Problem 1. Let $X, Y, Z, W$ be independent $U(0,1)$ random variables. Use a Monte Carlo method to compute $E[XY^2 + e^Z \cos(W)]$. How much computation should you do to be confident in your answer to three decimal places? (Turn in your code along with your answer!)

Solution: Before we do the numerical scheme, let us do an analytic computation. Since the four random variables are independent, we know that

$$E[XY^2 + e^Z \cos(W)] = E[X]E[Y^2] + E[e^Z]E[\cos(W)],$$

and we can compute these four numbers as follows:

$$E[X] = \int_0^1 x \, dx = \frac{1}{2},$$
$$E[Y^2] = \int_0^1 x^2 \, dx = \frac{1}{3},$$
$$E[e^Z] = \int_0^1 e^x \, dx = e - 1,$$
$$E[\cos(W)] = \int_0^1 \cos(x) \, dx = -\sin(1),$$

so

$$E[XY^2 + e^Z \cos(W)] = \frac{1}{6} + (e - 1) \sin(1) = 1.612550969 \cdots$$

Now, our Monte Carlo code will just be to grab four streams of $U(0,1)$’s, compute the objective function $Q = XY^2 + e^Z \cos(W)$, and then average. The question, of course, is how to estimate our error.

We know from the theorems in the notes that

$$\left| \frac{\hat{Q}_N}{N} - E[Q] \right| \approx \frac{\sigma}{\sqrt{N}} \mathcal{N}(0,1),$$

where $\mathcal{N}(0,1)$ is a Gaussian with mean zero and variance one. Let’s say that we want to be 99% confident of our result, which is (roughly) three standard deviations for the Gaussian, so we want

$$\text{error} < \frac{3\sigma}{\sqrt{N}},$$

or

$$N > 9 \times 10^6 \sigma^2.$$
We compute

\[ E[(e^Z \cos(W))^2] = \mathbb{E}[e^{2Z}]\mathbb{E}[^2(W)] = \left( \int_0^1 e^{2x} \, dx \right) \left( \int_0^1 \cos^2(x) \, dx \right) \]
\[ = e^2/2 \cdot \left( \frac{1}{2} + \sin(2)/4 \right) = e^2/8(\sin(2) + 2), \]

so

\[ V(e^Z \cos(W)) = \frac{e^2 - 1}{8}(\sin(2) + 2) - (e - 1)^2 \sin^2(1), \]

so

\[ V(Q) = \frac{7}{180} + \frac{e^2 - 1}{8}(\sin(2) + 2) - (e - 1)^2 \sin^2(1) = 0.2717655316 \cdots \]

This tells us that $3 \times 10^6$ samples should be good. Some code that produces the quantities of interest is:

```python
numSamples = 1e7;
X = rand(1,numSamples);
Y = rand(1,numSamples);
Z = rand(1,numSamples);
W = rand(1,numSamples);
QQ = X.* Y + exp(Z).*cos(W);
mean(QQ)
var(QQ)
```

**Problem 2.** We say that $T$ is an exponential random variable with parameter $\mu$ if $T > 0$ w. p. 1 and

\[ \mathbb{P}(T > t) = e^{-\mu t}, \quad \text{or} \quad \mathbb{P}(T \leq t) = 1 - e^{-\mu t}. \]

First, write a function that allows you to sample $T$ using a stream of $U(0,1)$ variables. Next, set $\mu = 2$ and compute $\mathbb{E}[T^p]$ for $p = 1, 2, 3, 4$. Do you see a pattern?

**Problem 3.** Suppose that $(X_n)_{n=0}^{\infty}$ is Markov($\lambda, P$). Define $Y_n = X_{kn}$ for some $k \geq 1$. Show that $(Y_n)_{n=0}^{\infty}$ is Markov($\lambda, P^k$).

**Solution:** We first want to show that $(Y_n)$ is a Markov chain. We compute:

\[ \mathbb{P}(Y_{n+1} = i_{n+1}|Y_0 = i_0, Y_1 = i_1, \ldots, Y_n = i_n) = \mathbb{P}(X_{k(n+1)} = i_{n+1}|X_0 = i_0, X_k = i_1, \ldots, X_{kn} = i_n) \]
\[ = \mathbb{P}(X_{k(n+1)} = i_{n+1}|X_{kn} = i_n) \]
\[ = \mathbb{P}(Y_{n+1} = i_{n+1}|Y_n = i_n), \]
so \( (Y_n) \) is a Markov chain. We can also compute
\[
\mathbb{P}(Y_1 = j | Y_0 = i) = \mathbb{P}(X_k = j | X_0 = i) = p_{ij}^{(3)},
\]
so the transition matrix for \( Y_n \) is \( P^k \).

**Problem 4.** We will consider a series of Markov chains which are given by a unidirectional ring with one escape point, i.e. pick \( N > 1 \) and \( 0 \leq p \leq 1 \), and consider the Markov chain with \( I = \{1, 2, \ldots, N + 1\} \) and transition probabilities
\[
\begin{align*}
p_{1,1} &= 0, & p_{1,2} &= p, & p_{1,N+1} &= 1 - p, \\
p_{i,i+1} &= p, & p_{i,i} &= 1 - p, & i &= 2, \ldots, N - 1, \\
p_{N,1} &= p, & p_{N,N} &= 1 - p, \\
p_{N+1,N+1} &= 1.
\end{align*}
\]

Some examples are:

(We are not drawing in the loops in these diagrams, since the weights are implied.)

Denote \( A = \{N + 1\} \).

Prove:

a. \( N + 1 \) is an absorbing state;

b. If \( 0 < p < 1 \), then \( h_i^A(p) = 1 \) for all \( i = 1, \ldots, N + 1 \).

c. If \( p = 0 \), then \( h_i^A(p) = 0 \) for all \( i = 2, \ldots, N \), and \( h_1^A(p) = h_{N+1}^A(p) = 1 \).

d. Compute \( k_i^A(p) \) for all \( i \) and all \( 0 \leq p \leq 1 \). For which \( i \) is this lowest? Highest? Does this make sense?

e. Show that \( k_1^A(\cdot) \) is discontinuous at 0, i.e. that
\[
k_1^A(0) \neq \lim_{p \to 0^+} k_1^A(p).
\]

Explain this paradox.

---

Solution:
a. Clearly \( \{N+1\} \) is a communicating class, and it is closed.

b. Let us choose \( i = 2, 3, \ldots, N - 1 \). Then

\[
h_i = ph_{i+1} + (1 - p)h_i,
\]

or

\[
ph_i = ph_{i+1}.
\]

Since \( p > 0 \), this means that \( h_i = h_{i+1} \), so that \( h_2 = h_3 = h_4 = \cdots = h_N \).

We also have

\[
h_N = ph_1 + (1 - p)h_N,
\]

so by a similar argument \( h_N = h_1 \). Finally, note that

\[
h_1 = ph_2 + (1 - p)h_{N+1} = ph_2 + (1 - p).
\]

Using the fact that \( h_2 = h_1 \), we obtain \( h_1 = 1 \), and therefore \( h_i = 1 \) for all \( i \).

c. If \( p = 0 \), then we can solve the Markov chain exactly. If \( X_0 = N + 1 \), then \( X_n = N + 1 \) for all \( n \). If \( X_0 = 1 \), then \( X_1 = N + 1 \) and thus \( X_n = N + 1 \) for all \( n \geq 1 \). If \( X_0 = i \) for \( i = 2, 3, \ldots, N - 1 \), then \( X_n = i \) for all \( n \). The result follows.

d. We approach these equations in a similar manner. For \( i = 2, \ldots, N - 1 \), we have

\[
k_i = pk_{i+1} + (1 - p)k_i + 1,
k_i = pk_{i+1} + 1,
k_i = k_{i+1} + \frac{1}{p},
\]

We obtain a similar equation for \( k_N \) and \( k_1 \), so we have \( k_N = k_1 + 1/p \) as well. We can recursively determine that

\[
k_2 = k_1 + \frac{N - 1}{p}.
\]

We now have that

\[
k_1 = pk_2 + (1 - p)k_{N+1} + 1,
k_1 = pk_2 + 1,
k_1 = pk_1 + (N - 1) + 1
\]

\[
k_1 = \frac{N}{1 - p}.
\]

We see from these formulas that \( k_1 \) is the lowest (except of course for \( k_{N+1} \)), and this makes sense, because all paths to \( N + 1 \) lead through state 1. Similarly, we see that \( k_2 \) is the highest, and again this makes sense from the topology: once we are at state 2, then we must go through the entire sequence of states to wrap around and get back to 1.
Thus we have
\[ \sum \]
Writing this out further, we have
and we are done. The argument is similar for the expectation.

This seems paradoxical at first glance: we expect that the average time to get somewhere should be continuous in the parameters governing the system. It is not here, as we have computed. To actually explain this paradox, let us consider in more detail the hitting time \( H^{(N+1)} \) when \( p \) is small. Notice that we have
\[
E_1[H^{(N+1)}] = E_1[H^{(N+1)}|X_k = N+1]P_1(X_k = N+1) + E_1[H^{(N+1)}|X_k = 2]P_1(X_k = 2)
\]
\[
= E_{N+1}[H^{(N+1)}]P_1(X_k = N+1) + E_2[H^{(N+1)}]P_1(X_k = 2)
\]
\[
= (1-p) \cdot 1 + pE_2[H^{(N+1)}].
\]
We write it like this (although of course we have already computed \( k_2 \) to point out that this is a situation where we have a random variable that has an overwhelming probability of being small, and a very small chance of being large, but when it is large, it is so large that it skews the mean significantly. So, with high probability, the two processes look the same — i.e., we jump to state \( N+1 \) immediately, but for \( p \ll 1 \), there is a small probability of going around the merry-go-round and taking forever.

---

**Problem 5.** Show that \( P_i, E_i \) satisfy the same formulas \( P, E \) do in the LTP, LTE. More specifically, if \( \{B_k\} \) is a partition of \( \Omega \), show that
\[
P_i(A) = \sum_k P_i(A|B_k)P_i(B_k), \quad E_i[A] = \sum_k E_i[A|B_k]P_i(B_k).
\]

**Solution:** Writing these expressions out, we have
\[
P_i(A) = P(A|X_0 = i), \quad P_i(A|B_k) = P(A|B_k, X_0 = i), \quad P_i(B_k) = P(B_k|X_0 = i).
\]
Writing this out further, we have
\[
P_i(A|B_k)P_i(B_k) = \frac{P(A \cap B_k \cap \{X_0 = i\})}{P(B_k \cap \{X_0 = i\})} \cdot \frac{P(B_k \cap \{X_0 = i\})}{P(X_0 = i)} = \frac{P(A \cap B_k \cap \{X_0 = i\})}{P(X_0 = i)}
\]
Since \( \{B_k\} \) is a partition of \( \Omega \), \( \{B_k \cap \{X_0 = i\}\} \) is a partition of \( \{X_0 = i\} \), and thus
\[
\sum_k P(A \cap B_k \cap \{X_0 = i\}) = P(A \cap \{X_0 = i\})
\]
Thus we have
\[
\sum_k \frac{P(A \cap B_k \cap \{X_0 = i\})}{P(X_0 = i)} = \frac{P(A \cap \{X_0 = i\})}{P(X_0 = i)},
\]
and we are done. The argument is similar for the expectation.