Chapter 5

Joint Probability Distributions and Random Samples

In Chapters 3 and 4, we developed probability models for a single random variable. Many problems in probability and statistics involve several random variables simultaneously. In this chapter, we discuss probability models for the joint (i.e., simultaneous) behavior of several random variables.

5.1 Jointly Distributed Random Variables

Two Discrete Random Variables

Suppose X and Y are both discrete random variables. The **joint probability mass function** $p(x, y)$ is defined for each pair (x, y) by

$$p(x, y) = P(X = x, Y = y)$$

satisfying $p(x, y) \geq 0$ and $\sum_x \sum_y p(x, y) = 1$.

Now let A be any set consisting of pairs of (x, y) values (e.g., $A = \{(x, y) : x + y = 5\}$ or $\{(x, y) : \max(x, y) \leq 3\}$).
Then the probability $P[(X, Y) \in A]$ is obtained by summing the joint pmf over pairs in A:

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

The **marginal probability mass function** of X and Y, denoted by $p_X(x)$ and $p_Y(y)$, respectively, are given by

$$p_X(x) = P(X = x) = \sum_y p(x, y) \quad \text{for each possible value } x$$

$$p_Y(y) = P(Y = y) = \sum_x p(x, y) \quad \text{for each possible value } y$$

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 the above condition is not satisfied for all (x, y), then X and Y are said to be **dependent**.

Example Suppose a fair coin is tossed three times. The sample space is $S = \{$HHH, HHT, HTT, HTH, TTT, TTH, THH, THT$\}$.

Let X denote the total number of heads and Y denote the number of head on the first toss.

(a) Write out the joint probability mass function $p(x, y)$ for each possible pair of x and y values.

(b) Calculate $P(X - Y = 1)$.

(c) What is the marginal probability $p_X(x)$ and $p_Y(y)$?

(d) Are X and Y independent?

<u>Exercise 5.2</u> When an automobile is stopped by a roving safety patrol, each tire is checked for tire wear, and each headlight is checked to see whether it is properly aimed. Let X denote the number of headlights that need adjustment, and let Y denote the number of defective tires.

(a) If X and Y are independent with $p_X(0) = .5$, $p_X(1) = .3$, $p_X(2) = .2$, and $p_Y(0) = .6$, $p_Y(1) = .1$, $p_Y(2) = p_Y(3) = .05$, $p_Y(4) = .2$, display the joint pmf of (X, Y) in a joint probability table.

(b) Compute $P(X \leq 1 \text{ and } Y \leq 1)$ from the joint probability table, and verify that it equals the product $P(X \leq 1) \cdot P(Y \leq 1)$.

(c) What is $P(X + Y = 0)$, the probability of no violations?

(d) Compute $P(X + Y \leq 1)$.

Two Continuous Random Variables

Let X and Y be continuous random variables. A **joint probability density function** $f(x,y)$ for these two variables is a function satisfying $f(x,y) \geq 0$ and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$.

Then for any two-dimensional set A

$$P[(X,Y) \in A] = \iint_A f(x,y)\,dx\,dy$$

In particular, if A is the two-dimensional rectangle $\{(x,y) : a \leq x \leq b, c \leq y \leq d\}$, then

$$\begin{aligned} P[(X,Y) \in A] &= P(a \leq X \leq b, c \leq Y \leq d) \\ &= \int_a^b \int_c^d f(x,y)\,dy\,dx \end{aligned}$$

The **marginal probability density function** of X and Y, denoted by $f_X(x)$ and $f_Y(y)$, respectively, are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \quad \text{for } -\infty < x < \infty$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx \quad \text{for } -\infty < y < \infty$$

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

$$f(x,y) = f_X(x) \cdot f_Y(y)$$

If the above condition is not satisfied for all (x,y), then X and Y are said to be **dependent**.

Exercises 5.9 Each front tire on a particular type of vehicle is supposed to be filled to a pressure of 26 psi. Suppose the actual air pressure in each tire is a random variable – X for the right tire and Y for the left tire, with joint pdf

$$f(x,y) = \begin{cases} K(x^2+y^2) & 20 \le x \le 30,\ 20 \le y \le 30 \\ 0 & \text{otherwise} \end{cases}$$

(a) What is the value of K?

(b) What is the probability that both tires are underfilled?

(c) Determine the (marginal) distribution of air pressure in the right tire.

(d) Are X and Y independent rv's?

Conditional Probability

If X and Y are two discrete random variables with joint pmf $p(x, y)$ and marginal X pmf $p_X(x) > 0$, then **conditional probability mass function of Y given that $X = x$** is

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

for any possible value y.

If X and Y are two continuous random variables with joint pdf $f(x, y)$ and marginal X pdf $f_X(x)$, then for any X value x for which $f_X(x) > 0$, the **conditional probability density function of Y given that $X = x$** is

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)} \qquad \text{for } -\infty < y < \infty$$

Example Suppose a fair coin is tossed three times. Let X denote the total number of heads and Y denote the number of head on the first toss.

(a) Write out the conditional probability $p_{Y|X}(y \mid x = 2)$.

(b) Write out the conditional probability $p_{X|Y}(x \mid y = 0)$.

Exercise 5.19 Given the two jointly distributed rv's X and Y in Exercise 5.9. X is the actual air pressure for the right tire and Y is the actual air pressure for the left tire.

$$f(x,y) = \begin{cases} \frac{3}{380000}(x^2+y^2) & 20 \le x \le 30,\ 20 \le y \le 30 \\ 0 & \text{otherwise} \end{cases}$$

(a) Determine the conditional pdf of Y given that $X = x$ and the conditional pdf of X given that $Y = y$.

(b) If the pressure in the right tire is found to be 22 psi, what is the probability that the left tire has a pressure of at least 25 psi? Compare this to $P(Y \ge 25)$.

5.2 Expected Values, Covariance, and Correlation

Let X and Y be jointly distributed random variables with pmf $p(x, y)$ or pdf $f(x, y)$ according to whether the variables are discrete or continuous. Then the **expected value of a function $h(X, Y)$**, denoted by $E[h(X, Y)]$ or $\mu_{h(X,Y)}$, is given by

$$E[h(X,Y)] = \begin{cases} \sum_x \sum_y h(x,y)p(x,y) & X, Y \text{ discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y)f(x,y)\,dx\,dy & X, Y \text{ continuous} \end{cases}$$

When two random variables X and Y are not independent, it is frequently of interest to assess how strongly they are related to one another. The extent to which two random variables vary together (co-vary) can be measured by their covariance.

The **covariance** between two random variables X and Y is

$$\text{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$
$$= \begin{cases} \sum_x \sum_y (x-\mu_x)(y-\mu_y)p(x,y) & X, Y \text{ discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y)f(x,y)\,dx\,dy & X, Y \text{ continuous} \end{cases}$$

Shortcut formula for Cov(X, Y)

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

Interpreting the Covariance

- positive covariance – indicates that higher than average values of one variable tend to be paired with higher than average values of the other variable.
- negative covariance – indicates that higher than average values of one variable tend to be paired with lower than average values of the other variable.
- zero covariance – If the two random variables are independent, the covariance will be zero. However, a covariance of zero does not necessarily mean that the variables are independent. A nonlinear relationship can exist that still would result in a covariance value of zero.

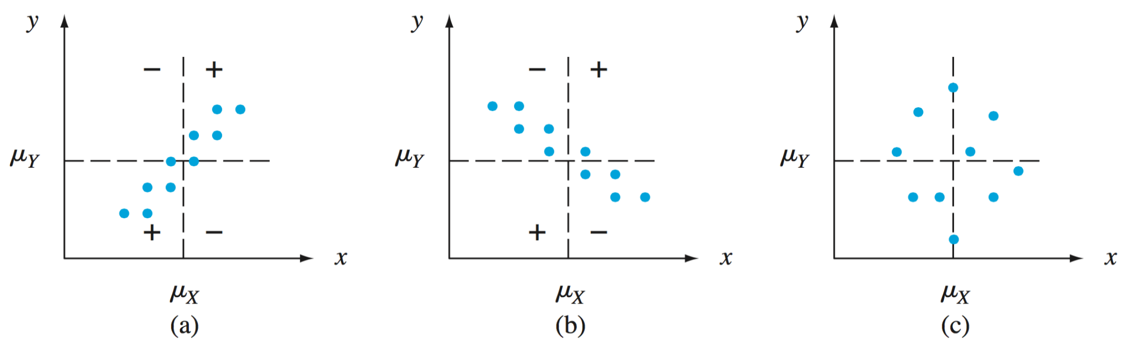

Figure 5.4 $p(x, y) = 1/10$ for each of ten pairs corresponding to indicated points: (a) positive covariance; (b) negative covariance; (c) covariance near zero

Because the value of covariance depends critically on the units of measurement, it is difficult to compare covariances among data sets having different scales. The correlation coefficient addresses this issue by creating a dimensionless quantity that facilitates this comparison.

The **correlation coefficient** between two random variables X and Y, denoted by Corr(X, Y) or $\rho_{X,Y}$ or just ρ, is defined by

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

Properties about Correlation

- If a and c are either both positive or both negative, Corr$(aX + b, cY + d)$ = Corr(X, Y)
- For any two rv's X and Y, $-1 \leq \text{Corr}(X, Y) \leq 1$.
- If X and Y are independent, then $\rho = 0$.
  But $\rho = 0$ does not imply independence.
- $\rho = 1$ or -1 iff $Y = aX + b$ for some numbers a and b with $a \neq 0$.
- Cov$(X, X) = E[(X - \mu_X)^2] = V(X)$ and Corr$(X, X) = 1$.

Example Suppose a fair coin is tossed three times. Let X denote the total number of heads and Y denote the number of head on the first toss.

(a) Calculate $E(X - Y)$.

(b) Calculate Cov(X, Y).

(c) Calculate $\rho_{X,Y}$.

Exercise 5.31 Given the two jointly distributed rv's X and Y in Exercise 5.9.

$$f(x,y) = \begin{cases} \frac{3}{380000}(x^2+y^2) & 20 \le x \le 30,\ 20 \le y \le 30 \\ 0 & \text{otherwise} \end{cases}$$

(a) Compute the covariance between X and Y.

(b) Compute the correlation coefficient ρ for X and Y.

5.3 Statistics and Their Distributions

General Problems of Statistics: A sample of size n is a collection of observations selected from a particular population distribution (with pmf $p(x)$ or pdf $f(x)$, for discrete or continuous cases, respectively). The data values recorded, $x_1, \cdots, x_n$, are the observed values of a set of n independent random variables $X_1, \cdots, X_n$, and each has the same probability distribution $p(x)$ or $f(x)$. The general problem is to estimate some unknown quantity about $p(x)$ or $f(x)$ based on the collected data $x_1, \cdots, x_n$.

A **statistic** 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.

Any statistic, being a random variable, has a probability distribution. For example, the sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ has a probability distribution, and the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ also has a probability distribution.

The probability distribution of a statistic is sometimes referred to as its **sampling distribution** to emphasize that it describes how the statistic varies in value across all samples that might be selected.

Statistics are random variables; parameters are constants.

Random Samples

The probability distribution of any particular statistic depends not only on the population distribution (normal, uniform, etc.) and the sample size n but also on the method of sampling.

The rv's $X_1, X_2, \cdots, X_n$ are said to form a **random sample** of size n if

- The X_i's are independent rv's.
- Every X_i has the same probability distribution.

In other words, we say that the X_i's are **independent and identically distributed (iid)**.

If sampling is either with replacement or from an infinite population, these conditions are satisfied exactly.

These conditions will be approximately satisfied if sampling is without replacement, yet the sample size n is much smaller than the population size N.

In practice, if $n/N \leq .05$ (at most 5% of the population is sampled), we can proceed as if the X_i's form a random sample. The virtue of this sampling method is that the probability distribution of any statistic can be more easily obtained than for any other sampling method.

Exercise 5.38 There are two traffic lights on a commuters route to and from work. Let X_1 be the number of lights at which the commuter must stop on his way to work, and X_2 be the number of lights at which he must stop when returning from work. Suppose that X_1 and X_2 are independent and each has the same distribution as given in the accompanying table (so that X_1, X_2 is a random sample of size $n = 2$).

x_i	0	1	2
$p(x_i)$	.2	.5	.3

$\mu = 1.1, \sigma^2 = .49$

(a) Determine the pmf of $T_o = X_1 + X_2$.

(b) Calculate μ_{T_o}. How does it relate to μ, the population mean?

(c) Calculate $\sigma^2_{T_o}$. How does it relate to σ^2, the population variance?

(d) Let X_3 and X_4 be the number of lights at which a stop is required when driving to and from work on a second day assumed independent of the first day. With T_o' = the sum of all four X_i's, what now are the values of $E(T_o')$ and $V(T_o')$?

5.4 The Distribution of the Sample Mean

If $X_1, \cdots, X_n$ are rv's and $a_1, \cdots, a_n$ and b are constants, then

$$E(a_1X_1 + \cdots + a_nX_n + b) = a_1E(X_1) + \cdots + a_nE(X_n) + b$$

If, in addition, $X_1, \cdots, X_n$ are independent, then

$$V(a_1X_1 + \cdots + a_nX_n + b) = a_1^2V(X_1) + \cdots + a_n^2V(X_n)$$

Let $X_1, \cdots, X_n$ be a random sample from a distribution with mean value μ and standard deviation σ.
Then for the sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, we have

- $\mu_{\overline{X}} = E(\overline{X}) = \mu$
- $\sigma_{\overline{X}}^2 = V(\overline{X}) = \sigma^2/n$ and $\sigma_{\overline{X}} = \sigma/\sqrt{n}$

In addition, for the sample total $T_o = \sum_{i=1}^{n} X_i = X_1 + \cdots + X_n$, we have

- $\mu_{T_o} = E(T_o) = n\mu$
- $\sigma_{T_o}^2 = V(T_o) = n\sigma^2$ and $\sigma_{T_o} = \sqrt{n}\sigma$.

Special case of Normal random sample

Let $X_1, \cdots, X_n$ be iid rv's and $X_i \sim N(\mu, \sigma^2)$, then for any n, $\overline{X} \sim N(\mu_{\overline{X}} = \mu, \sigma_{\overline{X}}^2 = \sigma^2/n)$.

The Central Limit Theorem (CLT)

Let $X_1, \cdots, X_n$ be a random variables from a distribution with mean μ and variance σ^2.

Then if n is sufficiently large, the distribution of the sample mean $\overline{X}$ can be approximated by a normal distribution with $\mu_{\overline{X}} = \mu$ and $\sigma^2_{\overline{X}} = \sigma^2/n$.

$$\overline{X} \overset{approx.}{\sim} N(\mu, \sigma^2/n)$$

Similarly, the distribution of sample total $T_o = X_1 + \cdots + X_n$, can be approximated by a normal distribution with $\mu_{T_o} = n\mu$ and $\sigma^2_{T_o} = n\sigma^2$.

$$T_o \overset{approx.}{\sim} N(n\mu, n\sigma^2)$$

The larger the value of n, the better the approximation.

Rule of the Thumb
If $n > 30$, the Central Limit Theorem can be used.

Exercise 5.50 The breaking strength of a rivet has a mean value of 10,000 psi and a standard deviation of 500 psi.

(a) What is the probability that the sample mean breaking strength for a random sample of 40 rivets is between 9900 and 10,200?

(b) If the sample size had been 15 rather than 40, could the probability requested in part (a) be calculated from the given information?

Exercise 5.52 The lifetime of a certain type of battery is normally distributed with mean value 10 hours and standard deviation 1 hour. There are four batteries in a package. What lifetime value is such that the total lifetime of all batteries in a package exceeds that value for only 5% of all packages?

5.5 The Distribution of a Linear Combination

If $X_1, \cdots, X_n$ 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.

In particular, the difference $X_1 - X_2$ between two independent, normally distributed variables is itself normally distributed with parameters $\mu_{X_1-X_2} = \mu_{X_1} - \mu_{X_2}$ and $\sigma^2_{X_1-X_2} = \sigma^2_{X_1} + \sigma^2_{X_2}$.

Exercise 5.62 Manufacture of a certain component requires three different machining operations. Machining time for each operation has a normal distribution, and the three times are independent of one another. The mean values are 15, 30, and 20 min, respectively, and the standard deviations are 1, 2, and 1.5 min, respectively. What is the probability that it takes at most 1 hour of machining time to produce a randomly selected component?