Joint Probability Distributions

Two discrete random variables

X: x_1, x_2, ..., x_m

Y: y_1, y_2, ..., y_n

Joint pmf

$$p(x_i, y_j) = P(X = x_i and Y = y_j)$$

Suppose $m = 3$ and $n = 4$

p	y_1	y_2	y_3	y_4	p_X
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	$p(x_1, y_3)$	$p(x_1, y_4)$	$p_X(x_1)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	$p(x_2, y_3)$	$p(x_2, y_4)$	$p_X(x_2)$
x_3	$p(x_3, y_1)$	$p(x_3, y_2)$	$p(x_3, y_3)$	$p(x_3, y_4)$	$p_X(x_2)$
p_Y	$p_Y(y_1)$	$p_Y(y_2)$	$p_Y(y_3)$	$p_Y(y_4)$	

Marginal pmfs

Example Exe 5.3

Independence of two random variables

X and Y are said to be independent if for every pair of x and y,

$$p(x, y) = p_X(x)p_Y(x).$$

If X and Y are known to be independent and their pmfs are p_1 and p_2, then the joint pmf for (X, Y) is $p_1(x)p_2(y)$.

In general, random variables X_1, X_2, ..., X_n are independent if and only if the joint pmf is the product of the pmfs of X_i's, i.e.,

$$p(x_1, x_2, \ldots, x_n) = p_1(x_1)p_2(x_2)\cdots p_n(x_n)$$

Example Exe, are X and Y independent?

Conditional pmf

$$P_{X|Y}(x_i \mid y_j) = P(X = x_i \mid Y = y_j)$$
$$= \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p(x_i, y_j)}{p_Y(y_j)}$$

Example Exe 5.3

Two continuous random variables
$X, Y \Rightarrow (X, Y)$
$P(X = x, Y = y) = 0$
Joint pdf of X and Y: $f(x, y)$
For any 2-dimensional rectangle $A = \{(x, y) : a \leq x \leq b, c \leq y \leq d\}$

$$P((X, Y) \in A) = P(a \leq X \leq b, c \leq Y \leq d)$$
$$= \int_a^b \int_c^d f(x, y)dxdy$$

Density Surface
Probability=Volume

Example Suppose the joint pdf of X and Y is as follows

$$f(x,y) = \begin{cases} \frac{3}{28}(x+y^2) & 0 \leq x \leq 2, 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Marginal pdfs of X and Y

pdf of X without reference to Y

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy$$

pdf of Y without reference to X

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx$$

Example

Independence of two random variables

X and Y are said to be independent if

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

If X and Y are known to be independent and their pdfs are $f_1(x)$ and $f_2(y)$, then the joint pdf for (X,Y) is

$$f(x,y) = f_1(x)f_2(y)$$

In general, random variables X_1, X_2, ..., X_n are independent if and only if the joint pdf is the product of the pdfs of X_i's, i.e.,

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

Example Are X and Y independent?

Conditional pdf

Conditional pdf of X given $Y = y$

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

Conditional pdf of Y given $X = x$

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

Example

Expected value

$$(X,Y) \sim \begin{cases} p(x,y) & \text{if } X \text{ and } Y \text{ are discrete} \\ f(x,y) & \text{if } X \text{ and } Y \text{ are continuous} \end{cases}$$

$$E(h(X,Y)) = \begin{cases} \Sigma_x \Sigma_y h(x,y)p(x,y) & \text{discrete} \\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y)f(x,y)dxdy & \text{continuous} \end{cases}$$

Covariance between X and Y

$$Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y))$$

$$= \begin{cases} \Sigma_x \Sigma_y (x-\mu_X)(y-\mu_Y)p(x,y) & \text{discrete} \\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_X)(y-\mu_Y)f(x,y)dxdy & \text{continuous} \end{cases}$$

$$Cov(X,Y) = E(X \cdot Y) - \mu_X\mu_Y$$

Correlation Coefficient

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

Properties:
1. $Corr(aX+b, cY+d) = Corr(X,Y)$, if $ac > 0$
2. $-1 \le Corr(X,Y) \le 1$
3. if X and Y are independent, $\rho_{X,Y} = 0$.
But $\rho_{X,Y} = 0$ doesn't imply independence.
4. $\rho = 1$ or -1, iff $Y = aX + b$ $(a \neq 0)$

Random Samples

The rv's X_1, X_2, ..., X_n are said to form a (simple) random sample of size n if
1. The X_i's are independent.

2. Every X_i has the same probability distribution.

Sample or an observed sample

Statistic
A statistic is any quantity whose value can be calculated from sample data.

A statistic is a random variable

Observed statistic

Sampling Distribution
The probability distribution of a statistic

Two ways to derive sampling distribution
1. Probability rules
2. Computer simulation

Example 5.20

Distribution of Sample Mean and Sum

From now on, random sample: X_1, X_2, …, X_n

Two formulas

$$E(a_1X_1+a_2X_2+\cdots+a_nX_n) = a_1E(X_1)+a_2E(X_2)+\cdots+a_nE(X_n)$$
$$V(a_1X_1+a_2X_2+\cdots+a_nX_n) = a_1^2V(X_1)+a_2^2V(X_2)+\cdots+a_n^2V(X_n)$$

Two statistics
Sample Sum

$$T_0 = X_1 + X_2 + \cdots X_n$$

Sample Mean

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Assume X_1, X_2, …, X_n from a distribution with mean μ and variance σ^2

1. $E(T_0) = n\mu$
2. $V(T_0) = n\sigma^2$
3. $E(\bar{X}) = \mu$
4. $V(\bar{X}) = \sigma^2/n$

Normal population distribution, $N(\mu, \sigma)$
Sampling distributions for T_0 **and** $\bar{X}$

$$T_0 \sim N(n\mu, \sqrt{n}\sigma)$$
$$\bar{X} \sim N(\mu, \sigma/n)$$

Example 5.25

Central Limit Theorem
X_1, X_2, $\ldots$, $X_n \sim$ **Distribution with** μ **and** σ^2

$$T_0 \textbf{ approximately } N(n\mu, \sqrt{n}\sigma)$$
$$\bar{X} \textbf{ approximately } N(\mu, \sigma/n)$$

When $n > 30$
Example 5.26