Due Monday August 1 at 1:59PM

1. **Beta Testing** (10 points: 4/6)

   Midterm 2 of CS70 consists of n questions. Before the test, m TAs are going to beta test the midterm. Each of them will solve a question correctly with probability p (independently of other TAs and independently of other questions). Let X be the number of distinct questions that no one solves correctly.

   (a) What is the expectation of X? What's the variance of X?

   Answer: Represent X as the sum of the indicator variables $X_1, X_2, \ldots, X_n$ such that $X_i = 1$ if no TA gets question i correctly. We have

   $$\mathbb{E}(X_i) = 1 \times \Pr[X_i = 1] + 0 \times \Pr[X_i = 0] = (1-p)^m,$$

   and from linearity of expectation we have

   $$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(X_i) = n(1-p)^m.$$

   Since the indicator variables are mutually independent, we have

   $$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

   $$\mathrm{Var}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = (1-p)^m(1-(1-p)^m)$$

   Therefore we have

   $$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = n(1-p)^m(1-(1-p)^m)$$

   (b) Now each TA is going to choose a question uniformly at random from the n questions to grade (independently of other TAs). Let Y be the number of distinct questions that no one chooses. What is the expectation of Y? What's the variance of Y?

   Answer: Similar to part a, represent Y as the sum of the indicator variables $Y_1, Y_2, \ldots, Y_n$, such that $Y_i = 1$ if no TA chooses question i. We have

   $$\mathbb{E}(Y_i) = 1 \times \Pr[Y_i = 1] + 0 \times \Pr[Y_i = 0] = \left(\frac{n-1}{n}\right)^m,$$

   and from linearity of expectation we have

   $$\mathbb{E}(Y) = \sum_{i=1}^{n} \mathbb{E}(Y_i) = n\left(\frac{n-1}{n}\right)^m.$$

To find the variance, we cannot simply sum the variances of our indicator variables, because they are not independent. However, we can still compute $\mathrm{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2$ using linearity of expectation, but how can we find $\mathbb{E}(Y^2)$? Recall that

$$\begin{aligned}
\mathbb{E}(Y^2) &= \mathbb{E}((Y_1 + \ldots + Y_n)^2) \\
&= \mathbb{E}(\sum_{i,j} Y_i Y_j) \\
&= \sum_{i,j} \mathbb{E}(Y_i Y_j) \\
&= \sum_{i}^{n} \mathbb{E}(Y_i^2) + \sum_{i \neq j} \mathbb{E}(Y_i Y_j).
\end{aligned}$$

The first term is simple to calculate: $\mathbb{E}(Y_i^2) = 1^2 \times \Pr[Y_i = 1] = \left(\frac{n-1}{n}\right)^m$, meaning that

$$\sum_{i=1}^{n} \mathbb{E}(Y_i^2) = n\left(\frac{n-1}{n}\right)^m.$$

$Y_i Y_j = 1$ when both Y_i and Y_j are 1, which means no one chooses question i and question j. This happens with probability

$$\Pr[Y_i = Y_j = 1] = \Pr[Y_i = 1 \cap Y_j = 1] = \left(\frac{n-2}{n}\right)^m.$$

Thus, we can now compute

$$\sum_{i \neq j} \mathbb{E}(Y_i Y_j) = n(n-1)\left(\frac{n-2}{n}\right)^m.$$

Finally, we plug in to see that

$$\mathrm{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = n\left(\frac{n-1}{n}\right)^m + n(n-1)\left(\frac{n-2}{n}\right)^m - \left(n\left(\frac{n-1}{n}\right)^m\right)^2.$$

2. **Distributions** (16 point: 2/2/2/2/2/2/2/2)

For the following scenarios, what distribution would best model each one of them individually? You may choose from Poisson, geometric, exponential, uniform or normal distribution. Provide a brief justification of your answer.

(a) Amount of time you need to wait until a fly enters through your window.

(b) Number of shooting stars seen on a given night.

(c) Average height of students in CS70.

(d) Angle between two needles when we spin them at random about a common point.

(e) Number of times you have to refresh a webpage until it loads.

(f) Average grade in a CS70 exam.

(g) Number of times a web server is accessed per minute.

(h) Amount of time until the next time your telephone rings.

Answer:

(a) Exponential distribution. A reasonable argument is that the amount of time until a fly enters through the window is a memoryless variable. If we wait 5 minutes and see no fly, the distribution on the amount of time we still have to wait should not change. Therefore the exponential distribution is the correct answer here.

(b) Poisson distribution. Moments in time where we see a shooting star form what is called a Poisson process. Very reasonably one can argue that whether or not we see a shooting star at any sufficiently small interval of time (like 1 second) is akin to a coin toss with very small probability of success. So if we divide the night into many many small intervals, the number of shooting stars becomes like a binomial distribution with a large number of trials and a small probability of success for each trial. And we know that these tend to the Poisson process as the mean is kept constant. The mean here is just the expected number of shooting stars, which does not depend on the number of intervals we break the night into.

(c) Normal distribution. The average height of students is an average of many many (turn your head around the class if you do not believe this) random variables. If the height of any particular student comes from a distribution $\mathscr{D}$, then we are taking the average of many independent and identically distributed random variables. The central limit theorem states that this average becomes more and more like a normal random variable as the number of students grows. It is noteworthy that even the height of a single student $\mathscr{D}$ very closely follows a normal distribution likely because of the fact that many independent factors affect the height (genetics, nutrition, etc.).

(d) Uniform distribution. Let us assume that the first needle stops parallel to the x-axis. Then the second needle can stop at any angle with equal probability. Therefore the angle it makes with the x-axis would follow a uniform distribution. This argument would not change if the first needle was stopped at any other angle, and the uniform distribution would remain the same. Therefore the angle between the two needles follows the uniform distribution.

(e) Geometric distribution. Each time we refresh, whether or not the webpage loads can be thought of as a coin flip. The number of times we flip the coin until we see heads (a page load) is the definition of the geometric distribution.

(f) Normal distribution. It's fair to assume that the grade of each student follows some distribution $\mathscr{D}$ fairly independently of the other students. Therefore when we take the average of many of these grades, according to the central limit theorem, we get closer and closer to the normal distribution.

(g) Poisson distribution. We can divide the time into minuscule intervals and see whether the web server was accessed in each interval. If the intervals become short enough, the probability of more than one access in one interval becomes negligible. Now in each interval whether or not the server was accessed can be thought of as a coin flip. Therefore we have a sum of coin flips (i.e. the binomial distribution) whose mean is fixed. As the intervals become shorter and shorter, the binomial distribution becomes more and more like a Poisson distribution.

(h) Exponential distribution. The amount of time until the telephone rings is a memoryless property. Knowing that no one has called yet, should not change the distribution on the amount of time we still have to wait. We can also reasonably model this using the following method: divide the time horizon into tiny intervals and in each interval flip a biased coin to see whether the phone rings. As the intervals become shorter, the probability of each coin flip has to become smaller too, to maintain the average rate of telephone calls per unit of time. This is exactly how the exponential distribution is defined.

3. **Poisson**(12 points: 6/6)

   (a) It is fairly reasonable to model the number of customers entering a shop during a particular hour as a Poisson random variable. Assume that this Poisson random variable X has mean λ. Suppose that whenever a customer enters the shop they leave the shop without buying anything with probability p. Assume that customers act independently, i.e. you can assume that they each simply flip a biased coin to decide whether to buy anything at all. Let us denote the number of customers that buy something as Y and the number of them that do not buy anything as Z (so $X = Y + Z$). What is the probability that $Y = k$ for a given k? How about $\Pr[Z = k]$? Prove that Y and Z are Poisson random variables themselves.
   Hint: you can use the identity $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.

   (b) Assume that you were given two independent Poisson random variables X_1, X_2. Assume that the first has mean λ_1 and the second has mean λ_2. Prove that $X_1 + X_2$ is a Poisson random variable with mean $\lambda_1 + \lambda_2$.

Answer:

(a) We consider all possible ways that the event $Y = k$ might happen: namely, $k + j$ people enter the shop ($X = k + j$) and then exactly k of them choose to buy something. That is,

$$
\begin{aligned}
\Pr[Y = k] &= \sum_{j=0}^{\infty} \Pr[X = k + j] \cdot \Pr[Y = k \mid X = k + j] \\
&= \sum_{j=0}^{\infty} \left(\frac{\lambda^{k+j}}{(k+j)!} e^{-\lambda} \right) \cdot \left(\binom{k+j}{k} p^j (1-p)^k \right) \\
&= \sum_{j=0}^{\infty} \frac{\lambda^{k+j}}{(k+j)!} e^{-\lambda} \cdot \frac{(k+j)!}{k!j!} p^j (1-p)^k \\
&= \frac{(\lambda(1-p))^k e^{-\lambda}}{k!} \cdot \sum_{j=0}^{\infty} \frac{(\lambda p)^j}{j!} \\
&= \frac{(\lambda(1-p))^k e^{-\lambda}}{k!} \cdot e^{\lambda p} \\
&= \frac{(\lambda(1-p))^k e^{-\lambda(1-p)}}{k!}
\end{aligned}
$$

Hence, Y follows the Poisson distribution with parameter $\lambda(1-p)$. The case for Z is completely analogous: $\Pr[Z = k] = \frac{(\lambda p)^k e^{-\lambda p}}{k!}$ and Z follows the Poisson distribution with parameter λp.

(b) Let X, Y, and Z be defined as in part (a), with $\lambda = \lambda_1 + \lambda_2$ and $p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$. Then Y and X_1 follow the Poisson distribution with parameter λ_1, and Z and X_2 follow the Poisson distribution with parameter λ_2. Moreover, Y and Z are independent because

$$
\begin{aligned}
\Pr[Y = k, Z = j] &= \Pr[X = k + j] \cdot \Pr[Y = k \mid X = k + j] \\
&= \frac{\lambda^{k+j}}{(k+j)!} e^{-\lambda} \cdot \frac{(k+j)!}{k!j!} p^j (1-p)^k \\
&= \frac{(\lambda(1-p))^k e^{-\lambda(1-p)}}{k!} \cdot \frac{(\lambda p)^j e^{-\lambda p}}{j!} \\
&= \Pr[Y = k] \cdot \Pr[Z = j].
\end{aligned}
$$

Hence, $Y+Z$ must have the same distribution as X_1+X_2. We complete the proof by observing that $X=Y+Z$ is a Poisson random variable with mean $\lambda=\lambda_1+\lambda_2$.

4. **Bound It** (6 points)

   A random variable X is always strictly larger than -100. You know that $E[X]=-60$. Give the best upper bound you can on $P(X\geq -20)$. (Hint: think of the information you are given and the information you require to compute certain bounds)

   Answer: Notice that we do not have the variance of X, so Chebyshev's bound is not applicable here. We know nothing else about it's distribution so we cannot evaluate $E[e^{sX}]$ and so Chernoff bounds are not available. Since X is also not a sum of other random variables, other bounds or approximations are not available. This leaves us with just Markov's Inequality. But Markov Bound only applies on a nonnegative random variable, whereas X can take on negative values.

   This suggests that we want to "shift" X somehow, so that we can apply Markov's Inequality on it. Define a random variable $Y=X+100$, which means Y is strictly larger than 0, since X is always strictly larger than -100. Then, $E[Y]=E[X+100]=E[X]+100=-60+100=40$. Finally, the upper bound on X that we want can be calculated via Y, and we can now apply Markov's Inequality on Y since Y is strictly positive.

$$P(X\geq -20)=P(Y\geq 80)\leq \frac{E[Y]}{80}=\frac{40}{80}=\frac{1}{2}$$

   Hence, the best upper bound on $P(X\geq -20)$ is $\frac{1}{2}$.

5. **LLN** (18 points: 3/5/5/5)

   The Law of Large Numbers holds if, for all $\varepsilon>0$,

$$\lim_{n\to\infty}\Pr\left[\left|\tfrac{1}{n}S_n-\mathbb{E}(\tfrac{1}{n}S_n)\right|>\varepsilon\right]=0.$$

   In class, we see the proof of the Law of Large Numbers for $S_n=X_1+\cdots+X_n$, where the X_i's are i.i.d. random variables. This problem explores if the Law of Large Numbers holds under other circumstances.

   You are spinning a roulette wheel to collect n coupons. On each spin, you have a winning probability of $1-p$ and the outcomes of different spins are independent. If you win, you get a certain amount of coupons; if you lose, you get nothing.

   For each of the following coupon rewarding schemes, determine whether the Law of Large Numbers holds when S_n is defined as the total number of coupons that you win out of the n coupons. Answer YES if the Law of Large Number holds, or NO if not, and give a brief justification of your answer. (Whenever convenient, you can assume that n is even.)

   (a) YES or NO: You spin the wheel n times. One each spin, if you win, you will get 1 coupon.

   Answer: Yes. Define X_i to be 1 if you win a coupon on the ith spin. Then X_i, $i=1,\dots,n$ is 0 with probability p and 1 otherwise. We have a total of n spins. The total number of coupons you get is hence $S_n=X_1+\cdots+X_n$. Since S_n is a sum of i.i.d. Bernoulli random variables, $S_n\sim\text{Binomial}(n,1-p)$.

Now similar to notation in the lecture slide, we define $Y_n = \frac{S_n}{n}$ to be the fraction of coupons you collected out of the n coupons. Moreover, for each X_i,

$$\mathbb{E}(X_i) = 1 - p$$

and

$$\mathrm{Var}(X_i) = p(1-p).$$

Using Chebyshev's inequality,

$$\begin{aligned}&\Pr[\,|Y_n - E[Y_n]| > \varepsilon]\\ =&\Pr[\,|Y_n - (1-p)| > \varepsilon] \le \frac{\mathrm{Var}(Y_n)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty.\end{aligned}$$

(b) YES or NO: You spin the wheel $\frac{n}{2}$ times. Each time you win, you get 2 coupons.

Answer: Yes. Now we have $\frac{n}{2}$ spins. Similarly to the previous question, we define $X_i, i = 1, \ldots, \frac{n}{2}$ to be 0 with probability p and 2 (coupons) otherwise. Now the total number of coupons collected is $S_n = X_1 + \cdots + X_{\frac{n}{2}}$ and the fraction of coupons collected is $Y_n = \frac{S_n}{n}$.

Now for each $i = 1, \ldots, \frac{n}{2}$

$$E[X_i] = 2(1-p)$$

and

$$Var[X_i] = 4p(1-p).$$

Thus,

$$E[Y_n] = \frac{E[X_1] + \ldots + E[X_{\frac{n}{2}}]}{n} = \frac{1}{n} \cdot \frac{n}{2} \cdot 2(1-p) = 1-p$$

and

$$\mathrm{Var}[Y_n] = \frac{1}{n^2}\left(\mathrm{Var}[X_1] + \ldots + \mathrm{Var}[X_{\frac{n}{2}}]\right) = \frac{1}{n^2} \cdot \frac{n}{2} 4p(1-p) = \frac{2p(1-p)}{n}.$$

Finally, we get

$$\begin{aligned}&Pr[\,|Y_n - E[Y_n]| > \varepsilon]\\ =&Pr[\,|Y_n - (1-p)| > \varepsilon] \le \frac{2p(1-p)}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty.\end{aligned}$$

(c) YES or NO: You spin the wheel 2 times. Each time you win, you get $\frac{n}{2}$ coupons.

Answer: No. In this situation, we have

$$X_i = \begin{cases} 0 & \text{with probability } p \\ \frac{n}{2} & \text{with probability } (1-p) \end{cases}$$

for $i = 1, 2$. Now $S_n = X_1 + X_2$ and $Y_n = \frac{X_1 + X_2}{2}$.
We have

$$E[X_i] = \frac{n}{2}(1-p)$$

and

$$Var[X_i] = \frac{n^2}{4}p(1-p).$$

Thus,

$$E[Y_n] = \frac{E[X_1]+E[X_2]}{n} = \frac{1}{n}n(1-p) = 1-p$$

and

$$\text{Var}[Y_n] = \frac{1}{n^2}(\text{Var}[X_1]+\text{Var}[X_2]) = \frac{1}{n^2}\cdot\frac{n^2}{2}p(1-p) = \frac{p(1-p)}{2}.$$

Finally, we get

$$Pr[|Y_n - E[Y_n]| > \varepsilon]$$
$$= Pr[|Y_n - (1-p)| > \varepsilon] \le \frac{p(1-p)}{2\varepsilon^2}$$

that does not converge to 0 as $n \to \infty$, so the Law of Large Numbers does not hold.

(d) YES or NO: You spin the wheel once. If you win, you will get n coupons.

Answer: No. $S_n = X_1$, where $X_1 = n$ with probability $1-p$ and $X_1 = 0$ with probability p. $Y_n = \frac{X_1}{n}$.

Thus,

$$E[Y_n] = \frac{E[X_1]}{n} = \frac{n(1-p)}{n} = 1-p$$

and

$$\text{Var}[Y_n] = \frac{1}{n^2}\text{Var}[X_1] = \frac{1}{n^2}\cdot n^2 p(1-p) = p(1-p).$$

The inequality results in

$$Pr[|Y_n - E[Y_n]| > \varepsilon]$$
$$= Pr[|Y_n - (1-p)| > \varepsilon] \le \frac{p(1-p)}{\varepsilon^2}.$$

Same as before, this does not converge to 0 as $n \to \infty$, and the LLN does not hold.

For problems (c) and (d), you should've had the intuition that since the you only spin the wheel once or twice, increasing n does not really help for LLN.

6. **Packets Sending** (8 points: 2/6)

Assume Alice is trying to send m packets across a noisy channel to her friend Bob. The channel has probability $1-p$ of dropping each packet. To account for the loss, Alice sends $n > m$ packets. Alice can send at most n packets, and she needs to ensure that Bob can receive at least m packets with probability at least r. She wants to figure out how big can m be.

(a) Modeling each successfully sent packet as a coin toss with probability p, what is the probability that Bob receive at least m packets?

Answer: Denoting X as the number of successfully sent packets in the n packets, and each packet is sent successfully with probability p. Then $X = \text{Binomial}(n,p)$ with $E[X] = np$ and $Var(X) = np(1-p)$. We want

$$\Pr[\mathrm{X} \ge \mathrm{m}] = \sum_{k=m}^{n}\binom{n}{k}p^k(1-p)^{n-k}.$$

(b) Assume $n = 100$, $r = 0.9$, and $p = 0.9$. What is the upperbound for m using the Chernoff bound? (use the Chernoff bound: $\Pr[X \leq (1-\delta)\mu] \leq e^{-\frac{\mu\delta^2}{2}}$)

Answer: We are interested in $\Pr[X < m]$. We know that $\Pr[X < m] = \Pr[X \leq m-1]$. To use the Chernoff bound, we set $(1-\delta)\mu = m-1$, we know that $\mu = E[X] = np$. Therefore we can solve for δ.

$$\begin{aligned}
(1-\delta)\mu &= m-1 \\
(1-\delta)np &= m-1 \\
\delta &= 1 - \frac{m-1}{np} \\
\delta &= \frac{np+1-m}{np}
\end{aligned}$$

Hence, from Chernoff bound we know that

$$\begin{aligned}
&\Pr[X \leq m-1] \\
&\leq e^{-\frac{\mu\delta^2}{2}} \\
&= e^{-\frac{(np+1-m)^2}{2np}} \\
&= e^{-\frac{(91-m)^2}{180}}
\end{aligned}$$

Since we want $\Pr[X \leq m-1]$ to be no greater than $1-r = 0.1$, we have

$$\begin{aligned}
e^{-\frac{(91-m)^2}{180}} &\leq 0.1 \\
10 &\leq e^{\frac{(91-m)^2}{180}} \\
180\ln 10 &\leq (91-m)^2 \qquad (0 \leq m \leq 100) \\
m &\leq 91 - \sqrt{180\ln 10} \\
m &\leq 70.642
\end{aligned}$$

So an upperbound on m is 70.

7. **Midterm Review** (10 points)

Alice and Bob are going to study for the upcoming midterm together. They agree to meet at time t this afternoon. Alice will show up X hours after t, where $X \in \text{Uniform}[0,2]$. Bob's arrival time is more unpredictable. He will be distracted by Pokemon Go and will show up Y hours after t, where $Y \in \text{Expo}(1)$. The person who shows up later is late for T hours. What is $E[T]$? (Hint: some useful integrals $\int x^2 e^{-x} dx = -e^{-x}(x^2+2x+2)+c$ and $\int -xe^{-x} dx = xe^{-x} + e^{-x} + c$)

Answer: First notice that $T = \max(X,Y)$. Therefore, for $u \in [0,2]$

$$F_T(u) = Pr[T \leq u] = Pr[X \leq u] \times Pr[Y \leq u] = \frac{u}{2}(1-e^{-u})$$

for $u \in (2,\infty)$

$$F_T(u) = Pr[T \leq u] = Pr[X \leq u] \times Pr[Y \leq u] = 1 - e^{-u}$$

The pdf of T is obtained by taking the derivative of cdf.

$$f_T(u) = \frac{u}{2}e^{-u} - \frac{e^{-u}}{2} + \frac{1}{2}, u \in [0,2]$$

$$f_T(u) = e^{-u}, u \in (2,\infty)$$

Hence,

$$\begin{aligned}
&E[T] \\
=&\int_0^\infty u f_t(u) \\
=&\int_0^2 \frac{u}{2}(ue^{-u} - e^{-u} + 1)du + \int_2^\infty ue^{-u}du \\
=&\frac{1}{2}[-u^2e^{-u} - ue^{-u} - e^{-u} + \frac{x^2}{2}]_0^2 + [-ue^{-u} - e^{-u}]_2^\infty \\
=&\frac{3}{2} - \frac{1}{2e^2} \\
\approx&1.43
\end{aligned}$$

8. **CIA** (8 points)

Jason Bourne has been held captive in a prison from which there are three possible routes to escape: an air duct, a sewer pipe and the door (which happens to be unlocked). The air duct leads him on a three hour trip whereupon he falls through a trap door onto his head. The sewer pipe is similar but takes two hours to traverse. Each fall produces amnesia and he is returned to the cell immediately after each fall. Assume that he always immediately chooses one of the three exits from the cell with probability $\frac{1}{3}$. On average, how long does it take before he opens the unlocked door and escapes?

Answer:

Solution: Due to the memorylessness of the scenario, i.e., what Bourne chose in the previous attempt has nothing to do with his current attempt and all the following attempts, it is like starting all over again, as every attempt is just like his first attempt. Let X be the random variable of the time Jason Bourne needs to escape. We can start by considering the outcome of Bourne's first attempt, and see how the expected time to escape, $E[X]$, changes as a result of this. We have

$$E[X] = E[X|A]Pr[A] + E[X|S]Pr[S] + E[X|D]Pr[D] = \frac{1}{3}(E[X|A] + E[X|S] + E[X|D])$$

where $E[X|A]$ means the expected time to escape given that Bond went through the AC-duct in his first attempt, and similarly for $E[X|S]$ and $E[X|D]$. Now we take a closer look at these conditional expectations. Clearly, $E[X|D] = 0$ because Bourne escapes immediately. Meanwhile, we have $E[X|A] = 3 + E[X]$ because given that Bourne chose AC-duct in his first attempt, he wastes 3 hours and have to try again (which triggers another completely fresh attempt with no memory and thus takes E[X] hours to escape). Similarly, we have $E[X|S] = 2 + E[X]$. So now, we have derived a recursive relation for the expected time to escape:

$$E[X] = \frac{1}{3}(3 + E[X] + 2 + E[X] + 0)$$

Solving for $E[X]$, we get that $E[X] = 5$

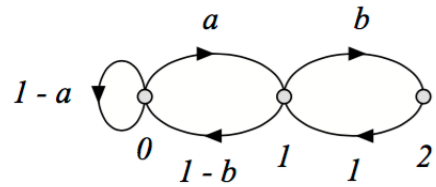

Figure 1: State diagram

9. **Markov Chain** (12 points: 4/3/5)

Consider the Markov chain $X(n)$ with the state diagram shown below, where $a, b \in (0,1)$.

(a) Is this Markov chain irreducible? Is it aperiodic? Briefly justify your answers.

Answer: The Markov chain is irreducible because $a, b(0,1)$. Also, $P(0,0) > 0$, so that g.c.d.$n > 0|Pn(0,0) > g.c.d.1,2,3,... = 1$, which shows that the Markov chain is aperiodic.

(b) Calculate $\Pr[X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 1 | X_0 = 0]$.

Answer: We see that the probability is $P(0,1)P(1,0)P(0,0)P(0,1) = a(1-b)(1-a)a$.

(c) Calculate the invariant distribution. **Answer:** The balance equations are

$$\pi(0) = (1-a)\pi(0) + (1-b)\pi(1)$$
$$\pi(1) = a\pi(0) + \pi(2).$$

After some simple manipulations, we see that they imply the following equations:

$$a\pi(0) = (1-b)\pi(1)$$
$$b\pi(1) = \pi(2)$$

These equations express the equality of the probability of a jump from i to $i+1$ and from $i+1$ to i, for $i = 0$ and $i = 1$, respectively. These relations are called the detailed balance equations. From these equations we find successively that

$$\pi(1) = \frac{a}{1-b}\pi(0) \text{ and } \pi(2) = b\pi(1) = \frac{ab}{1-b}\pi(0)$$

The normalization equation is

$$\begin{aligned} 1 &= \pi(0) + \pi(1) + \pi(2) \\ &= \pi(0)\left(1 + \frac{a}{1-b} + \frac{ab}{1-b}\right) \\ &= \pi(0)\frac{1-b+a+ab}{1-b} \end{aligned}$$

so that

$$\pi(0) = \frac{1-b}{1-b+a+ab}$$

Thus,

$$\pi = \frac{1}{1-b+a+ab}[1-b \quad a \quad ab].$$