MTH 102: Probability and Statistics

Quiz 4

26/07/2022

Sanjit K. Kaul

No books, notes, or devices are allowed. Just a pen and eraser. Any exchange of information related to the quiz with a human or machine will be deemed as cheating. Institute rules will apply. You have about 60 minutes.

Question 1. 45 **marks** Student performance in a course is a random variable that takes values in $(0, 100)$. Student attendance (percentage) is also a random variable that takes values in $(0, 100)$. Studies in a university show that student attendance is well modeled by the continuous uniform distribution. Also, a key finding is that student performance and attendance are dependent. Specifically, for a student with attendance a, the conditional distribution of performance is uniform over $(a, 100)$. Denote performance by the RV M and attendance by the RV A. Answer the following questions.

(a) (5 +5 marks) Write down the conditional PDF of performance M, given attendance.

(b) (10 marks) Derive the joint PDF and joint CDF of performance and attendance.

Consider two students. Suppose one of the students has an attendance of a_1. Let X be the random variable that models the performance of this student. Suppose the other student has an attendance of a_2. Let Y be the random variable that models the performance of this student. Assume that students attend lectures independently of others and their performance is also independent of others' performance.

(c) (5 +5 marks) Derive the expected value of the difference in performance of the first and the second student.

(d) (10 +5 marks) Derive $P[X > Y]$ for $a_2 > a_1$.

~~(e) (15 marks) Derive $E[X|X > Y]$ for $a_2 > a_1$~~.

Question 2. 25 **marks** You need to evaluate three estimators that estimate $E[X]$. The first estimator $M_n^{(1)}(X)$ adds noise W_i to the i^{th} sample X_i, $i = 1, \ldots, n$. The first estimator calculates

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

We know that $E[W_i] = 0$ and $\text{Var}[W_i] = \sigma^2$, $i = 1, \ldots, n$. The second estimator calculates

$$M_n^{(2)}(X) = W + \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Here W is noise that is identically distributed as the W_i above. The third estimator is the most complex. It calculates

$$M_n^{(3)}(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i + I_i W_i).$$

Here the random variables I_i, $i = 1, \ldots, n$, are independent and identically distributed and take a value of -1 or 1 with equal probability. Note that W, W_i, X_i, I_i, $i = 1, \ldots, n$ are independent random variables. Derive the mean squared error of the three estimators. Which of these has the smallest error? Which of these estimators are unbiased and why? Which of these estimators are consistent and why?

Question 3. 30 **marks** A local coffee store announces an offer for 7 days. Customers would receive one coupon for every two days they purchased coffee at the store. Let N be the number of days a customer purchases coffee and C be the number of coupons the customer earns over the 7 days. Suppose a customer purchases coffee on any day with probability p independently of other days. Answer the following questions.

(a) (12.5 marks) Derive the joint PMF of N and C.

(b) (12.5 marks) Derive the average number of coupons a customer earns.

(c) (5 marks) Derive the average number of days a customer purchases coffee.

Q1

$$f_A(y) = \begin{cases} \frac{1}{100}, & y \in (0,100) \\ 0, & \text{otherwise} \end{cases}$$

(a) $f_{M|A}(x|y) = \frac{1}{100-y}, \quad x \in (y, 100).$

10

(b) Joint PDF

$$f_{M,A}(x,y) = \left(\frac{1}{100}\right)\left(\frac{1}{100-y}\right),$$

Perf Attendance

$$0 \le y \le x \le 0$$

attendance performance

5

Joint CDF is

$$F_{M,A}(x,y) = P[M \le x, A \le y]$$

$$= \begin{cases} \int_0^y \int_y^x f_{M,A}(u,v)\,du\,dv & 0 < y \le x < 100, \\ 0 & \text{otherwise.} \end{cases}$$

0

$$\int_0^y \int_y^x f_{M,A}(u,v)\, du\, dv$$

Correct limits (3)

Correct integrand (2)

$$= \int_0^y \int_y^x \left(\frac{1}{100}\right)\left(\frac{1}{100-v}\right) du\, dv$$

$$= \int_0^y \left(\frac{1}{100}\right) \frac{(x-y)}{100-v}\, dv = \frac{(x-y)}{100} \int_0^y \frac{dv}{100-v}$$

$w = 100 - v \Rightarrow dw = -dv.$ Substitution.

$v = 0 \Rightarrow w = 100$

$v = y \Rightarrow w = 100 - y$

$$= \frac{(x-y)}{100} \int_{100-y}^{100} \frac{dw}{w} = \frac{(x-y)}{100}\left[\ln\left(\frac{100}{100-y}\right)\right]$$

$0 < y \leq x < 100.$

(c) $E[X-Y] = E[X] - E[Y]$

$$\underbrace{E[M \mid A = a_1]}_{\substack{\text{The corresponding} \\ \text{conditional distribution} \\ \text{is } \frac{1}{100-a_1} \text{ for } x \in (a_1, 100).}} - \underbrace{E[M \mid A = a_2]}_{\substack{\text{Corr distribution} \\ \text{is } \frac{1}{100-a_2}, \\ y \in (a_2, 100)}}$$ 5

The difference in expectations is:

$$\frac{100+a_1}{2} - \frac{100+a_2}{2}$$ 5

$$= \frac{a_1 - a_2}{2}.$$

(d) $P[X > Y] = ?$

X and Y are given to be independent.

$$f_X(x) = \frac{1}{100-a_1} \qquad x \in (a_1, 100)$$

$$f_Y(y) = \frac{1}{100-a_2} \qquad y \in (a_2, 100)$$

$$f_{X,Y}(x,y) = \left(\frac{1}{100-a_1}\right)\left(\frac{1}{100-a_2}\right), \quad \begin{matrix} x \in (a_1, 100) \\ y \in (a_2, 100) \end{matrix}; \ a_2 > a_1$$

$$P[X>Y] = \int_{a_2}^{100}\int_{y}^{100} \left(\frac{1}{100-a_1}\right)\left(\frac{1}{100-a_2}\right) dx\, dy$$

Note that $y \geq a_2$ and $a_2 > a_1$.

Correct Limits 5

$$= \left(\frac{1}{100-a_1}\right)\left(\frac{1}{100-a_2}\right) \int_{a_2}^{100}\int_{y}^{100} dx\, dy$$

Correct Integrand 5

Rest 5.

$$= \left(\frac{1}{100-a_1}\right)\left(\frac{1}{100-a_2}\right) \int_{a_2}^{100} (100-y)\, dy$$

$$= \frac{1}{(100-a_1)(100-a_2)}\left(100(100-a_2) - \frac{1}{2}\left(100^2 - a_2^2\right)\right)$$

$$= \left(\frac{1}{100-a_1}\right)\left(100 - \frac{1}{2}(100+a_2)\right)$$

$$= \left(\frac{1}{100-a_1}\right)\frac{(100-a_2)}{2} = \frac{1}{2}\frac{(100-a_2)}{(100-a_1)}$$

[Consider $a_1 = a_2$, the result is $1/2$, which makes sense]

Q2 Rewriting the estimators:

$$M_n^{(1)}(X) = \frac{\overset{Y_1 \triangleq}{(X_1+W_1)} + \overset{Y_2 \triangleq}{(X_2+W_2)} + \cdots + \overset{Y_n \triangleq}{(X_n+W_n)}}{n}$$

$$M_n^{(2)}(X) = \frac{X_1 + X_2 + \cdots + X_n}{n} - W.$$

$$M_n^{(3)}(X) = \frac{Z_1 + Z_2 + \cdots + Z_n}{n}$$

where $Z_i := X_i + I_i W_i$

Note that all of them are unbiased!

$$E[M_n^{(1)}(X)] = \frac{E[Y_1] + \cdots + E[Y_n]}{n}$$

$$= \frac{E[X_1] + \cdots + E[X_n]}{n}$$

$$= E[X].$$

$$E[M_n^{(2)}(X)] = E[X] - E[W]$$
$$= E[X]$$

$$E[M_n^{(3)}(X)] = \frac{1}{n}\sum_{i=1}^{n} E[Z_i]$$
$$= \frac{1}{n}\sum_{i=1}^{n} E[X_i + I_i W_i]$$
$$= \frac{1}{n}\sum_{i=1}^{n} E[X_i] + \underbrace{E[I_i]}_{0}\,\underbrace{E[W_i]}_{0}$$
$$= E[X]$$

Calculating three Means: 5

MSE of an unbiased estimator is simply its variance.

$$\mathrm{Var}[M_n^{(1)}(X)] = \mathrm{Var}\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right]$$
$$= \frac{1}{n^2}\left(\mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n]\right) \quad \left[X_i, X_j \text{ are independent}\right]$$

$Y_i \sim Y$

$X_i \sim X.$

$$= \frac{1}{n^2}\left(n \operatorname{Var}[Y]\right)$$

$$= \frac{n}{n^2} \operatorname{Var}[X + W]$$

$$= \frac{1}{n} \operatorname{Var}[X + W]$$

$$= \frac{1}{n} \left(\operatorname{Var}[X] + \operatorname{Var}[W]\right)$$

$$= \frac{1}{n}\left(\operatorname{Var}[X] + \sigma^2\right)$$

$$\operatorname{Var}\left[M_n^{(2)}(X)\right] = \left(\frac{1}{n} \operatorname{Var}[X]\right) + \operatorname{Var}[W]$$

$$= \frac{\operatorname{Var}[X]}{n} + \sigma^2$$

$Z_i \sim Z$

Z_i, Z_j are independent

$$\operatorname{Var}\left[M_n^{(3)}(X)\right] = \frac{1}{n}\left[\operatorname{Var}[Z]\right]$$

$$= \frac{1}{n}\left[\operatorname{Var}[X + IW]\right]$$

$$= \frac{1}{n}\left[\operatorname{Var}[X] + \operatorname{Var}[IW]\right]$$

$\mathrm{Var}[IW]$

$$= E[(IW)^2] - (E[IW])^2$$

$$= E[I^2W^2] - (E[I]E[W])^2$$

$$= E[I^2]\,E[W^2]$$

$$= (1)\,E[W^2].$$

Calculating Variances (Mean squared errors)

$3 \times 5 = 15.$

$$\therefore \mathrm{Var}[M_n^{(2)}(X)] = \frac{\mathrm{Var}[X]}{n} + \frac{\sigma^2}{n}$$

Note $E[W] = 0$

1

All three are unbiased.

1 & 3 have smallest mean squared

2

error. In both we have $\frac{\sigma^2}{n}$ instead of just σ^2 in 2.

2

For unbiased estimators, if they are consistent, Variance goes to 0 as $n \to \infty$.

$\therefore$ 1 & 3 are consistent & 2 is not.

Q1

$S_N = \{0, 1, 2, \dots, 7\}$

$S_C = \{0, 1, 2, 3\}$

(a)

$P[N=0, C=0] = P[N=0]$

$P[N=0, C=c] = 0, \quad c = 1, 2, 3$

2

$P[N=1, C=0] = P[N=1]$

$P[N=1, C=c] = 0, \quad c = 1, 2, 3$

$P[N=2, C=1] = P[N=2]$

$P[N=2, C=c] = 0, \quad c = 0, 2, 3$

2

$P[N=3, C=1] = P[N=3]$

$P[N=3, C=c] = 0, \quad c = 0, 2, 3$

$P[N=4, C=2] = P[N=4]$

$P[N=4, C=c] = 0, \quad c = 0, 1, 3.$

$P[N=5, C=2] = P[N=5]$

$P[N=5, C=c] = 0, \quad c = 0, 1, 3.$

②

$P[N=6, C=3] = P[N=6]$

$P[N=6, C=c] = 0, \quad c = 0, 1, 2.$

$P[N=7, C=3] = P[N=7]$

$P[N=7, C=c] = 0, \quad c = 0, 1, 2$

②

Getting the Binomial Right

Here $P[N=n] = {}^7C_n \, p^n \, (1-p)^{7-n}$,

$n = 0, 1, 2, \ldots, 7.$

4.5

(b) $E[C] = 1 \, P[C=1] + 2 \, P[C=2] + 3 \, P[C=3]$

5.5

$P[C=1] = P[N=2, C=1] + P[N=3, C=1]$
$= P[N=2] + P[N=3].$

$P[C=2] = P[N=4] + P[N=5]$

$P[C=3] = P[N=6] + P[N=7]$

$$E[C] = 1\left(P[N=2] + P[N=3]\right) + 2\left(P[N=4] + P[N=5]\right) + 3\left(P[N=6] + P[N=7]\right).$$

Calc of
$P[C=1]$
$P[C=2]$
$P[C=3]$
$\Rightarrow$
$E[C]$

7

(c) $E[N] =$

$$\sum_{n=0}^{7} n \, {}^{7}C_{n} \, p^{n} (1-p)^{7-n}$$

5

$$= 7p.$$

[Just the answer is fine]