MTH 102: Probability and Statistics

End Semester Exam

03/08/2022

Sanjit K. Kaul

No books, notes, or devices are allowed. Just a pen and eraser. You have about 120 minutes.

Question 1. 35 **marks** A toddler is attempting to stand. An attempt fails with probability p. Answer the following questions.

(a) (5 marks) Calculate the average number of attempts up to and including the n^{th} time the toddler stands successfully, where $n \geq 1$.

(b) (10 marks) Calculate the variance of the number of attempts up to and including the n^{th} time the toddler stands successfully, where $n \geq 1$.

(c) (10 marks) Suppose you observe the toddler up to and including the toddler's N^{th} successful attempt at standing. Rewrite the average number of attempts in terms of N. Calculate the expected value of the resulting function for when N is a Poisson RV with PMF $P[N = n] = e^{-1}/n!$, $n \geq 0$.

[Hint: For (a) and (b), you may want to write the number of attempts as a sum of n iid random variables.]

Now suppose that the toddler learns from its attempts to stand. Specifically, the failure probability is halved after a successful attempt to stand. Till the first time the toddler stands, the failure probability is p as before. However, it becomes $p/2$ after the toddler stands successfully the first time and $p/4$ after the toddler stands successfully the second time, and so on.

(aa) (10 marks) Let N_2 be the RV that counts the number of attempts up to and including the attempt in which the toddler stands successfully the second time. Derive the PMF of N_2.

Question 2. 40 **marks** Two babies are born at the same time. Baby 1 is born into a chatty family and Baby 2 is born into a family that prefers quiet. Babies that are born in chatty families are expected to learn how to speak in a number of years uniformly distributed over $(0, 3)$. Babies born to quiet families are expected to learn to speak in a number of years uniformly distributed over $(0, 5)$. Answer the following questions.

1) (0 marks) For both the babies, name the random variables that count the number of years to speaking.
2) (20 marks) Calculate the probability that the baby born to the quiet family learns to speak sooner than the baby born to the chatty family.
3) (20 marks) Suppose it is known that the baby born to the quiet family learns to speak in greater than 2 years. Calculate the probability that the baby born to the quiet family learns to speak sooner than the baby born to the chatty family. [Hint: What is the PDF of interest? Marginal? Joint? Conditional? Some of the above?]

Question 3. 15 **marks** Consider the estimator $J_n(X)$ of the parameter $E[X]$. We have

$$J_n(X) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i + I_i W_i).$$

Here the random variables I_i, $i = 1, \ldots, n$, are independent and identically distributed and take a value of 0 or 1 with equal probability. We know that $E[W_i] = 0$ and $\text{Var}[W_i] = \sigma^2$, $i = 1, \ldots, n$. The random variables W_i, X_i, I_i, $i = 1, \ldots, n$ are independent random variables. Answer the following questions

1) (2.5 marks) Is the estimator unbiased? Arrive at your answer by calculating the expected value of the estimator.
2) (10 marks) Derive the mean squared error of the estimator. [Hint: Approach using first principles. If you need to calculate the second moment, first calculate the variance and use it to calculate the second moment.]
3) (2.5 marks) Is the estimator unbiased in the limit as $n \to \infty$? Calculate the mean squared error in the limit as $n \to \infty$.

Question 4. 10 **marks** Let X be a Binomial RV with number of trials $n = 100$ and the probability of success 0.4. Approximate the probability $P[X < 50]$ using the central limit theorem. Leave your answer as a function of the CDF $\Phi_Z(z)$ of the standard Normal.

(Q1)

Let X_1 be the no. of attempts to the first successful attempt to stand.

Let X_2 be the no. of attempts to the second successful attempt after the first succ attempt.

$\vdots$

Let X_i be the no. of attempts to the ith succ attempt after the $(i-1)^{th}$ succ attempt.

We have X_i, $i = 1, 2, \dots$

$$X_i \sim \text{Geom}(q), \text{ where } q = 1-p.$$

No of attempts up to & including the n^{th} succ attempt is

$$\sum_{i=1}^{n} X_i.$$

(a). We want

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

4 for appropriate sum.

$$= \sum_{i=1}^{n}\left(\frac{1}{q}\right)$$

$$= \frac{n}{q}.$$

5

Correct values 1.

If a student recognizes the Pascal distr & states $\frac{n}{p}$, that is fine too.

(b) We want

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right]$$

$$= \sum_{i=1}^{n} \mathrm{Var}[X_i]$$

6

2

X_i are independent

$\mathrm{Var}[X_i]$ is the variance of a geometric RV.

2

$$\mathrm{Var}[X_i] = \left(\frac{1-q}{q^2}\right) \quad \therefore$$

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \frac{n(1-q)}{q^2} \;//.$$

Recall $q = 1-p$.

(c) The average no. of attempts for a random N number of successes is

$$\frac{N}{q}.$$

5

$$E\left[\frac{N}{q}\right] = \frac{1}{q} E[N],$$

where $E[N]$ is the expected value of the Poisson RV with PMF $p[N=n] = \frac{e^{-1}}{n!}$, $n \geq 0$.

$$E[N] = \sum_{n=0}^{\infty} n \frac{e^{-1}}{n!} = e^{-1} + \frac{e^{-1}}{2!} + \cdots$$

$$= e^{-1} e = 1$$

4

$$\therefore E\left[\frac{N}{q}\right] = \frac{1}{q}.$$

1

(aa) We want the PMF of N_2.

Consider the RV N_1. Let N_1 be the no. of attempts to the first successful standing.

$$S_{N_1} = \{1, 2, \dots\}$$

$$S_{N_2} = \{2, 3, \dots\}$$

$$P[N_2 = 2] = P[N_1 = 1, N_2 = 2]$$

$$= (1-p)(1-p/2)$$

$$P[N_2 = 3] = P[N_1 = 1, N_2 = 3] + P[N_1 = 2, N_2 = 3]$$

$$= (1-p)(p/2)(1-p/2)$$

$$+ p(1-p)(1-p/2)$$

$\vdots$

$$P(N_2 = k) = \sum_{n_1=1}^{k-1} P[N_1 = n_1, N_2 = k]$$

Writing as a suitable sum 4

$$= \sum_{n_1=1}^{k-1} p^{n_1-1}(1-p)\,(p/2)^{k-n_1-1}(1-p/2)$$

Proper substitution 4

$\vdots$

So the PMF of N_2 is:

$$P\{N_2=k\} = \begin{cases} \displaystyle\sum_{n_1=1}^{k-1} p^{n_1-1}(1-p)\left(p/2\right)^{k-n_1-1}\left(1-p/2\right), & k=2,3,\dots \\ 0 & \text{otherwise.} \end{cases}$$

Final Answer. (2)

Q2

(1) Let X_1 be the no. of years to speaking for baby born to the chatty family.

Let X_2 be the no. of years to speaking for the baby born to the quiet family.

$$X_1 \sim \text{Unif}(0,3)$$

$$X_2 \sim \text{Unif}(0,5)$$

0 marks.

(2) The probability of interest is

$P[X_2 < X_1]$. To calculate this, we need the joint PDF $f_{X_1,X_2}(x_1,x_2)\ \forall x_1,x_2$

Assuming X_1 & X_2 are independent, we have:

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)\, f_{X_2}(x_2)$$

10

$$= \begin{cases} \left(\frac{1}{3}\right)\left(\frac{1}{5}\right) & 0 \le x_1 \le 3, \\ & 0 \le x_2 \le 5, \\ 0 & \text{otherwise.} \end{cases}$$

$$P\{X_2 < X_1\}$$

$$= \int_0^3 \int_0^{x_1} f_{X_1,X_2}(x_1,x_2)\, dx_2\, dx_1$$

10

$$= \frac{1}{15} \int_0^3 \left(\int_0^{x_1} dx_2 \right) dx_1$$

$$= \frac{1}{15} \int_0^3 x_1\, dx_1 = \frac{1}{15} \left(\frac{x_1^2}{2} \right]_0^3$$

$$= \frac{1}{15} \left(\frac{9}{2} \right)$$

$$= \frac{3}{10} \,//$$

(3) It is given that $X_2 > 2$.

We want to calculate

$$P[X_2 < X_1 \mid X_2 > 2].$$ Identifying the correct probability 2.5

We need the conditional joint PDF

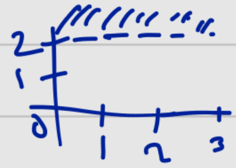
5

$$f_{X_1, X_2 \mid X_2 > 2}(x_1, x_2) = \begin{cases} \dfrac{f_{X_1, X_2}(x_1, x_2)}{P[X_2 > 2]} & 0 \le x_1 \le 3 \\ & 2 \le x_2 \le 5, \\ 0 & \text{otherwise.} \end{cases}$$

2.5 $$P[X_2 > 2] = \int_0^3 \int_2^5 f_{X_1, X_2}(x_1, x_2)\, dx_2\, dx_1$$

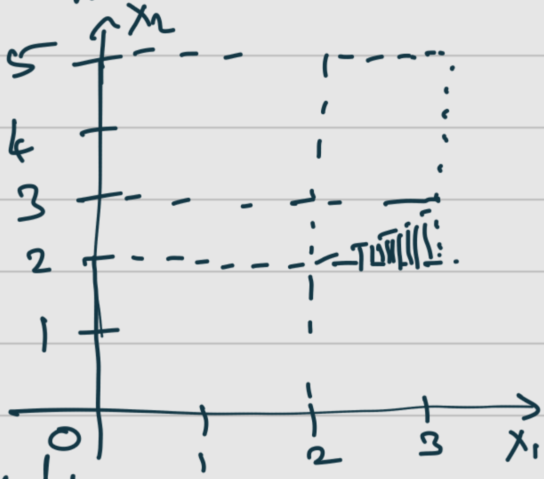
$$= \frac{1}{15} \int_0^3 3\, dx_1 = \frac{9}{15} = \frac{3}{5}.$$

A simpler way is to recognize that $X_2 \sim \text{Unif}(0, 5)$. You could use the marginal PDF.



2.5

$$f_{X_1,X_2|X_2>2}(x_1,x_2) = \begin{cases} \dfrac{(1/15)}{(3/5)} = \dfrac{1}{9} & 0 \le x_1 \le 3, \quad 2 \le x_2 \le 5, \\ 0 & \text{otherwise.} \end{cases}$$



7.5

$$P[X_2 < X_1 \mid X_2 > 2] = \int_2^3 \int_{x_2}^3 f_{X_1,X_2|X_2>2}(x_1,x_2)\, dx_1\, dx_2$$

$$= \frac{1}{9} \int_2^3 (3 - x_2)\, dx_2 = \frac{1}{9}\left(3 - \frac{1}{2}(9-4)\right)$$

$$= \frac{1}{9}(3 - 2.5) = \frac{1}{18} \;//$$

Q3 Given estimate

$$J_n(x) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i + I_i W_i)$$

(1)

$$E[J_n(x)] = E\left[\frac{1}{n-1} \sum_{i=1}^{n} (X_i + I_i W_i)\right]$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} E[X_i + I_i W_i]$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} E[X_i] + E[I_i W_i]$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} E[X_i] + E[I_i] E[W_i]$$

Since I_i & W_i are independent.

$$= \frac{1}{n-1} n E[X]$$

Since $E[W_i] = 0 \ \forall i$ & $X_i \sim X \ \forall i$.

2

We have

$$E[J_n(X)] = \frac{n}{n-1} E[X].$$ 0.5

Thus $J_n(X)$ is ~~an~~ ~~un~~biased estimator.

2 MSE is

$$E\left[\left(J_n(X) - E[X]\right)^2\right]$$ 2

$$= E\left[J_n^2(X) + E^2[X] - 2J_n(X)E[X]\right]$$

$$= E\left[J_n^2(X)\right] + E^2[X] - 2E[X]\,E[J_n(X)]$$

Calc of 6

$$= \boxed{E\left[J_n^2(X)\right]} + \boxed{E^2[X] - 2E^2[X]\frac{n}{n-1}}$$

$$= \mathrm{Var}[J_n(X)] + E^2[J_n(X)] + E^2[X] - 2E^2[X]\frac{n}{n-1}.$$

$$= Var\left[J_n(X)\right] + \frac{n^2}{(n-1)^2} E^2[X] + E^2[X]$$
$$- 2 E^2[X] \frac{n}{n-1}$$

$$Var\left[J_n(X)\right]$$
$$= Var\left[\frac{1}{n-1} \sum_{i=1}^{n} (X_i + I_i W_i)\right]$$
$$= \left(\frac{1}{n-1}\right)^2 \sum_{i=1}^{n} Var\left[X_i + I_i W_i\right]$$

(Since the RV(s) are mutually independent)

$$= \left(\frac{1}{n-1}\right)^2 \sum_{i=1}^{n} \left(Var[X_i] + Var[I_i W_i]\right)$$
$$= \frac{n\, Var[X_i]}{(n-1)^2} + \frac{1}{(n-1)^2} \underbrace{Var[I_i W_i]}_{\left(E[I_i^2 W_i^2] - E^2[I_i W_i]\right)}$$
$$= \frac{n\, Var[X]}{(n-1)^2} + \frac{1}{(n-1)^2} (0.5)\sigma^2.$$

$\therefore$ MSE is

$$\mathrm{Var}\left[J_n(X)\right] + \frac{n^2}{(n-1)^2} E^2[X] + E^2[X]$$
$$- 2E^2[X]\frac{n}{n-1}$$

$$= \frac{n\,\mathrm{Var}[X]}{(n-1)^2} + \frac{1}{(n-1)^2}(0.5)\sigma^2.$$
$$+ E^2[X]\left(\frac{n^2 + (n-1)^2 - 2n(n-1)}{(n-1)^2}\right)$$

$$= \frac{n\,\mathrm{Var}(X)}{(n-1)^2} + \frac{0.5\sigma^2}{(n-1)^2}$$
$$+ E^2[X]\left(\frac{1}{(n-1)^2}\right)$$

Final answer 2

(3) We calculated:

$$E\left[J_n(X)\right] = \frac{n}{n-1} E[X].$$

$$\lim_{n\to\infty} E\left[J_n(X)\right] = \lim_{n\to\infty} \frac{n}{n-1} E[X] = E[X]. \quad 1$$

The estimate is unbiased in the limit as $n \to \infty$.

The MSE is:

$$\frac{n \operatorname{Var}[X]}{(n-1)^2} + \frac{0.5\sigma^2}{(n-1)^2}$$

1.5

$$+ E^2[X]\left(\frac{1}{(n-1)^2}\right)$$

In the limit as $n \to \infty$, the MSE $\to 0$.

Q4

2

$E[X] = 40$; $\sigma_X^2 = (100)(0.4)(0.6) = 24$

$$P[X < 50] = P\left[X - E[X] < 50 - E[X]\right]$$

$$= P\left[\frac{X - E[X]}{\sigma_X} < \frac{50 - E[X]}{\sigma_X}\right]$$

5

P

$$\approx \Phi\left(\frac{50 - E[X]}{\sigma_X}\right)$$

3

$$= \Phi\left(\frac{10}{\sqrt{24}}\right)$$

Alternately,

$$P[X < 50] = P[0 \le X < 50]$$

$$3 = P\left[\frac{-E(X)}{\sigma_X} \le \frac{X - E(X)}{\sigma_X} < \frac{50 - E(X)}{\sigma_X}\right]$$

$$3 = \Phi_Z\left(\frac{50 - E(X)}{\sigma_X}\right) - \Phi_Z\left(\frac{-E(X)}{\sigma_X}\right)$$

8