Joint Probability Distributions and Random Samples (5.1 - 5.5)

1. Probability models for joint behavior of several random variables

   - Previously we considered the individual properties of different types of random variables. Now we want to consider the joint properties of several random variables simultaneously.
   - We will consider both the case where random variables are independent of each other and where they are not.
   - For a single discrete rv, X, the pmf describes how much probability mass is placed on each possible value of X. The joint pmf for two discrete random variables, X and Y, describes how probability mass is placed on each possible pair of (x, y) values.

2. Jointly Distributed Random Variables

   (a) Joint Probability Mass Function for Two Discrete Random Variables

   - Let X and Y be two discrete rv's defined on the sample space of an experiment. The **joint probability mass function p(x,y)** is defined for each pair of numbers (x, y) by

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

   - If A is a set consisting of pairs of (x, y) values, then

$$P\left[(X, Y) \in A\right] = \sum_{(x,y)} \sum_{\in A} p(x, y)$$

Example: An insurance company offers both homeowner and car insurance to its customers. For each type of policy a deductable amount needs to be specified. For car insurance, the options are \$100 and \$250. For homeowner insurance, the deductable options are \$0, \$100 and \$200. Let X = {car deductable} and Y = {homeowner deductable} for a randomly selected customer that has both car and homeowner policies.

   - Possible (x, y) pairs are enumerated as follows:
     (100,0), (100,100), (100,200), (250,0), (250,100), (250,200)
   - Show the joint pmf of X and Y in the form of a **joint probability table**

$p(x,y)$			y	
		0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

- To be a legitimate joint pmf, $p(x, y) \geq 0$ for all x and y, and $\sum_x \sum_y p(x, y) = 1$.

- What is the probability that a randomly selected customer has \$100 deductable for both his/her policies?
  - $p(100, 100) = P(X = 100 \; and \; Y = 100) = 0.10$
- What is the probability that a randomly selected customer has homeowner deductable of at least \$100 ?
  - To determine $P(Y \geq 100)$, sum all the probabilities for which $Y \geq 100$.
  - $P(Y \geq 100) = p(100, 100) + p(250, 100) + p(100, 200) + p(250, 200) = 0.75$

(a) Marginal Probability Mass Function

- The **marginal probabililty mass functions** of X and Y, denoted $p_X(x)$ and $p_Y(y)$ are given by

$$p_X(x) = \sum_y p(x, y) \qquad\qquad p_Y(y) = \sum_x p(x, y)$$

Example: Looking again at the insurance company which offers both homeowner and car insurance to its customers, where $X = \{\text{car deductable}\}$ and $Y = \{\text{homeowner deductable}\}$.

$p(x, y)$		y 0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

- The marginal pmf of X, $p_X(x)$, evaluated at $x = 100$ is determined by summing probabilities, $p(100, y)$ over all values of y.

$$p_X(100) = \sum_y p(100, y) = p(100, 0) + p(100, 100) + p(100, 200) = 0.50$$

- $p_X(x)$, evaluated at $x = 250$ is determined by summing probabilities, $p(250, y)$ over all values of y.

$$p_X(250) = \sum_y p(250, y) = p(250, 0) + p(250, 100) + p(250, 200) = 0.50$$

so that:

100, 250

$$p_X(x) = \begin{cases} 0.5 & x = \cancel{0, 100} \\ 0 & Otherwise \end{cases}$$

- The marginal pmf of Y, $p_Y(y)$, evaluated at $y = 100$ is determined by summing probabilities, $p(x, 100)$ over all values of x.

$$p_Y(100) = \sum_x p(x, 100) = p(100, 100) + p(250, 100) = 0.25$$

- $p_Y(y)$, evaluated at $y = 200$ is determined by summing probabilities, $p(x, 200)$ over all values of x.

$$p_Y(200) = \sum_x p(x, 200) = p(100, 200) + p(250, 200) = 0.50$$

so that:

$$p_Y(y) = \begin{cases} 0.25 & y = 0, 100 \\ 0.50 & y = 200 \\ 0 & Otherwise \end{cases}$$

(a) Independent Random Variables

- Two discrete random variables, X and Y, are said to be **independent** if for every pair of x and y values:

$$p(x, y) = p_X(x) \cdot p_Y(y)$$

- If this condition is not satisfied for all pairs of x and y values, then X and Y, are **dependent**.

- Example: Check for independence of discrete rv's X and Y in the insurance deductable example above.

$$p(100, 100) = 0.10 \neq p_X(100) \cdot p_Y(100) = 0.5(0.25) = 0.125$$

- X and Y, are not independent.

(b) More than Two Random Variables

- If random variables, $X_1, X_2, ..., X_n$ are discrete, then their joint pmf is:

$$p(x_1, x_2, ..., x_n) = P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$$

- Many of the ideas discussed above are readily extended to more than 2 discrete random variables, but we will not consider such cases further.

(c) Conditional Distributions

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

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

Example: Looking again at the insurance company, where $X = \{\text{car deductable}\}$ and $Y = \{\text{homeowner deductable}\}$.

$p(x, y)$		y: 0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

- Earlier we determined that X and Y are not independent.
- Find the conditional probability that $Y = 200$, given that $X = 250$.

$$p_{Y|X}(200|250) = \frac{p(250, 200)}{p_X(250)} = \frac{0.3}{0.05 + 0.15 + 0.30} = 0.60$$

- This probability could also have been computed as we did earlier in Chapter 2.

$$P(Y = 200|X = 250) = \frac{P(Y = 200 \ and \ X = 250)}{P(X = 250)} = \frac{0.3}{0.05 + 0.15 + 0.30} = 0.60$$

- Find the conditional probability that $X = 100$, given that $Y = 0$.

$$p_{X|Y}(100|0) = \frac{p(100, 0)}{p_Y(0)} = \frac{0.20}{0.20 + 0.05} = 0.80$$

3. Expected Values, Covariance, and Correlation

(a) Expected Value

- Let X and Y be jointly distributed discrete rv's with pmf, $p(x, y)$, then the expected value of a function, $h(X, Y)$, denoted $E[h(X, Y)]$, is

$$E[h(X, Y)] = \sum_x \sum_y h(x, y) \cdot p(x, y)$$

Example: Looking again at the insurance company, where $X = \{\text{car deductable}\}$ and $Y = \{\text{homeowner deductable}\}$. Let $h(X, Y) = X + Y$, the total deductable.

$p(x,y)$		y 0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

- Determine the expected value of $h(X, Y)$.

$$E[h(X, Y)] = \sum_x \sum_y (x + y) \cdot p(x, y)$$

x	y	$p(x,y)$	$h(x,y)$	$h(x,y)p(x,y)$
100	0	0.20	100	20.0
100	100	0.10	200	20.0
100	200	0.20	300	60.0
250	0	0.05	250	12.5
250	100	0.15	350	52.5
250	200	0.30	450	135.0
				300.0

- The expected value, $E[h(X,Y)] = 300.0$

(b) Covariance

- When two rv's are not independent, it is of interest to assess how strongly they are related to each other.
- The covariance between two discrete rv's, X and Y, provides a measure of their relationship.

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x \sum_y (x-\mu_X)(y-\mu_Y)p(x,y)$$

Example: Looking again at the insurance company, where $X = \{\text{car deductable}\}$ and $Y = \{\text{homeowner deductable}\}$.

$p(x,y)$		y 0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

x	100	250
$p_X(x)$	0.50	0.50

y	0	100	200
$p_Y(y)$	0.25	0.25	0.50

- $\mu_X = E(X) = \sum x p_X(x) = 0.5(100) + 0.5(250) = 175$
- $\mu_Y = E(Y) = \sum y p_Y(y) = 0.25(0) + 0.25(100) + 0.5(200) = 125$

$$Cov(X,Y) = \sum_{(x,} \sum_{y)} (x-175)(y-125)p(x,y)$$

$$= (100-175)(0-125)(0.20) + ... + (250-175)(200-125)(0.30) = 1875$$

- This shows a rather strong positive relationship between X and Y.
- Shortcut formula for computation of $Cov(X,Y)$

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

- Where $E(XY) = \sum_x \sum_y x \cdot y \cdot p(x,y)$
- Note that when X and Y are independent: $E(XY) = E(X) \cdot E(Y)$
- Also, $Cov(X,X) = Var(X)$
- A great weakness in using covariance to measure the strength of association between two random variables is that the magnitude of the covariance varies enormously with the units of measurement used for X and Y.

4. Correlation

- The above noted weakness of the covariance is remedied by use of the correlation coefficient, which normalizes the covariance by dividing it by the two rv standard deviations.
- The correlation coefficient of two discrete rv's, X and Y, provides a more trustworthy measure of the linear relationship between them.

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

- For any two rv's X and Y:
$$-1 \leq Corr(X,Y) \leq 1$$
- Correlation coefficient values always range between +1 and -1.
- $\rho = +1$ indicates the strongest possible + relation between X and Y
- $\rho = -1$ indicates the strongest possible - relation between X and Y
- $\rho = 0$ indicates the a lack of linear relation between X and Y
- Strong relation: $|\rho| \geq 0.8$
- Moderate relation: $0.5 < |\rho| < 0.8$
- Weak relation: $|\rho| \leq 0.5$

Example: Looking again at the insurance company, where $X = \{\text{car deductable}\}$ and $Y = \{\text{homeowner deductable}\}$.

$p(x,y)$		y 0	100	200	x	100	250	y	0	100	200
x	100	0.20	0.10	0.20	$p_X(x)$	0.50	0.50	$p_Y(y)$	0.25	0.25	0.50
	250	0.05	0.15	0.30							

- $Cov(X,Y) = 1875$; $\mu_X = 175$; $\mu_Y = 125$
- $E(X^2) = 100^2(0.5) + 250^2(0.5) = 36,250$
- $E(Y^2) = 100^2(0.25) + 200^2(0.5) = 22,500$
- $Var(X) = 36,250 - 175^2 = 5625$; $\sigma_X = \sqrt{Var(X)} = \sqrt{5625} = 75$
- $Var(Y) = 22,500 - 125^2 = 6875$; $\sigma_Y = \sqrt{Var(YX)} = \sqrt{6875} = 82.92$

$$Corr(X,Y) = \rho_{X,Y} = \frac{1875}{(75)(82.92)} = 0.301$$

- This shows a weak positive relationship between X and Y.

- If X and Y are independent, then $\rho_{X,Y} = 0$, however $\rho = 0$ does not imply independence.
- Only when $Y = aX + b$, for some numbers a and b with $a \neq 0$ is $\rho = \pm 1$
- If a and c are either both positive or both negative constants, then

$$Corr(aX + b, cY + d) = Corr(X, Y)$$

1. Statistics and their Distributions

(a) **Statistical Inference:** Statistical Inference often involves making decisions about a population based on a sample. Usually the values of population parameters are unknown and are estimated from the values of sample statistics. Here, population parameters are considered to be numbers, while sample statistics are considered to be random variables whose values change from one sample to another.

(b) **Population:** A population is a well defined collection of objects. The distribution of values of the attribute of interest is called the **population distribution**. An example would be the weights of all registered Dalhousie students. When information is available for the entire population we have a **census**.

- **Parameter:** A numerical descriptive measure of a population or its distribution. Examples are the population mean, population variance, standard deviation or proportion. These values are typically unknown. They are represented by Greek symbols μ, σ^2, σ, and p, respectively.

(c) **Sample:** A subset of the population is a sample. A random selection of objects from a population.

- **Sample Statistic:** A sample statistic is a numerical descriptive measure of a sample. It is calculated from the observations in the sample. Examples are the sample mean, sample variance, sample standard deviation or sample proportion. These values are calculated directly from the sample and are known. They are represented by the symbols $\bar{x}$, s^2, s, and $\hat{p}$, respectively.

- The sample mean, a measure of center of the data, is computed as,

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- The sample variance, a measure of spread of the data, is computed as,

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{S_{xx}}{n}$$

- The sample standard deviation, also a measure of spread of the data, is computed as the positive square root of the sample variance,

$$s = \sqrt{s^2}$$

- **Sampling Distribution:** Since a sample statistic is a random variable, it is represented by a probability distribution that describes the frequency with which it takes on particular values. This distribution is called the sampling distribution for the sample statistic and whose properties depend not only on the population distribution and sample size, but also on the method of sampling.

(d) **Random Samples**

i. Random variables $X_1, ..., X_n$ are said to form a simple random sample of size n if

- The X_i's are independent rv's.
- Each X_i has the same probability distribution.
- The two above conditions are summarized as **independent and identically distributed, (iid)**.

(e) Distribution of the Sample Mean

i. Normally Distributed Population

A. Statistical inference is based on the sampling distributions of sample statistics, such as the sample mean.

B. When a population is normally distributed, as say $X \sim N(\mu, \sigma)$, under repeated random sampling the sample mean, $\bar{x}$, is also normally distributed, according to normal theory, as $\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$. This distribution is called the sampling distribution for $\bar{x}$.

C. Normally distributed $\bar{x}$ can be transformed to its standard normal form as $Z \sim N(0, 1)$ by subtracting its mean and dividing by its standard deviation.

D. Inferences for μ, the population mean are made based on estimator $\bar{x}$ in its standard normal form as a Z statistic when the value of σ is known and as a t statistic when σ is unknown.

E. **Simulation:** Using simulation, we can empirically determine the sampling distribution of $\bar{x}$, at least approximately, and compare it with that predicted by normal theory. The simulation is carried out as follows:

- Generate 10,000 random samples of size n from population distribution, $N(\mu, \sigma)$.
- Compute the sample mean, $\bar{x}$, for each of the 10,000 random samples.
- Make a histogram of these 10,000 sample mean values.
- Overlay plots of both the above histogram of $\bar{x}$ values and the sampling distribution for $\bar{x}$ predicted by normal theory, $\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$. How well do they match up?

F. Central Limit Theorem

G. Under normal theory, when a population is normally distributed as $X \sim N(\mu, \sigma)$, under repeated random sampling, $\bar{x}$ is also normally distributed as $\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$. This distribution serves as the basis for inferences about the population mean, μ.

H. The central limit theorem extends the normal theory result outlined above to situations where the population distribution is not normally distributed, provided the sample size, n, is large enough.

I. Under the central limit theorem, $\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ even when the population distribution is not normal, provided the sample size is larger than about 30.

J. Confidence intervals and hypothesis tests, for μ, using both Z and t statistics can then be used as discussed earlier for large samples, $n > 30$.

ii. Distribution of a Linear Combination

A. Given a collection of n continuous random variables $X_1, X_2, ..., X_n$ and n constants $a_1, a_2, ..., a_n$, random variable Y, below, is called a linear combination of the X_i's.

$$Y = a_1X_1 + a_2X_2 + ... + a_nX_n = \sum_{i=1}^{n} a_iX_i$$

B. Expected Value of a Linear Combination

C. When random variables $X_1, X_2, ..., X_n$ have expected values $\mu_1, \mu_2, ..., \mu_n$ respectively, the expected value for the linear combination is as shown below whether or not the X_i's are independent.

$$E(Y) = E(a_1X_1 + a_2X_2 + ... + a_nX_n)$$

$$E(Y) = a_1E(X_1) + a_2E(X_2) + ... + a_nE(X_n)$$

$$E(Y) = a_1\mu_1 + a_2\mu_2 + ... + a_n\mu_n$$

D. Variance of a Linear Combination

E. When random variables $X_1, X_2, ..., X_n$ are independent and variances $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$ respectively, the variance for the linear combination is:

$$Var(Y) = Var(a_1X_1 + a_2X_2 + ... + a_nX_n)$$

$$Var(Y) = a_1^2Var(X_1) + a_2^2Var(X_2) + ... + a_n^2Var(X_n)$$

$$Var(Y) = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + ... + a_n^2\sigma_n^2$$

F. For any random variables $X_1, X_2, ..., X_n$, independent or not, with variances $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$ respectively, the variance for the linear combination is:

$$Var(Y) = Var(a_1X_1 + a_2X_2 + ... + a_nX_n)$$

$$Var(Y) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_jCov(X_i, X_j)$$

G. For any random variables $X_1, X_2, ..., X_n$, independent or not, with variances $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$ respectively, the standard deviation for the linear combination is:

$$SD(Y) = \sqrt{Var(a_1X_1 + a_2X_2 + ... + a_nX_n)}$$

iii. Example: Let random variable X represent the height of female Dalhousie students and random variable Y represent the height of female St. Marys students. Assume that X and Y are independent and normally distributed with means of 62 in. and 63 in. respectively, and standard deviations of 3 in. and 2 in. respectively. Suppose one woman is randomly chosen from each school population.

A. What is the probability that the St. Marys woman's height is less than 65 in.?

- Given: $Y \sim N(63, 4)$, find $P(Y < 65)$.
- $P(Y < 65) = P(Z < \frac{65-63}{2}) = P(Z < 1) = 0.8413$

B. What is the probability that the Dalhousie woman's height is greater than 65 in.?

- Given: $X \sim N(62, 9)$, find $P(X > 68)$.
- $P(X > 68) = 1 - P(Z < \frac{68-62}{3}) = 1 - P(Z < 2) = 1 - 0.9772 = 0.0228$

C. What is the probability that the St. Marys woman's height is between 59 and 65 in. inclusive?

- Given: $Y \sim N(63, 4)$, find $P(59 \le Y \le 65)$.
- $P(59 \le Y \le 65) = P(\frac{59-63}{2} \le Z \le \frac{65-63}{2}) = P(-2 \le Z \le 1) = \Phi(1.0) - \Phi(-2.0) = 0.8413 - (1 - 0.9772) = 0.8185$

D. What is the probability that the St. Marys woman's is taller than the Dalhousie woman?

- Given: $Y \sim N(63, 4)$, $X \sim N(62, 9)$, find $P(Y - X > 0)$.
- let $W = Y - X$, then $E(W) = E(Y) - E(X) = 63 - 62 = 1$ and $Var(W) = Var(Y) + (-1)^2Var(X) = 4 + 9 = 13$
- $W \sim N(1, 13)$
- $P(W > 0) = 1 - P(W < 0) = 1 - P(Z < \frac{0-1}{\sqrt{13}}) = 1 - P(Z < -0.28) = 1 - (1 - .6103) = 0.3897$