One page of notes, front and back. Closed book. 60 minutes.
1. True or false. (2 points if correct, 1 point if left blank.)
   (If you wish, write an explanation of your thinking.)
   (a) T ← F If \( u \) is a random number, then the function \( x = 6 + 4u^{0.3} \) is an inverse cdf transformation.
   (b) T ← F If the output data \( y_1, \ldots, y_m \) are indicator variables, then \( \hat{p} = \bar{y} \) is a point estimate for \( P(Y = 1) \).
   (c) T ← F If \( \hat{p} = 0.1234567 \) and \( \text{sfe}(\hat{p}) = 0.01191111 \), then the meaningful digits of the point estimator are \( \hat{p} = 0.1234 \).
   (d) T ← F If the Monte Carlo sample size is increased by a factor of ten, then the point estimator will have one addition meaningful digit.
   (e) T ← F In our world view of Monte Carlo simulation experiments, the logic model has random variates as inputs and output data with unknown distribution.
   (f) T ← F If the inverse cdf \( F_X^{-1} \) is used to generate random variates, then \( E(X) = F_X^{-1}(0.5) \).
   (g) T ← F The Box-Muller method for generating normal random variates is based on the concept of composition.

2. Consider the sentences
   "The performance measure \( \theta \) is estimated by the point estimator \( \hat{\theta} \), which has standard error \( \text{sfe}(\hat{\theta}) \). The point estimator is, for example, a sample mean \( \bar{Y} \) or a sample variance \( S_Y^2 \).

   Suppose that you rerun a simulation experiment, keeping everything the same except the random-number seeds. Using the above two sentences for context, circle "constant", "random", and "undefined" for each part below.
   (a) \( S_Y \) constant random ← undefined
   (b) \( \bar{Y} > 6 \) constant random ← undefined
   (c) \( E(S_Y^2) \) constant ← random undefined
   (d) \( P(S_Y) \) constant random undefined ←
   (e) \( \text{sfe}(\hat{\theta}) \) constant random ← undefined
   (f) \( \theta \) constant ← random undefined
3. If random variates from the cdf $F_X$ can be generated using the inverse cdf transformation $x = 3u^2$, write the cdf. (Remember that the cdf is defined for all real numbers $x$.)

The range of $X$ is $F_X^{-1}(0) = 0$ to $F_X^{-1}(1) = 3$.
Therefore the cdf is zero for negative values of $x$ and one for $x > 3$.
In the range $[0, 3]$, solve for $u$ in terms of $x$ to obtain

$$u = F_X(x) = (x / 3)^{1/2}$$

4. In acceptance/rejection, we are given a density or mass function $f$ from which we wish to generate a random variate $x$. A majorizing function $t$ is chosen so that $t(x) \geq f(x)$ for every real number $x$. To generate a random variate, generate an $x$ from the density or mass function $r$ that is proportional to $t$; then accept $x$ with probability $f(x) / t(x)$.

Consider this situation. Three coffee drinkers wish to choose randomly which of the three will pay for the coffee. Let $X$ denote the chosen drinker, with $f(x) = \frac{1}{3}$ for $x = 1, 2, 3$ and zero elsewhere. That is, if $X = i$, the $i$th drinker pays.

They devise this method of choosing the drinker. Independently they flip two labeled coins and look at the result in binary. In particular, Drinker 1 pays if the first coin is tail and the second is head (binary 1), drinker 2 pays if the first coin is head and the second is tail (binary 2), drinker 3 pays if both are heads (binary 3), and they flip again if both coins are tails (binary 0).

(a) State the mass function $r$.

The coin flipping generates the values $z = 0, 1, 2, 3$ with equal likelihood. Therefore, the mass function is

$$r(z) = \frac{1}{4} \quad \text{for} \quad z = 0, 1, 2, 3$$

and zero elsewhere. (The answer doesn’t depend on the choice of dummy variable.)

(b) What is the expected number of times that the coins are flipped until the winner (also known as the loser) is determined?

$$\frac{4}{3} \leftarrow$$

Method 1.
Let $p$ be the probability that any iteration chooses one of the drinkers.
Then $p = P(X > 0) = \frac{3}{4}$.
The number of iterations is geometric, with mean $1/p = \frac{4}{3} \leftarrow$

Method 2.
The majorizing function is $t(x) = \frac{1}{3}$ for $x = 0, 1, 2, 3$.
The "area" under $t$ is then $k = \sum_{x=0}^{3} t(x) = \frac{4}{3} \leftarrow$
5. Consider a full-period pseudorandom-number generator with \( m \) values.

(a) Is the generator uniform in one dimension. Explain your answer.

If the generator is linear congruential, then "yes", as much as the discreteness allows, since the values will be equally spaced: \( u = i / m \) for \( i = 0, 1, \ldots, m - 1 \).

If \( m \) is "small", or if the generator is of some different structure, then the answer might be "no".

(b) In \( k \) dimensions, there are \( m^k \) grid points where \( k \)-tuples \((u_i, u_{i+1}, \ldots, u_{i+k-1})\) might fall. Use this information to argue that pseudorandom-numbers are, or are not, uniformly distributed in high dimensions.

For any dimension \( k \), there are only \( m \) \( k \)-tuples. Therefore, for large values of \( k \), almost all grid points are not covered.

Because Marsaglia’s hyperplanes are equally spaced, they help to guarantee that the \( m \) points are spread out over the \( k \)-dimensional cube.

6. Consider the Lehmer generator \( X_{i+1} = (2X_i + 2)(\text{mod5}) \). Show that it is, or is not, full period.

\[
\begin{align*}
X_0 &= 1 \\
X_1 &= 4 \\
X_2 &= 0 \\
X_3 &= 2 \\
X_4 &= 1 \\
X_5 &= 3 \\
X_6 &= 1 \\
X_7 &= 3 \\
X_8 &= 1 \\
X_9 &= 3 \\
X_{10} &= 1 \\
X_{11} &= 3 \\
X_{12} &= 1 \\
X_{13} &= 3 \\
X_{14} &= 1 \\
X_{15} &= 3 \\
X_{16} &= 1 \\
X_{17} &= 3
\end{align*}
\]

Therefore, there are two cycles and the generator is not full period.
7. Consider this probability problem. A dart board is circular with radius 12 inches. One dart is thrown, landing at location \((X_1, X_2)\), where \((0, 0)\) is the center of the board and the \(X\)s are measured in inches. Assume that the location of the dart is uniformly distributed over the dart board. We wish to know the variance of the distance to the center of the board, \(V[(X_1^2 + X_2^2)^{1/2}]\).

Consider solving this problem using Monte Carlo simulation experiment in which \(m\) darts are thrown independently. Denote the \(m\) dart locations by \((X_{1,i}, X_{2,i})\) for \(i = 1, 2, \ldots, m\).

(Comment: Nowhere here are you asked to solve this probability problem analytically.)

(a) What is the input model?

\((X_1, X_2)\) uniformly distributed over \(\{(x_1, x_2) \mid (x_1^2 + x_2^2)^{1/2} \leq 12\}\).

(There is no independence assumption, because there is only one dart.)

(b) What is the logic model?

Measuring the distance from the dart to the board’s center.

That is, \(Y = (X_1^2 + X_2^2)^{1/2}\).

(c) What is the performance measure, \(\theta\)?

\(\theta = V[(X_1^2 + X_2^2)^{1/2}] = V(Y)\).

(d) What are the output data?

The \(m\) distances, \(Y_i = (X_{1i}^2 + X_{2i}^2)^{1/2}\) for \(i = 1, 2, \ldots, m\).

(e) What is the point estimator, \(\hat{\theta}\)?

\[\hat{\theta} = \hat{S}_Y^2 = \frac{\sum_{i=1}^{m} Y_i^2 - m \bar{Y}^2}{m - 1}\]

where \(Y = \sum_{i=1}^{m} Y_i / m\).