Problem 1: Short Questions

(a) The rectangle below represents the entire event space. We have depicted two events, $A$ and $B$ as portions of the event space. Notice that events $A$ and $B$ do not overlap. Assume they are not empty.

![Diagram of event space with events A and B]

True or False: Events $A$ and $B$ are independent. Explain your answer.

**Answer: False.** $A$ and $B$ cannot be independent as $P(B|A) = 0$ and $P(B) \neq 0$ so $P(B|A) \neq P(B)$

(b) True or False: If $P(A|C) > P(B|C)$ and $P(A|C^c) > P(B|C^c)$, then $P(A) > P(B)$. (Remember: $C^c$ is the complement of $C$.) Explain your answer.

**Answer: True.** We start with $P(A|C) > P(B|C)$

$$P(A|C) > P(B|C)$$

$$\frac{P(C|A)P(A)}{P(C)} > \frac{P(C|B)P(B)}{P(C)}$$

$$P(C|A)P(A) > P(C|B)P(B) \quad (1)$$

Now using $P(A|C^c) > P(B|C^c)$ we have

$$P(A|C^c) > P(B|C^c)$$

$$\frac{P(C^c|A)P(A)}{P(C^c)} > \frac{P(C^c|B)P(B)}{P(C^c)}$$

$$(1 - P(C|A))P(A) > (1 - P(C|B))P(B) \quad (2)$$

Now adding inequalities (1) and (2) together we have

$$P(A) > P(B)$$

(c) True or false: If $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$, then $X$ and $Y$ are independent. Explain your answer.

**Answer: False.** From the HW we know $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(X) + 2 \cdot \text{Cov}(X,Y)$. So the assertion will be valid whenever $\text{Cov}(X,Y) = 0$ which does not implies independence (rather it implies uncorrelatedness).
(d) Let \( X \) be a continuous random variable with a probability density function \( f_X \), and let \( Y = X - 3 \). True or false: The probability density function of \( Y \) is \( f_X(y - 3) \). Explain your answer.

**Answer: False** As per usual we start by using the CDF.

\[ F_Y(y) = P(Y \leq y) \]
\[ = P(X - 3 \leq y) \]
\[ = P(X \leq y + 3) \]
\[ = F_X(y + 3) \]

And performing the differentiation we get immediately that

\[ f_Y(y) = f_X(y + 3) \]

(e) Which of these properties can/must a valid probability density function \( f_X \) for a continuous random variable \( X \) have? You can assume \( f_X \) is a continuous function. Explain all your answers.

("Can" if the property is possible, but not mandatory. "Must" if the property is mandatory. "Impossible" if no valid PDF has that property.)

(i) \( f_X(x) > 1 \) at some point \( x \). CAN / MUST / IMPOSSIBLE?

**Answer: CAN.** Consider the random variable distributed \( Exp(2) \) whose PDF evaluated at 0 is 2.

(ii) \( f_X(x) < 0 \) at some point \( x \). CAN / MUST / IMPOSSIBLE?

**Answer: IMPOSSIBLE.** If you consider some small interval \( \delta x \) around a point where the PDF is negative then you could achieve a negative probability, which is inconsistent with our probability axioms.

(iii) \( \int_{-\infty}^{+\infty} f_X(x)dx = 1 \). CAN / MUST / IMPOSSIBLE?

**Answer: MUST.** The integral of the PDF over an interval gives us the probability the random variable takes on a value in that interval. From the probability axioms, this must be normalized to 1.

(f) Let \( X_1, X_2, \ldots \), be a Markov chain. Let \( p_4, p_{3,4}, p_{4,7} \) be respectively the probabilities of error of the MAP estimate of \( X_5 \) given \( X_4 \), given \( X_3 \) and \( X_4 \) and given \( X_4 \) and \( X_7 \).

(i) Is it always true that \( p_{3,4} < p_4 \), \( p_{3,4} > p_4 \) or \( p_{3,4} = p_4 \)? Or does the answer depend on the Markov chain in question? Explain your answers.

**Answer: \( p_{3,4} = p_4 \).** Since a Markov chain is conditionally independent given the previous state, knowing state 4 entirely summarizes all information we can have about state 5 so knowing state 3 adds no new information.

(ii) Is it always true that \( p_{4,7} < p_4 \), \( p_{4,7} > p_4 \) or \( p_{4,7} = p_4 \)? Or does the answer depend on the Markov chain in question? Explain your answers.

**Answer: This will depend on the particular structure of the Markov chain.**
Problem 2: Take-home Exams

There are 100 students in a probability class. The time needed to finish a take-home is exponentially distributed with mean \( \mu \) hours, and is independent from student to student. The instructor can design the level of difficulty of the exam to control \( \mu \).

(a) The class will sue the instructor for inhumane treatment if the average number of hours taken by the whole class exceeds 10 hours. Find a (positive) value of \( \mu \) for which the instructor can guarantee that the chance of being sued is less than 1%.

**Answer:** Let \( T_i \) be the random variable which is the time it takes student \( i \) to finish the exam. We know from the problem statement that \( T_i \sim \text{Exp}(1/\mu) \). We want to ensure

\[
P \left( \frac{1}{100} \sum_{i=1}^{100} T_i \geq 10 \right) \leq 0.01
\]

We can use the Chebyshev Inequality to bound this probability above as

\[
P \left( \frac{1}{100} \sum_{i=1}^{100} T_i - \mu \geq 10 - \mu \right) \leq \frac{\text{Var} \left( \frac{1}{100} \sum_{i=1}^{100} T_i \right)}{(10 - \mu)^2}
\]

Setting the quantity on the right equal to 0.1 we solve for \( \mu \) as follows

\[
\frac{\text{Var} \left( \frac{1}{100} \sum_{i=1}^{100} T_i \right)}{(10 - \mu)^2} = 0.01
\]

\[
\frac{1}{100} \text{Var}(T_i) = 0.01 \cdot (10 - \mu)^2
\]

\[
\mu^2 = (10 - \mu)^2
\]

Where we use the fact that the variance of an exponential random variable with parameter \( \lambda \) is \( \lambda^{-2} \). Solving we find \( \mu = 5 \).

(b) Find a value of \( \mu \) for which the chance of the instructor being sued is approximately 1%. Compare this value with the value you found in part (a).

**Answer:** Again we are concerned with \( P(\sum_{i=1}^{100} T_i \geq 1000) = 0.01 \). Since we have the sum of i.i.d random variables we can invoke the CLT approximation. We start by noting that \( \mathbb{E} \left[ \sum_{i=1}^{100} T_i \right] = 100\mu \) and \( \text{Var}(\sum_{i=1}^{100} T_i) = 100\mu^2 \)

\[
P \left( \sum_{i=1}^{100} T_i \geq 1000 \right) = P \left( \frac{\sum_{i=1}^{100} T_i - 100\mu}{10\mu} \geq \frac{1000 - 100\mu}{10\mu} \right)
\]

Using the z table provided we find that in order for this probability to be at most 1% we need \( z = 2.33 \). Therefore,

\[
\frac{1000 - 100\mu}{10\mu} = 2.33
\]

Solving for \( \mu \) we find we need to choose \( \mu \approx 8.11 \). Compared with part (A) this value is considerably larger indicating the professor could make the exam much more difficult without being sued!
(c) The exam is supposed to be finished in 24 hours. Using the value of $\mu$ you found in part (a), what is a good approximation to the distribution of the number of students who cannot finish their exam in the 24 hours?

\textbf{Answer:} The approximate probability that a particular student finishes the exam is 

$$P(T \leq 24) = \int_0^{24} \frac{e^{-t/5}}{5} \, dt \approx 0.9918$$

So the probability that the student \textit{cannot} finish the exam is $\approx 0.00823$. Since each student’s time to complete the exam is independent of the others, the number of student who cannot finish is distributed $\text{Binom}(100, 0.00823)$. Since we the probability of success is quite small and $n$ is reasonably large we can approximate the distribution as $\text{Poiss}(0.823)$.

(d) Our class only has 15 students. Do you think the approaches you used to solve parts (a),(b) and (c) remain valid?

\textbf{Answer:} Part (A) remains valid but the CLT does not hold for only $n = 15$ so the approximation would be very poor. Additionally, since $n$ is not very large, we would likely not be justified in using the Poisson approximation.
Problem 3: Total Expectation

(a) Let $X$ and $Y$ be two discrete random variables. The expectation of $Y$ conditional on $X = a$ is defined to be:

$$E[Y|X = a] = \sum_b bP(Y = b|X = a).$$

Show that

$$E[Y] = \sum_a P_X(a)E[Y|X = a].$$

(Hint: which basic rule in probability does this formula remind you of?)

**Answer:** Starting from the right hand side of the assertion and using the definition of conditional expectation we have

$$\sum_a P_X(X = a)E[Y|X = a] = \sum_a P_X(X = a) \left( \sum_b b \cdot P(Y = b|X = a) \right)$$

$$= \sum_a \sum_b b \cdot P_X(X = a)P_Y|X(Y = b|X = a)$$

$$= \sum_a \sum_b b \cdot P_{X,Y}(X = a, Y = b)$$

$$= \sum b \sum_a P_{X,Y}(X = a, Y = b)$$

The inner sum is just *marginalizing* the random variable $X$ to give $P(Y = b)$

$$\sum_a P_X(X = a)E[Y|X = a] = \sum_b b \cdot P_Y(Y = b)$$

$$= E[Y]$$

(b) Let $U_i$’s be i.i.d.random variables each with mean $\mu_U$ and variance $\sigma_U^2$. Let $N$ be a random variable which takes on only positive integer values with mean $\mu_N$ and variance $\sigma_N^2$, and independent of the $U_i$’s. Let

$$S = \sum_{i=1}^N U_i.$$

(i) Assume that $\mu_U = 0$. Compute the mean and the variance of $S$ in terms of $\sigma_U^2, \mu_N$ and $\sigma_N^2$.

**Answer:** Using the more general result in the next part we find

$$E[S] = 0$$

$$\text{Var}(X) = \sigma_U^2 \mu_N$$

(ii) Repeat part (b)(i) for general $\mu_U$. 


Answer: We start by finding the expectation of $S$ using iterated expectation as proven above

$$
\mathbb{E}[S] = \sum_n P_N(N = n) \mathbb{E}[S|N = n]
$$

$$
= \sum_n P_N(N = n) \mathbb{E} \left[ \sum_{i=1}^N U_i | N = n \right]
$$

$$
= \sum_n P_N(N = n) \mathbb{E} \left[ \sum_{i=1}^n U_i \right]
$$

$$
= \sum_n P_N(N = n) (n \mu_U)
$$

$$
= \mu_U \sum_n n \cdot P_N(N = n)
$$

$$
= \mu_U \mu_N
$$

We can now find the variance of the random variable by computing it’s second moment.

$$
\mathbb{E}[S^2] = \sum_n P_N(N = n) \mathbb{E}[S^2|N = n]
$$

$$
= \sum_n P_N(N = n) \mathbb{E} \left[ \left( \sum_{i=1}^n U_i \right)^2 | N = n \right]
$$

$$
= \sum_n P_N(N = n) \mathbb{E} \left[ \sum_{i=1}^n \sum_{j=1}^n U_i U_j \right]
$$

$$
= \sum_n P_N(N = n) \mathbb{E} \left[ \sum_{i=1}^n U_i^2 + \sum_{i \neq j} U_i U_j \right]
$$

Now using the fact that the second term has $n^2 - n$ terms (not necessarily unique) and using the independence of $U_i$ and $U_j \ \forall i \neq j$ we find

$$
\mathbb{E}[S^2] = \sum_n P_N(N = n) \left[ n \cdot \mathbb{E}[U_i^2] + (n^2 - n) \cdot \mathbb{E}[U_i U_j] \right]
$$

$$
= \sum_n P_N(N = n) \left[ n \cdot (\sigma_U^2 + \mu_U^2) + (n^2 - n) \cdot \mu_U^2 \right]
$$

$$
= \sigma_U^2 \cdot \sum_n n \cdot P_N(N = n) + \mu_U^2 \cdot \sum_n n^2 \cdot P_N(N = n) - \mu_U^2 \cdot \mathbb{E}[N^2]
$$

$$
= \sigma_U^2 \mu_N + \mu_U^2 \left( \mu_N^2 + \sigma_N^2 \right)
$$

Now we can compute the variance as

$$
\text{Var}(S) = \mathbb{E}[S^2] - (\mathbb{E}[S])^2
$$

$$
= \sigma_U^2 \mu_N + \mu_U^2 \left( \mu_N^2 + \sigma_N^2 \right) - \mu_U^2 \mu_N^2
$$

$$
= \sigma_U^2 \mu_N + \mu_U^2 \sigma_N^2
$$
Problem 4: Markov Chains

Consider the two-state Markov chain \( X_1, X_2, \ldots, \) with \( X_i \in \{0, 1\} \) and symmetric transition probabilities \( P(1|0) = P(0|1) = \alpha \) (0 < \( \alpha \) < 0.5). Suppose the initial distribution (distribution of \( X_1 \)) is \( \pi(0) = \pi(1) = 0.5 \).

(a) Draw the state transition diagram.

\[ 1 - \alpha \quad \xrightarrow{\alpha} \quad 0 \quad \xrightarrow{\alpha} \quad 1 \quad \xrightarrow{1 - \alpha} \]

(b) Compute the distribution of \( X_2 \).

\[
P(X_2 = 0) = P(X_2 = 0|X_1 = 0)P(X_1 = 0) + P(X_2 = 0|X_1 = 1)P(X_1 = 1) \\
= (1 - \alpha)P(X_1 = 0) + \alpha P(X_1 = 1) \\
= (1 - \alpha) \times 0.5 + \alpha \times 0.5 \\
= 0.5, 
\]

and \( P(X_2 = 1) = 0.5 \) as well.

(c) What is the distribution of \( X_n \) for general \( n \)? Answer: Since the transition probabilities do not depend on time, we can repeat the argument with \( X_2 \) replacing \( X_1 \) and \( X_3 \) replacing \( X_2 \) and obtain \( P(X_3 = 0) = P(X_3 = 1) = 0.5 \). Repeating this argument, we see that \( P(X_n = 1) = P(X_n = 0) = 0.5 \) for all \( n \).

(d) Compute the joint distribution of \( X_1 \) and \( X_3 \). Are these two random variables independent?

Answer: By the total probability rule,

\[
P(X_1 = 1, X_3 = 0) = P(X_1 = 1, X_3 = 0, X_2 = 0) + P(X_1 = 1, X_3 = 0, X_2 = 1) \\
= \frac{1}{2} (P(X_3 = 0, X_2 = 0|X_1 = 1) + P(X_3 = 0, X_2 = 1|X_1 = 1)) \\
= \frac{1}{2} ((1 - \alpha)\alpha + (1 - \alpha)\alpha) \\
= \alpha(1 - \alpha). 
\]

Similarly, \( P(X_1 = 0, X_3 = 1) = \alpha(1 - \alpha) \).

\[
P(X_1 = 0, X_3 = 0) = \frac{1}{2} (P(X_3 = 0, X_2 = 0|X_1 = 0) + P(X_3 = 0, X_2 = 1|X_1 = 0)) \\
= \frac{1}{2} ((1 - \alpha)^2 + \alpha^2) 
\]

and \( P(X_1 = 0, X_3 = 0) = P(X_1 = 1, X_3 = 1) \).
Using part (c) we know that
\[ P(X_1 = 1)P(X_3 = 0) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} \neq \alpha(1 - \alpha) \]
since \( \alpha < 0.5 \). Therefore the two random variables are not independent.

(e) Show that for \( k \geq 2 \),
\[ P(X_k = 0|X_1 = 0) = c_1 + c_2\beta^{k-1} \]
for some constants \( \beta, c_1, c_2 \) that do not depend on \( k \), with \( 0 < \beta < 1 \). Determine the constants in terms of \( \alpha \).
(Hint: develop the formula for \( P(X_k = 0|X_1 = 0) \) recursively in \( k \) starting with \( k = 2 \).)

**Answer:** We can write down a formula for this recursion on the conditional probabilities as
\[
P(X_n = 1|X_m = 1) = P(X_n = 1, X_{n-1} = 1|X_m = 1) + P(X_n = 1, X_{n-1} = 0|X_m = 1)
= P(X_n = 1|X_{n-1} = 1)P(X_{n-1} = 1|X_m = 1) + P(X_n = 1|X_{n-1} = 0)P(X_{n-1} = 0|X_m = 1)
= (1 - \alpha)P(X_{n-1} = 1|X_m = 1) + \alpha(1 - P(X_{n-1} = 1|X_m = 1))
= \alpha + (1 - 2\alpha)P(X_{n-1} = 1|X_m = 1)
\]
Since there is no distinction between state 0 and 1, then by symmetry \( P(X_n = 1|X_1 = 1) = P(X_n = 0|X_1 = 0) \).
The general form of the recursion is then
\[
a_n = \alpha + (1 - 2\alpha)a_{n-1} \quad a_1 = 1
\]
where the initial condition comes from the fact that \( P(X_1 = 0|X_1 = 0) = 1 \). This can be solved using \( z \)-transforms or plugging into software such as WolframAlpha. However, since we were given the general form of the solution (a so-called ansatz), we need only plug into the difference equation to find the constants.
\[
c_1 + c_2\beta^{n-1} = (c_1 + c_2\beta^{n-2})(1 - 2\alpha) + \alpha
= \alpha + c_1(1 - 2\alpha) + c_2(1 - 2\alpha)\beta^{n-2}
\]
Now we can equate the additive constants that don’t depend on \( n \).
\[
c_1 = c_1(1 - 2\alpha) + \alpha
= 1/2
\]
Now equating the portions that *do* depend on \( n \)
\[
c_2\beta^{n-1} = c_2(1 - 2\alpha)\beta^{n-2}
\]
\[
c_2\beta^{n-1} - c_2(1 - 2\alpha)\beta^{n-2} = 0
\]
\[
c_2\beta^{n-1}(1 - (1 - 2\alpha)\beta^{-1}) = 0
\implies (1 - (1 - 2\alpha)\beta^{-1}) = 0
\]
So we conclude \( \beta = 1 - 2\alpha \). Now using the initial condition \( a_1 = 1 \).
\[
1/2 + c_1(1 - 2\alpha)^{1-1} = 1
\]
\[
c_1 = 1/2
\]
So the solution to the recursion is given by
\[
P(X_n = 0|X_1 = 0) = 1/2 + 1/2(1 - 2\alpha)^{n-1}
\]
(f) Give an expression for \( P(X_n = 0|X_m = 0) \) for general \( m, n \) with \( n > m \), and hence compute the joint distribution of \( X_m \) and \( X_n \).

**Answer:** We can write the recursion for second case as

\[
P(X_n = 1|X_1 = 0) = P(X_n = 1, X_{n-1} = 1|X_1 = 0) + P(X_n = 1, X_{n-1} = 0|X_1 = 0)
= P(X_n = 1|X_{n-1} = 1)P(X_{n-1} = 1|X_1 = 0) + P(X_n = 1|X_{n-1} = 0)P(X_{n-1} = 0|X_1 = 0)
= (1 - \alpha)P(X_{n-1} = 1|X_1 = 0) + \alpha (1 - P(X_{n-1} = 1|X_1 = 0))
= \alpha + (1 - 2\alpha)P(X_{n-1} = 1|X_1 = 0)
\]

Again by symmetry \( P(X_n = 1|X_m = 0) = P(X_n = 0|X_m = 1) \). The recursion is of the form

\[a_n = \alpha + (1 - 2\alpha)a_{n-1} \quad a_1 = 0\]

Note that even though the functional form of this difference equation is the same as the previous part, we have a different initial condition. This only changes the \( c_1 \) constant however and leaves the others unchanged. Solving for it in the same manner as above will yield

\[
P(X_n = 1, X_1 = 0) = 1/2 - 1/2(1 - 2\alpha)^{n-1}
\]

Again by symmetry \( P(X_n = 0, X_1 = 1) = P(X_n = 1, X_1 = 0) \). In the general case our conditionals are for \( n \geq m \)

\[
P(X_n = 1|X_m = 1) = P(X_n = 0|X_m = 0) = 1/2 + 1/2(1 - 2\alpha)^{n-m}

P(X_n = 0|X_m = 1) = P(X_n = 1, X_m = 0) = 1/2 - 1/2(1 - 2\alpha)^{n-m}
\]

Since the prior probability \( P(X_m = 1) = P(X_n = 0) = 1/2 \) our joint probability mass function is given by

\[
P(X_n = 1|X_m = 1) = P(X_n = 0|X_m = 0) = 1/4(1 + (1 - 2\alpha)^{n-m})

P(X_n = 0|X_m = 1) = P(X_n = 1, X_m = 0) = 1/4(1 - (1 - 2\alpha)^{n-m})
\]

(g) Let

\[A_n = \frac{X_1 + X_2 + \ldots + X_n}{n}\]

be the *time-average* of the Markov chain. Using part (f) or otherwise, show that the variance of \( A_n \) approaches 0 as \( n \to \infty \).

**Answer:** Since the random variables are not independent we cannot write the variance of the sum as the sum
of the variances. Instead we use the rule derived in homework

$$\text{Var}(A_n) = \text{Var} \left( \frac{X_1 + X_2 + \cdots + X_n}{n} \right)$$

$$= \frac{1}{n^2} \left( \sum_{i=1}^{n} \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j) \right)$$

$$= \frac{1}{n^2} \left( n \cdot \left( \frac{1}{2} \right) \left( 1 - \frac{1}{2} \right) + \sum_{i \neq j} \left[ \text{E}(X_i X_j) - \text{E}(X_i) \text{E}(X_j) \right] \right)$$

$$= \frac{1}{n^2} \left( n/4 - (n^2 - n)/4 + \sum_{i \neq j} \text{E}(X_i X_j) \right)$$

$$= \frac{1}{n^2} \left( n/2 - n^2/4 + \sum_{i \neq j} \sum_{k=0}^{1} \sum_{\ell=0}^{1} k\ell \cdot \text{P}(X_i = k, X_j = \ell) \right)$$

$$= \frac{1}{n^2} \left( n/2 - n^2/4 + \sum_{i \neq j} \text{P}(X_i = 1, X_j = 1) \right)$$

Now using the joint PMF we derived in the previous part.

$$= \frac{1}{n^2} \left( n/2 - n^2/4 + \sum_{i \neq j} \left( 1/4 + 1/4(1 - 2\alpha)^{|i-j|} \right) \right)$$

$$= \frac{1}{n^2} \left( n/2 - n^2/4 + (n^2 - n)/4 + 1/4 \sum_{i \neq j} \left( (1 - 2\alpha)^{|i-j|} \right) \right)$$

$$= \frac{1}{n^2} \left( n/4 + 1/2 \sum_{i=2}^{n} \sum_{j=1}^{i-1} \left( 1 - 2\alpha \right)^{(i-j)} \right)$$

Though in principle we could carry out this summation using the standard formula for the sum of a geometric series, we are only considered with the limit of large $n$. The first term goes as $1/n$ and since $|1 - 2\alpha| < 1$, the summation will be finite and will decay as $1/n$ as well. Therefore the variance tends to 0 for large $n$. If $\alpha = 0$ however, the variance will not tend 0 as the summation will no longer be finite but will instead grow as $n^2$.

(h) State the law of large numbers for i.i.d. random variables. Using part (g) or otherwise, state and prove an extension of the law of large numbers for this Markov chain.

**Answer:** If $Y_1, Y_2, \ldots, Y_n$ are i.i.d random variables with mean $\mu$ then the law of large numbers states $S_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$ will have the following behavior

$$\text{P}(|S_n - \mu| \geq \epsilon) \xrightarrow{n \to \infty} 0$$

Using Chebyshev’s bound we have

$$\text{P}(|A_n - 1/2| > \epsilon) \leq \frac{\text{Var}(A_n)}{\epsilon^2}$$

As $n$ grows large then the variance tends to 0 (provided $\alpha \neq 0$) demonstrating

$$\text{P}(|A_n - 1/2| \geq \epsilon) \xrightarrow{n \to \infty} 0$$
(i) Simulate the Markov chain for $\alpha = 0.1, 0.01, 0.001$ and plot $A_n$ as a function of $n$ to empirically verify the law of large number you stated and proved in part (h). Does the time duration for which you have to simulate the Markov chain to see the law of large number effect depend on $\alpha$? Explain.

**Answer:** The plot below is one realization of the random process for the various values of $\alpha$ over 10,000 trials. Yes the time duration is dependent on $\alpha$ since for $\alpha$ close to 1 or 0, the state will remain constant for long periods of time so the averaging will not be seen until after transitions have been made.

```python
# EE178 Final 2015 Problem 4 Part(H)
from __future__ import division
import numpy as np
from random import randint, random
import matplotlib.pyplot as plt

def simulate(state, alpha):
    if (random() <= alpha):
        return (state + 1) % 2
    else:
        return state

N = 10000
time_avg = np.zeros((4,N)) # alpha indxs the row and num steps indxs column
for indx, alpha in enumerate([0,0.001,0.01,1]) : # Loop over transition probs
    total = 0 # Keep track of running sum
    x = randint(0, 1) # Initialize state uniformly at random
    time_avg[indx, 0] = x
    for n in range(1,N):
        x = simulate(x, alpha)
        total += x
        time_avg[indx, n] = total/n
```
(j) Does the law of large numbers you stated in part (g) hold if $\alpha$ were 0? Justify your answer.

**Answer:** No the law of large numbers does not hold for $\alpha = 0$. Intuitively, once the initialization is performed, the state will never change and the average will never converge to $1/2$ as the state will always be either 0 or 1 for all time.
Problem 5: Convolutional Codes

Consider the convolutional code shown in the figure below:

Each box is a delay element and the addition is modulo 2, exactly as in the figure for the convolution code studied in Lab 3.

(a) What is the rate of this code in terms of number of information bits per coded bit?

Answer: The rate of the code is 1/2 just as in the convolutional code for the lab since there are two output bits for every one input bit.

(b) Suppose this code is fed with i.i.d. Bernoulli(0.5) information bits.

(a) Write down the state transition diagram for the Markov chain description of the code. Be explicit about how the states are defined in terms of the information bits, how many states there are in the Markov chain, and what are the transition probabilities

Answer: Since there are 4 shift registers, this code has a memory of 4 so our state will be of the form $q_1^{(i)} q_2^{(i)} q_3^{(i)} q_4^{(i)}$
where \( q_k^{(i)} \) is the signal leaving the \( k^{th} \) delay block at time \( i \). The state is updated according to

\[
\begin{align*}
q_1^{(i)} &= d_{i-1} \oplus q_3^{(i-1)} \oplus q_4^{(i-1)} \\
q_2^{(i)} &= q_1^{(i-1)} \\
q_3^{(i)} &= q_2^{(i-1)} \\
q_4^{(i)} &= q_3^{(i-1)}
\end{align*}
\]

Since each bit could be 1 or 0 we have \( 2^4 = 16 \) possible states. Since the bits are I.I.D the transition probabilities are 1/2 for all transitions.

(b) Write down the output coded bits at time \( n \) in terms of the state of your Markov chain at time \( n \) and the input information bit at time \( n \).

**Answer:** The FSM is shown below (though it’s not pretty) with the input bits over the arrow. Though not drawn explicitly on the state transition diagram, the output bits can be calculated using the formula in the subsequent part.

\[
\begin{align*}
S &= d_n \\
P &= ((q_{n-4} \oplus q_{n-3}) \oplus d_n) \oplus q_{n-4} = q_{n-3} \oplus d_n
\end{align*}
\]

(c) Suppose the coded bits are passed through a noisy channel where the bits are flipped with probability \( p \). Roughly by how many times do you expect the computational complexity of the Viterbi algorithm to increase by for this code compared to the convolutional code in Lab 3?

**Answer:** In the convolutional code in lab 3, we needed to do one comparison for each node in each layer, so a total of four comparisons per layer. Since we have 16 states and still only two edges incident to each node, we need a total of 16 comparisons per layer so the complexity increases by a factor of 4.
Problem 6: Buffon Needle on Tiles

Suppose the floor is tiled into unit squares by infinite parallel horizontal and vertical lines. A needle of unit length is dropped randomly onto the floor. What is the probability that the needle intersects exactly two lines?

**Answer:** We start by defining the random variables of the problem. Let $\Theta \sim Uni(0, \pi/2)$ be the acute angle the needle makes with the x-axis. Let $X$ and $Y$ be the $x$ and $y$ distances of the center of the needle from the closest line. Note that $X, Y \sim Uni(0, 1/2)$. Since these 3 random variables are independent we can write their joint density as

$$f(\Theta = \theta, X = x, Y = y) = f_\Theta(\Theta = \theta) f_X(X = x) f_Y(Y = y) = \frac{2}{\pi} \cdot 2 \cdot 2$$

From class we know that the needle will intersect the **horizontal** grid lines if $Y \leq 1/2 \sin(\theta)$. Similarly, we can see that the needle will intersect the **vertical** lines if $X \leq 1/2 \cos(\theta)$. We can simply integrate the joint density over the region as follows

$$\int_0^{\pi/2} \int_0^{1/2 \sin \theta} \int_0^{1/2 \cos \theta} \frac{8}{\pi} \, dx \, dy \, d\theta = \frac{8}{\pi} \int_0^{\pi/2} \frac{1}{4} \sin(\theta) \cos(\theta) \, d\theta$$

$$= \frac{2}{\pi} \int_0^{\pi/2} \frac{1}{2} \sin(2\theta) \, d\theta$$

$$= \frac{1}{\pi} \left[ -\frac{1}{2} \cos(2\theta) \right]_0^{\pi/2}$$

$$= \frac{1}{\pi} \left[ \frac{1}{2} + \frac{1}{2} \right]$$

$$= \frac{1}{\pi}$$