Lecture 13

Joint Probability Distributions

Definition: The function $f(x, y)$ is a **joint density function** of the continuous random variables X and Y if

1. $f(x, y) \geq 0$, for all (x, y),
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$,
3. $P[(X, Y) \in A] = \iint_A f(x,y)\,dx\,dy$, for any region A in the xy plane.

Definition: The **marginal distributions** of X alone and of Y alone are

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{and} \quad h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

for the continuous case.

Statistical Independence

Definition: Let X and Y be two random variables, discrete or continuous, with joint probability distribution $f(x, y)$ and marginal distributions $g(x)$ and $h(y)$, respectively.
The random variables X and Y are said to be **statistically independent** if and only if

$$f(x, y) = g(x)h(y)$$

for all (x, y) within their range.

Definition: Let $X_1, X_2, \ldots, X_n$ be n random variables, discrete or continuous, with joint probability distribution $f(x_1, x_2, \ldots, x_n)$ and marginal distribution $f_1(x_1), f_2(x_2), \ldots, f_n(x_n)$, respectively. The random variables $X_1, X_2, \ldots, X_n$ are said to be **mutually statistically independent** if and only if

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1)f_2(x_2)\cdots f_n(x_n)$$

for all $(x_1, x_2, \ldots, x_n)$ within their range.

Example 12.2: A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x, y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify that $f(x, y)$ is a joint density function.

(b) Find $P[(X, Y) \in A]$, where $A = \{(x, y) \mid 0 < x < \frac{1}{2},\ \frac{1}{4} < y < \frac{1}{2}\}$.

(c) Find the marginal distribution of X alone and of Y alone $g(x)$ and $h(y)$.

Solution: (a) The integration of $f(x, y)$ over the whole region is

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = \int_0^1\int_0^1 \frac{2}{5}(2x+3y)dxdy$$

$$= \int_0^1 \left(\frac{2x^2}{5} + \frac{6xy}{5}\right)\Big|_{x=0}^{x=1} dy$$

$$= \int_0^1 \left(\frac{2}{5} + \frac{6y}{5}\right)dy = \left(\frac{2y}{5} + \frac{3y^2}{5}\right)\Big|_0^1 = \frac{2}{5} + \frac{3}{5} = 1$$

(b) To calculate the probability, we use

$$P[(X, Y) \in A] = P\left(0 < X < \frac{1}{2}, \frac{1}{4} < Y < \frac{1}{2}\right)$$

$$= \int_{1/4}^{1/2}\int_0^{1/2} \frac{2}{5}(2x+3y)dxdy$$

$$= \int_{1/4}^{1/2} \left(\frac{2x^2}{5} + \frac{6xy}{5}\right)\Big|_{x=0}^{x=1/2} dy = \int_{1/4}^{1/2} \left(\frac{1}{10} + \frac{3y}{5}\right)dy$$

$$= \left(\frac{y}{10} + \frac{3y^2}{10}\right)\Big|_{1/4}^{1/2} = \frac{13}{160}.$$

(c) By definition,

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^1 \frac{2}{5}(2x+3y)dy = \left(\frac{4xy}{5} + \frac{6y^2}{10}\right)\Big|_{y=0}^{y=1} = \frac{4x+3}{5},$$

for $0 \le x \le 1$, and $g(x) = 0$ elsewhere. Similarly,

$$h(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^1 \frac{2}{5}(2x+3y)dx = \frac{2(1+3y)}{5},$$

for $0 \le y \le 1$, and $h(y) = 0$ elsewhere.

Example 12.6: Suppose that the shelf life, in years, of a certain perishable food product packaged in cardboard containers is a random variable whose probability density function is given by

$$f(x) = \begin{cases} e^{-x}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Let X_1, X_2, and X_3 represent the shelf lives for three of these containers selected independently and find $P(X_1 < 2,\ 1 < X_2 < 3,\ X_3 > 2)$.

Solution: Since the containers were selected independently, we can assume that the random variables X_1, X_2, and X_3 are statistically independent, having the joint probability density

$$f(x_1, x_2, x_3) = f(x_1)f(x_2)f(x_3) = e^{-x_1}e^{-x_2}e^{-x_3} = e^{-x_1-x_2-x_3},$$

for $x_1 > 0, x_2 > 0, x_3 > 0$, and $f(x_1, x_2, x_3) = 0$ elsewhere. Hence

$$P(X_1 < 2,\ 1 < X_2 < 3, X_3 > 2) = \int_2^{\infty}\int_1^{3}\int_0^{2} e^{-x_1-x_2-x_3}\,dx_1dx_2dx_3$$

$$= (1-e^{-2})(e^{-1}-e^{-3})e^{-2} = 0.0372.$$

Definition: Let X and Y be random variables with joint probability distribution $f(x, y)$. The mean, or expected value, of the random variable $g(X, Y)$ is

$$\boxed{\mu_{g(X,Y)} = E[g(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy} \qquad \text{if } X \text{ and } Y \text{ are continuous.}$$

Note that if $g(X, Y) = X$, we have

$$\boxed{E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dydx = \int_{-\infty}^{\infty} xg(x)dx \qquad \text{(continuous case),}}$$

where $g(x)$ is the marginal distribution of X.

Similarly, we define

$$\boxed{E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dxdy = \int_{-\infty}^{\infty} yh(y)dy \qquad \text{(continuous case),}}$$

where $h(y)$ is the marginal distribution of the random variable Y.

Definition: Let X and Y be random variables with joint probability distribution $f(x, y)$. The covariance of X and Y is

$$\boxed{\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)f(x,y)dxdy}, \quad X \text{ and } Y \text{ are continuous.}$$

Theorem: The covariance of two random variables X and Y with means μ_X and μ_Y, respectively, is given by

$$\boxed{\mathrm{Cov}(X, Y) = \sigma_{XY} = E(XY) - \mu_X\mu_Y = E(XY) - E(X)E(Y).}$$

Theorem: Let X and Y be two independent random variables. Then

$$\boxed{E(XY) = E(X)E(Y).}$$

Corollary: Let X and Y be two independent random variables. Then $\boxed{\sigma_{XY} = 0.}$

Definition: Let X and Y be random variables with covariance σ_{XY} and standard deviations σ_X and σ_Y, respectively. The correlation coefficient of X and Y is

$$\boxed{\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}. \qquad -1 \le \rho_{XY} \le 1}$$

$$\boxed{\sigma_X = \sqrt{E(X^2) - E^2(X)}} \qquad \boxed{\sigma_{XY} = E(XY) - E(X)E(Y)} \qquad \boxed{\sigma_y = \sqrt{E(Y^2) - E^2(Y)}}$$

Example 12.10: The fraction X of male runners and the fraction Y of female runners who compete in marathon races are described by the joint density function

$$f(x, y) = \begin{cases} 8xy, & 0 \le y \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the correlation coefficient between X and Y.

Solution: We first compute the marginal density functions. They are

$$g(x) = \begin{cases} 4x^3, & 0 \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

and

$$h(y) = \begin{cases} 4y(1-y^2), & 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

From these marginal density functions, we compute

$$\mu_X = E(X) = \int_0^1 4x^4 dx = \frac{4}{5} \text{ and } \mu_Y = E(Y) = \int_0^1 4y^2(1-y^2)dy = \frac{8}{15}.$$

$$E(X^2) = \int_0^1 4x^5 dx = \frac{2}{3} \text{ and } E(Y^2) = \int_0^1 4y^3(1-y^2)dy = 1 - \frac{2}{3} = \frac{1}{3},$$

$$\sigma_X^2 = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{75} \text{ and } \sigma_Y^2 = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{11}{225}.$$

From the joint density function given above, we have

$$E(XY) = \int_0^1 \int_y^1 8x^2y^2 dxdy = \frac{4}{9}.$$

Then
$$\sigma_{XY} = E(XY) - \mu_X \mu_Y = \frac{4}{9} - \left(\frac{4}{5}\right)\left(\frac{8}{15}\right) = -\frac{4}{225}.$$

Hence,
$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{-4/225}{\sqrt{(2/75)(11/225)}} = \frac{-4}{\sqrt{66}}.$$

Theorem: If X and Y are random variables with joint probability distribution $f(x, y)$ and a, b, and c are constants, then

$$\sigma_{aX+bY+c}^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}.$$

Corollary 1: Setting $b = 0$, we see that

$$\sigma_{aX+c}^2 = a^2\sigma_X^2 = a^2\sigma^2.$$

Corollary 2: Setting $a = 1$ and $b = 0$, we see that

$$\sigma_{X+c}^2 = \sigma_X^2 = \sigma^2.$$

Corollary 3: Setting $b = 0$ and $c = 0$, we see that

$$\sigma_{aX}^2 = a^2\sigma_X^2 = a^2\sigma^2.$$

Corollary 4: If X and Y are independent random variables, then

$$\sigma_{aX+bY}^2 = a^2\sigma_X^2 + b^2\sigma_Y^2.$$

Corollary 5: If X and Y are independent random variables, then

$$\sigma_{aX-bY}^2 = a^2\sigma_X^2 + b^2\sigma_Y^2.$$

Example: The continuous random variables X & Y have joint density function given by:

$$f(x, y) = \begin{cases} C(6-x-y), & 0<x<2 \; \& \; 2<y<4, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Calculate C.

(b) $P(X+Y>3)$

(c) Find the marginal distribution of X alone and of Y alone $g(x)$ and $h(y)$.

(d) Check independence of X & Y.

(e) Find Cov($2X$, $5Y$).

(f) Find the correlation coefficient between X and Y.

Solution: (a) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = \int_2^4\int_0^2 C(6-x-y)dxdy = 1$

$$\Rightarrow C\int_2^4 (6x - \frac{x^2}{2} - xy)\Big|_{x=0}^{x=2} dy = 1$$

$$\Rightarrow C\int_2^4 (10-2y)dy = 1$$

$$\Rightarrow C[(10y - y^2)\Big|_{y=2}^{y=4}] = 1$$

$$\Rightarrow C(20-12) = 1 \Rightarrow C = \frac{1}{8}$$

(b) To calculate the probability, we use

$P[(X, Y) \in A] = P(X+Y>3)$

$$\int_2^3\int_{3-y}^2 \frac{1}{8}(6-x-y)dxdy + \int_3^4\int_0^2 \frac{1}{8}(6-x-y)dxdy$$

$$\Rightarrow \frac{1}{8}\int_2^3 (6x - \frac{x^2}{2} - xy)\Big|_{x=3-y}^{x=2} dy + \frac{1}{8}\int_3^4 (6x - \frac{x^2}{2} - xy)\Big|_{x=0}^{x=2} dy$$

$$\Rightarrow \frac{-1}{16}\int_2^3 (y^2 - 8y + 7)dy + \frac{1}{8}\int_3^4 (10-2y)dy = \frac{19}{24}$$

<u>Method 2:</u>

$$P(X+Y>3) = 1 - \int_2^3\int_0^{3-y} \frac{1}{8}(6-x-y)dxdy$$

$$= 1 - \int_2^3 \frac{1}{8}(6x - \frac{x^2}{2} - xy)\Big|_0^{3-y} dy = \frac{19}{24}$$

(c) By definition,

$$g(x)=\int_{-\infty}^{\infty} f(x,y)dy=\int_{2}^{4}\frac{1}{8}(6-x-y)dy=\frac{1}{4}(3-x),$$

for $0\le x\le 2$, and $g(x)=0$ elsewhere. Similarly,

$$h(y)=\int_{-\infty}^{\infty} f(x,y)dx=\int_{0}^{2}\frac{1}{8}(6-x-y)dx=\frac{1}{4}(5-y),$$

for $2\le y\le 4$, and $h(y)=0$ elsewhere.

(d) $\boxed{f(x,y)=\frac{1}{8}(6-x-y)\neq\frac{1}{4}(3-x).\frac{1}{4}(5-y)=g(x)h(y)}$

$\Rightarrow X\ \&\ Y$ are not independent.

(e) $\boxed{\mathrm{Cov}(2X,5Y)=\sigma_{2X,5Y}=E(2X*5Y)-\mu_{2X}\mu_{5Y}=10[E(XY)-\mu_X\mu_Y]}$

Solution: from the marginal density functions:

$$g(x)=\begin{cases}\frac{1}{4}(3-x), & 0\le x\le 2,\\ 0, & \text{elsewhere.}\end{cases}$$

and

$$h(y)=\begin{cases}\frac{1}{4}(5-y), & 2\le y\le 4,\\ 0, & \text{elsewhere.}\end{cases}$$

we compute $\sigma_{2X,5Y}=E(2X*5Y)-\mu_{2X}\mu_{5Y}=10[E(XY)-\mu_X\mu_Y]$

$$\mu_X=E(X)=\int_0^2\frac{x}{4}(3-x)dx=\frac{5}{6}\text{ and }\mu_Y=E(Y)=\int_2^4\frac{y}{4}(5-y)dy=\frac{17}{6}.$$

$$E(X^2)=\int_0^2\frac{x^2}{4}(3-x)dx=1\text{ and }E(Y^2)=\int_2^4\frac{y^2}{4}(5-y)dy=\frac{25}{3},$$

$$\sigma_X^2=1-\left(\frac{5}{6}\right)^2=\frac{11}{36}\text{ and }\sigma_Y^2=\frac{25}{3}-\left(\frac{17}{6}\right)^2=\frac{11}{36}.$$

From the joint density function given above, we have

$$E(XY)=\int_2^4\int_0^2\frac{xy}{8}(6-x-y)dxdy=\frac{56}{24}$$

Then

$$\sigma_{XY}=E(XY)-\mu_X\mu_Y=\frac{56}{24}-\left(\frac{5}{6}\right)\left(\frac{17}{6}\right)=-\frac{1}{36}.$$

$$\Rightarrow\sigma_{2X,5Y}=10\sigma_{XY}=-\frac{5}{18}.$$

Hence,

$$\rho_{XY}=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}=\frac{-1/36}{\sqrt{(11/36)(11/36)}}=\frac{-1}{11}.$$

Try $\rho_{2X,5Y}$!!!!!!!