

# STAT 110 HW 10

Moses Stewart

November 2022

## 0.1 Problem 1

We are given that  $P(L(X) < \theta < U(X)) = 0.95$ , and that there are  $n$  frequentists where the  $j^{\text{th}}$  frequentist observes data  $X_j$  and makes a confidence interval for the parameter  $\theta_j$ . We want to show that as  $n \rightarrow \infty$ , the fraction of the intervals which contain the parameter  $\rightarrow 0.95$ .

Now, let:

$$I_j = I(L(X_j) < \theta_j < U(X_j))$$

for the  $j^{\text{th}}$  frequentist, where  $P(I_j) = 0.95$  based on the condition given in the problem.

Additionally, let:

$$R_n = I_1 + I_2 + \dots + I_n$$

count the number of frequentists whose confidence interval contained  $\theta_j$ . Since the indicators are bernoulli with a probability of success  $P(I_j) = 0.95$ , we know that  $R_n \sim \text{Bin}(n, 0.95)$ .

This means that by definition, the expectation of  $R_n$  is given by:

$$E(R_n) = 0.95n$$

and the sample mean of each  $I_1, I_2, \dots, I_n$  is given by:

$$\bar{I}_j = \frac{R_n}{n}$$

Because  $R_n$  is the sum of  $I_1, I_2, \dots, I_n$  i.i.d. indicators with  $P(I_j) = 0.95$ , the law of large numbers tells us that with probability 1, the sample mean for each  $\bar{I}_j = \frac{R_n}{n} \rightarrow 0.95 = P(I_j)$  as  $n \rightarrow \infty$ .

## 0.2 Problem 2

### 0.2.1 Part a

We are given that a hedge fund manager always invests half of her fortune into the stock each day.  $Y_n$  is her fortune after  $n$  days, starting from an initial fortune of  $Y_0 = 100$ . We are asked what happens to  $Y_n$  as  $n \rightarrow \infty$ .

Let  $U_n$  be the number of times the stock increases in a period of  $n$  days. Because the hedge fund manager invests half of her fortune each day, on day  $n = 1$  if the stock price increases:

$$Y_1 = \frac{Y_0}{2} + \frac{Y_0}{2}(1.7) = 1.35Y_0$$

If the stock price decreases:

$$Y_1 = \frac{Y_0}{2} + \frac{Y_0}{2}(0.5) = 0.75Y_0$$

This means that every time the stock goes up, her fortune increases by 1.35 fold, and every time the stock goes down her fortune increases by 0.75 fold. Therefore, we can write a general equation:

$$Y_n = Y_0(1.35)^{U_n}(0.75)^{n-U_n}$$

We can simplify this using exponent rules to find that:

$$Y_n = Y_0 \left[ \left( \frac{1.35}{0.75} \right)^{\frac{U_n}{n}} 0.75 \right]^n = Y_0 \left[ (1.8)^{\frac{U_n}{n}} 0.75 \right]^n$$

To find what happens to  $Y_n$  as  $n \rightarrow \infty$ , we can take the limit of each side:

$$\lim_{n \rightarrow \infty} Y_n = \lim_{n \rightarrow \infty} Y_0 \left[ (1.8)^{\frac{U_n}{n}} 0.75 \right]^n$$

By the law of large numbers, because  $\frac{U_n}{n}$  is the sample mean of how many times the stock goes up, as  $n \rightarrow \infty$ ,  $\frac{U_n}{n} \rightarrow \frac{1}{2}$

$$\Rightarrow \lim_{n \rightarrow \infty} Y_n = \lim_{n \rightarrow \infty} Y_0 \left[ (1.8)^{\frac{1}{2}} 0.75 \right]^n$$

$$\lim_{n \rightarrow \infty} Y_n = \lim_{n \rightarrow \infty} Y_0 [1.006]^n$$

$$\lim_{n \rightarrow \infty} Y_n = \infty$$

### 0.2.2 Part b

We are given that the hedge fund manager invests a fraction  $\alpha$  of her current fortune each day, and are asked to find the function  $g(\alpha)$  such that:

$$\frac{\log Y_n}{n} \rightarrow g(\alpha)$$

with probability 1 as  $n \rightarrow \infty$ .

First, to find a function for what  $\frac{\log Y_n}{n}$  approaches as  $n \rightarrow \infty$  in terms of  $\alpha$ , we first want to find a function  $Y_n$  in terms of  $\alpha$ .

By definition, if her stock goes up, the hedge fund manager will gain  $\alpha \times 1.7$  her current fortune, and the remaining  $(1 - \alpha)$  portion of her fortune that she didn't invest will remain. Therefore, every time her stock increases her fortune will increase by:

$$(1.7)\alpha + (1 - \alpha) = 0.7\alpha + 1$$

By the same logic, if her stock goes down, the hedge fund manager will lose the  $\alpha \times 0.5$  her current fortune, and the remaining  $(1 - \alpha)$  portion of her fortune that she didn't invest will remain. Therefore, every time her stock increases her fortune will decrease by:

$$(0.5)\alpha + (1 - \alpha) = -0.5\alpha + 1$$

Plugging this into our original equation for  $Y_n$  in place of the ratios we had for part a, we find that:

$$Y_n = Y_0(0.7\alpha + 1)^{U_n}(-0.5\alpha + 1)^{n - U_n}$$

$$\log Y_n = \log Y_0 + U_n \log(0.7\alpha + 1) + n \log(-0.5\alpha + 1) - U_n \log(-0.5\alpha + 1)$$

$$\log Y_n = \log Y_0 + U_n \log\left(\frac{0.7\alpha + 1}{-0.5\alpha + 1}\right) + n \log(-0.5\alpha + 1)$$

$$\frac{\log Y_n}{n} = \frac{\log Y_0}{n} + \frac{U_n}{n} \log\left(\frac{0.7\alpha + 1}{-0.5\alpha + 1}\right) + \log(-0.5\alpha + 1)$$

Now, to find  $g(\alpha)$  as  $n \rightarrow \infty$ , we can take the limit of both sides:

$$\lim_{n \rightarrow \infty} \frac{\log Y_n}{n} = \lim_{n \rightarrow \infty} \frac{\log Y_0}{n} + \lim_{n \rightarrow \infty} \frac{U_n}{n} \log\left(\frac{0.7\alpha + 1}{-0.5\alpha + 1}\right) + \log(-0.5\alpha + 1)$$

By the law of large numbers, because  $\frac{U_n}{n}$  is the sample mean of how many times the stock goes up, as  $n \rightarrow \infty$ ,  $\frac{U_n}{n} \rightarrow \frac{1}{2}$

$$\Rightarrow \lim_{n \rightarrow \infty} \frac{\log Y_n}{n} = 0 + \frac{1}{2} \log\left(\frac{0.7\alpha + 1}{-0.5\alpha + 1}\right) + \log(-0.5\alpha + 1)$$

$$\lim_{n \rightarrow \infty} \frac{\log Y_n}{n} = \frac{1}{2} \log(0.7\alpha + 1) + \frac{1}{2} \log(-0.5\alpha + 1)$$

$$\lim_{n \rightarrow \infty} \frac{\log Y_n}{n} = \frac{1}{2} \log[(0.7\alpha + 1)(-0.5\alpha + 1)]$$

Therefore:

$$\frac{\log Y_n}{n} \rightarrow g(\alpha) = \frac{1}{2} \log[(0.7\alpha + 1)(-0.5\alpha + 1)]$$

With probability 1 as  $n \rightarrow \infty$

A log function is strictly increasing over its entire domain, so to find where  $g(\alpha)$  is maximized, we can calculate where  $(0.7\alpha + 1)(-0.5\alpha + 1)$  is maximized. Doing so, we find:

$$g(\alpha) = \frac{1}{2} \log\left[-\frac{7}{20}\alpha^2 + \frac{1}{5}\alpha + 1\right]$$

$$g(\alpha) = \frac{1}{2} \log \left[ -\frac{7}{20} \left( \alpha^2 - \frac{4}{7} \alpha + \frac{4}{49} - \frac{144}{49} \right) \right]$$

$$g(\alpha) = \frac{1}{2} \log \left[ -\frac{7}{20} \left( \alpha - \frac{2}{7} \right)^2 + \frac{36}{35} \right]$$

The function  $-\frac{7}{20}(\alpha - \frac{2}{7})^2 + \frac{36}{35}$  is clearly a concave quadratic with a vertex at  $\frac{2}{7}$ , so we can conclude that that is where  $g(\alpha)$  is maximized.

### 0.3 Problem 3

#### 0.3.1 Part a

We are trying to estimate  $\theta$  based on data  $X_1, X_2, \dots, X_n$ , where  $T_1$  is a function of  $X_1, X_2, \dots, X_n$  used to estimate  $\theta$  and  $T_2 = E(T_1|R)$ . We are trying to show that:

$$E(T_2 - \theta)^2 < E(T_1 - \theta)^2$$

We can simplify the right side of the equation to find that:

$$E(T_1 - \theta)^2 = E(T_1^2 - 2T_1\theta + \theta^2)$$

$$E(T_1 - \theta)^2 = E(T_1^2) - E(2T_1\theta) + E(\theta^2)$$

In this problem, both 2 and  $\theta$  are treated as constants. Therefore:

$$E(T_1 - \theta)^2 = E(T_1^2) - 2\theta E(T_1) + \theta^2$$

Using Adam's law, we find:

$$E(T_1 - \theta)^2 = E(E(T_1^2|R)) - 2\theta E(E(T_1|R)) + \theta^2$$

We are given that  $T_2 = E(T_1|R)$ , so we know that  $T_2^2 = E(T_1|R)^2$ . Now,  $T_1^2$  is a convex function. Therefore, by Jensen's inequality, we can conclude that:

$$T_2^2 = E(T_1|R)^2 \leq E(T_1^2|R)$$

Therefore:

$$E(T_1 - \theta)^2 \geq E(T_2^2) - 2\theta E(T_2) + \theta^2$$

Again,  $\theta$  is just a constant, so we can pull it inside the inequality.

$$E(T_1 - \theta)^2 \geq E(T_2^2 - 2\theta T_2 + \theta^2)$$

$$E(T_1 - \theta)^2 \geq E(T_2 - \theta)^2$$

### 0.3.2 Part b

We are given that the *bias* of an estimator  $T$  for  $\theta$  is defined to be  $b(T) = E(T) - \theta$ . We also know that the *bias-variance tradeoff* is given by:

$$E(T - \theta)^2 = \text{Var}(T) + (b(T))^2$$

To give an alternative proof for part a, it suffices to show that the mean squared error  $E(T_1 - \theta)^2$  is greater than  $E(T_2 - \theta)^2$  using the equation above. First, notice that using Adam's law:

$$b(T_1) = E(T_1) - \theta$$

$$b(T_1) = E(E(T_1|R)) - \theta$$

$$b(T_1) = E(T_2) - \theta$$

$$b(T_1) = b(T_2)$$

Therefore,  $(b(T_1))^2 = (b(T_2))^2$ . This means to prove that  $E(T_1 - \theta)^2$  is greater than  $E(T_2 - \theta)^2$ , using the equation above, it suffices to show that:

$$\text{Var}(T_1) > \text{Var}(T_2)$$

Using EVE's law, we find that:

$$\text{Var}(T_1) = E(\text{Var}(T_1|R)) + \text{Var}(E(T_1|R))$$

By definition, we are given that  $E(T_1|R) = T_2$ , therefore:

$$\text{Var}(T_1) = E(\text{Var}(T_1|R)) + \text{Var}(T_2)$$

$$\text{Var}(T_1) - E(\text{Var}(T_1|R)) = \text{Var}(T_2)$$

By definition, variance is always positive, therefore  $E(\text{Var}(T_1|R))$  must be positive. Thus, for all  $T_1, T_2$ :

$$\text{Var}(T_1) > \text{Var}(T_2)$$

Now, because the mean squared error:  $E(T - \theta)^2 = \text{Var}(T) + (b(T))^2$ , and  $(b(T_1))^2 = (b(T_2))^2$ , this means that:

$$E(T_2 - \theta)^2 < E(T_1 - \theta)^2$$

### 0.3.3 Part c

We are given that  $X_1, X_2, \dots$  are i.i.d. with mean  $\theta$ ,  $T_1 = X_1$ , and  $R = \sum_{j=1}^n X_j$ . We are asked to find  $T_2$ , check that it has a lower mean squared error than  $T_1$  for  $n \geq 2$ , and say what happens to  $T_1$  and  $T_2$  as  $n \rightarrow \infty$ .

We are given that  $T_2 = E(T_1|R)$ , so to find  $T_2$  we need to find  $E(T_1|R)$ . We are also given that  $T_1 = X_1$ , therefore:

$$T_2 = E(T_1|R) = E(X_1|R)$$

To find  $E(X_1|R)$ , notice that  $R = \sum_{j=1}^n X_j$ . Therefore, we know by LOTUS that:

$$E(R) = \sum_{j=1}^n E(X_j)$$

$$R = \sum_{j=1}^n E(X_j|R)$$

Because  $X_j$  are i.i.d., this means that:

$$R = nE(X_1|R)$$

$$E(X_1|R) = \frac{R}{n} = T_2$$

Now that we found  $T_2$ , we want to check that it has a lower mean squared error than  $T_1$  for  $n \geq 2$ . Using the equation from part b, we find that:

$$E(T_2 - \theta)^2 = \text{Var}(T_2) + (E(T_2) - \theta)^2$$

We showed in part b that  $(E(T_2) - \theta)^2 = (E(T_1) - \theta)^2$ , therefore we can simplify this to:

$$E(T_2 - \theta)^2 = \text{Var}\left(\frac{R}{n}\right) + (E(T_1) - \theta)^2$$

$T_1$  is an estimator for  $\theta$ , therefore by definition  $E(T_1) = \theta$ , so  $(E(T_1) - \theta)^2 = 0$ . Thus, we find:

$$E(T_2 - \theta)^2 = \text{Var}\left(\frac{R}{n}\right)$$

$$E(T_1 - \theta)^2 = \text{Var}(T_1) = \text{Var}(X_1)$$

Now,  $n$  is a constant, so we can simplify the top equation to:

$$E(T_2 - \theta)^2 = \frac{\text{Var}(R)}{n^2}$$

Again, recall that  $R = \sum_{j=1}^n X_j$  where  $X_j$  are i.i.d. Therefore, using the properties of variance and covariance:

$$\text{Var}(R) = \sum_{j=1}^n \text{Var}(X_j)$$

$$\text{Var}(R) = n\text{Var}(X_1)$$

$$\Rightarrow E(T_2 - \theta)^2 = \frac{n\text{Var}(X_1)}{n^2}$$

$$E(T_2 - \theta)^2 = \frac{\text{Var}(X_1)}{n}$$

$n$  is always a positive constant, therefore:

$$\frac{\text{Var}(X_1)}{n} < \text{Var}(X_1)$$

Therefore because  $E(T_2 - \theta)^2 = \frac{\text{Var}(X_1)}{n}$  and  $E(T_1 - \theta)^2 = \text{Var}(X_1)$  we can conclude that:

$$E(T_2 - \theta)^2 < E(T_1 - \theta)^2$$

Finally,  $T_1$  is defined to be  $X_1$ , so as  $n \rightarrow \infty$   $T_1 \rightarrow X_1$  because  $X_1$  does not depend on  $n$ .

However,  $T_2 = \frac{R}{n}$ , which is the sample mean for  $\sum_{j=1}^n X_j$ . Therefore, by the law of large numbers, as  $n \rightarrow \infty$ ,  $T_2 \rightarrow \mu$ , where  $\mu$  is the true mean of  $X_1, \dots, X_n$ .

## 0.4 Problem 4

### 0.4.1 Part a

We are given that each page of an  $n$ -age book has a  $\text{Pois}(\lambda)$  number of typos. The number of typos on each page are i.i.d.  $X_1, \dots, X_n \sim \text{Pois}(\lambda)$ . The probability that a page has no typos is given by  $P(X_j = 0) = \theta = e^{-\lambda}$ .

We are asked to show that  $T_n = e^{-\bar{X}_n}$  is biased for estimating  $\lambda$ , or that  $E(T_n) \neq \theta$ . By definition, we know that:

$$E(T_n) = E(e^{-\bar{X}_n})$$

Notice that  $E(e^{-\bar{X}_n})$  is a function of  $\bar{X}_n$ , and that  $\frac{d^2}{dx}e^{-x} = e^{-x}$ , therefore  $e^{-\bar{X}_n}$  is a convex function. Therefore, by Jensen's inequality, this tells us that:

$$E(T_n) = E(e^{-\bar{X}_n}) \geq e^{-E(\bar{X}_n)}$$

Because  $\bar{X}_n$  is the sample mean for i.i.d. variables with mean  $\lambda$ , we know that  $E(\bar{X}_n) = \lambda$ , therefore this gives us that:

$$E(T_n) \geq e^{-\lambda}$$

which shows that in general  $T_n$  is biased for estimating  $\theta$ .

### 0.4.2 Part b

$\bar{X}_n$  is the sample mean for i.i.d.  $X_i$ , and is given by:

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

By the law of large numbers, as  $n \rightarrow \infty$ ,  $\bar{X}_n \rightarrow \lambda = E(X_i)$ . Therefore, because  $T_n$  is given by:

$$T_n = e^{-\bar{X}_n}$$

By the law of large numbers, we can conclude that with probability 1, as  $n \rightarrow \infty$ :

$$T_n \rightarrow e^{-\lambda} = \theta$$

### 0.4.3 Part c

To show that  $W = \frac{1}{n}(I(X_1 = 0) + \dots + I(X_n = 0))$  is unbiased for  $\lambda$ , we want to show that  $E(W) = \lambda$ . Additionally, we are given that  $X_1|(X_1 + \dots + X_n = s) \sim \text{Bin}(s, \frac{1}{n})$ . Calculating  $E(W)$ :

$$E(W) = E\left(\frac{1}{n}(I_1 + \dots + I_n)\right)$$

$$E(W) = \frac{1}{n}E(I_1 + \dots + I_n)$$

$$E(W) = \frac{1}{n}[E(I_1) + \dots + E(I_n)]$$

$$E(W) = \frac{1}{n}[nP(I_1)]$$

We know that  $I_1$  is the event that  $X_1 = 0$ , therefore:

$$P(I_1) = P(X_1 = 0)$$

$$P(I_1) = e^{-\lambda}$$

Thus, we find:

$$E(W) = \frac{1}{n}[ne^{-\lambda}] = e^{-\lambda}$$

Therefore,  $E(W) = \theta$ , so  $W$  is unbiased for estimating  $\lambda$ .

Now, we are asked to find  $E(W|(X_1 + \dots + X_n = s))$ . Solving, we find:

$$E(W|(X_1 + \dots + X_n = s)) = E(I_1 + \dots + I_n|(X_1 + \dots + X_n = s))$$

$$E(W|(X_1 + \dots + X_n = s)) = E(I_1|(X_1 + \dots + X_n = s)) + \dots + E(I_n|(X_1 + \dots + X_n = s))$$

$$E(W|(X_1 + \dots + X_n = s)) = nP(I_1|(X_1 + \dots + X_n = s))$$

Here, we are given that  $X_1|(X_1 + \dots + X_n = s) \sim \text{Bin}(s, \frac{1}{n})$ .

$$P(X_1 = x|(X_1 + \dots + X_n = s)) = \binom{s}{x} \left(\frac{1}{n}\right)^x \left(\frac{n-1}{n}\right)^{s-x}$$

$$P(X_1 = 0|(X_1 + \dots + X_n = s)) = \left(\frac{n-1}{n}\right)^s$$

$$P(X_1 = 0|(X_1 + \dots + X_n = s)) = \left(\frac{n-1}{n}\right)^s$$

Recall that  $I_i$  is the indicator that page  $i$  has zero typos. Therefore:

$$P(I_1|(X_1 + \dots + X_n = s)) = \left(\frac{n-1}{n}\right)^s$$



Plugging this back in, we find:

$$E(W|(X_1 + \dots + X_n = s)) = \left(\frac{n-1}{n}\right)^s$$

Now, we are asked if  $\tilde{W} = E(W|(X_1 + \dots + X_n = s))$  is unbiased for  $\theta$ . Taking the expected value of both sides:

$$E(\tilde{W}) = E(E(W|(X_1 + \dots + X_n = s)))$$

This is precisely defined by Adam's law, so we can easily conclude that:

$$E(\tilde{W}) = E(E(W|(X_1 + \dots + X_n = s))) = E(W) = e^{-\lambda}$$

Therefore,  $E(\tilde{W}) = \theta$ , so  $\tilde{W}$  is also unbiased for  $\theta$ .

#### 0.4.4 Part d

We are asked to show that  $\tilde{W}$  has a lower variance than  $W$ .

Using Eve's law to calculate the Variance of  $W$ , we find:

$$\text{Var}(W) = E(\text{Var}(W|(X_1 + \dots + X_n = s))) + \text{Var}(E(W|(X_1 + \dots + X_n = s)))$$

By definition, we know that  $\tilde{W} = E(W|(X_1 + \dots + X_n = s))$ , thus we find:

$$\text{Var}(W) = E(\text{Var}(W|(X_1 + \dots + X_n = s))) + \text{Var}(\tilde{W})$$

By definition, we know that variance is always positive. Also,  $W|(X_1 + \dots + X_n = s)$  is a random variable without a certain outcome, so we can conclude that  $\text{Var}(W|(X_1 + \dots + X_n = s)) > 0$ . Thus,  $E(\text{Var}(W|(X_1 + \dots + X_n = s))) > 0$ .

Rewriting our equation above, we showed that:

$$\text{Var}(\tilde{W}) = \text{Var}(W) - k$$

Where  $k$  is a positive number,  $k = E(\text{Var}(W|(X_1 + \dots + X_n = s)))$ . Therefore, we can conclude that:

$$\text{Var}(\tilde{W}) < \text{Var}(W)$$

In the last question, we showed that the mean squared bias is lower for a variable  $\tilde{X} = E(X|R)$  than it is for  $X$ , and here we defined  $\tilde{W}$  and  $W$  in the same relationship.

Because of this, we know that  $\tilde{W}$  and  $W$  must share the same squared bias as well (from our results in 3.b), and can therefore conclude that the mean squared error for  $\tilde{W}$  must also be lower than the mean squared error for  $W$ . We would also intuitively expect the relationship from #3 to hold, because the variance is simply the average squared distance from the *true* mean.

## 0.5 Problem 5

### 0.5.1 Part a

We are given that  $X_1, X_2, \dots, X_n$  are i.i.d.  $\text{Expo}(1)$ , and that  $N = \min\{n : X_n \geq 1\}$ . We are asked to find the distribution of  $N - 1$ .

Intuitively,  $P(N = 1)$  is telling us the probability that the first variable is greater than 1:

$$P(N = 1) = P(X_1 \geq 1)$$

Furthermore,  $P(N = 2)$  is the same as the probability that the variable is less than 1 *and* the second variable is greater than 1:

$$P(N = 2) = P(X_1 \leq 1 \cap X_2 \geq 1)$$

$$P(N = 2) = P(X_1 \leq 1)P(X_2 \geq 1)$$

Similarly, for all  $P(N = n)$ , this is telling us the probability that all the  $X_{i < n}$  were less than 1, and  $X_n$  was greater than 1:

$$P(N = n) = P(X_1, X_2, \dots, X_{n-1} \leq 1 \cap X_n \geq 1)$$

$$P(N = n) = P(X_1 \leq 1) \times P(X_2 \leq 1) \times \dots \times P(X_{n-1} \leq 1) \times P(X_n \geq 1)$$

Because  $X_i \sim \text{Expo}(1)$ ,  $P(X_i \leq 1) = F_X(1) = 1 - e^{-1}$ , and similarly  $P(X_n \geq 1) = e^{-1}$ .

Therefore, we can think of all the  $X_i \leq 1$  as failures with probability  $1 - e^{-1}$ , and the first  $X_n \geq 1$  as a success with probability  $p = e^{-1}$ . Therefore,  $N - 1$  is simply a geometric distribution with probability  $p = e^{-1}$ .  $N - 1 \sim \text{Geom}(e^{-1})$ .

Thus,  $N \sim \text{FS}(e^{-1})$ , which has a mean of:  $E(N) = e$  and variance of:  $\text{Var}(N) = e^2 - e$ .

### 0.5.2 Part b

Letting  $M = \min\{n : X_1 + X_2 + \dots + X_n \geq 10\}$ , we are asked to find the distribution of  $M - 1$ .

Because  $X_1, X_2, \dots, X_n$  are i.i.d.  $\text{Expo}(1)$ ,  $M$  is the sum of  $n$   $\text{Expo}(1)$  distributions. Asking for the number of  $X_j$ 's we observe until their sum exceeds 10 is the same as asking at time  $t = 10$ , how many  $X_j$ 's did we observe and adding 1.

The adding 1 comes from the fact that if we observed  $k$   $X_j$ 's at time 10, then we needed  $k + 1$   $X_j$ 's for our sum to exceed 10.

By the definition of the Poisson process, the number of "successes," or  $X_j$ 's we see at time  $t$  is distributed as  $M - 1 \sim \text{Pois}(10)$ . Therefore,  $E(M) = 11$ .

### 0.5.3 Part c

For very large  $n$ , by the central limit theorem, the distribution of our sample mean will be approximately normal. Therefore, for large  $n$ :

$$\sqrt{n} \left( \frac{\bar{X}_n - \mu}{\sigma} \right) \sim \mathcal{N}(0, 1)$$

$$\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

$$\bar{X}_n \sim \mathcal{N}(1, \frac{1}{n})$$

To find the exact distribution of  $\bar{X}_n$ , notice that  $\bar{X}_n$  is the sum of  $n$  different  $\frac{X_i}{n}$  variables.

We know that:  $X_i \sim \text{Expo}(1)$ , therefore:

$$\frac{X_i}{n} \sim \text{Expo}(n)$$

This is because the exponential distribution tells us how long it takes for an event to occur, where on average  $\lambda$  events occur per unit time. Therefore, dividing the crystallized time by a number is the *same* as asking how much time would it take if an average of  $n\lambda$  events occurred per unit time.

Therefore,  $\bar{X}_n$  is the sum of  $n$  unique  $\text{Expo}(n)$  random variables. By the story of the gamma, the sum of  $n$  i.i.d. exponential functions is given by  $\text{Gamma}(n, \lambda)$ , therefore we can conclude that the exact distribution is:

$$\bar{X}_n \sim \text{Gamma}(n, n)$$

## 0.6 Problem 6

### 0.6.1 Part a

For a king, queen, rook, bishop, and knight, we are to determine if their movements on a chess board are irreducible, and if their movements are periodic.

#### 0.6.1.1 King

**Irreducible** Yes, clearly the king can move in every direction, so it can reach any square on the board.

**Aperiodic** Yes, it can reach its original square in 2 moves by moving left and right. It can reach it in 3 moves by moving left/right once, then moving up/down, then moving back to its original square.  $\text{GCD}(2, 3) = 1$ , so every square is aperiodic.

#### 0.6.1.2 Queen

**Irreducible** Yes, the queen is even more buff than the king.

**Aperiodic** Yes, the queen can do the same process as the king.

### 0.6.1.3 Rook

**Irreducible** Yes; the rook can reach every tile on a chessboard.

**Aperiodic** Yes, the rook can move reach its original square in 2 moves by moving left and right. It can reach it in 3 moves by moving sideways two squares, then moving back 1 square, then moving back 1 square again. Thus,  $\text{GCD}(2, 3) = 1$  for every square.

### 0.6.1.4 Bishop

**Irreducible** No; a bishop can't switch from a white square to a black square or vice-versa.

**Aperiodic** Yes, by doing the same thing as the Rook but diagonally.

### 0.6.1.5 Knight

**Irreducible** Yes; the Wikipedia link showed the various Knight's tours.

**Aperiodic** No; a knight always alternates between light and dark squares. This means that if you start on a dark square, you can only return in an even number of moves, so the gcd of each square is 2.

## 0.6.2 Part b

For this a rook is placed with an initial position chosen uniformly at random, and we are asked to find the distribution of where the rook is after  $n$  moves.

Because the Rook is being placed at random, its distribution after  $n$  moves will be determined by its steady state.

The rook is choosing *uniformly randomly* where to go, so its movements are an undirected network. This means that the steady state is proportional to the degree of each tile on the chess board.

For every tile on the chessboard, the rook can move to any of 7 tiles side-to-side, and any of 7 tiles up-and-down. This means every tile has degree 14. This tells us that each tile's stationary probability is:

$$14a \text{ for some } a$$

There are 64 tiles on a chessboard, so we know that:

$$\sum s_i = 64 * 14a$$

Therefore, for the sum to equal 1, we must have  $a = \frac{1}{64*7}$ . Thus, we find that  $a = \frac{1}{896}$ , which tells us:

$$s_i = \frac{14}{896} = \frac{1}{64}$$

for all  $s_i \in s$ .

Therefore,  $s = (\frac{1}{64}, \frac{1}{64}, \dots, \frac{1}{64})$ . Thus, we can easily conclude that if the rooks' position is chosen uniformly randomly, it has a uniform  $\frac{1}{64}$  chance of being on any tile on the chess board.

### 0.6.3 Part c

We know that if  $r_0$  is the expected time it takes the chain to return to A1, given that it starts at A1. Then:

$$r_0 = \frac{1}{s_0}$$

The kings movements is an undirected network, with degrees:

**Corners** = 3

**Sides** = 5

**Middle** = 8

Therefore:

$$s \propto (3a \times 4, 5a \times 24, 8a \times 36)$$

$$\sum s_i = 12a + 120a + 288a = 420a$$

$$\Rightarrow a = \frac{1}{420}$$

Then  $s_{A1} = \frac{3}{420} = \frac{1}{140}$ . Therefore:

$$r_0 = 140$$

### 0.6.4 Part d

Given that we have a wandering king, that:

1. Generates a *proposal square*  $j$  by picking uniformly at random among the legal moves from  $i$
2. Flips a coin with probability  $\min(d_i/d_j, 1)$  of Heads. If the coin lands Heads, goes to  $j$ . Otherwise, stays at  $i$ .

We are asked to show that it has a uniform stationary distribution.

Now, we can calculate the transition probabilities  $q_{ij}$  of our matrix  $Q$  by breaking the problem into two cases. If the degree of  $i$  is greater than  $j$ , then we simply have a uniform  $\frac{1}{d_i}$  chance of moving to  $j$ .

In the other case, if the degree of  $i$  is less than  $j$ , then there is a  $\frac{1}{d_i}$  chance of selecting tile  $j$ , then a  $\frac{d_i}{d_j}$  chance of the coin landing heads. Therefore, we find:

$$q_{ij} = \left\{ \begin{array}{ll} \frac{1}{d_i} & \text{if } d_i \geq d_j \\ \frac{1}{d_i} \times \frac{d_i}{d_j} & \text{if } d_i < d_j \end{array} \right\} \quad (1)$$

$$q_{ij} = \left\{ \begin{array}{ll} \frac{1}{d_i} & \text{if } d_i \geq d_j \\ \frac{1}{d_j} & \text{if } d_i < d_j \end{array} \right\} \quad (2)$$

Now, notice that in either case, if  $d_i \geq d_j$  or  $d_i < d_j$ , we can find that  $q_{ij} = q_{ji}$ :

**If**  $d_i \geq d_j$

$$q_{ij} = \frac{1}{d_i} \quad (3)$$

$$= \frac{1}{d_j} \times \frac{d_j}{d_i} \quad (4)$$

$$= q_{ji} \quad (5)$$

**If**  $d_i < d_j$

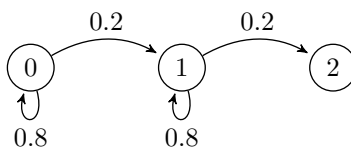
$$q_{ij} = \frac{1}{d_j} \quad (6)$$

$$= q_{ji} \quad (7)$$

Therefore, because  $q_{ij} = q_{ji}$ , any uniform vector  $s$  that is steady over the state space (i.e.  $s_i = s_j$  for all  $s, j$ ) will satisfy the reversibility condition. This means that  $s$  is a stationary uniform distribution.

## 0.7 Problem 7

We're asked to find the expectation and variance of the following markov chain:



The story of this markov chain is that once we progress from node 0 to node 1, we can't move back to node 0. Similarly, once we're at node 1, we stay at node 1 until we reach node 2.

Let  $T_i$  be the time spend at time  $i$ . It is easy to see that:

$$T = T_1 + T_2$$

Now, we can think of progressing from one node to the next as a "success" with probability  $p = 0.2$ . The amount of time, or trials, it takes to progress to the next node is the number of failures + the final success. Therefore,  $T_1, T_2 \sim \text{FS}(0.2)$ .

Thus, we can easily calculate  $E(T)$ :

$$E(T) = E(T_1 + T_2)$$

$$E(T) = E(T_1) + E(T_2)$$

$$E(T) = 5 + 5 = 10$$

Similarly, notice that  $T_1$  and  $T_2$  are independent; the number of times it took to get to node 1 doesn't affect how long it gets to node 2. Therefore:

$$\text{Var}(T) = \text{Var}(T_1 + T_2)$$

$$\text{Var}(T) = \text{Var}(T_1) + \text{Var}(T_2)$$

$$\text{Var}(T) = 20 + 20 = 40$$

