Two Discrete Random Variables

Joint Probability Distributions

The **joint probability mass function** of the discrete random variables X and Y, denoted as $f_{XY}(x, y)$, satisfies

(1) $f_{XY}(x, y) \geq 0$

(2) $\sum_x \sum_y f_{XY}(x, y) = 1$

(3) $f_{XY}(x, y) = P(X = x, Y = y)$ (5-1)

Two Discrete Random Variables

Marginal Probability Distributions

• The individual probability distribution of a random variable is referred to as its **marginal probability distribution**.

• In general, the marginal probability distribution of X can be determined from the joint probability distribution of X and other random variables.

Two Discrete Random Variables

Definition: Marginal Probability Mass Functions

If X and Y are discrete random variables with joint probability mass function $f_{XY}(x, y)$, then the **marginal probability mass functions** of X and Y are

$$f_X(x) = P(X = x) = \sum_y f_{XY}(x, y) \quad \text{and} \quad f_Y(y) = P(Y = y) = \sum_x f_{XY}(x, y) \tag{5-2}$$

where the first sum is over all points in the range of (X, Y) for which $X = x$ and the second sum is over all points in the range of (X, Y) for which $Y = y$

Two Discrete Random Variables

Conditional Probability Distributions

Given discrete random variables X and Y with joint probability mass function $f_{XY}(x, y)$ the **conditional probability mass function** of Y given $X = x$ is

$$f_{Y|x}(y) = f_{XY}(x, y)/f_X(x) \qquad \text{for } f_X(x) > 0 \tag{5-3}$$

Two Discrete Random Variables

Conditional Probability Distributions

Because a conditional probability mass function $f_{Y|x}(y)$ is a probability mass function for all y in R_x, the following properties are satisfied:

(1) $f_{Y|x}(y) \geq 0$

(2) $\sum_{y} f_{Y|x}(y) = 1$

(3) $P(Y = y | X = x) = f_{Y|x}(y)$ (5-4)

Two Discrete Random Variables

Definition: Conditional Mean and Variance

The **conditional mean** of Y given $X = x$, denoted as $E(Y|x)$ or $\mu_{Y|x}$, is

$$E(Y|x) = \sum_y y f_{Y|x}(y) \tag{5-5}$$

and the **conditional variance** of Y given $X = x$, denoted as $V(Y|x)$ or $\sigma^2_{Y|x}$, is

$$V(Y|x) = \sum_y (y - \mu_{Y|x})^2 f_{Y|x}(y) = \sum_y y^2 f_{Y|x}(y) - \mu^2_{Y|x}$$

Example 1

In a group of 100 sports car buyers, 40 bought alarm systems, 30 purchased bucket seats, and 20 purchased an alarm system and bucket seats. If a car buyer chosen at random bought an alarm system, what is the probability they also bought bucket seats?

Two Discrete Random Variables

Independence

For discrete random variables X and Y, if any one of the following properties is true, the others are also true, and X and Y are **independent.**

(1) $f_{XY}(x, y) = f_X(x) f_Y(y)$ for all x and y

(2) $f_{Y|x}(y) = f_Y(y)$ for all x and y with $f_X(x) > 0$

(3) $f_{X|y}(x) = f_X(x)$ for all x and y with $f_Y(y) > 0$

(4) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any sets A and B in the range of X and Y, respectively. (5-6)

Problem

The joint PMF of two random variables X and Y is given by

$$P_{XY}(x,y)= \begin{cases} k(2x+y) & x=1,2;\, y=1,2 \\ 0 & otherwise \end{cases}$$

where k is constant

a) What is the value of k?

b) Find the marginal PMFs of X and Y

c) Are X and Y independent?

Problem

- A fair coin is tossed three times. Let X be a random variable that takes the value 0 if the first toss is a tail and the value 1 if the first toss is a head. Also, let Y be a random variable that defines the total number of heads in the three tosses.

a) Determine the joint PMF of X and Y
b) Are X and Y independent?

Problem

The joint PMF of two random variables X and Y is given by

$$P_{XY}(x,y) = \begin{cases} \frac{1}{18}(2x+y) & x = 1,2; y = 1,2 \\ 0 & otherwise \end{cases}$$

where k is constant

a) What is the conditional PMF of Y given X?

b) What is the conditional PMF of X given Y?

Two Continuous Random Variables

Joint Probability Distribution

Definition

A **joint probability density function** for the continuous random variables X and Y, denoted as $f_{XY}(x, y)$, satisfies the following properties:

(1) $f_{XY}(x, y) \geq 0$ for all x, y

(2) $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)\, dx\, dy = 1$

(3) For any region R of two-dimensional space

$$P((X, Y) \in R) = \iint_R f_{XY}(x, y)\, dx\, dy \qquad (5\text{-}14)$$

Two Continuous Random Variables

Marginal Probability Distributions

Definition

If the joint probability density function of continuous random variables X and Y is $f_{XY}(x, y)$, the **marginal probability density functions** of X and Y are

$$f_X(x) = \int_y f_{XY}(x, y)\, dy \quad \text{and} \quad f_Y(y) = \int_x f_{XY}(x, y)\, dx \tag{5-15}$$

where the first integral is over all points in the range of (X, Y) for which $X = x$ and the second integral is over all points in the range of (X, Y) for which $Y = y$

Two Continuous Random Variables

Conditional Probability Distributions

Definition

Given continuous random variables X and Y with joint probability density function $f_{XY}(x, y)$, the **conditional probability density function** of Y given $X = x$ is

$$f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_X(x)} \quad \text{for} \qquad f_X(x) > 0 \tag{5-16}$$

Two Continuous Random Variables

Independence

Definition

For continuous random variables X and Y, if any one of the following properties is true, the others are also true, and X and Y are said to be **independent.**

(1) $f_{XY}(x, y) = f_X(x) f_Y(y)$ for all x and y

(2) $f_{Y|x}(y) = f_Y(y)$ for all x and y with $f_X(x) > 0$

(3) $f_{X|y}(x) = f_X(x)$ for all x and y with $f_Y(y) > 0$

(4) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any sets A and B in the range of X and Y, respectively. (5-19)

Problem

Let X and Y two continuous random variables whose joint PDF is given by

$$f_{XY}(x,y)= \begin{cases} x + Cy^2 & 0 \le x \le 1\ ;\ 0 \le y \le 1 \\ 0 & otherwise \end{cases}$$

(a) Find the constant C

(b) Find $P(0 \le x \ \le \frac{1}{2}\ , 0 \le y \ \le \frac{1}{2})$

Problem

X and Y two continuous random variables whose joint PDF is given by

$$f_{XY}(x,y) = \begin{cases} e^{-(x+y)} & 0 \le x < \infty\,;\; 0 \le x < \infty \\ 0 & otherwise \end{cases}$$

Are X and Y independent?

Problem

The joint PMF of two random variables X and Y is given by

$$f_{XY}(x,y)= \begin{cases} xe^{-x(y+y)1} & 0 \le x < \infty\,;\ 0 \le x < \infty \\ 0 & otherwise \end{cases}$$

a) What is the conditional PMF of Y given X?

b) What is the conditional PMF of X given Y?

Covariance

The *covariance* between two rv's X and Y is

$$\text{Cov}(X, Y) = E\left[(X - \mu_X)(Y - \mu_Y)\right]$$

$$= \begin{cases} \displaystyle\sum_x \sum_y (x - \mu_X)(y - \mu_Y)\, p(x, y) & \text{discrete} \\ \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)\, f(x, y)\, dx\, dy & \text{continuous} \end{cases}$$

Short-cut Formula for Covariance

$$\boxed{\operatorname{Cov}(X,Y) = E(XY) - \mu_X \cdot \mu_Y}$$

Theorem: X and Y indep implies $\mathrm{Cov}(X, Y) = 0$.

Proof:

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] \\ &= \mathrm{E}[X]\mathrm{E}[Y] - \mathrm{E}[X]\mathrm{E}[Y] \quad (X,\ Y \text{ indep}) \\ &= 0.\end{aligned}$$

Cov(X, Y) = 0 *does not imply* X and Y are independent!!

Definition: The **correlation** between X and Y is

$$\rho = \mathsf{Corr}(X,Y) \equiv \frac{\mathsf{Cov}(X,Y)}{\sqrt{\mathsf{Var}(X)\mathsf{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}.$$

Remark: Cov has "square" units; corr is unitless.

Corollary: X, Y indep implies $\rho = 0$.

Correlation Proposition

1. If a and c are either both positive or both negative, $\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$

2. For any two rv's X and Y,

   $-1 \le \text{Corr}(X, Y) \le 1.$

Theorem: $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$, *whether or not* X and Y are indep.

Remark: If X, Y are indep, the Cov term goes away.

Proof: By the work we did on a previous proof,

$$\begin{aligned}\text{Var}(X+Y) &= \text{E}[X^2] - (\text{E}[X])^2 + \text{E}[Y^2] - (\text{E}[Y])^2 \\ &\qquad + 2(\text{E}[XY] - \text{E}[X]\text{E}[Y]) \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y).\end{aligned}$$

Theorem: $\mathsf{Cov}(aX, bY) = ab\mathsf{Cov}(X, Y)$.

Proof:

$$
\begin{aligned}
\mathsf{Cov}(aX, bY) &= \mathsf{E}[aX \cdot bY] - \mathsf{E}[aX]\mathsf{E}[bY] \\
&= ab\mathsf{E}[XY] - ab\mathsf{E}[X]\mathsf{E}[Y] \\
&= ab\mathsf{Cov}(X, Y).
\end{aligned}
$$

Example: $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y)$.

Example:

$$\begin{aligned} &\text{Var}(X - 2Y + 3Z) \\ &= \text{Var}(X) + 4\text{Var}(Y) + 9\text{Var}(Z) \\ &\qquad - 4\text{Cov}(X, Y) + 6\text{Cov}(X, Z) - 12\text{Cov}(Y, Z). \end{aligned}$$

The proportions X and Y of two chemicals found in samples of an insecticide have the joint probability density function

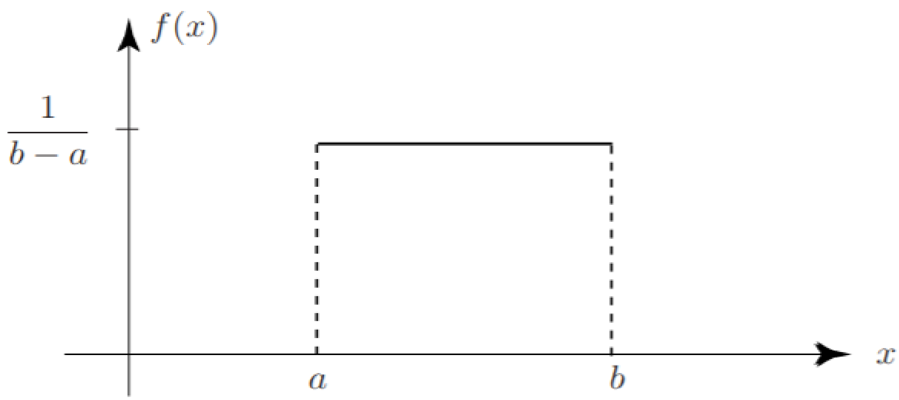
$$f(x,y)=\begin{cases}2, & 0\le x\le 1, 0\le y\le 1, 0\le x+y\le 1\\ 0, & \textit{elsewhere}\end{cases}$$

The random variable Z=X + Y denotes the proportion of the insecticide due to both chemicals combined.

1)Find E(Z) and V(Z)

Uniform Distribution (continuous)

The Uniform or Rectangular distribution has random variable X restricted to a finite interval $[a, b]$ and has $f(x)$ a constant over the interval.



The function $f(x)$ is defined by:

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

$$
\begin{aligned}
\mathsf{E}(X) &= \int_{-\infty}^{\infty} x f(x)\,dx = \int_a^b x\frac{1}{b-a}\,dx = \frac{1}{2(b-a)}\left[x^2\right]_a^b \\
&= \frac{b^2-a^2}{2(b-a)} \\
&= \frac{b+a}{2}
\end{aligned}
$$

Using the formula for the variance, we may write:

$$
\begin{aligned}
\mathsf{V}(X) &= \mathsf{E}(X^2) - [\mathsf{E}(X)]^2 \\
&= \int_a^b x^2.\frac{1}{b-a}\,dx - \left(\frac{b+a}{2}\right)^2 = \frac{1}{3(b-a)}\left[x^3\right]_a^b - \left(\frac{b+a}{2}\right)^2 \\
&= \frac{b^3-a^3}{3(b-a)} - \left(\frac{b+a}{2}\right)^2 \\
&= \frac{b^2+ab+a^2}{3} - \frac{b^2+2ab+a^2}{4} \\
&= \frac{(b-a)^2}{12}
\end{aligned}
$$

Problem

You arrive into a building and are about to take an elevator to your floor. Once you call the elevator it will take between 0 and 40 seconds to arrive to you. We will assume that the elevator arrives uniformly between the 0 and 40 seconds after you press the button. To calculate the probability that takes less than 15 seconds to arrive.

The Discrete Uniform Distribution

Suppose the possible values of a random variable from an experiment are a set of integer values occurring with the same frequency. That is, the integers 1 through k occur with equal probability. Then the probability of obtaining any particular integer in that range is 1/k and the probability distribution can be written

$$p(y) = 1 / k, \quad y = 1, 2, \ldots, k.$$

k depending on the range of existing values of the variable

The mean and variance

$$\mu = (k + 1) / 2$$

and

$$\sigma^2 = (k^2 - 1) / 12.$$

Problem

Let Y be the random variable describing the number of spots on the top face of the die. Then

$$p(y) = 1/6, \quad y = 1, 2, \ldots, 6,$$

which is the discrete uniform distribution with $k = 6$. The mean of Y is

$$\mu = (6 + 1)/2 = 3.5,$$

and the variance is

$$\sigma^2 = (36 - 1)/12 = 2.917.$$