5 Multivariate Random Variables

Chapter Preview

This set of notes will examine pairs of random variables that are potentially related. To describe the relationship between the random variables, we will begin by defining joint probability distributions and then look at the covariance and correlation between two random variables.

5.1: Jointly Distributed Random Variables

Joint Probability Functions

Def: Let X and Y be two **discrete** random variables defined on the sample space Ω of an experiment. The joint probability mass function $p(x, y)$ is defined for each pair of numbers (x,y) by: $p(x, y) = P(X = x \text{ and } Y = y)$

Note: $p(x, y) \geq 0$ and $\sum_x \sum_y p(x, y) = 1$

Def: Let X and Y be two **continuous** random variables defined on the sample space Ω of an experiment. The joint probability density function $f(x, y)$ for variables X and Y is a function satisfying:

(i) $f(x, y) \geq 0$
(ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dydx = 1$

Note: For any two-dimensional set A, $P((X, Y) \in A) = \int \int_A f(x, y) dydx$.

Note: If A is a two-dimensional rectangle $\{(x, y) : a \leq x \leq b, c \leq y \leq d\}$, then:

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

Example: Let $f(x, y) = cxy$, $\cancel{0 \leq x, y \leq 1}$ $0 \leq x \leq 1, 0 \leq y \leq 1$. Find c so that $f(x, y)$ is a valid joint pdf.

↑ joint pdf for two Continuous RVs, X and Y.

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dydx = \int_0^1 \underbrace{\int_0^1 cxy\,dy}\,dx = \int_0^1 cx\,dx \int_0^1 y\,dy = \int_0^1 cx\,dx \cdot \left[\frac{y^2}{2}\right]_0^1$$

First, we integrate with respect to y.

$$= \underbrace{\int_0^1 \frac{1}{2} cx\,dx} = \left[\frac{cx^2}{4}\right]_0^1 = \frac{c}{4} = 1 \rightarrow \boxed{c=4}$$

Now, integrate with respect to x.

Marginal Probability Functions

Def: The marginal probability mass function of a discrete random variable X with joint pmf $p(x, y)$ is denoted by $p_X(x)$ and is given by:

$$\boxed{p_X(x) = \sum_y p(x, y)}$$

sum the joint pdf over "y" which just leaves "x"

Def: The marginal probability density function of a continuous random variable X with joint pdf $f(x, y)$ is denoted by $f_X(x)$ and is given by:

$$\boxed{f_X(x) = \int_y f(x, y)dy}$$

integrate out "y" which just leaves "x"

Example: An apartment complex has the following joint distribution of X bedrooms and Y bathrooms. Determine the marginal pmfs and calculate the probability a randomly selected apartment has more bedrooms than bathrooms (i.e., $P(X > Y)$).

# of bedrooms

# of bathrooms

$Y \setminus X$	1	2	3	Total
1	0.3	0.3	0	0.6
2	0	0.3	0.1	0.4
Total	0.3	0.6	0.1	1.0

joint pmf

$$\boxed{p_Y(y) = \begin{cases} 0.6 & y=1 \\ 0.4 & y=2 \end{cases}}$$

(marginal pmf of Y)

$$\boxed{p_X(x) = \begin{cases} 0.3 & x=1 \\ 0.6 & x=2 \\ 0.1 & x=3 \end{cases}}$$

(marginal pmf of X)

ex) $P(X>Y) = p(2,1) + p(3,1) + p(3,2)$

$= 0.3 + 0 + 0.1$

$= \boxed{0.4 = P(X>Y)}$

Example: Let X and Y have the joint pdf, $f(x,y)$, as follows. Confirm $f(x,y)$ is a valid pdf, determine the marginal pdfs, and calculate $P(X < 1/2, Y < 1/2)$.

"and"

joint pdf

$$f(x,y) = \begin{cases} \frac{4}{3}(xy+x) & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

① Confirm $f(x,y)$ is a valid (joint) pdf.

(i) $f(x,y) \ge 0$ for all x,y? Yes ✓

(ii) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dy\,dx = 1$?

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dy\,dx = \int_0^1\int_0^1 \frac{4}{3}(xy+x)\,dy\,dx = \frac{4}{3}\int_0^1 \left[\frac{xy^2}{2} + xy\right]_0^1 dx$$

$$= \frac{4}{3}\int_0^1 \left(\frac{x}{2} + x\right)dx = \frac{4}{3}\int_0^1 \frac{3}{2}x\,dx = \frac{4}{\cancel{3}}\cdot\frac{\cancel{3}}{2}\cdot\int_0^1 x\,dx = \left[2\cdot\frac{x^2}{2}\right]_0^1 = 1 \checkmark$$

• By (i) and (ii), $f(x,y)$ is a valid (joint) pdf.

② Find the marginal pdfs.

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \frac{4}{3}\int_0^1 (xy+x)\,dy = \frac{4}{3}\cdot\left[\frac{xy^2}{2} + xy\right]_0^1 = 2x \quad (0 \le x \le 1)$$

$$\boxed{f_X(x) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \frac{4}{3}\int_0^1 (xy+x)\,dx = \frac{4}{3}\cdot\left[\frac{x^2y}{2} + \frac{x^2}{2}\right]_0^1 = \frac{2}{3}y + \frac{2}{3} \quad (0 \le y \le 1)$$

$$\boxed{f_Y(y) = \begin{cases} \frac{2}{3}y + \frac{2}{3} & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}}$$

③ $P(X < \frac{1}{2} \text{ and } Y < \frac{1}{2}) = \int_0^{1/2}\int_0^{1/2} \frac{4}{3}(xy+x)\,dy\,dx = \frac{4}{3}\int_0^{1/2}\left[\frac{xy^2}{2} + xy\right]_0^{1/2} dx$

$$= \frac{4}{3}\int_0^{1/2} \frac{5}{8}x\,dx = \frac{4}{3}\cdot\frac{5}{8}\int_0^{1/2} x\,dx = \left[\frac{20}{24}\cdot\frac{x^2}{2}\right]_0^{1/2} = \boxed{\frac{5}{48}}$$

Independent Random Variables

Def: Two discrete random variables are independent if $p(x, y) = p_X(x) \cdot p_Y(y)$.

Def: Two continuous random variables are independent if $f(x, y) = f_X(x) \cdot f_Y(y)$.

Def: If two random variables are not independent, they are dependent.

Note: You can also check independence by seeing if the conditional distribution equals the marginal distribution, i.e., $p_{Y|X}(y|x) = p_Y(y)$ OR $p_{X|Y}(x|y) = p_X(x)$.

Example: Let X and Y have the joint pdf, $f(x)$, as follows. Are X and Y independent?

$$f(x, y) = \begin{cases} \frac{4}{3}(xy + x) & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Example: An apartment complex has the following joint distribution of X bedrooms and Y bathrooms. Are X and Y independent?

$Y \setminus X$	1	2	3
1	0.3	0.3	0
2	0	0.3	0.1

Conditional Distributions

Def: Let X and Y be two discrete distributions with joint pmf $p(x, y)$ and marginal pmf of X $p_X(x)$. Then for any X value of x for which $p_X(x) > 0$, the conditional probability mass function of Y given that $X = x$ is:

$$p_{Y|X}(y|x) = \frac{p(x, y)}{p_X(x)}$$

Def: Let X and Y be two continuous distributions with joint pdf $f(x, y)$ and marginal pdf of X $f_X(x)$. Then for any X value of x for which $f_X(x) > 0$, the conditional probability density function of Y given that $X = x$ is

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}$$

Example: For the apartment example, determine the conditional pmf $p_{Y|X}(y|2)$. The joint pmf, $p_{X,Y}(x, y)$, is provided below.

$Y \setminus X$	1	2	3
1	0.3	0.3	0
2	0	0.3	0.1

Example: Let X and Y have the joint pdf, $f(x, y)$, as follows. Determine the conditional pdf, $f_{Y|X}(y|x)$.

$$f(x, y) = \begin{cases} \frac{4}{3}(xy + x) & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

5.2: Expected Values, Covariance, and Correlation

Theorem: Let X and Y be jointly distributed pmf p(x,y) (if X and Y are discrete) or pdf f(x,y) (if X and Y are continuous). The expected value of the function h(X,Y), E[h(X,Y)], is given by:

$$E[h(X,Y)] = \sum_x \sum_y h(x,y) \cdot p(x,y) \text{ (if X and Y are discrete)}$$

$$E[h(X,Y)] = \int_x \int_y h(x,y) \cdot f(x,y) dy dx \text{ (if X and Y are continuous)}$$

Def: The covariance between two random variables is given by:

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

Def: The correlation coefficient between two random variables is given by:

$$Corr(X,Y) = \rho_{XY} = \rho = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

Example: For the apartment example, determine $E[X]$, $E[Y]$, $E[XY]$, and $Cov(X,Y)$. The joint pmf, $p_{X,Y}(x,y)$, is provided below.

$Y \setminus X$	1	2	3
1	0.3	0.3	0
2	0	0.3	0.1

Theorem: For any two random variables X and Y, $-1 \leq \rho_{XY} \leq 1$.

Theorem: If X and Y are independent then $\rho_{XY} = 0$

Note: $\rho_{XY} = 0$ does not imply that two random variables are independent!

Some illustrations of correlation:

Linear Combinations of Random Variables

Let $Z = aX + bY$, where X and Y are random variables and a and b are constants. We are interested in determining the mean and variance of Z, the linear combination of X and Y.

Let X and Y be random variables and a and b be constants.

Theorem: $E[aX + bY] = aE[X] + bE[Y]$

Theorem: $Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2ab \cdot Cov(X,Y)$

Theorem: $Var(aX + bY) = a^2Var(X) + b^2Var(Y)$, if X and Y are independent

Proof:

Bivariate Normal Distribution

The <u>bivariate normal distribution</u> of (X,Y) has the following joint probability distribution function:

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)\right]$$

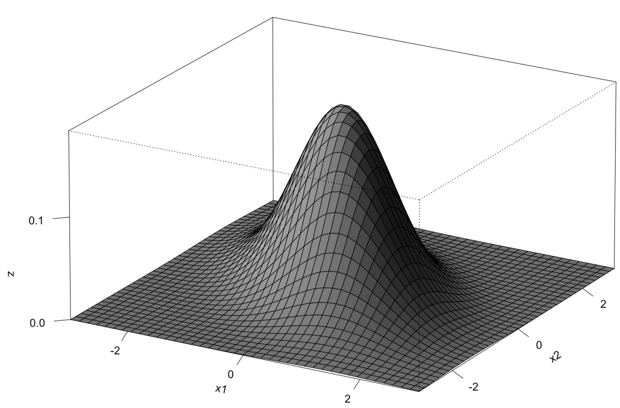
$$-\infty < x < \infty, -\infty < y < \infty$$

where $E[X] = \mu_X$, $Var(X) = \sigma_X^2$, $E[Y] = \mu_Y$, $Var(Y) = \sigma_Y^2$, $\rho = Corr(X,Y)$

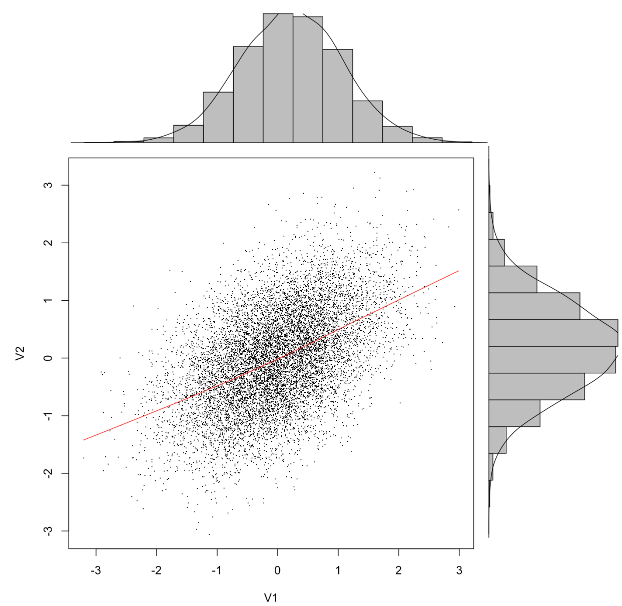
We won't work directly with the multivariate normal distribution, but there are some useful properties of normal random variables which we will use.

<u>Theorem:</u> If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, then:
$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2 + 2Cov(X,Y))$, that is, the sum of normal random variables is a normal random variable.

<u>Theorem:</u> If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ and X and Y are independent, then:
$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.



(a) Bivariate Normal Distribution ($\rho = 0.5$)



(b) 1,000 random samples from a bivariate normal distribution ($\rho = 0.5$)