Three pages of hand-written notes, front and back. Closed book. 120 minutes.

For all true/false questions, three points if correct, two points if blank;
if you wish, write an explanation of your thinking.

Score ____________________________
1. The variance-reduction techniques of common random numbers, antithetic variates, and (external) control variates require "correlation induction".

(a) T F To support correlation induction is why SimLib has multiple random-number streams.

(b) T F The correlation being induced is (i) a property of the simulation experiment and (ii) not a property of the model being analyzed.

(c) T F The induced correlation should always be positive.

(d) T F The use of NORTA (normal to anything) is an example of such correlation induction.

2. Consider also output data $Y_1, Y_2, \ldots, Y_n$ from one long simulation run. If the data are steady state, then

$$V(\bar{Y}) = \frac{V(Y)}{n} \left[ 1 + 2 \sum_{h=1}^{n} \left[ 1 - \frac{n}{h} \rho_h \right] \right],$$

where $\rho_h$ is the lag-$h$ autocorrelation. To estimate the standard error, we discussed the formula

$$\hat{V}(\bar{Y}) = \frac{\hat{V}(Y)}{n} \left[ 1 + 2 \sum_{h=1}^{m} \left[ 1 - \frac{m}{h} \hat{\rho}_h \right] \right].$$

(a) T F Despite the output data being dependent, a reasonable interpretation of $\hat{V}(Y)$ is that it is the sample variance

$$S_Y^2 = \frac{\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2}{n-1}.$$

(b) T F If output data are independent, then $\rho_0 = \rho_{-1} = 0$.

(c) T F If the output data are the times in the system for the clean-and-patch model, then $\rho_h$ is negative for every positive $h$.

(d) T F If $m = 1$ when the autocorrelations are positive, $E[\hat{V}(\bar{Y})] < V(\bar{Y})$; that is, standard-error estimates tend to be too small.
3. Input modeling. Recall: The $k$th standardized moment is $\alpha_k = \mathbb{E} \left[ \left( \frac{X - \mu}{\sigma} \right)^k \right]$ for positive integers $k$.

(a) T  F The fourth standardized moment, $\alpha_4$, is a measure of distribution skewness.

(b) T  F The first and second standardized moments, $\alpha_1$ and $\alpha_2$, can be used to hypothesize a family of distributions.

(c) T  F A goodness-of-fit test is sometimes used to decide whether a particular input model is adequate. Your instructor argued against such tests for simulation experiments because they consider only statistical significance, ignoring practical significance.

(d) T  F The software package PRIME (PRobability Input Modeling Environment) models univariate distributions by approximating the cdf by a polynomial.

4. Importance Sampling. For input model $f$ and logic model $g$, we can choose a function $h(x)$ to rewrite the integral

$$
\mathbb{E}(g(X)) = \int g(x) f(x) \, dx = \int \left[ g(x) f(x) / h(x) \right] h(x) \, dx
$$

to provide an alternate simulation experiment that is (it is hoped) either easier to implement or statistically more efficient. (If the random variables $X$ are discrete, replace the integral with a sum.)

(a) T  F The alternate experiment might be easier to implement because generating random variates from $h$ might be easier than generating from $f$.

(b) T  F For statistical efficiency, ideally $h(x) = g(x) f(x)$ for every value of $x$, since then $\text{ste} \left[ \mathbb{E}(g(X)) \right] = 0$.

5. Law and Kelton’s SimLib.

(a) T  F The time of the next event and the type of the next event are stored as "attributes" in a "list".

(b) T  F The average size, over time, of every list is collected automatically.

(c) T  F The average size of a list is an observational statistic, maintained through the routine "sampst".

(d) T  F The variance of the size of a list is a time-based statistic.
6. Consider the nonhomogeneous Poisson process (nhpp) with rate function $\lambda(\tau)$ for $0 \leq \tau < \infty$. The cdf for a nhpp is

$$F_{T_i | T_{i-1}}(t) = 1 - \exp(-\left(\Lambda(t) - \Lambda(t_{i-1})\right))$$

where $\Lambda(t) = \int_0^t \lambda(\tau) d\tau$.

Suppose that $\lambda(\tau) = 2$ for $\tau < 5$ and $\lambda(\tau) = 4$ otherwise.

(a) (Five points) Consider using thinning to generate the times from the rate function $\lambda$. What is the best constant rate function, $\lambda^*$, to use for the thinning. State criteria that underlie your answer.

(b) (Five points) Consider using the inverse transformation to generate times from the rate function $\lambda$. Suppose that the current time is $t_{i-1} = 2.5$ and that $w = 0.5$ is a random variate from the exponential distribution with a mean of one. What is the next generated time?

(c) T F With thinning, but not the inverse transformation, it is possible that the next generated time occurs before the current time.
7. Recall: Stochastic simulation experiments, as studied in this class, are used to analyze a given probability model. In particular, a performance measure $\theta$ is estimated by generating random variates $X$ from a known input model, transforming them to output data $Y$ using a known logic model, and from $Y$ computing a point estimator $\hat{\theta}$. The standard error of the point estimator depends upon how long the model is simulated.

Consider the output report, from the clean-and-patch model, on Page 6. The standard errors are estimated using $k = 20$ independent replications of $m = 1,000$ minutes each. No initial data were deleted.

(a) T  F The "macro average" for "arrival overflows" is the performance measure $\theta = 0.0267$.

(b) T  F The standard error "0.0018" associated with "arrival overflows" is computed using $\left[ \rho \left( 1 - \rho \right) / k \right]^{1/2}$, where $\rho = 0.0267$.

(c) T  F Consider "macro stdev" for "time at the patching operation". All digits in "3.7716" are meaningful, since the standard error "0.3505" is given to four digits after the decimal.

(d) T  F If the experiment were changed only by setting the number of replications to $k = 80$, then the "macro stdev" values would (approximately) decrease by a factor of two.

(e) T  F If cleaning times were autocorrelated with lag-1 autocorrelation $\rho$, the value of $\rho$ would be listed as an experiment parameter, because we would want to experiment with the value of $\rho$ to determine its effect on the performance measures.

(g) T  F If the experiment is changed to run $k = 10$ replications of $m = 2,000$ minutes each, so that the total time simulated remains at 20,000 minutes, the standard errors would all increase by a factor of about $\sqrt{2}$.

(h) T  F If the experiment is changed to run $k = 10$ replications of $m = 2,000$ minutes each, so that the total time simulated remains at 20,000 minutes, the quality of the standard-error estimators would be harmed.
8. Let $N_j(t)$ denote the number of jobs at the patcher at time $t$ during replication $j$. Consider the last line of the clean-and-patch output report on Page 6.

(a) (Five points) Write the formula to compute the sample average "macro average" (4.0119).

(b) (Five points) Write the formula to compute the "standard error" (0.3222) for the "macro average".

(c) (Five points) Write the formula to compute the standard deviation "macro stdev" (4.2281).

Clean-and-patch model using simlib
Purdue University, IE581, HW5
Schmeiser, April 2004

Model parameters...
Mean interarrival time = 1.000 minutes
Mean cleaning time = 0.700 minutes
Mean patching time = 0.800 minutes
Max jobs waiting to clean = 3 jobs
Corr(cleantime, patchtime) = 0.000

Experiment parameters...
Simulation end time = 1000.000 minutes
Number of macro reps = 20 replications
Initial Seed, Stream 1 = 1973272912
Initial Seed, Stream 2 = 281629770
Initial Seed, Stream 3 = 20006270

Arrival overflows:
<table>
<thead>
<tr>
<th></th>
<th>macro average</th>
<th>standard error</th>
<th>macro standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0267</td>
<td>0.0018</td>
<td>0.1596</td>
</tr>
</tbody>
</table>

Time at the cleaning operation, in minutes:
<table>
<thead>
<tr>
<th></th>
<th>macro average</th>
<th>standard error</th>
<th>macro standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.6909</td>
<td>0.0071</td>
<td>0.6801</td>
</tr>
</tbody>
</table>

Time at the patching operation, in minutes:
<table>
<thead>
<tr>
<th></th>
<th>macro average</th>
<th>standard error</th>
<th>macro standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.0661</td>
<td>0.3011</td>
<td>3.7716</td>
</tr>
</tbody>
</table>

Time in the system, in minutes:
<table>
<thead>
<tr>
<th></th>
<th>macro average</th>
<th>standard error</th>
<th>macro standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.7573</td>
<td>0.3006</td>
<td>3.7580</td>
</tr>
</tbody>
</table>

Number at the cleaner:
<table>
<thead>
<tr>
<th></th>
<th>macro average</th>
<th>standard error</th>
<th>macro standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.6778</td>
<td>0.0075</td>
<td>0.9920</td>
</tr>
</tbody>
</table>

Number at the patcher:
<table>
<thead>
<tr>
<th></th>
<th>macro average</th>
<th>standard error</th>
<th>macro standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.0119</td>
<td>0.3222</td>
<td>4.2281</td>
</tr>
</tbody>
</table>