Chapter 5.1 and 5.2

Joint Probability Distributions

DALHOUSIE UNIVERSITY

Outline

- Jointly Distributed Random Variables

- Expected Values, Covariance and Correlation

Joint Distribution for Discrete Random Variables

- Let X and Y be two discrete random variables defined on the sample space $\mathbf{S}$, the joint probability mass function (joint pmf) of X and Y, $p(x, y)$, is defined for each pair (x, y) by

$$p(x, y) = P(X = x \text{ and } Y = y).$$

- The joint pmf satisfies the following:
  - $0 \leq p(x, y) \leq 1$
  - $\sum_x \sum_y p(x, y) = 1$
  - $P[(X, Y) \in A] = \sum\sum_{(x,y)\in A} p(x, y)$

- More than two discrete random variables: If $X_1, X_2, \ldots, X_n$ are all discrete random variables, the joint pmf is

$$p\left(x_1, x_2, \ldots, x_n\right) = P\left(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\right)$$

Example

- An insurance company offers both homeowner and car insurance to its customers. For each type of policy a deductible amount needs to be specified. For car insurance, the options are $100 and $250. For homeowner insurance, the deductible options are $0, $100 and $200.
- Let X = car deductible and Y = homeowner deductible for a randomly selected customer that has both car and homeowner policies.Given the joint pmf table of X and Y as below,

p(x,y)		Y		
		0	100	200
X	100	0.2	0.1	0.2
	250	0.05	0.15	0.3

  - 1) what is the probability that a randomly selected customer has $100 deductible for both policies.
  - 2) what is the probability that a randomly selected customer has homeowner deductible of at least $100?

Marginal probability mass function

- The marginal pmf of X:

$$p_X(x) = \sum_y p(x, y) \quad \text{(fix a value of X and sum over Y)}$$

- The marginal pmf of Y:

$$p_Y(y) = \sum_x p(x, y) \quad \text{(fix a value of Y and sum over X)}$$

- Example: given the joint pmf table as below, find the marginal pmf of X ? the marginal pmf of Y?

p(x,y)		Y		
		0	100	200
X	100	0.2	0.1	0.2
	250	0.05	0.15	0.3

DALHOUSIE UNIVERSITY

Conditional pmf

- the conditional pmf of $X = x$, given $Y = y$, is

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

- the conditional pmf of $Y = y$, given $X = x$, is

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

Independence

- If X, Y are independent $\Leftrightarrow$ $p(x, y) = p_X(x) \cdot p_Y(y)$

equivalently, $p_{X|Y} = p_X(x)$ or $p_{Y|X} = P_Y(y)$

Otherwise, they are not independent.

Example

- Given the joint pmf table from the first example,

p(x,y)		Y		
		0	100	200
X	100	0.2	0.1	0.2
	250	0.05	0.15	0.3

1) Find the conditional pmf of X given Y=100

2) Find the conditional pmf of Y given X=250

3) Are X and Y independent?

Expected Value

- Expected Value of X: $E[X] = \sum_{x} x p_X(x)$
- Expected Value of Y: $E[Y] = \sum_{Y} y p_Y(y)$
- Let $h(X, Y)$ be a random function of two discrete random variables X and Y,

  expected value: $E[h(x, y)] = \sum_{x} \sum_{y} h(x, y) p(x, y)$

  Variance: $Var[h(X, Y)] = \sum_{x} \sum_{y} (h(x, y) - E[h(x, y)])^2 p(x, y)$

  shortcut formula: $Var[h(X, Y)] = E[h(X, Y)^2] - E[h(X, Y)]^2$

Covariance

- The covariance is a measure of how much random variables vary together. Positive covariance indicates both variables move in the same direction; Negative covariance indicates they move in opposite directions.
- The covariance of any two random variables X and Y is

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

where $\mu_X = E(X)$ and $\mu_Y = E(Y)$

- **Properties:**
  - $Cov(X, X) = Var(X)$
  - $Cov(aX + b, cY + d) = acCov(X, Y)$
  - if X and Y are independent, then $Cov(X, Y) = 0$. The converse is wrong: covariance is 0 does not imply independence
- We can put variances and covariance of more than one variable in a matrix form, named variance-covariance matrix.

$$\Sigma = \begin{pmatrix} Cov(X, X) & Cov(X, Y) \\ Cov(Y, X) & Cov(Y, Y) \end{pmatrix} = \begin{pmatrix} Var(X) & Cov(X, Y) \\ Cov(X, Y) & Var(Y) \end{pmatrix}$$

  - $Cov(X, Y) = Cov(Y, X)$ so it's symmetric

Correlation Coefficient

- The correlation standardized the covariance so that the correlation doesn't depend on the scale and unit of the random variables.
- The correlation coefficient of X and Y is

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

where $\sigma_X = \sqrt{Var(X)}$ and $\sigma_Y = \sqrt{Var(Y)}$

- **Properties**
  - $-1 \leq Corr(X, Y) \leq 1$
  - $\text{Corr}(aX + b, cY + d) = \begin{cases} \text{Corr}(X, Y) \text{ if } ac > 0 \text{ (a,c have same sign)} \\ -\text{Corr}(X, Y) \text{ if } ac < 0 \text{ (a,c have different signs)} \end{cases}$
  - If X and Y are independent, then $\rho = Corr(X, Y) = 0$, but $\rho = 0$ does not imply independence

DALHOUSIE UNIVERSITY

Example

- Let X and Y be two random variables with following joint distribution,

p(x,y)		Y	
		0	1
X	1	0.2	0.1
	2	0.3	0.4

- 1) Find $P(X+Y \geq 2)$
- 2) Find $P(Y=1 \mid X=1)$
- 3) Calculate $E[X]$ and $Var(X)$
- 4) Calculate $E\left(\frac{Y}{X}\right)$ and $Var\left(\frac{Y}{X}\right)$
- 5) Calculate $Cov(X,Y)$ and $Corr(X,Y)$

Some useful rules

- $Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$
  - $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
  - $Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$
- $Cov(aX + b, cY + d) = acCov(X, Y)$
  - $Cov(X, -Y) = -Cov(X, Y)$
- If X and Y are independent, then $Cov(X, Y) = 0$. Thus,
  - $\rho = Corr(X, Y) = 0$, but $\rho = 0$ does not imply independence
  - $E[XY] = E[X]E[Y]$
  - $Var(X + Y) = Var(X) + Var(Y)$
  - $Var(X - Y) = Var(X) + Var(Y)$
- $\rho = 1$ or -1 iff (if and only if) $Y = aX + b$ with $a \neq 0$.

Continuous Random Variables (not required)

- If X and Y are two continuous random variables, the joint probability density function (joint pdf) is denoted by f(x, y), which satisfies the following:
  - $f(x, y) \geq 0$, for all possible (x, y)
  - $\int\int f(x, y)dxdy = 1$
  - $P[(X, Y) \in A] = P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y)dydx$
- Marginal pdf: $f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$ and $f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$
- Conditional pdf : $f_{Y|X}(y\,|\,x) = \dfrac{f(x, y)}{f_X(x)}$ and $f_{X|Y}(x\,|\,y) = \dfrac{f(x, y)}{f_Y(y)}$
- If X and Y are independent: ⇔ $f(x, y) = f_X(x) \cdot f_Y(y)$

  equivalently $f_{X|Y} = f_X(x)$ or $f_{Y|X} = f_Y(y)$
- $E[h(X, Y)] = \int\int h(x, y)f(x, y)dxdy$

Chapter 5.3, 5.4 and 5.5

Statistics and Their Distributions

The Distribution of a Linear Combination

Jing Zhang
Department of Mathematics and Statistics

DALHOUSIE UNIVERSITY

Random Samples

Definition: The rv's $X_1, X_2, \ldots, X_n$ are said to form a (simple) random sample of size n if

1. The X_i's are independent rv's.
2. Every X_i has the same probability distribution (from the same population)

Note: Condition 1 and 2 can be paraphrased by saying that the X_i's are *independent* and *identically distributed* (iid).

Statistic

Definition:

A statistics is any quantity whose value can be calculated from sample data. Prior to obtaining data, there is uncertainty as to what value of any particular statistic will result. Therefore, a statistic is a random variable and will be denoted by an uppercase letter; a lowercase letter is used to represent the calculated or observed value of the statistic.

Ex: Sample mean is denoted by $\bar{X}$ (before a sample has been selected or an experiment carried out).The calculated value of this statistic is $\bar{x}$.

The Distribution of the Sample Mean

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean value μ and standard deviation σ. Then

1. $E(\bar{X}) = \mu_{\bar{X}} = \mu$
2. $V(\bar{X}) = {\sigma^2}_{\bar{X}} = \sigma^2/n$ and $\sigma_{\bar{X}} = \sigma/\sqrt{n}$

The Case of a Normal Population Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a **normal** distribution with mean value μ and standard deviation σ. Then for any n, $\bar{X}$ is normally distributed (with mean μ and standard deviation $\sigma/\sqrt{n}$.

The Central Limit Theorem (CLT)

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean value μ and standard deviation σ. Then if n is sufficiently large, $\bar{X}$ has approximately a normal distribution with $\mu_{\bar{X}} = \mu$ and $\sigma^2{}_{\bar{X}} = \sigma^2/n$.

Note: Rule of Thumb

The Central Limit Theorem can generally be used if

$n > 30$.

Example

The amount of a particular impurity in a batch of a certain chemical product is a random variable with mean value 4.0 g and standard deviation 1.5 g. If 50 batches are independently prepared, what is the probability that the sample average amount of impurity $\bar{X}$ is between 3.5 and 3.8 g?

Solutions:

According to the rule of thumb, n=50 is large enough for the CLT to be applicable.

$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ with $\mu = 4.0\ and\ \frac{\sigma}{\sqrt{n}} = \frac{1.5}{\sqrt{50}} = 0.2121$

$P(3.5 \leq \bar{X} \leq 3.8) \approx P(\frac{3.5-4.0}{0.2121} \leq Z \leq \frac{3.8-4.0}{0.2121})$

$=\Phi(-0.94) - \Phi(-2.36) = 0.1645$

The Distribution of a Linear Combination

Definition:

Given a collection of n random variables $X_1, X_2, \ldots, X_n$ and n numerical constants $a_1, a_2, \ldots, a_n$, the rv

$$Y = a_1 X_1 + \cdots + a_n X_n = \sum_{i=1}^{n} a_i X_i$$

is called a Linear Combination of the $X_i's$.

Proposition

Let $X_1, X_2, \dots, X_n$ have mean values $\mu_1, \mu_2, \dots, \mu_n$, respectively, and variances $\sigma_1^2, \sigma_2^2, \dots \sigma_n^2$, respectively.

1. Whether or not the $X_i's$ are independent,

$$E(a_1X_1 + \cdots + a_nX_n)$$
$$= a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n)$$
$$= a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$$

2. If $X_1, X_2, \dots, X_n$ are independent,

$$V(a_1X_1 + \cdots + a_nX_n)$$
$$= a_1^2V(X_1) + a_2^2V(X_2) + \cdots + a_n^2V(X_n)$$
$$= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2$$

and $\sigma_{a_1X_1 + \cdots + a_nX_n} = \sqrt{a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2}$

3. For any $X_1, X_2, \dots, X_n$,

$$V(a_1X_1 + \cdots + a_nX_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j Cov(X_i, X_j)$$

Corollary

- $E(X_1 - X_2) = E(X_1) - E(X_2)$ for any two rv's $X_1$and X_2.
- $V(X_1 - X_2) = V(X_1) + V(X_2)$ if X_1 and X_2 are independent.
- If $X_1, X_2, \dots, X_n$ are independent, normally distributed rv's (with possibly different means and/or variances), then any linear combination of the $X_i's$ also has a normal distribution.

Example

A gas station sells three grades of gasoline: regular, extra, and super. These are priced at $3.00, $3.20 and $3.40 per gallon, respectively. Let X_1, X_2, and X_3 denote the amounts of these grades purchased (gallons) on a particular day. Suppose the X_i's are independent and normally distributed with $\mu_1 = 1000, \mu_2 = 500$, $\mu_3 = 300$, $\sigma_1 = 100$, $\sigma_2 = 80$,and $\sigma_3 = 50$. Find the probability that total revenue exceeds $4500.

Solutions:

The total revenue from sales is

$$Y = 3.0X_1 + 3.2X_2 + 3.4X_3$$

$$E(Y) = 3.0\mu_1 + 3.2\mu_2 + 3.4\mu_3 = \$5620$$

$$V(Y) = 3.0^2\sigma_1^2 + 3.2^2\sigma_2^2 + 3.4^2\sigma_3^2 = 184436$$

$$\sigma_Y = \sqrt{184436} = \$429.46$$

$$P(Y > 4500) = P\left(Z > \frac{4500 - 5620}{429.46}\right)$$

$$= P(Z > -2.61) = 1 - \Phi(-2.61) = 0.9955$$