LECTURE: MIDTERM 2 – REVIEW

1. Covariance

Example 1:

Let X and Y be a pair of random variables which take on values

$$(1,0),(0,1),(-1,0) \text{ and } (0,-1) \text{ each with probability } 1/4$$

Show that $\text{Cov}(X,Y) = 0$ but X and Y are not independent

$$\text{Cov}(X,Y) = E(XY) - E(X)E(Y)$$

STEP 1: Let $p(x,y)$ be the joint pmf of (X,Y) then

$$\begin{aligned}
E[XY] =& \sum_{(x,y)} xy\, p(x,y) \\
=&(1)(0)p(1,0) + (0)(1)p(0,1) + (-1)(0)p(-1,0) + (0)(-1)p(0,-1) \\
=&0
\end{aligned}$$

STEP 2: To calculate $E(X)$ calculate the marginal $p_X(x)$

Notice X takes on values $-1, 0, 1$

$$p_X(-1) = \sum_y p(-1, y) = p(-1, 0) = \frac{1}{4}$$
$$p_X(0) = \sum_y p(0, y) = p(0, 1) + p(0, -1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$
$$p_X(1) = \sum_y p(1, y) = p(1, 0) = \frac{1}{4}$$

$$E[X] = \sum_x x\, p_X(x) = -1p_X(-1) + 0p_X(0) + 1p_X(1) = -\frac{1}{4} + 0 + \frac{1}{4} = 0$$

STEP 3: To calculate $E(Y)$ calculate the marginal p_Y

Notice Y takes on values $-1, 0, 1$

$$p_Y(-1) = \sum_x p(x, -1) = p(0, -1) = \frac{1}{4}$$
$$p_Y(0) = \sum_x p(x, 0) = p(-1, 0) + p(1, 0) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$
$$p_Y(1) = \sum_x p(x, 1) = p(0, 1) = \frac{1}{4}$$

$$E[Y] = \sum_y y\, p_Y(y) = -1p_Y(-1) + 0p_Y(0) + 1p_Y(1) = -\frac{1}{4} + 0 + \frac{1}{4} = 0$$

STEP 4: Hence

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0 - 0 = 0$$

STEP 5: X and Y are not independent because for example

$$p(0,1) = \frac{1}{4} \neq p_X(0)p_Y(1) = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$$

2. Conditional Density

Example 2:

A stick of length 1 is broken in two places.

The first break point is chosen uniformly at random along the length of the stick from $[0,1]$

The second break point is chosen uniformly at random from 0 to the first break point.

(a) Find the joint pdf of the two break points.

(b) Find the expectation of the product of the two break points.

(a) **STEP 1:** Let X be the random variable of the first point and Y the random variable of the second.

Intuitively we want to use

$$\text{Conditional} = \frac{\text{Joint}}{\text{Marginal}} \Rightarrow \text{Joint} = \text{Conditional} \times \text{marginal}$$

Then from above we have

$$f(x,y) = f(y|x)f_X(x)$$

STEP 2: By assumption

$$f_X(x) \text{ is uniform over the interval } [0,1] \text{ so } f_X(x) = \frac{1}{1-0} = 1$$

$$f(y|x) \text{ is uniform over the interval } [0,x] \text{ so } f(y|x) = \frac{1}{x-0} = \frac{1}{x}$$

$$\text{Hence } f(x,y) = f_X(x)f(y|x) = 1\left(\frac{1}{x}\right) = \frac{1}{x}$$

Warning: Don't forget about the bounds! Since the second break point is before the first, we have $y \leq x$ and so

$$f(x,y) = \frac{1}{x} \text{ where } 0 \leq y \leq x \leq 1$$

(b)

$$E[XY] = \int_0^1 \int_0^x xyf(x,y)\,dydx = \int_0^1 \int_0^x xy\left(\frac{1}{x}\right)\,dydx = \frac{1}{2}\int_0^1 x^2\,dx = \frac{1}{6}$$

3. Bias and MSE

Example 3:

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from Unif $[0, L]$

Let $Y_{\max} = \max\{X_1, X_2, \ldots, X_n\}$ whose density is:

$$f(y) = \begin{cases} \frac{n}{L^n}\, y^{n-1} & 0 < y < L \\ 0 & \text{otherwise} \end{cases}$$

(a) Compute the bias of $Y_{\max}$ as an estimator for L

(b) Give an unbiased estimator for $\hat{L}$ in terms of $Y_{\max}$

(a)

$$\begin{aligned} E(Y_{\max}) &= \int_0^L y f(y) dy = \int_0^L y \left(\frac{n}{L^n} y^{n-1}\right) dy = \int_0^L \frac{n}{L^n} y^n dy \\ &= \left[\frac{n}{L^n}\left(\frac{y^{n+1}}{n+1}\right)\right]_{y=0}^{y=L} = \left(\frac{n}{L^n}\right)\frac{L^{n+1}}{n+1} = \left(\frac{n}{n+1}\right) L \end{aligned}$$

$$\text{Bias}\left(Y_{\max}\right) = E\left(Y_{\max}\right) - L = \left(\frac{n}{n+1}\right) L - L$$

(b) To convert this into an unbiased estimator, just use

$$\hat{L} = \left(\frac{n+1}{n}\right) Y_{\max}$$

$$\begin{aligned} \text{Bias}(\hat{L}) =& E\left(\hat{L}\right) - L \\ =& E\left(\left(\frac{n+1}{n}\right) Y_{\max}\right) - L \\ =& \left(\frac{n+1}{n}\right) E\left(Y_{\max}\right) - L \\ =& \left(\frac{n+1}{n}\right)\left(\frac{n}{n+1}\right) L - L = 0 \checkmark \end{aligned}$$

4. Inequalities

Example 4:

The average height of a raccoon is 10 inches.

(a) Given an upper bound on the probability that a given raccoon is at least 15 inches tall.

(b) The standard deviation of the height distribution is 2 inches. Find a lower bound on the probability that a raccoon is between 5 and 15 inches tall.

(c) Repeat (b), this time assuming the distribution is normal

(a) Let Y be the height of the raccoon.

Since we don't know the distribution of Y and we know $E(Y) = 10$, we use Markov, which says

$$P(Y \geq 15) \leq \frac{E(Y)}{15} = \frac{10}{15} = \frac{2}{3}$$

(b) This time we *also* know $\sigma = 2$ and so we use Chebyshev:

$$P(|Y - E(Y)| \geq a) \leq \frac{\text{Var}(Y)}{a^2} \Rightarrow P(|Y - 10| \geq a) \leq \frac{4}{a^2}$$

In this case we want

$$5 < Y < 15 \Rightarrow -5 \leq Y - 10 < 5 \Rightarrow |Y - 10| < 5$$

And so we choose $a = 5$ to get

$$P(|Y - 10| \geq 5) \leq \frac{4}{5^2} = \frac{4}{25}$$

Hence, by the complement rule,

$$P(|Y - 10| < 5) \geq 1 - \frac{4}{25} = \frac{21}{25}$$

(c) This time we know $Y \sim N(\mu, \sigma^2)$ where $\mu = 10$ and $\sigma = 2$

So converting to the standard normal variable, we get

$$\begin{aligned} P(5 \leq Y \leq 15) =& P\left(\frac{5-10}{2} \leq \frac{Y-10}{2} \leq \frac{15-10}{2}\right) \\ =& P(-2.5 \leq Z \leq 2.5) \\ =& F(2.5) - F(-2.5) \\ =& .9938 - 0.0062 = 0.9876 \end{aligned}$$

Example 5:

Let X be a random variable which takes on values

$$-1 \text{ and } 1 \text{ each with probability } 1/2$$

Show that in this case, equality holds in Chebyshev's inequality (with $a = 1$)

STEP 1: In this case we need to show

$$P(|X - E(X)| \geq 1) = \frac{\text{Var}(X)}{1^2} \Rightarrow P(|X - E(X)| \geq 1) = \text{Var}(X)$$

STEP 2:

$$E(X) = (-1)P(X = -1) + (1)P(X = 1) = -\frac{1}{2} + \frac{1}{2} = 0$$

$$E\left(X^2\right) = (-1)^2 P(X = -1) + (1)^2 P(X = 1) = \frac{1}{2} + \frac{1}{2} = 1$$

$$\text{Var}\left(X^2\right) = E\left(X^2\right) - (E(X))^2 = 1 - 0 = 1$$

STEP 3: Hence we need to show that

$$P(|X - 0| \geq 1) = 1 \Rightarrow P(|X| = 1) = 1$$

But this is true because $|X|$ is always 1

5. Confidence Intervals

Example 6:

A random sample of 120 students from a large university yields mean GPA 2.71 with sample standard deviation 0.51.

Construct a 90% confidence interval for the mean GPA of all students at the university.

We know that $\bar{Y} = 2.71$ and $S = 0.51$

STEP 1: Since the population is large and we don't know σ, we use the standard normal distribution, along with

$$\hat{\sigma} = \frac{S}{\sqrt{n}} = \frac{0.51}{\sqrt{120}}$$

STEP 2: Now $1 - \alpha = 0.9 \Rightarrow \alpha = 0.1 \Rightarrow \frac{\alpha}{2} = 0.05$

Moreover from the $Z-$table, we get $z_{0.05} = 1.645$

STEP 3: Therefore our 90% confidence interval for μ is

$$
\begin{aligned}
[\hat{L}, \hat{U}] &= \left[\overline{Y} - z_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right), \overline{Y} + z_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)\right] \\
&= \left[2.71 - (1.645)\left(\frac{0.51}{\sqrt{120}}\right), 2.71 + (1.645)\left(\frac{0.51}{\sqrt{120}}\right)\right] \\
&= [2.63, 2.79]
\end{aligned}
$$