STA2001 Probability and Statistics (I)

Lecture 11

Tianshi Chen

The Chinese University of Hong Kong, Shenzhen

Review of the last lecture

Key concepts and/or techniques:

- Bivariate RV: (X, Y) or X and Y with range $\overline{S} \subseteq \overline{S_X} \times \overline{S_Y} \subseteq \mathbb{R}^2$
- Joint pmf $f(x, y) : \overline{S} \to (0, 1]$
- How to derive Marginal pmf from the joint pmf

$$f_X(x) = P_X(X = x) \triangleq P\left(\{X = x, Y \in \overline{S_Y}(x)\}\right) = \sum_{y \in \overline{S_Y}(x)} f(x, y)$$

- Trinomial distribution: $(X, Y) \sim \text{Trinomial}(n, p_X, p_Y)$

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

- X and Y are independent if $f(x, y) = f_X(x) f_Y(y)$

Review of the last lecture

Definition[Joint pmf]

The function $f(x,y) : \overline{S} \to (0,1]$ is called the joint probability mass function (joint pmf) of X and Y or (X,Y), if

1. $f(x,y) > 0$ for $(x,y) \in \overline{S}$,
2. $\sum_{(x,y)\in\overline{S}} f(x,y) = 1$,
3. For $A \subseteq \overline{S}$,

   $$P[(X,Y) \in A] \stackrel{\Delta}{=} P(\{(X,Y) \in A\}) = \sum_{(x,y)\in A} f(x,y)$$

   which defines the probability function for a set A. In particular, taking $A = \{(x,y)\}$ yields the probability of $X = x$ and $Y = y$, i.e.,

   $$P(X = x, Y = y) = f(x,y)$$

Review of the last lecture

Definition[Marginal pmf]

Let (X, Y) be a bivariate RV, or X and Y be two RVs, and have the joint pmf $f(x, y) : \overline{S} \to (0, 1]$. Sometimes, we are interested in the pmf of X or Y alone, which is called the marginal pmf of X or Y and described by

For $x \in \overline{S_X}$,

$$
\begin{aligned}
f_X(x) = P_X(X = x) &\stackrel{\Delta}{=} P\left(\{X = x, Y \in \overline{S_Y}(x)\}\right) \\
&= \sum_{y \in \overline{S_Y}(x)} f(x, y)
\end{aligned}
$$

where

$$\overline{S_Y}(x) = \{y | (x, y) \in \overline{S}\} \text{ for the given } x \in \overline{S_X}.$$

Review of the last lecture

It is crucial to understand the following definitions

$\overline{S}, \overline{S_X}, \overline{S_Y}, \overline{S_X}(y), \overline{S_Y}(x)$

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

$$\begin{aligned}
\overline{S_X}(y) &= \left\{x | (x, y) \in \overline{S}\right\} \text{ for a given y} \in \overline{S_Y} \\
\overline{S_Y}(x) &= \left\{y | (x, y) \in \overline{S}\right\} \text{ for a given x} \in \overline{S_X}
\end{aligned}$$

Review of the last lecture

Definition

The random variables X and Y are said to be independent if for every $x \in \overline{S_X}$ and $y \in \overline{S_Y}$

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

or equivalently,

$$P(X = x, Y = y) = P_X(X = x) P_Y(Y = y).$$

X and Y are said to be dependent if otherwise.

When X and Y are independent,

$$\overline{S} = \overline{S_X} \times \overline{S_Y}, \quad \overline{S} \text{ is said to be rectangular}$$

which is a necessary condition for independence of X and Y.

Section 4.2 The correlation coefficient

Motivation

Study the relation between two RVs (random phenomena)

Covariance of X and Y

Definition

Let X and Y be RVs with joint pmf $f(x,y) : \overline{S} \to (0,1]$
Take

$$
\begin{aligned}
g(X,Y) &= (X - E(X))(Y - E(Y)) \\
Cov(X,Y) &\triangleq E[(X - E(X))(Y - E(Y))] \\
&= \sum_{(x,y)\in\overline{S}} (x - E(X))(y - E(Y)) f(x,y)
\end{aligned}
$$

Moreover, we have

$$Cov(X,Y) \triangleq E(XY) - E(X)E(Y)$$

Covariance of X and Y

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

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

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

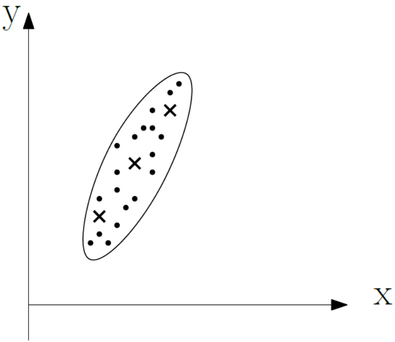
  When $Cov(X, Y) < 0$, X and Y are negatively correlated.

- 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.

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

Example 1 [Positively Correlated and Negatively Correlated RVs]

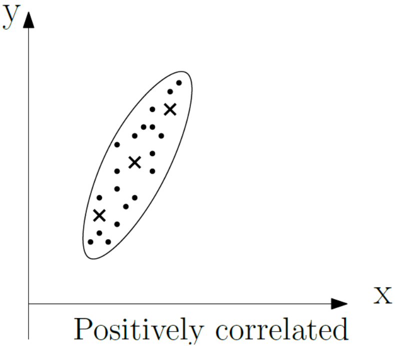
Assume that X and Y are uniformly distributed over the ellipses.



Question: which figure shows that X and Y are positively correlated?

Example 1 [Positively Correlated and Negatively Correlated RVs]

Assume that X and Y are uniformly distributed over the ellipses.



Independence $\Rightarrow$ Uncorrelation

- If X and Y are independent, we have

$$f(x,y) = f_X(x)f_Y(y) \Rightarrow \overline{S} = \overline{S_X} \times \overline{S_Y}$$

$$E(XY) = \sum_{(x,y)\in\overline{S}} xyf(x,y) = \sum_{x\in\overline{S_X}} \sum_{y\in\overline{S_Y}} xyf_X(x)f_Y(y)$$

$$= \sum_{x\in\overline{S_X}} xf_X(x) \left[\sum_{y\in\overline{S_Y}} yf_Y(y) \right] = E(X)E(Y)$$

Independence $\Rightarrow$ Uncorrelation

Therefore

$$Cov(X, Y) = E(XY) - E(X)E(Y) = 0$$

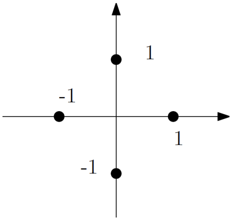
Independence of two RVs $\Rightarrow$ uncorrelation of two RVs.

However, the converse is not true , i.e, there exist X and Y which are uncorrelated but not independent.

Example 2 [Uncorrelation $\not\Rightarrow$ Independence]

Question

Let (X, Y) be a bivariate RV that takes values $(1, 0), (0, 1)$, $(-1, 0)$,$(0, -1)$, each with probability $\frac{1}{4}$, as shown in the figure below



Question 1

What are the marginal pmf of X and Y?

Question 2

What is $Cov(X, Y)$?

Question 3

Are X and Y independent?

Example 2 [Uncorrelation $\not\Rightarrow$ Independence]

To find marginal pmf of X and Y, $\overline{S_X} = \overline{S_Y} = \{-1, 0, 1\}$

$$f_X(x) = \begin{cases} \frac{1}{4}, & x = 1 \\ \frac{1}{2}, & x = 0 \\ \frac{1}{4}, & x = -1 \end{cases} \quad f_Y(y) = \begin{cases} \frac{1}{4}, & y = 1 \\ \frac{1}{2}, & y = 0 \\ \frac{1}{4}, & y = -1 \end{cases}$$

$$Cov(X, Y) = E(XY) - E(X)E(Y) = 0 - 0 \times 0 = 0$$

which shows that X and Y are uncorrelated.

$$f_X(0)f_Y(1) = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8} \neq f(0, 1) = \frac{1}{4}$$

which shows that X and Y are NOT independent.

Correlation Coefficient

Definition

The correlation coefficient of X and Y that have nonzero variance is defined as

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

- Interpretation : $\rho > 0$ (or $\rho < 0$) indicate the values of $X - E[X]$ and $Y - E[Y]$ "tend" to have the same (or negative, respectively) sign.

Properties of the Correlation Coefficient

- It is a normalized version of $Cov(X, Y)$ and in fact $-1 \leq \rho(X, Y) \leq 1$.

- $\rho = 1$ (resp. $\rho = -1$) if and only if there exists a positive (resp. negative) constant c such that

$$Y - E(Y) = c(X - E(X)),$$

and the size of $|\rho|$ provides a normalized measure of the extent to which this is true.

Proof for Properties of Correlation Coefficient (1/3)

To prove $|\rho(X, Y)| \leq 1$ is equivalent to prove that

$$\text{Cov}(X, Y)^2 \leq Var(X)Var(Y)$$

Proof for Properties of Correlation Coefficient (1/3)

To prove $|\rho(X, Y)| \leq 1$ is equivalent to prove that

$$\mathrm{Cov}(X, Y)^2 \leq Var(X)Var(Y)$$

To this goal, we consider

$$E((V + tW)^2) \geq 0,$$

where $t \in \mathbb{R}$, $V = X - E(X)$, $W = Y - E(Y)$. Then we have

$$\begin{aligned} E((V + tW)^2) &= E(V^2 + 2tVW + W^2) \\ &= E(V^2) + 2tE(VW) + t^2E(W^2) \geq 0 \end{aligned}$$

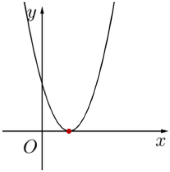
Proof for Properties of Correlation Coefficient (2/3)

Noting that

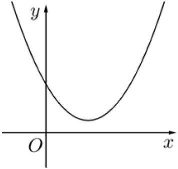
$$E(V^2) = Var(X), E(W^2) = Var(Y), E(VW) = \mathrm{Cov}(X, Y)$$

yields that for $t \in \mathbb{R}$,

$$E((V + tW)^2) = Var(X) + 2\mathrm{Cov}(X, Y)t + Var(Y)t^2 \geq 0$$



One Solution



No Real Solution

Since the above equation is true for any $t \in \mathbb{R}$, it must hold that

$4\mathrm{Cov}(X, Y)^2 - 4Var(X)Var(Y) \leq 0$, i.e., $\mathrm{Cov}(X, Y)^2 \leq Var(X)Var(Y)$

which implies that $|\rho(X, Y)| \leq 1$.

Proof for Properties of Correlation Coefficient (3/3)

When $\text{Cov}(X,Y)^2 - Var(X)Var(Y) = 0$, $|\rho(X,Y)| = 1$ implying

$$
\begin{aligned}
E((V+tW)^2) &= Var(X) + 2\text{Cov}(X,Y)t + Var(Y)t^2 = 0 \\
&= Var(X) \pm 2\sqrt{Var(X)}\sqrt{Var(Y)}t + Var(Y)t^2 = 0 \\
&= (\sqrt{Var(X)} \pm \sqrt{Var(Y)}t)^2 = 0
\end{aligned}
$$

$$t^* = \mp \frac{\sqrt{Var(X)}}{\sqrt{Var(Y)}}.$$

Inserting the above t^* back in $E((V+tW)^2)$ yields

$$E((V+t^*W)^2) = 0 \implies V = -t^*W$$

$$X - E(X) = \pm \frac{\sqrt{Var(X)}}{\sqrt{Var(Y)}}(Y - E(Y)),$$

where $\rho = 1$ (resp. $\rho = -1$) corresponds to $+$ (resp. $-$).

Example

Question

Consider n independent tosses of a coin with probability of a head equal to p. Let X and Y be the number of heads and of tails in the n tosses, respectively. Calculate the correlation coefficient of X and Y.

Example

Question

Consider n independent tosses of a coin with probability of a head equal to p. Let X and Y be the number of heads and of tails in the n tosses, respectively. Calculate the correlation coefficient of X and Y.

$$X + Y = n \quad \Rightarrow E[X] + E[Y] = n \Rightarrow X - E(X) = -[Y - E(Y)]$$

Example

Question

Consider n independent tosses of a coin with probability of a head equal to p. Let X and Y be the number of heads and of tails in the n tosses, respectively. Calculate the correlation coefficient of X and Y.

$$X + Y = n \quad \Rightarrow E[X] + E[Y] = n \Rightarrow X - E(X) = -[Y - E(Y)]$$

$$\begin{aligned}
Cov(X, Y) &= E[(X - E(X))(Y - E(Y))] \\
&= -E[(Y - E(Y))^2] = -Var(Y) \\
Var(X) &= E[(X - E(X))^2] = E[(Y - E(Y))^2] = Var(Y) \\
\rho(X, Y) &= \frac{-Var(Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{-Var(Y)}{\sqrt{Var(Y)}\sqrt{Var(Y)}} = -1
\end{aligned}$$

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 $A \subseteq \overline{S_Y}(x)$

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

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 $P(1 \le X \le 2 | Y = 1)$?

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

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

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

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