^2 = -7/90 \approx -0.0778$ under AV. This covariance is negative, as suggested in the plot in (*a*) where the curve is monotone, so AV should yield a variance reduction.

Example (2): $E(X_j^{(1)}) = E(X_j^{(2)}) = E[4(U-0.5)^2] = 4E(U^2 - U + 1/4) = 4[E(U^2) - E(U) + 1/4] = 4(1/3 - 1/2 + 1/4) = 1/3$. Under AV, $E(X_j^{(1)} X_j^{(2)}) = E\{[4(U-0.5)^2][4(1-U-0.5)^2]\} = 16E(U^4 - 2U^3 + 3U^2/2 - U/2 + 1/16) = 16[E(U^4) - 2E(U^3) + 3E(U^2)/2 - E(U)/2 + 1/16] = 16(1/5 - 2/4 + 3/6 - 1/4 + 1/16) = 1/5$. Thus, $\text{Cov}(X_j^{(1)}, X_j^{(2)}) = 1/5 - (1/3)^2 = 4/45 \approx 0.0889$ under AV. This covariance is positive, as suggested in the plot in (*a*) where the curve indicates larger X_j values for small *and* large values of U, so AV should backfire and yield a variance increase.

(*c*) In general, $\text{Var}[(X_j^{(1)} + X_j^{(2)})/2] = (1/4)[\text{Var}(X_j^{(1)}) + \text{Var}(X_j^{(2)}) + 2\text{Cov}(X_j^{(1)}, X_j^{(2)})]$. Also, $\text{Var}(A) = E(A^2) - [E(A)]^2$ for any random variable A.

Example (1): $\text{Var}(X_j^{(1)}) = \text{Var}(X_j^{(2)}) = \text{Var}(U^2) = E(U^4) - [E(U^2)]^2 = 1/5 - (1/3)^2 = 4/45$, so $\text{Var}[(X_j^{(1)} + X_j^{(2)})/2]$ would be $(1/4)(4/45 + 4/45) = 2/45 \approx 0.0444$ if we did independent sampling, and would be $(1/4)[4/45 + 4/45 + 2(-7/90)] = 1/180 \approx 0.0056$ under AV. Thus, AV would yield a variance reduction of some 88 percent.

Example (2): $\text{Var}(X_j^{(1)}) = \text{Var}(X_j^{(2)}) = \text{Var}[4(U-0.5)^2] = E[16(U-0.5)^4] - [E(X_j^{(1)})]^2 = 16E(U^4 - 2U^3 + 3U^2/2 - U/2 + 1/16) - (1/3)^2 = 16[E(U^4) - 2E(U^3) + 3E(U^2)/2 - E(U)/2 + 1/16] - 1/9 = 16(1/5 - 2/4 + 1/2 - 1/4 + 1/16) - 1/9 = 4/45$, so $\text{Var}[(X_j^{(1)} + X_j^{(2)})/2]$ would be $(1/4)(4/45 + 4/45) = 2/45 \approx 0.0444$ if we did independent sampling, and would be $(1/4)[4/45 + 4/45 + 2(4/45)] = 4/45 \approx 0.0889$ under AV. Thus, AV would backfire and yield a variance increase of 100 percent.

(*d*) For each example we sampled $n = 1000$ pairs of $(X_j^{(1)}, X_j^{(2)})$'s using both independent sampling and AV; we used stream 1 of the random-number generator in App. 7A for the $X_j^{(1)}$'s and the AV-based $X_j^{(2)}$'s, and stream 2 for the independently sampled $X_j^{(2)}$'s. We estimated $\text{Cov}(X_j^{(1)}, X_j^{(2)})$ by

$$\widehat{\text{Cov}}\left(X_j^{(1)}, X_j^{(2)}\right) = \frac{\sum_{j=1}^{n}[X_j^{(1)} - \bar{X}^{(1)}(n)][X_j^{(2)} - \bar{X}^{(2)}(n)]}{n-1}$$

and used the variance estimator $S^2(n)$ from Eq. (4.4) applied to the values $X_j = (X_j^{(1)} + X_j^{(2)})/2$. We got

	Example (1)		Example (2)	
	Independent	AV	Independent	AV
$\widehat{\text{Cov}}(X_j^{(1)}, X_j^{(2)})$	0.0017	−0.0778	0.0015	0.0908
$S^2(n)$	0.0462	0.0057	0.0462	0.0908

which are all very close to their analytically computed counterparts in (*b*) and (*c*).

11.4. We used simlib from Chap. 2 to code this model, but used a single stream of the linear congruential generator defined by $Z_i = 16{,}807Z_{i-1} \bmod (2^{31} - 1)$ with $Z_0 = 12{,}345{,}678$ for all sources of randomness. Synchronization for CRN and AV was facilitated by generating 4 random numbers immediately upon each customer's initial arrival; these four random numbers were respectively used to generate the next interarrival time, this customer's service time at the first server, his or her service time at the second server (even though it might not be used), and to decide where to send this customer after he or she leaves the first server. The last three random variates were stored as attributes in a record that went through the system from list to list as the customer progressed. Also, it sometimes happened that delays 100 and 101 were actually completed simultaneously; this occurs if a customer leaves the first server and goes to an idle second server, the first person in queue in front of the first server does not need to visit the second server, and there have previously been 99 delays completed. In this case, we took the average of 101 delays.

(*a*) When using CRN, the run of the second model was preceded by resetting the seed of the random-number generator to its value just prior to the first system's run. To protect against overlapping if the second system's run of a pair did not use as many random numbers as did the first system's run, we "wasted" 400 random numbers after each run of the second system. (Each simulation should require about 400 random numbers, so this seemed safe enough to prevent overlapping.) From $n = 10$ runs of each configuration, the estimated variances of the Z_j's were 0.70 for independent sampling and 0.43 for CRN, a reduction of some 39 percent. (This probably understates the actual variance reduction; in a separate run we made $n = 200$ replications of each configuration and obtained an estimated variance reduction of 66 percent. Thus, $n = 10$ appears not to have been a big enough "pilot" study in this case.)

(*b*) We took care of synchronization for AV and overlap protection similarly to (*a*). From $n = 5$ pairs of runs of the model, the estimated variances of the X_j's were 0.99 for independent sampling and 0.78 for AV, a reduction of 21 percent. (Again, in a separate run with $n = 200$ pairs, the reduction in estimated variance was 27 percent. Other sets of $n = 5$ pairs actually led to estimated variance *increases* of several hundred percent when using AV; $n = 5$ is clearly much too small here to judge the efficacy of AV.)

11.5. We consider each of the three types of input random variates separately, and analyze the effect of a "small" random number U; A "large" U should have the opposite effect in each case.

Interdemand times: If U is small, then $-\ln U$ will be large, resulting in a large interdemand time, i.e., fewer demands. This in turn should lead to less frequent ordering, lowering the number of times the setup cost for ordering is incurred, which should lower the ordering cost. Fewer demands should also lead to a generally higher inventory, which would raise holding costs but lower backlog costs.

Demand sizes: Here, a small U implies that the quantity demanded will be less, i.e., the total amount demanded is less. This should lead to less frequent ordering (lowering the ordering cost), and generally higher inventory levels (raising the holding cost but lowering the backlog cost).

Delivery lags: If U is small, the delivery lag will be short. This should lead once again to higher inventory levels, and thus higher holding costs but lower backlog costs. Since the next ordering decision is never made until an order arrives, and since all random variables in the simulation are independent, there should be no effect on the ordering cost.

11.6. We modified the simlib code of Sec. 2.6 to generate a customer's service requirement upon arrival and store it as a second attribute of the record that goes through the queue(s) with this customer. We used stream 1 for interarrival times and stream 2 for service times for synchronization. (See Prob. 11.16 for another way to implement CRN for this model.) We made 50 replications of each of the three variants (5, 6, and 7 tellers), and to measure the success of CRN we also made a separate set of 50 replications of each variant using independent sampling throughout. For $i = 5$, 6, and 7, let μ_i be the expected average delay in queue for the model with i servers, and let X_{ij} be the observed average delay in queue on the jth replication of the model with i servers; for $i = 6$ and 7 only, let $Z_{ij} = X_{ij} - X_{5j}$. Using the sample variance $S^2(n)$ from Eq. (4.4) applied to the Z_{ij}'s and the paired-t approach (see Sec. 10.3.1) to form 95 percent confidence intervals for $\mu_6 - \mu_5$ and $\mu_7 - \mu_5$, we got

	Independent		CRN	
Parameter	$\mu_6 - \mu_5$	$\mu_7 - \mu_5$	$\mu_6 - \mu_5$	$\mu_7 - \mu_5$
$S^2(n)$	35.08	38.51	26.96	34.81
95% confidence interval	-5.82 ± 1.68	-6.70 ± 1.76	-5.65 ± 1.48	-6.65 ± 1.68

indicating that the variance reduction is fairly weak here. The problem is not that CRN is failing to induce the desired positive correlation; in fact, we used the sample correlation formula given in Example 11.2 to estimate that the correlation under CRN between X_{6j} and X_{5j} is 0.89, and that between X_{7j} and X_{5j} is 0.85. Recall that $\mathrm{Var}(Z_{6j}) = \mathrm{Var}(X_{5j}) + \mathrm{Var}(X_{6j}) - 2\,\mathrm{Cov}(X_{5j}, X_{6j})$; from our data we estimated $\mathrm{Var}(X_{5j})$ to be 37.79, $\mathrm{Var}(X_{6j})$ to be 1.22, and $\mathrm{Cov}(X_{5j}, X_{6j})$ to be 6.02 and putting these estimates in place of their corresponding parameters we get $37.79 + 1.22 - 2(6.02) = 26.97$, matching up with our direct estimate 26.96 of $\mathrm{Var}(Z_{6j})$. [This works out similarly for $\mathrm{Var}(Z_{7j})$.] What is hindering CRN here is just that $\mathrm{Var}(X_{5j})$ is evidently very large, so that the information we get from this model is quite noisy and limits how successful CRN can be. However, the cost of implementing CRN for this model is almost nothing, so we would probably use it anyway.

11.7. (*a*) Since large interarrival times should tend to decrease congestion levels (and thus decrease harbor times of ships and all utilizations), they should be generated antithetically. Also, large unload times should increase congestion, so they should also be generated antithetically. Synchronization could be maintained by dedicating separate streams to each source of randomness or by generating a single-crane unload time for a ship immediately upon its arrival and storing it as an additional attribute of the ship.

(*b*) For the same reasons given in (*a*), both interarrival and unload times should be generated using CRN across the two systems. Synchronization could be maintained by either of the methods mentioned in (*a*).

(*c*) We made n = 100 pairs of replications of both configurations using independent sampling, and then 100 pairs with CRN. Stream dedication (1 for interarrival times and 2 for unloading times) was used to synchronize. Let X_{1j} and X_{2j} be a given performance measure (there are 6 performance measures, as described in Prob. 2.19) from configurations 1 and 2, respectively, on the jth replication, and $Z_j = X_{2j} - X_{1j}$; we estimated Var(Z_j) using Eq. (4.4) applied to the Z_j's and used the paired-*t* approach (see Sec. 10.2.1) to form 90 percent confidence intervals for $E(Z_j)$. We also computed the sample correlation given in Example 11.2:

		Performance measure					
		Total time in harbor			Utilizations		
		Minimum	Maximum	Average	Berth 1	Berth 2	Cranes
Indep.	$S^2(n)$	0.0003	0.2530	0.0093	0.0031	0.0024	0.0036
	90% c.i.	–0.16±0.003	–1.01±0.084	–0.37±0.016	–0.13±0.009	–0.10±0.008	–0.15±0.010
	Correl.	–0.06	0.19	0.03	–0.12	–0.03	–0.10
CRN	$S^2(n)$	0.0001	0.1203	0.0032	0.0003	0.0007	0.0003
	90% c.i.	–0.16±0.002	–1.06±0.058	–0.38±0.009	–0.14±0.003	–0.11±0.005	–0.16±0.003
	Correl.	0.58	0.81	0.86	0.96	0.88	0.99

Thus, CRN seems to be working well for all performance measures; for instance, it reduces the variance of the average time in the harbor to about a third of what it was under independent sampling. It was virtually no work to use CRN instead of independent sampling, so it clearly pays off here.

11.8. Small interarrival times will tend to increase congestion and so increase the average delay in queue of a job, the proportion of jobs delayed more than 5 minutes, and the maximum number of jobs in queue. The same is true of large maximum processing times and large actual processing times. Thus, it makes sense to use CRN for all types of input random variables. To implement CRN, both systems should see jobs arriving at the same times with the same maximum and actual processing times. Synchronization could be maintained by generating the maximum and actual processing times of a job immediately upon its arrival, and storing them as attributes of this job to be carried along with it wherever it might be placed in the queue. Also, we could dedicate random-number streams to the three individual sources of randomness.

11.9. We made $n = 100$ pairs of replications of both configurations using independent sampling, and then 100 pairs with CRN. Stream dedication (as specified in Prob. 2.22) was used to synchronize. We took advantage of knowing the probabilities that a job is of a given class and used them to weight the average delays in queue from each individual class, as for the job-shop model in Sec. 2.7. Let X_{1j} and X_{2j} be this weighted average delay in queue from the no-preemption case and from the preemption case, respectively, on the jth replication, and then let $Z_j = X_{2j} - X_{1j}$; we estimated $\text{Var}(Z_j)$ using Eq. (4.4) applied to the Z_j's and used the paired-t approach (see Sec. 10.2.1) to form 90 percent confidence intervals for $E(Z_j)$. We also computed the sample correlation given in Example 11.2:

	Independent	CRN
$S^2(n)$	0.62	0.52
90% confidence interval	-0.34 ± 0.13	-0.26 ± 0.12
Correlation estimate	0.11	0.41

While CRN does induce the desired positive correlation, the variance reduction is not strong (about 16 percent). But it was virtually no work to use CRN instead of independent sampling, so we would probably do it anyway.

From the confidence intervals above (which lie entirely below zero), it appears that the preemption rule would reduce the weighted average delay in queue by perhaps 10 to 25 seconds. As noted in our solution to Prob. 2.22, though, service to low-priority jobs would probably suffer (although we did not look at this criterion in the present analysis).

11.10. The principal problem here is generating the gamma service times in the second system. For them to be as positively correlated as possible with the exponential service times of the first queue, the inverse-transform method should be used for both service-time generations. For the exponential service times, we should thus use the formula $X = -\beta \ln(1 - U)$ (rather than $X = -\beta \ln U$), since this is the literal inverse-transform formula (see Example 8.1). For the gamma service times, the inverse-transform method can be implemented with a numerical algorithm to invert the gamma distribution function, as discussed at the end of Sec. 8.3.4. Finally, the exponential interarrival times can easily be made to be positively correlated by using the same random numbers to generate corresponding interarrival times in both systems.

11.11. (*a*) Let $\mu_i = E(X_i^{(l)})$ and $W_i^{(l)} = X_i^{(l)} - \mu_i$, so $E(W_i^{(l)}) = 0$. Then since ζ $(= \mu_1 - \mu_2)$ is a constant (i.e., nonrandom),

$$
\begin{aligned}
\operatorname{Var}(Z) &= \operatorname{Var}(Z-\zeta)\\
&= \operatorname{Var}\left\{\left[\left(X_1^{(1)} + X_1^{(2)}\right)/2 - \mu_1\right] - \left[\left(X_2^{(1)} + X_2^{(2)}\right)/2 - \mu_2\right]\right\}\\
&= \operatorname{Var}\left\{\left[\left(X_1^{(1)} - \mu_1 + X_1^{(2)} - \mu_1\right)/2 - \mu_1\right] - \left[\left(X_2^{(1)} - \mu_1 + X_2^{(2)} - \mu_1\right)/2 - \mu_2\right]\right\}\\
&= \operatorname{Var}\left[\left(W_1^{(1)} + W_1^{(2)}\right)/2 - \left(W_2^{(1)} + W_2^{(2)}\right)/2\right]
\end{aligned}
$$

so that the $X_i^{(l)}$'s in the definition of Var(Z) can be replaced by the corresponding $W_i^{(l)}$'s. Furthermore, since $E(W_i^{(l)}) = 0$,

$$
\operatorname{Cov}\left(W_{i_1}^{(l_1)}, W_{i_2}^{(l_2)}\right) = E\left(W_{i_1}^{(l_1)} W_{i_2}^{(l_2)}\right) = E\left[\left(X_{i_1}^{(l_1)} - \mu_{i_1}\right)\left(X_{i_2}^{(l_2)} - \mu_{i_2}\right)\right] = \operatorname{Cov}\left(X_{i_1}^{(l_1)}, X_{i_2}^{(l_2)}\right)
$$

so the (centered) $W_i^{(l)}$'s have the same variance/covariance structure as do the original $X_i^{(l)}$'s. So if we let

$$
C_{i_1 i_2}^{(l_1 l_2)} = \operatorname{Cov}\left(X_{i_1}^{(l_1)}, X_{i_2}^{(l_2)}\right) = \operatorname{Cov}\left(W_{i_1}^{(l_1)}, W_{i_2}^{(l_2)}\right) = E\left(W_{i_1}^{(l_1)} W_{i_2}^{(l_2)}\right)
$$

then

$$
\begin{aligned}
\operatorname{Var}(Z) &= \operatorname{Var}\left[\left(W_1^{(1)} + W_1^{(2)}\right)/2 - \left(W_2^{(1)} + W_2^{(2)}\right)/2\right]\\
&= \tfrac{1}{4}\operatorname{Var}\left(W_1^{(1)} + W_1^{(2)} - W_2^{(1)} - W_2^{(2)}\right)\\
&= \tfrac{1}{4}E\left[\left(W_1^{(1)} + W_1^{(2)} - W_2^{(1)} - W_2^{(2)}\right)^2\right] \qquad \left(\text{since } E\left(W_i^{(l)}\right) = 0\right)\\
&= \tfrac{1}{4}E\Big[\left(W_1^{(1)}\right)^2 + \left(W_1^{(2)}\right)^2 + \left(W_2^{(1)}\right)^2 + \left(W_2^{(2)}\right)^2\\
&\qquad +2\left(W_1^{(1)}W_1^{(2)} - W_1^{(1)}W_2^{(1)} - W_1^{(2)}W_2^{(2)} + W_2^{(1)}W_2^{(2)}\right)\\
&\qquad +2\left(-W_1^{(2)}W_2^{(1)} - W_1^{(1)}W_2^{(2)}\right)\Big]\\
&= \tfrac{1}{4}\Big\{E\left[\left(W_1^{(1)}\right)^2\right] + E\left[\left(W_1^{(2)}\right)^2\right] + E\left[\left(W_2^{(1)}\right)^2\right] + E\left[\left(W_2^{(2)}\right)^2\right] +\\
&\qquad +2\left[E\left(W_1^{(1)}W_1^{(2)}\right) - E\left(W_1^{(1)}W_2^{(1)}\right) - E\left(W_1^{(2)}W_2^{(2)}\right) + E\left(W_2^{(1)}W_2^{(2)}\right)\right]\\
&\qquad +2\left[-E\left(W_1^{(2)}W_2^{(1)}\right) - E\left(W_1^{(1)}W_2^{(2)}\right)\right]\Big\}\\
&= \tfrac{1}{4}\Big\{\operatorname{Var}\left(W_1^{(1)}\right) + \operatorname{Var}\left(W_1^{(2)}\right) + \operatorname{Var}\left(W_2^{(1)}\right) + \operatorname{Var}\left(W_2^{(2)}\right) +\\
&\qquad +2\left[\operatorname{Cov}\left(W_1^{(1)}, W_1^{(2)}\right) - \operatorname{Cov}\left(W_1^{(1)}, W_2^{(1)}\right) - \operatorname{Cov}\left(W_1^{(2)}, W_2^{(2)}\right) + \operatorname{Cov}\left(W_2^{(1)}, W_2^{(2)}\right)\right]\\
&\qquad +2\left[-\operatorname{Cov}\left(W_1^{(2)}, W_2^{(1)}\right) - \operatorname{Cov}\left(W_1^{(1)}, W_2^{(2)}\right)\right]\Big\}\\
&= \tfrac{1}{4}\left[\operatorname{Var}\left(X_1^{(1)}\right) + \operatorname{Var}\left(X_1^{(2)}\right) + \operatorname{Var}\left(X_2^{(1)}\right) + \operatorname{Var}\left(X_2^{(2)}\right)\right] +\\
&\qquad +\tfrac{1}{2}\left[C_{11}^{(12)} - C_{12}^{(11)} - C_{12}^{(22)} + C_{22}^{(12)}\right]\\
&\qquad +\tfrac{1}{2}\left[-C_{12}^{(21)} - C_{12}^{(12)}\right]
\end{aligned}
$$

as required.

(Problem 11.11, continued from previous page)

(*b*) Note that from (*a*), under independent sampling throughout we would have

$$\mathrm{Var}(Z) = \frac{1}{4}\left[\mathrm{Var}\left(X_1^{(1)}\right) + \mathrm{Var}\left(X_1^{(2)}\right) + \mathrm{Var}\left(X_2^{(1)}\right) + \mathrm{Var}\left(X_2^{(2)}\right)\right]$$

and that this term is not affected by whether or not we use either CRN or AV. Now CRN's "working" on its own implies that both $C_{12}^{(11)}$ and $C_{12}^{(22)}$ are positive. Likewise, AV's working on its own means that $C_{11}^{(12)}$ and $C_{22}^{(12)}$ are both negative. Thus, the contribution of the term

$$\frac{1}{2}\left[C_{11}^{(12)} - C_{12}^{(11)} - C_{12}^{(22)} + C_{22}^{(12)}\right]$$

in the final expression for Var(*Z*) in (*a*) is negative, reducing Var(*Z*) from what it would be under completely independent sampling. But if $X_1^{(1)}$ is positively correlated with $X_2^{(1)}$, which in turn is negatively correlated with $X_2^{(2)}$, it seems likely that $X_1^{(1)}$ could be negatively correlated with $X_2^{(2)}$ as well, i.e., $C_{12}^{(12)} < 0$. Symmetrically, we would suspect that $C_{12}^{(21)} < 0$. In this case, these negative cross covariances would bring a *positive* contribution to Var(*Z*) through the positive term

$$\frac{1}{2}\left[-C_{12}^{(21)} - C_{12}^{(12)}\right]$$

in the final expression for Var(*Z*) in (*a*). Thus it is not clear whether the immediate (and beneficial) effect of CRN and AV on their own will more than offset their potentially harmful interaction through the effects of these cross covariances.

Can the numbers work out so that this backfiring actually occurs? That is, is it possible to have a legitimate (see below) set of covariances that cause the CRN-with-AV scheme to increase Var(*Z*) above what it would be with completely independent sampling throughout? Yes. Define the 4-dimensional random vector

$$\mathbf{X} = \left(X_1^{(1)}, X_1^{(2)}, X_2^{(1)}, X_2^{(2)}\right)^T$$

(The superscript *T* denotes transposition) and suppose it has covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} \mathrm{Var}\left(X_1^{(1)}\right) & C_{11}^{(12)} & C_{12}^{(11)} & C_{12}^{(12)} \\ C_{11}^{(12)} & \mathrm{Var}\left(X_1^{(2)}\right) & C_{12}^{(21)} & C_{12}^{(22)} \\ C_{12}^{(11)} & C_{12}^{(21)} & \mathrm{Var}\left(X_2^{(1)}\right) & C_{22}^{(12)} \\ C_{12}^{(12)} & C_{12}^{(22)} & C_{22}^{(12)} & \mathrm{Var}\left(X_2^{(2)}\right) \end{bmatrix} = \begin{bmatrix} 5 & -1 & 1 & -4 \\ -1 & 5 & -4 & 1 \\ 1 & -4 & 5 & -1 \\ -4 & 1 & -1 & 5 \end{bmatrix}$$

(By "legitimate" above we mean that the covariance matrix Σ is positive definite, as all covariance matrices must be. One way to verify that this particular Σ is positive definite is to note that its eigenvalues, being 11, 7, 1, and 1, are all positive.) Reading the appropriate elements from Σ, we would get Var(*Z*) = 5 under completely independent sampling throughout, but would get Var(*Z*) = 5 + (–2) + 4 = 7 under the combined CRN-with-AV scheme, so that it does backfire. Now this example, which demonstrates that it is possible for backfiring to occur, is admittedly contrived and did not arise from any simulation context. We do not know how likely backfiring is to occur in an actual simulation project, or by how much Var(*Z*) would typically be increased if it did occur, or decreased if it did not occur.

11.12. Suppose we allow for a random number N of Y_i's to be averaged to obtain a CV Y, and that there is indeed bias in Y with respect to $E(Y_i)$, i.e., $E(Y) \neq E(Y_i)$. Then the controlled estimator X_C defined near the beginning of Sec. 11.4 might be biased for $E(X)$, even with a nonrandom value of the weight a, since the expectation of the second term in the definition of X_C is not 0. Now in practice, since we have to estimate the optimal weight(s), there may be bias introduced anyway in the CV estimator that we actually use, and the additional bias owing to the fact that $E(Y) \neq E(Y_i)$ may or may not be important. In general, though, it is probably good advice to avoid bias wherever possible, and it *is* possible in this case to avoid it just by being careful about the definition of the control variate Y.

11.13. Let $v_l = \text{Var}(Y_l)$, $C_l = \text{Cov}(X, Y_l)$, $C_{l_1 l_2} = \text{Cov}(Y_{l_1}, Y_{l_2})$, and $g_m(a_1, a_2, \ldots, a_m) = \text{Var}(X_C)$ as given in Eq. (11.3), viewed as a function of the a_l's. Also, let H_m be the Hessian matrix of g_m, i.e., H_m is a symmetric $m \times m$ matrix whose (l_1, l_2)th entry is the second partial derivative $\partial^2 g_m / \partial a_{l_1} \partial a_{l_2}$.

(*a*) For $m = 2$, $g_2(a_1, a_2) = \text{Var}(X) + a_1^2 v_1 + a_2^2 v_2 - 2(a_1 C_1 + a_2 C_2) + 2a_1 a_2 C_{12}$; setting

$$\frac{\partial g_2}{\partial a_1} = 2a_1 v_1 - 2C_1 + 2a_2 C_{12} = 0$$

$$\frac{\partial g_2}{\partial a_2} = 2a_2 v_2 - 2C_2 + 2a_1 C_{12} = 0$$

and solving simultaneously for a_1 and a_2 yields

$$a_1^* = \frac{C_1 v_2 - C_2 C_{12}}{v_1 v_2 - C_{12}^2}, \qquad a_2^* = \frac{C_2 v_1 - C_1 C_{12}}{v_1 v_2 - C_{12}^2}$$

To check the second-order sufficient conditions for this solution to maximize g_2 (as opposed to minimizing it or being a saddle point), note that

$$H_2 = 2\begin{bmatrix} v_1 & C_{12} \\ C_{12} & v_2 \end{bmatrix}$$

which is twice the covariance matrix of the random vector $(Y_1, Y_2)^T$ (the superscript T denotes transposition), so is automatically positive definite; this is sufficient for maximization.

For $m = 3$,

$$g_3(a_1, a_2, a_3) = \text{Var}(X) + a_1^2 v_1 + a_2^2 v_2 + a_3^2 v_3 - 2(a_1 C_1 + a_2 C_2 + a_3 C_3)$$
$$+2(a_1 a_2 C_{12} + a_1 a_3 C_{13} + a_2 a_3 C_{23})$$

Setting

$$\frac{\partial g_3}{\partial a_1} = 2a_1 v_1 - 2C_1 + 2(a_2 C_{12} + a_3 C_{13}) = 0$$

$$\frac{\partial g_3}{\partial a_2} = 2a_2 v_2 - 2C_2 + 2(a_1 C_{12} + a_3 C_{23}) = 0$$

$$\frac{\partial g_3}{\partial a_3} = 2a_3 v_3 - 2C_3 + 2(a_1 C_{13} + a_2 C_{23}) = 0$$

and solving for a_1, a_2, and a_3 yields

$$a_1^* = \left(C_2 C_{13} C_{23} + C_3 C_{12} C_{23} - C_1 C_{23}^2 - C_3 C_{13} v_2 - C_2 C_{12} v_3 + C_1 v_2 v_3\right)/d$$

$$a_2^* = \left(C_1 C_{13} C_{23} + C_3 C_{12} C_{13} - C_2 C_{13}^2 - C_3 C_{23} v_1 - C_1 C_{12} v_3 + C_2 v_1 v_3\right)/d$$

$$a_3^* = \left(C_1 C_{12} C_{23} + C_2 C_{12} C_{13} - C_3 C_{12}^2 - C_2 C_{23} v_1 - C_1 C_{13} v_2 + C_3 v_1 v_2\right)/d$$

where $d = v_1 v_2 v_3 - v_1 C_{23}^2 - v_2 C_{13}^2 - v_3 C_{12}^2 + 2C_{12} C_{13} C_{23}$. The Hessian matrix is

$$H_3 = 2\begin{bmatrix} v_1 & C_{12} & C_{13} \\ C_{12} & v_2 & C_{23} \\ C_{13} & C_{23} & v_3 \end{bmatrix}$$

which is twice the covariance matrix of the random vector $(Y_1, Y_2, Y_3)^T$ so is automatically positive definite, satisfying the second-order sufficient condition for a maximum.

(*b*) This assumption means that the $C_{l_1 l_2}$'s are all 0, in which case

$$g_m(a_1, a_2, \ldots, a_m) = \text{Var}(X) + \sum_{l=1}^{m} a_l^2 v_l - 2\sum_{l=1}^{m} a_l C_l$$

so for all l

$$\frac{\partial g_m}{\partial a_l} = 2a_l v_l - 2C_l$$

Setting all m of these first partials to 0 yields $a_l^* = C_l / v_l$. The Hessian matrix here is diagonal with the v_l's on the diagonal, so is certainly positive definite.

(*c*) From n replications of the simulation, let Y_{lj} and X_j be the observations on Y_l and X, respectively, from the jth replication. Compute the estimators

$$\hat{v}_l = \frac{1}{n-1}\sum_{j=1}^{n}\left[Y_{lj} - \overline{Y}_l(n)\right]^2$$

$$\hat{C}_l = \frac{1}{n-1}\sum_{j=1}^{n}\left[X_j - \overline{X}(n)\right]\left[Y_{lj} - \overline{Y}_l(n)\right]$$

$$\hat{C}_{l_1 l_2} = \frac{1}{n-1}\sum_{j=1}^{n}\left[Y_{l_1 j} - \overline{Y}_{l_1}(n)\right]\left[Y_{l_2 j} - \overline{Y}_{l_2}(n)\right]$$

all of which are strongly consistent for the corresponding parameters (without the hats). Then just replace the v_l's, C_l's, and $C_{l_1 l_2}$'s in the formulas for the optimal a_l's in (*a*) and (*b*) with their behatted counterparts. If we know the v_l's, then they should be used instead of estimates of them, and similarly for the $C_{l_1 l_2}$'s. For more on the effect of knowing variances and covariances, see the papers referenced in Sec. 11.4.

11.14. We made 100 sets of $n = 10$ replications each (using the external CV as described), and estimated $\mathrm{Var}[\overline{X}_C^*(10)]$ to be 0.095; from another 100 sets of $n = 10$ replications each (without using CV, and separate from the first 100 sets for purposes of getting timing information, as discussed below) our estimate of $\mathrm{Var}[\overline{X}(10)]$ was 0.422. Thus, CV reduced the estimated variance by some 78 percent here.

To recommend whether this external CV approach is advisable overall, though, we must consider the extra work required to do it; we will use the framework and notation of Prob. 11.1. The first 100 sets of runs (using CV) required 1679 seconds using THINK C 4.0 on an Apple Macintosh SE/30, and the second 100 sets (without using CV) took 1091 seconds, so there is certainly noticeable extra computation to use the VRT, involving the separate simulations for the *M/M/*1 queue to get the external control variate. (We wrote the code as efficiently as we could, and used the recursion from Example 4.19 to generate the delays; overhead, such as input/output and computing the estimate of $\mathrm{Var}[\overline{X}_C^*(10)]$ across the 100 sets of n replications, was excluded from the timings.) Since the CV technique requires n (= 10) replications to compute the estimate $\hat{a}^*(10)$, we view $n = 10$ replications as the smallest simulation "unit." Using run time as the cost measure, the unit cost of using this CV VRT is thus approximately $C_1 = 16.79$, and the unit cost if we don't use CV is about $C_0 = 10.91$, since we generated 100 simulation units in each case. We estimated above that $V_1 = 0.095$ and $V_0 = 0.422$, so the efficiency ratio (see our solution to Prob. 11.1) of this external CV is $(C_0V_0)/(C_1V_1) = 2.89$, meaning that the CV technique *would* be advisable (since the efficiency ratio exceeds 1) and in fact that our overall efficiency would be nearly tripled by using CV in this case. This appealing result is in part due to the fact that generating the external control variate in this case is quite simple, and does not increase run time very much.

Finally, we consider the potential bias in $\overline{X}_C^*(10)$ (as an estimator of μ = the expected average delay in the *M/G/*1 queue with Weibull demands, which is unknown in the present terminating-simulation model) introduced by the need to estimate the optimal weight a^*. From the experiments above, an unbiased estimate of μ (obtained from the runs without CV) was 3.032, and our estimate of $E[\overline{X}_C^*(10)]$ (obtained from the runs using CV) was 3.076; the question is whether these two results differ significantly from each other. Since we certainly cannot count on equality of variances of these two estimators (we're consciously trying to make one of them smaller than the other, and seem to have succeeded), we can use the Welch approach to form a confidence interval for $\mu - E[\overline{X}_C^*(10)]$ (see Sec. 10.2.2). The estimated df work out to be $\hat{f} = 12.9$ (in the notation of Sec. 10.2.2, $n_1 = n_2 = 100$), so for a 90 percent interval we interpolate in Table T.1 to get $t_{\hat{f},0.95} \approx 1.772$; the interval is thus $(3.032 - 3.076) \pm 1.772\sqrt{0.422/100 + 0.095/100}$, or -0.044 ± 0.127. Since this interval contains 0, we have no reason to think that the controlled estimator is biased for μ. (We could have used the paired-t confidence-interval approach here as well.)

11.15. This problem is different from standard queueing models, since jobs need not return to their terminals in the order in which they were sent to the CPU.

For use of AV with a fixed number of terminals, we could dedicate two random-number streams to each terminal (if there are enough streams available), one for generating the think times at the terminal and the other for generating the service requirements of jobs sent to the computer from the terminal; this does not necessarily induce any kind of antithetic behavior on the times of arrival of successive jobs to the computer.

To use CRN to compare, for example, configurations with 30 and 35 terminals (and that are the same otherwise), the think times and service requirements of the *first* 30 terminals in the second system could be matched up with those from the 30 terminals in the first system via random-number stream dedication.

Neither of the above ideas guarantees a variance reduction, and care should be taken in complex models to make sure that AV or CRN make sense, and that proper synchronization is maintained. Making pilot experiments to investigate empirically whether a VRT is working with complicated models is also advisable.

11.16. We carried out three sets of 50 replications each. In the first set, configurations 1 and 2 were simulated independently (to provide a benchmark variance estimate against which the variances from the two different versions of CRN can be measured). In the second set we simulated the two configurations using CRN as described in (*a*), and in the third set we used CRN as in (*b*). For $i = 1$ and 2, let X_{ij} be the average delay in queue(s) of customers from the jth replication of configuration i, and let $Z_j = X_{2j} - X_{1j}$. Using the sample variance $S^2(n)$ from Eq. (4.4) applied to the Z_j's and the paired-t approach (see Sec. 10.2.1) to form 90 percent confidence intervals for $E(Z_j)$, we got

	Independent	CRN (*a*)	CRN (*b*)
$S^2(n)$	57.07	0.21	0.12
90% confidence interval	0.93 ± 1.80	1.15 ± 0.11	1.07 ± 0.08

First, note the dramatic variance reductions achieved by either version of CRN; the confidence intervals resulting from using either CRN approach clearly miss 0, but the interval from independent sampling indicates that we would not have been able to conclude that there is any difference between the configurations since it covers 0. In comparing (*a*) against (*b*), it appears that the (*b*) may be a little better, i.e., run the simulations so that the ordered sequence of service times begun for both configurations is identical, rather than forcing the "same" customers to arrive to both systems. (Since we used streams 1 and 2 from their beginnings for all three sets of simulations, we cannot do the F test for equality of variances using the statistic 0.21/0.12 since the two variance estimates were not obtained independently.)

Incidentally, we note that since there appears to be good evidence that $E(Z_j)$ is about 1 minute (or a little more), prohibiting jockeying would increase average delays in queue by this amount.

11.17. Only seven streams are required if we use stream 1 for interdemand times and stream 2 for demand sizes for all replications, and we use streams 3, 4, …, 7 for delivery lags on replications 1, 2, …, 5, respectively. Streams 1 and 2 are NOT reset after each replication. On a particular replication, both inventory policies will generate the same number of interdemand times and the same number of demand sizes.

We can use streams 1 and 2 as above. However, we can use stream 3 for delivery lags on all replications. We generate a delivery lag at the beginning of each month whether it is needed or not. Thus, we will need 120 random numbers from stream 3 for each of the two policies on a particular replication. If both policies order in a particular month, then the same delivery lag will be used.

11.18. In the independent case, the correlation estimator is a random variable with mean 0, and -0.17 is just one observation of this random variable.

11.19.

Synchronize	Sample variance	Variance reduction (percent)
None	0.375	-
Interarrival times of new parts	0.148	60.53
Times to failure and repair times for the machine	0.313	16.53
Processing times and inspection times of a part on the first pass	0.372	0.83

Clearly, synchronizing the external interarrival times gives the largest benefit.

11.20. Use one-random number stream for both interdemand times and demand sizes. When an order arrives, generate the demand size at that time.