Math 461 Spring 2024

Renming Song

University of Illinois Urbana-Champaign

April 24, 2024

Outline

Outline

HW11 is due on Friday, 04/26, before the end of class.

In the rest of the semester, I will review and answer questions. If there are certain topics or questions you want me to go over in the lecture, please send me emails.

HW11 is due on Friday, 04/26, before the end of class.

In the rest of the semester, I will review and answer questions. If there are certain topics or questions you want me to go over in the lecture, please send me emails.

Definition

The covariance $\mathrm{Cov}(X, Y)$ of two random variables X and Y is defined by

$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

One can easily check that

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

If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$. But the converse is not true. When $\mathrm{Cov}(X, Y) = 0$, we say that X and Y are uncorrelated. Independence implies uncorrelated, but not the other way around.

Definition

The covariance $\text{Cov}(X, Y)$ of two random variables X and Y is defined by

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

One can easily check that

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

If X and Y are independent, then $\text{Cov}(X, Y) = 0$. But the converse is not true. When $\text{Cov}(X, Y) = 0$, we say that X and Y are uncorrelated. Independence implies uncorrelated, but not the other way around.

Definition

The covariance $\mathrm{Cov}(X, Y)$ of two random variables X and Y is defined by

$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

One can easily check that

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

If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$. But the converse is not true. When $\mathrm{Cov}(X, Y) = 0$, we say that X and Y are uncorrelated. Independence implies uncorrelated, but not the other way around.

Proposition

(i) $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.

(ii) $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.

(iii) $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$.

(iv) $\mathrm{Cov}(\sum_{i=1}^{m} X_i, \sum_{j=1}^{n} Y_j) = \sum_{i=1}^{m} \sum_{j=1}^{n} \mathrm{Cov}(X_i, Y_j)$.

Combining (ii) and (iv) above, we get

$$\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{i<j} \mathrm{Cov}(X_i, X_j).$$

Proposition

(i) $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.

(ii) $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.

(iii) $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$.

(iv) $\mathrm{Cov}(\sum_{i=1}^{m} X_i, \sum_{j=1}^{n} Y_j) = \sum_{i=1}^{m} \sum_{j=1}^{n} \mathrm{Cov}(X_i, Y_j)$.

Combining (ii) and (iv) above, we get

$$\begin{aligned}\mathrm{Var}(\sum_{i=1}^{n} X_i) &= \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j) \\ &= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{i<j} \mathrm{Cov}(X_i, X_j).\end{aligned}$$

Let X and Y be discrete random variables with joint mass function $p(\cdot, \cdot)$. If y is a possible value of Y (i.e, $p_Y(y) > 0$), then

$$P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p(x, y)}{p_Y(y)}.$$

The function $x \mapsto \frac{p(x,y)}{p_Y(y)}$ is a mass function. It is called the conditional mass function of X given $Y = y$.

The function

$$p_{X|Y}(x|y) = \begin{cases} \frac{p(x,y)}{p_Y(y)}, & p_Y(y) > 0, \\ 0, & \text{otherwise} \end{cases}$$

is called the conditional mass function of X given Y.

Let X and Y be discrete random variables with joint mass function $p(\cdot,\cdot)$. If y is a possible value of Y (i.e, $p_Y(y) > 0$), then

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p(x,y)}{p_Y(y)}.$$

The function $x \mapsto \frac{p(x,y)}{p_Y(y)}$ is a mass function. It is called the conditional mass function of X given $Y = y$.

The function

$$p_{X|Y}(x|y) = \begin{cases} \frac{p(x,y)}{p_Y(y)}, & p_Y(y) > 0, \\ 0, & \text{otherwise} \end{cases}$$

is called the conditional mass function of X given Y.

If X and Y are independent, then for any possible value y of Y,

$$p_{X|Y}(x|y) = p_X(x), \quad x \in \mathbb{R}.$$

We always have

$$p(x,y) = p_Y(y)p_{X|Y}(x|y), \quad x,y \in \mathbb{R}.$$

We can similarly define the conditional mass function of Y given X:

$$p_{Y|X}(x|y) = \begin{cases} \frac{p(x,y)}{p_X(x)}, & p_X(x) > 0, \\ 0, & \text{otherwise} \end{cases}$$

We also have

$$p(x,y) = p_X(x)p_{Y|X}(y|x), \quad x,y \in \mathbb{R}.$$

If X and Y are independent, then for any possible value y of Y,

$$p_{X|Y}(x|y) = p_X(x), \quad x \in \mathbb{R}.$$

We always have

$$p(x, y) = p_Y(y)p_{X|Y}(x|y), \quad x, y \in \mathbb{R}.$$

We can similarly define the conditional mass function of Y given X:

$$p_{Y|X}(x|y) = \begin{cases} \frac{p(x,y)}{p_X(x)}, & p_X(x) > 0, \\ 0, & \text{otherwise} \end{cases}$$

We also have

$$p(x, y) = p_X(x)p_{Y|X}(y|x), \quad x, y \in \mathbb{R}.$$

Suppose $p_Y(y) = P(Y = y) > 0$. The conditional expectation of X given $Y = y$ is defined to be

$$E[X|Y = y] = \sum_x x p_{X|Y}(x|y).$$

The conditional expectation of $\phi(X)$ given $Y = y$ is defined to be

$$E[\phi(X)|Y = y] = \sum_x \phi(x) p_{X|Y}(x|y).$$

If $p_Y(y) = P(Y = y) = 0$, we define

$$E[X|Y = y] = 0, \quad E[\phi(X)|Y = y] = 0.$$

Suppose $p_Y(y) = P(Y = y) > 0$. The conditional expectation of X given $Y = y$ is defined to be

$$E[X|Y = y] = \sum_x x p_{X|Y}(x|y).$$

The conditional expectation of $\phi(X)$ given $Y = y$ is defined to be

$$E[\phi(X)|Y = y] = \sum_x \phi(x) p_{X|Y}(x|y).$$

If $p_Y(y) = P(Y = y) = 0$, we define

$$E[X|Y = y] = 0, \quad E[\phi(X)|Y = y] = 0.$$

Suppose that X and Y are jointly absolutely continuous with joint density $f(\cdot,\cdot)$. For any y with $f_Y(y) > 0$, the function

$$x \mapsto \frac{f(x,y)}{f_Y(y)}, \qquad x \in \mathbb{R}$$

is a probability density function. It is called the conditional density of X given $Y = y$.

More generally, the function

$$f_{X|Y}(x|y) = \begin{cases} \frac{f(x,y)}{f_Y(y)}, & f_Y(y) > 0, \\ 0, & \text{otherwise} \end{cases}$$

is called the conditional density of X given Y.

Suppose that X and Y are jointly absolutely continuous with joint density $f(\cdot,\cdot)$. For any y with $f_Y(y) > 0$, the function

$$x \mapsto \frac{f(x,y)}{f_Y(y)}, \qquad x \in \mathbb{R}$$

is a probability density function. It is called the conditional density of X given $Y = y$.

More generally, the function

$$f_{X|Y}(x|y) = \begin{cases} \frac{f(x,y)}{f_Y(y)}, & f_Y(y) > 0, \\ 0, & \text{otherwise} \end{cases}$$

is called the conditional density of X given Y.

If X and Y are independent, then for any y with $f_Y(y) > 0$,

$$f_{X|Y}(x|y) = f_X(x), \quad x \in \mathbb{R}.$$

We always have

$$f(x,y) = f_Y(y) f_{X|Y}(x|y), \quad x, y \in \mathbb{R}.$$

We can similarly define the conditional density of Y given X:

$$f_{Y|X}(x|y) = \begin{cases} \frac{f(x,y)}{f_X(x)}, & f_X(x) > 0, \\ 0, & \text{otherwise} \end{cases}$$

We also have

$$f(x,y) = f_X(x) f_{Y|X}(y|x), \quad x, y \in \mathbb{R}.$$

If X and Y are independent, then for any y with $f_Y(y) > 0$,

$$f_{X|Y}(x|y) = f_X(x), \quad x \in \mathbb{R}.$$

We always have

$$f(x,y) = f_Y(y) f_{X|Y}(x|y), \quad x, y \in \mathbb{R}.$$

We can similarly define the conditional density of Y given X:

$$f_{Y|X}(x|y) = \begin{cases} \frac{f(x,y)}{f_X(x)}, & f_X(x) > 0, \\ 0, & \text{otherwise} \end{cases}$$

We also have

$$f(x,y) = f_X(x) f_{Y|X}(y|x), \quad x, y \in \mathbb{R}.$$

For any y with $f_Y(y) > 0$, the conditional density $f_{X|Y}(x|y)$ allows us to define the conditional probability $P(X \in A|Y = y)$. For example, for any $a < b$,

$$P(X \in (a,b)|Y = y) = \int_a^b f_{X|Y}x|y)dx.$$

Suppose $f_Y(y) > 0$. The conditional expectation of X given $Y = y$ is defined to be

$$E[X|Y = y] = \int_{\infty}^{\infty} x f_{X|Y}(x|y)dx.$$

The conditional expectation of $\phi(X)$ given $Y = y$ is defined to be

$$E[\phi(X)|Y = y] = \int_{\infty}^{\infty} \phi(x) f_{X|Y}(x|y)dx.$$

If $f_Y(y) = 0$, we define

$$E[X|Y = y] = 0, \quad E[\phi(X)|Y = y] = 0.$$

For any y with $f_Y(y) > 0$, the conditional density $f_{X|Y}(x|y)$ allows us to define the conditional probability $P(X \in A|Y = y)$. For example, for any $a < b$,

$$P(X \in (a,b)|Y = y) = \int_a^b f_{X|Y}x|y)dx.$$

Suppose $f_Y(y) > 0$. The conditional expectation of X given $Y = y$ is defined to be

$$E[X|Y = y] = \int_\infty^\infty x f_{X|Y}(x|y)dx.$$

The conditional expectation of $\phi(X)$ given $Y = y$ is defined to be

$$E[\phi(X)|Y = y] = \int_\infty^\infty \phi(x) f_{X|Y}(x|y)dx.$$

If $f_Y(y) = 0$, we define

$$E[X|Y = y] = 0, \quad E[\phi(X)|Y = y] = 0.$$

Define
$$E[X|Y] = \varphi(Y),$$
where
$$\varphi(y) = E[X|Y = y].$$
Then $E[X|Y]$ is a random variable.

Theorem

$$E[E[X|Y]] = E[X].$$

This theorem can be used to be find the expectation of some complicated random variables.

Define

$$E[X|Y] = \varphi(Y),$$

where

$$\varphi(y) = E[X|Y = y].$$

Then $E[X|Y]$ is a random variable.

Theorem

$$E[E[X|Y]] = E[X].$$

This theorem can be used to be find the expectation of some complicated random variables.

Define

$$E[X|Y] = \varphi(Y),$$

where

$$\varphi(y) = E[X|Y = y].$$

Then $E[X|Y]$ is a random variable.

Theorem

$$E[E[X|Y]] = E[X].$$

This theorem can be used to be find the expectation of some complicated random variables.

Example

At a party n men throw their hats into the center of the room. The hats are mixed up and each randomly selects one. Suppose that those choosing their own hats depart, while others (those without a match) put their selected hats into the center of the room, mix them up and then re-select. Suppose that this process continues until each individual has his own hat. Let R_n be the number of rounds needed (when n individual are initially present). Find $E[R_n]$.

We know that, no matter how many people remain, there will, on average, be one match per round. Hence, we might guess that $E[R_n] = n$. This is indeed the case. We will show this by induction and conditioning on X_n, the number of matches in the first round.

Example

At a party n men throw their hats into the center of the room. The hats are mixed up and each randomly selects one. Suppose that those choosing their own hats depart, while others (those without a match) put their selected hats into the center of the room, mix them up and then re-select. Suppose that this process continues until each individual has his own hat. Let R_n be the number of rounds needed (when n individual are initially present). Find $E[R_n]$.

We know that, no matter how many people remain, there will, on average, be one match per round. Hence, we might guess that $E[R_n] = n$. This is indeed the case. We will show this by induction and conditioning on X_n, the number of matches in the first round.

Obviously $E[R_1] = 1$. Assume $E[R_k] = k$ for $k = 1, \ldots, n-1$. To find $E[R_n]$, we conditioning on X_n:

$$E[R_n] = E[E[R_n|X_n]] = \sum_{i=0}^{n} E[R_n|X_n = i]P(X_n = i).$$

Now, given a total of i matches in the first round, the number of rounds needed will equal to 1 plus the number of rounds needed when $n - i$ persons are initially present. Thus

$$E[R_n] = \sum_{i=0}^{n}(1 + E[R_{n-i}])P(X_n = i)$$

$$= 1 + E[R_n]P(X_n = 0) + \sum_{i=1}^{n} E[R_{n-i}]P(X_n = i).$$

Obviously $E[R_1] = 1$. Assume $E[R_k] = k$ for $k = 1, \ldots, n-1$. To find $E[R_n]$, we conditioning on X_n:

$$E[R_n] = E[E[R_n|X_n]] = \sum_{i=0}^{n} E[R_n|X_n = i]P(X_n = i).$$

Now, given a total of i matches in the first round, the number of rounds needed will equal to 1 plus the number of rounds needed when $n - i$ persons are initially present. Thus

$$E[R_n] = \sum_{i=0}^{n}(1 + E[R_{n-i}])P(X_n = i)$$

$$= 1 + E[R_n]P(X_n = 0) + \sum_{i=1}^{n} E[R_{n-i}]P(X_n = i).$$

By the induction hypothesis,

$$E[R_n] = 1 + E[R_n]P(X_n = 0) + \sum_{i=1}^{n}(n - i)P(X_n = i)$$
$$= 1 + E[R_n]P(X_n = 0) + n(1 - P(X_n = 0)) - E[X_n]$$
$$= E[R_n]P(X_n = 0) + n(1 - P(X_n = 0)).$$

From this we immediately get $E[R_n] = n$.

By the induction hypothesis,

$$\begin{aligned} E[R_n] =& 1 + E[R_n]P(X_n = 0) + \sum_{i=1}^{n}(n-i)P(X_n = i) \\ =& 1 + E[R_n]P(X_n = 0) + n(1 - P(X_n = 0)) - E[X_n] \\ =& E[R_n]P(X_n = 0) + n(1 - P(X_n = 0)). \end{aligned}$$

From this we immediately get $E[R_n] = n$.

If you are asked to find probabilities involving two independent discrete random random variables, try to break things up and then use independence.

Example

Let X and Y be independent geometric random variables with parameters p_1 and p_2 respectively. Find (a) $P(X \geq Y)$; (b) $P(X = Y)$.

$$P(X \geq Y) = \sum_{i=1}^{\infty} P(X \geq Y, Y = i) = \sum_{i=1}^{\infty} P(X \geq i, Y = i)$$

$$= \sum_{i=1}^{\infty} P(X \geq i) P(Y = i) = \sum_{i=1}^{\infty} (1 - p_1)^{i-1} (1 - p_2)^{i-1} p_2$$

$$= \frac{p_2}{1 - (1 - p_1)(1 - p_2)}.$$

If you are asked to find probabilities involving two independent discrete random random variables, try to break things up and then use independence.

Example

Let X and Y be independent geometric random variables with parameters p_1 and p_2 respectively. Find (a) $P(X \geq Y)$; (b) $P(X = Y)$.

$$
\begin{aligned}
P(X \geq Y) &= \sum_{i=1}^{\infty} P(X \geq Y, Y = i) = \sum_{i=1}^{\infty} P(X \geq i, Y = i) \\
&= \sum_{i=1}^{\infty} P(X \geq i) P(Y = i) = \sum_{i=1}^{\infty} (1 - p_1)^{i-1} (1 - p_2)^{i-1} p_2 \\
&= \frac{p_2}{1 - (1 - p_1)(1 - p_2)}.
\end{aligned}
$$

If you are asked to find probabilities involving two independent discrete random random variables, try to break things up and then use independence.

Example

Let X and Y be independent geometric random variables with parameters p_1 and p_2 respectively. Find (a) $P(X \geq Y)$; (b) $P(X = Y)$.

$$\begin{aligned} P(X \geq Y) &= \sum_{i=1}^{\infty} P(X \geq Y, Y = i) = \sum_{i=1}^{\infty} P(X \geq i, Y = i) \\ &= \sum_{i=1}^{\infty} P(X \geq i) P(Y = i) = \sum_{i=1}^{\infty} (1 - p_1)^{i-1} (1 - p_2)^{i-1} p_2 \\ &= \frac{p_2}{1 - (1 - p_1)(1 - p_2)}. \end{aligned}$$

$$
\begin{aligned}
P(X = Y) &= \sum_{i=1}^{\infty} P(X = Y, Y = i) = \sum_{i=1}^{\infty} P(X = i, Y = i) \\
&= \sum_{i=1}^{\infty} P(X = i)P(Y = i) = \sum_{i=1}^{\infty} (1 - p_1)^{i-1} p_1 (1 - p_2)^{i-1} p_2 \\
&= \frac{p_1 p_2}{1 - (1 - p_1)(1 - p_2)}.
\end{aligned}
$$