Review of the last lecture

Key concepts and/or techniques:

- Mathematical Expectation:

$$E(g(X,Y)) = \sum_{(x,y)\in \overline{S}} g(x,y)f(x,y)$$

- Covariance and correlation coefficient:

  To study the relation between 2 RVs

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}, \quad Var(X) > 0, Var(Y) > 0.$$

- Interpretation and properties of covariance and correlation coefficient

Review of the last lecture

- $Cov(X, Y) = E(XY) - E(X)E(Y)$

  When $Cov(X, Y) = 0$, X and Y are uncorrelated.

  When $Cov(X, Y) > 0$, X and Y are positively correlated.

  When $Cov(X, Y) < 0$, X and Y are negatively correlated.

- Interpretation: Roughly speaking, a positive or negative covariance indicate that the values of $X - E(X)$ and $Y - E(Y)$ obtained in a single experiment "tend" to have the same or the opposite sign respectively.

- Independence of X and Y $\Rightarrow$ uncorrelation of X and Y, but the converse is in general not true.

Review of the last lecture

Correlation coefficient

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

- It is a normalized version of $Cov(X, Y)$ and in fact $-1 \leq \rho(X, Y) \leq 1$ and the size of $|\rho|$ provides a normalized measure of the extent to which this is true.

- $\rho = 1$ *or* $(\rho = -1)$ if and only if there exists a positive. (or negative,respectively) constant c such that
$$Y - E(Y) = c(X - E(X))$$

STA2001 Probability and Statistics (I)

Lecture 16

Tianshi Chen

The Chinese University of Hong Kong, Shenzhen

Section 4.3 Conditional distribution

Conditional Distribution

Motivation: the conditional probability distribution is a probability distribution that describes the distribution of probability of events of a RV given the occurrence of a particular event

Assume that X and Y have a joint pmf $f(x,y) : \overline{S} \to (0,1]$.

The marginal pmf of X and Y are

$$f_X(x) : \overline{S_X} \to (0,1] \qquad f_Y(y) : \overline{S_Y} \to (0,1]$$

$$\overline{S_X} = \{\text{all possible values of } X \text{ in } \overline{S}\}$$
$$\overline{S_X}(y) = \{x | (x,y) \in \overline{S}\} \text{ for } y \in \overline{S_Y}$$

$$\overline{S_Y} = \{\text{all possible values of } Y \text{ in } \overline{S}\}$$
$$\overline{S_Y}(x) = \{y | (x,y) \in \overline{S}\} \text{ for } x \in \overline{S_X}$$

Conditional Distribution

By definition,

$$
\begin{aligned}
f(x,y) &= P(X=x, Y=y) \\
&\triangleq P(\{X=x, Y=y\}), (x,y) \in \overline{S} \\
f_X(x) &= P_X(X=x) \\
&\triangleq P(\{X=x, Y \in \overline{S_Y}(x)\}) = \sum_{y \in \overline{S_Y}(x)} f(x,y) \\
f_Y(y) &= P_Y(Y=y) \\
&\triangleq P(\{X \in \overline{S_X}(y), Y=y\}) = \sum_{x \in \overline{S_X}(y)} f(x,y)
\end{aligned}
$$

Conditional Distribution

Let

$$A = \{X = x, Y \in \overline{S_Y}(x)\}$$
$$B = \{X \in \overline{S_X}(y), Y = y\}$$

Then for $(x, y) \in \overline{S}$,

$$A \cap B = \{X = x, Y = y\}$$

and recall the conditional probability of event A given event B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{f(x, y)}{f_Y(y)}$$

under the assumption $P(B) > 0$, i.e., $f_Y(y) > 0$.

Conditional pmf

Definition

Conditional pmf of X given $Y = y$ is defined by

$$g(x|y) = \frac{f(x,y)}{f_Y(y)}, \quad x \in \overline{S_X}(y)$$

provided that $f_Y(y) > 0$.

Similarly, the conditional pmf of Y given that $X = x$ is defined by

$$h(y|x) = \frac{f(x,y)}{f_X(x)}, \quad y \in \overline{S_Y}(x)$$

provided that $f_X(x) > 0$.

- What is the interpretation of the conditional pmf $g(x|y)$?
- What if X and Y are independent?

Some Remarks

Conditional pmf is a well-defined pmf:

1. $h(y|x) > 0$
2. $\sum_{y \in \overline{S_Y}(x)} h(y|x) = 1$

$$\sum_{y \in \overline{S_Y}(x)} h(y|x) = \sum_{y \in \overline{S_Y}(x)} \frac{f(x,y)}{f_X(x)} = \frac{\sum_{y \in \overline{S_Y}(x)} f(x,y)}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1$$

3. for $A \subseteq \overline{S_Y}(x)$

$$\begin{aligned} P(Y \in A|X = x) &= \frac{P(X = x, Y \in A)}{P(X = x)} \\ &= \frac{\sum_{y \in A} f(x,y)}{f_X(x)} = \sum_{y \in A} h(y|x) \end{aligned}$$

Therefore, $h(y|x)$ (resp. $g(x|y)$) determines the distribution of probability of events of Y (resp. X) given $X = x$ (resp. $Y = y$).

Some Remarks

If X and Y are independent, then $f(x, y) = f_X(x)f_Y(y)$ and thus

$$g(x|y) = f_X(x), \text{ and } h(y|x) = f_Y(y),$$

which implies

- the occurrence of the event $Y = y$ does not change the probability of the occurrence of events of X
- the occurrence of the event $X = x$ does not change the probability of the occurrence of events of Y

Now, the implication of independent RVs becomes clear.

Example 1

Question

Let X and Y have the joint pmf

$$f(x, y) = \frac{x + y}{21}, \quad x = 1, 2, 3; \quad y = 1, 2.$$

We have showed

$$f_X(x) = \frac{2x + 3}{21}, \quad x = 1, 2, 3$$
$$f_Y(y) = \frac{y + 2}{7}, \quad y = 1, 2.$$

Q1: What is the conditional pmf of X given $Y = y$?

Q2: What is the conditional pmf of Y given $X = x$?

Q3: What is $P(1 \leq X \leq 2 | Y = 1)$?

Example 1

Q1: The conditional pmf of X given $Y = y$ is

$$g(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{x+y}{21} \Big/ (\frac{y+2}{7}) = \frac{x+y}{3(y+2)},$$
$$x = 1,2,3; \qquad y = 1,2.$$

Q2: The conditional pmf of Y given $X = x$ is

$$h(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{x+y}{21} \Big/ (\frac{2x+3}{21}) = \frac{x+y}{2x+3}$$
$$x = 1,2,3; \qquad y = 1,2.$$

Q3:

$$P(1 \le X \le 2 | Y = 1) = \sum_{x=1}^{2} g(x|1) = \sum_{x=1}^{2} \frac{x+1}{3(1+2)} = \frac{5}{9}$$

Conditional Mathematical Expectation

- Let $g(Y)$ be a function of Y.

  Then the conditional expectation of $g(Y)$ given $X = x$

$$E(g(Y)|X = x) = \sum_{y \in \overline{S_Y}(x)} g(y)h(y|x)$$

- When $g(Y) = Y$,

$$E(Y|X = x) = \sum_{y \in \overline{S_Y}(x)} yh(y|x) \rightarrow \text{conditional mean}$$

Conditional Mathematical Expectation

- When $g(Y) = [Y - E(Y|X = x)]^2$

$$
\begin{aligned}
Var(Y|X = x) &\triangleq E\{[Y - E(Y|X = x)]^2 | X = x\} \\
&= \sum_{y \in \overline{S_Y}(x)} [y - E(Y|X = x)]^2 h(y|x) \\
&= E(Y^2|X = x) - [E(Y|X = x)]^2 \\
&\rightarrow \text{conditional variance}
\end{aligned}
$$

Example 1, continued

Question

Let X and Y have the joint pmf

$$f(x,y) = \frac{x+y}{21}, \quad x = 1,2,3; \quad y = 1,2.$$

We have showed

$$f_X(x) = \frac{2x+3}{21}, \quad x = 1,2,3$$
$$f_Y(y) = \frac{y+2}{7}, \quad y = 1,2.$$

Q1: What is the expectation of Y given $X = 3$?

Q2: What is the variance of Y given $X = 3$?

Example 1, continued

$$Q1: E(Y|X=3) = \sum_{y \in \overline{S_Y}(3)} yh(y|3) = \sum_{y=1}^{2} y\left(\frac{3+y}{9}\right) = \frac{14}{9}$$

$$Q2: Var(Y|X=3) = \sum_{y \in \overline{S_Y}(3)} [y - E(Y|X=3)]^2 h(y|3)$$

$$= \sum_{y=1}^{2} \left(y - \frac{14}{9}\right)^2 \frac{3+y}{9} = \frac{20}{81}$$

Example 2: Trinomial Distribution

Assume that $(X, Y) \sim \text{Trinomial}(n, p_X, p_Y)$. Then what is the conditional pmf of Y given $X = x$, i.e., $h(y|x)$?

Example 2: Trinomial Distribution

Assume that $(X, Y) \sim \text{Trinomial}(n, p_X, p_Y)$. Then what is the conditional pmf of Y given $X = x$, i.e., $h(y|x)$?

First, we recall that

$$f(x,y) = \frac{n!}{x!y!(n-x-y)!} p_X^x p_Y^y (1 - p_X - p_Y)^{n-x-y}, (x,y) \in \overline{S},$$
$$\overline{S} = \{(x,y) | x + y \leq n, x = 0, 1, \cdots, n, y = 0, 1, \cdots, n\}$$

and $f_X(x) = \frac{n!}{x!(n-x)!} p_X^x (1 - p_X)^{n-x}$, $x \in \overline{S_X} = \{0, 1, \cdots, n\}$.

Example 2: Trinomial Distribution

Assume that $(X, Y) \sim \text{Trinomial}(n, p_X, p_Y)$. Then what is the conditional pmf of Y given $X = x$, i.e., $h(y|x)$?

First, we recall that

$$f(x,y) = \frac{n!}{x!y!(n-x-y)!} p_X^x p_Y^y (1-p_X-p_Y)^{n-x-y}, (x,y) \in \overline{S},$$

$$\overline{S} = \{(x,y)|x+y \leq n, x = 0, 1, \cdots, n, y = 0, 1, \cdots, n\}$$

and $f_X(x) = \frac{n!}{x!(n-x)!} p_X^x (1-p_X)^{n-x}$, $x \in \overline{S_X} = \{0, 1, \cdots, n\}$.

Then we have $\overline{S_Y}(x) = \{0, \cdots, n-x\}$, and for any $y \in \overline{S_Y}(x)$,

$$h(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{\frac{n!}{x!y!(n-x-y)!} p_X^x p_Y^y (1-p_X-p_Y)^{n-x-y}}{\frac{n!}{x!(n-x)!} p_X^x (1-p_X)^{n-x}}$$

$$= \frac{(n-x)!}{y!(n-x-y)!} \frac{p_Y^y (1-p_X-p_Y)^{n-x-y}}{(1-p_X)^y (1-p_X)^{n-x-y}}$$

$$= \frac{(n-x)!}{y!(n-x-y)!} \left(\frac{p_Y}{1-p_X}\right)^y \left(1 - \frac{p_Y}{1-p_X}\right)^{n-x-y}$$

Example 2: Covariance for Trinomial Distribution

What is the conditional distribution of Y given $X = x$?

$$Y|X = x \sim b(n - x, \frac{p_Y}{1 - p_X}).$$

Example 2: Covariance for Trinomial Distribution

What is the conditional distribution of Y given $X = x$?

$$Y|X = x \sim b(n - x, \frac{p_Y}{1 - p_X}).$$

Then what is $Cov(X, Y)$?

Example 2: Covariance for Trinomial Distribution

What is the conditional distribution of Y given $X = x$?

$$Y|X = x \sim b(n - x, \frac{p_Y}{1 - p_X}).$$

Then what is $Cov(X, Y)$?

$$\begin{aligned} Cov(X, Y) &= \sum_{(x,y)\in\overline{S}} (x - E[X])(y - E[Y])f(x, y) \\ &= \sum_{x\in\overline{S_X}} (x - E[X]) \sum_{y\in\overline{S_Y}(x)} (y - E[Y])f(x, y) \\ &= \underbrace{\sum_{x\in\overline{S_X}} (x - E[X])f_X(x)}_{\text{expectation of function of } X} \underbrace{\sum_{y\in\overline{S_Y}(x)} (y - E[Y])h(y|x)}_{\text{expectation of function of } Y|X=x} \end{aligned}$$

The calculation of $Cov(X, Y)$ is converted to that of mathematical expectation of functions of two univariate RV, which could be easier if the distributions of two univariate RV are known.

Section 4.4 Bivariate Distribution of Continuous Type

Bivariate Continuous RV

Definition

Let X and Y be two continuous random variables and (X, Y) be a pair of RVs with their range denoted by $\overline{S} \subseteq R^2$. Then (X, Y) or X and Y is said to be a bivariate continuous RV.

Moreover, let $\overline{S_X} \subseteq R$ and $\overline{S_Y} \subseteq R$ denote the range of X and Y, respectively.

$$\begin{aligned} \overline{S} &= \{\text{all possible values of } (X, Y)\} \\ \overline{S_X} &= \{\text{all possible values of } X\} = \{x | (x, y) \in \overline{S}\} \\ \overline{S_Y} &= \{\text{all possible values of } Y\} = \{y | (x, y) \in \overline{S}\} \end{aligned}$$

Then, it holds that

$$\overline{S} \subseteq \overline{S_X} \times \overline{S_Y} = \{(x, y) | x \in \overline{S_X}, y \in \overline{S_Y}\}$$

Roadmap for bivariate continuous random distributions

To study the bivariate continuous random variable

discrete RV $\longrightarrow$	continuous RV	Mathematical expectations
pmf $\longrightarrow$	pdf	mean
joint pmf $\longrightarrow$	joint pdf	variance
marginal pmf $\longrightarrow$	marginal pdf	covariance
conditional pmf $\longrightarrow$	conditional pdf	correlation coefficient

Joint pdf

Definition

The joint pdf of two continuous RVs X and Y is a function $f(x,y) : \overline{S} \to (0,\infty)$ with the following properties:

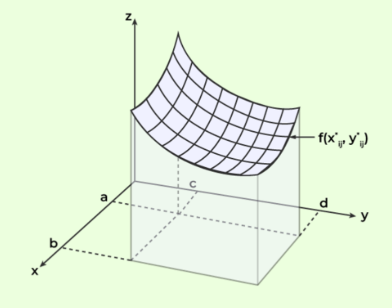
1. $f(x,y) > 0, (x,y) \in \overline{S}$
2. $\iint_{\overline{S}} f(x,y)dxdy = 1$
3. $$\begin{aligned} P((X,Y) \in A) &\triangleq P(\{(X,Y) \in A\}) \\ &= \iint_A f(x,y)dxdy, A \subseteq \overline{S} \end{aligned}$$

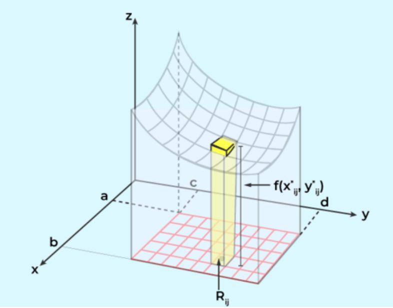
Remarks

Recall the geometric interpretation of double integral:

$$P((X,Y) \in A) = \iint_A f(x,y)\,dxdy$$

calculates the volume of the solid under the surface $z = f(x,y)$ over the region A in the the xy-plane.

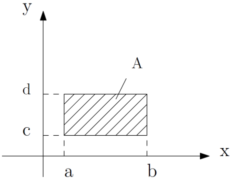




Remarks

- Very often, we extend the definition domain of $f(x, y)$ from $\overline{S}$ to $R \times R$ by letting $f(x, y) = 0$, for $(x, y) \notin \overline{S}$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dxdy = 1$

- Bottom line:



If the set A is rectangular with its line segments parallel to the coordinate axes, i.e.,

$$A = \{(x, y) | a \leq x \leq b, c \leq y \leq d\},$$

then the double integral becomes

$$P((X, Y) \in A) = \int_a^b \int_c^d f(x, y) dydx$$