Chapter 8: 
Multivariable Distributions

Very often, questions in probability theory involve the value of more than one random variable at the same time. For example, in a course with 3 exams, we could ask for the probability that a randomly selected student will get an A on all three exams. If we let $X_1$ be the student’s score on the 1st exam, $X_2$ be the score on the 2nd exam, and $X_3$ be the score on the 3rd exam, then the probability of that the student gets an A on all three is $P(X_1 \geq 90 \text{ and } X_2 \geq 90 \text{ and } X_3 \geq 90)$.

Computing this probability can be very involved. In the simplest case, if $X_1$, $X_2$, and $X_3$ were independent, then we could simply compute the 3 probabilities separately and then multiply them:

$$P(X_1 \geq 90 \text{ and } X_2 \geq 90 \text{ and } X_3 \geq 90) = P(X_1 \geq 90) \cdot P(X_2 \geq 90) \cdot P(X_3 \geq 90).$$

However, in this scenario, the variables $X_1$, $X_2$, and $X_3$ are clearly not independent – students who do very well on one exam tend to also do well on other exams in the same class, so the methods involved in the computations will be more complicated.

This chapter provides an introduction to techniques for dealing with probabilities involving more than one random variable. For simplicity, the focus will primarily be on probabilities involving two random variables.

Distributions of Two Random Variables

Example 1: Two pills are selected at random from a bottle containing 3 aspirin, 2 sedative, and 4 laxative pills. Let $X$ denote the number of aspirin pills among the 2 pills chosen. Clearly the possible values of $X$ are 0, 1, or 2. We are selecting a random sample of size $n = 2$ from a population of size $N = 9$ which contains a total of $M = 3$ aspirin pills (which we can treat as a success) and $N - M = 6$ non-aspirin pills (failures). The random variable $X$ is simply the number of successes found in this sample, so $X$ has a hypergeometric distribution and the probability mass function for $X$ is given by

$$f_X(x) = \frac{\binom{M}{x} \binom{N-M}{2-x}}{\binom{N}{2}}, \quad x = 0, 1, 2.$$ 

In particular,

$$P(X = 0) = \frac{\binom{3}{0} \binom{6}{2}}{36} = \frac{5}{12}, \quad P(X = 1) = \frac{\binom{3}{1} \binom{6}{1}}{36} = \frac{1}{2}, \quad P(X = 2) = \frac{\binom{3}{2} \binom{6}{0}}{36} = \frac{1}{12}. $$

Next, let $Y$ be the number of sedative pills among the 2 pills chosen. Then $Y$ also has a hypergeometric distribution.

Exercise: Find the p.m.f. for $Y$ and compute the probabilities of all possible values of $Y$. 

8-1
So far, this example has only involved looking at probabilities involving the two random variables separately. What if we want to deal with probabilities involving both $X$ and $Y$ at the same time?

Find the probabilities associated with all possible pairs of $X$ and $Y$.

Note that in the example above, we could also consider a third random variable: let $W$ be the number of laxative pills among the 2 pills chosen. However, $W$ is redundant; once we know the values of $X$ and $Y$, the value of $W$ is completely determined since $W = 2 - X - Y$.

**Definition:** Let $X$ and $Y$ be two discrete random variables. For each pair of values $(x,y)$ within the range of $X$ and $Y$, the probability that $X = x$ and $Y = y$ is denoted by

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

The function $f(x,y)$ is called the **joint probability mass function** (joint p.m.f.) of $X$ and $Y$. The **space** or **support** of $f$ is the set of all pairs $(x,y)$ for which $f(x,y) > 0$.

Since the joint p.m.f. represents a probability, it must satisfy $0 \leq P(A) \leq 1$ for every event $A$ and $P(S) = 1$, where $S$ is the sample space (all possible outcomes). In terms of the joint p.m.f., this means that

1. $0 \leq f(x,y) \leq 1$ for all pairs $(x,y)$, and
2. $\sum_x \sum_y f(x,y) = 1$ where the double sum extends over all possible pairs of $(x,y)$. 

8-2
**Definition:** If $X$ and $Y$ are two discrete random variables with joint p.m.f. $f(x, y)$, then the function $f_X(x) = \sum_y f(x, y)$ (where the sum extends over all possible values of $y$) is called the **marginal probability mass function of $X$**.

Similarly, the function $f_Y(y) = \sum_x f(x, y)$ (where the sum extends over all possible values of $x$) is called the **marginal probability mass function of $Y$**.

In fact, the function $f_X$ is simply the probability mass function of $X$ alone, and $f_Y$ is the p.m.f. of $Y$ alone.

**Example 2:** Let the joint p.m.f. of $X$ and $Y$ be defined by

$$f(x, y) = c(x + y) \text{ for } x = 0, 1, 2, 3; \ y = 0, 1, 2.$$  

a. What is the value of $c$?

b. Find $f_X(x)$, the marginal p.m.f. of $X$.

c. Find $f_Y(y)$, the marginal p.m.f. of $Y$.
d. Find \( P(X = Y) \).

e. Find \( P(X + Y > 2) \)

f. Find \( P(X = 2Y) \)

We can also state definitions for the joint distribution of two continuous random variables.

**Definition:** Let \( X \) and \( Y \) be two continuous random variables. The **joint probability density function** (joint p.d.f.) of \( X \) and \( Y \) is an integrable function \( f(x, y) \) defined for all pairs of real numbers \((x, y)\) which satisfies

a. \( f(x, y) \geq 0 \) for all pairs of numbers \((x, y)\),

b. \( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = 1 \), and

c. If \( A \) is any set of points in the plane, then
   \[ P((X, Y) \in A) = \int_A f(x, y) \, dx \, dy. \]

The **space** or **support** of \( f \) is the set of all pairs \((x, y)\) for which \( f(x, y) > 0 \).

The **marginal probability density functions** for \( X \) and \( Y \) are defined by

\[ f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) \, dx. \]

**Example 3:** Let \( X \) and \( Y \) have joint p.d.f. given by

\[ f(x, y) = 1 - \frac{1}{2}x - \frac{1}{2}y \quad \text{for} \quad 0 < x < 1 \quad \text{and} \quad 0 < y < 2. \]

a. Verify that \( f \) is in fact a joint density.
b. Find $P\left(X \leq \frac{1}{2} \text{ and } Y \leq \frac{1}{2}\right)$

c. Find the marginal densities of $X$ and $Y$.

d. Find $P(X + Y < 1)$.

Example 4: Let $f(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$, $-\infty < x < \infty$, $-\infty < y < \infty$ be the joint p.d.f. of $X$ and $Y$.

Find the marginal densities $f_x$ and $f_y$.

Note that in this example, $X$ and $Y$ each have the standard normal distribution since their marginal densities are both the density for $N(0,1)$. The joint distribution of $X$ and $Y$ is said to be a **bivariate normal distribution**. Also note that $f(x, y) = f_x(x) \cdot f_y(y)$. When this occurs, $X$ and $Y$ are said to be **independent** random variables.
For discrete random variables, this is equivalent to \( P(X = x \text{ and } Y = y) = P(X = x) \cdot P(Y = y) \) for all pairs \((x, y)\). This matches the definition of independence of two events that was given in chapter 2.

It is easily verified that the random variables in examples 1, 2, and 3 are not independent, but the random variables in example 4 are independent.

**Definition:** Let \( X \) and \( Y \) be two discrete random variables with joint p.m.f. \( f(x, y) \) and let \( u(X, Y) \) be a function of these two random variables. If the sum exists, then the expected value of \( u(X, Y) \) is given by \( E(u(X, Y)) = \sum_x \sum_y u(x, y) \cdot f(x, y) \), where the double sum extends over all pairs \((x, y)\) in the space of \( f \).

Let \( X \) and \( Y \) be two continuous random variables with joint p.d.f. \( f(x, y) \) and let \( u(X, Y) \) be a function of these two random variables. If the integral exists, then the expected value of \( u(X, Y) \) is given by \( E(u(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x, y) \cdot f(x, y) \, dx \, dy \).

So, for example, if \( X \) and \( Y \) are discrete random variables, then the expected value of their sum would be given by
\[
E(X + Y) = \sum_y \sum_x (x + y) \cdot f(x, y).
\]

In fact, mean (expected value) of the sum of two random variables is just the sum of the means of the two separate random variables. That is, \( E(X + Y) = E(X) + E(Y) \). This fact is proven as our first theorem of the chapter:

**Theorem 8.1:** Let \( X \) and \( Y \) be random variables with means \( \mu_X \) and \( \mu_Y \). Then \( \mu_{X+Y} = \mu_X + \mu_Y \).

Equivalently, \( E(X + Y) = E(X) + E(Y) \).

**Proof:** We will prove each part for discrete random variables; the proofs for continuous random variables are very similar with sums replaced by integrals. Assume that \( X \) and \( Y \) have joint p.m.f. \( f(x, y) \). Then
**Distributions of Sums of Independent Random Variables**

Theorem 8.1 provides us with a simple method for determining the mean of a sum of random variables. In general, it is quite difficult to determine the variance and the moment generating function of a sum of random variables. However, when the random variables are *independent*, then there is a simple method for computing these items as well. In many cases, the following theorem will allow us to determine the distribution of the sum of independent random variables.

**Theorem 8.1:** Let $X$ and $Y$ be *independent* random variables with variances $\sigma_X^2$ and $\sigma_Y^2$. Furthermore, assume that the moment-generating functions $M_X$ and $M_Y$ both exist. Then

1. $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$. Equivalently, $Var(X + Y) = Var(X) + Var(Y)$.
2. $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$.

**Proof:** Once again, we prove each part for discrete random variables; the proofs for continuous random variables are very similar with sums replaced by integrals. Assume that $X$ and $Y$ have joint p.m.f. $f(x, y)$.

1. First note that $Var(X + Y) = E[(X + Y)^2] - [E(X + Y)]^2$. From Theorem 8.1, $E(X + Y) = E(X) + E(Y)$, so

   $[E(X + Y)]^2 = [E(X) + E(Y)]^2 = [E(X)]^2 + 2E(X)E(Y) + [E(Y)]^2$.

   We now need to compute $E[(X + Y)^2]$. 

---

8-7
Putting it all together, we now have
\[ Var(X + Y) = E \left[ (X + Y)^2 \right] - [E(X + Y)]^2 \]
\[ = \left\{ E(X^2) + 2E(Y)E(X) + E(Y^2) \right\} - \left\{ E(X)^2 + 2E(X)E(Y) + E(Y)^2 \right\} \]
\[ = \left\{ E(X^2) - [E(X)]^2 \right\} + \left\{ E(Y^2) - [E(Y)]^2 \right\} \]
\[ = Var(X) + Var(Y) \]

2. The proof is left as an exercise.

Note that Theorem 8.2 is only valid if the random variables are independent. Through the use of moment generating functions, this theorem provides us with a powerful tool for determining the distribution of a sum of independent random variables, as illustrated in the example below:

**Example 4:** Let \( X \) and \( Y \) be independent random variables both having the gamma distribution with parameters \( \lambda = \frac{1}{r} \) and \( r = 5 \).

Recall that a gamma distribution with parameters \( \lambda \) and \( r \) has probability mass function
\[ f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \quad \text{for} \quad x \geq 0, \quad \text{mean} \quad \mu = \frac{r}{\lambda}, \quad \text{variance} \quad \sigma^2 = \frac{r}{\lambda^2}, \quad \text{and moment generating function} \quad M_X(t) = \frac{\lambda^r}{(\lambda - t)^r}, \quad t < \lambda. \]

a. Find the mean and variance of \( W = X + Y \).

b. Find the m.g.f. for \( W \).

c. Identify the distribution of \( W \) and find the pmf of \( W \).
The definitions of joint p.d.f.’s and p.m.f.’s can easily be extended to more than two variables. The joint p.m.f. of \(n\) random variables \(X_1, X_2, \ldots, X_n\) would be a function \(f(x_1, x_2, \ldots, x_n)\) of \(n\) variables. In the absence of independence, computing probabilities for such multivariate distributions is extremely difficult (for example, in the continuous case, it would involve computing \(n\)-dimensional integrals. Fortunately, for most applications in statistics, we are able to assume independence, which simplifies the situation considerably.

**Definition:** A collection of random variables \(X_1, X_2, \ldots, X_n\) are independent if and only if
\[
f(x_1, x_2, \ldots, x_n) = f_1(x_1) f_2(x_2) \cdots f_n(x_n)
\]
for all \(n\)-tuples \((x_1, x_2, \ldots, x_n)\).

The following theorem extends our results regarding the distribution of the sum of two independent random variables to the distribution of the sum of \(n\) independent random variables. The proof is omitted since it is completely analogous to the proof in the two variable case.

**Theorem 8.3:** Let \(X_1, X_2, \ldots, X_n\) be random variables. Assume that the means, variances, and moment-generating functions all exist for \(X_1, X_2, \ldots, X_n\). Then

1. \(E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)\).

Furthermore, if \(X_1, X_2, \ldots, X_n\) are independent, we also have

2. \(Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)\).

and

3. \(M_{X_1 + X_2 + \cdots + X_n}(t) = M_{X_1}(t) M_{X_2}(t) \cdots M_{X_n}(t)\).

**The Distribution of a Sample Mean**

In applied statistics, we often select random samples from a large population.

**Definition:** A collection of random variables \(X_1, X_2, \ldots, X_n\) is called a random sample provided that they are independent and they all have the same distribution. (This is also often referred to as independent and identically distributed and is abbreviated i.i.d.)

Since each of the r.v.’s \(X_1, X_2, \ldots, X_n\) in a random sample has the same distribution, they all must have the same mean and variance. That is, \(E(X_1) = E(X_2) = \cdots = E(X_n)\) and \(Var(X_1) = Var(X_2) = \cdots = Var(X_n)\). If we call this common mean \(\mu\) and the common variance \(\sigma^2\), then we say that \(X_1, X_2, \ldots, X_n\) are a random sample of size \(n\) from a population with mean \(\mu\) and variance \(\sigma^2\).

In practice, the mean \(\mu\) and the common variance \(\sigma^2\) are typically unknown. One of the most common goals of statistics is to use the information from a random sample to estimate the value...
of \( \mu \). We will usually use \( \overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) \) to estimate \( \mu \). From a theoretical perspective, \( \overline{X} \) is simply a new random variable which is defined as a function of the random variables \( X_1, X_2, \ldots, X_n \). One of the most critical questions of probability theory is to obtain information about the distribution of the random variable \( \overline{X} \).

**Theorem 8.4:** Let \( X_1, X_2, \ldots, X_n \) be a random sample of size \( n \) from a population with mean \( \mu \), variance \( \sigma^2 \), and moment generating function \( M(t) \). Then

1. \( E(\overline{X}) = \mu \),
2. \( Var(\overline{X}) = \frac{\sigma^2}{n} \), and
3. \( M_{\overline{X}}(t) = \left[M\left(\frac{t}{n}\right)\right]^n \).

**Proof:**
Theorem 8.5: Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n$ from a population which is normally distributed with mean $\mu$ and variance $\sigma^2$. Then the distribution of $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is $N(\mu, \sigma^2/n)$.

Proof:

Example 5: The gas mileage (in miles per gallon) of a compact car selected at random is normally distributed with mean $\mu = 29.5$ and variance $\sigma^2 = 9$.

a. What is the probability that such a car selected at random will get at least 25 miles per gallon?

b. If 9 such cars are selected at random, what is the probability that the mean gas mileage for the 9 cars will be at least 25 miles per gallon?
Exercises

1. Let \( f(x, y) = cy^2, x = 1, 2, 3, \) and \( y = 1, 2 \) be the joint p.m.f. for \( X \) and \( Y \).
   a. What is the value of the constant \( c \)?
   b. Find the marginal probability mass functions \( f_X \) and \( f_Y \).
   c. Are \( X \) and \( Y \) independent? Explain your answer.
   d. Find \( P(X = Y) \).
   e. Find \( P(X > Y) \).

2. Draw 5 cards at random from an ordinary deck of 52 cards. Among these 5 cards, let \( X \) be the number of hearts, and let \( Y \) be the number of spades.
   a. What kind of distribution does \( X \) have? What is the p.m.f. of \( X \)?
   b. Give a formula for the joint p.m.f. of \( X \) and \( Y \). Then make a table showing the values of the joint p.m.f. for all possible pairs \( (x, y) \).
   c. Use the marginal totals of the table from part (b) to find the values of the marginal p.m.f. of \( X \). Verify that these values match the values of the p.m.f. given in part (a).
   d. Are \( X \) and \( Y \) independent? Explain your answer.

3. Let \( X \) and \( Y \) have joint probability density function \( f(x, y) = x + y, 0 \leq x \leq 1, 0 \leq y \leq 1 \).
   a. Find the marginal probability density functions \( f_X \) and \( f_Y \).
   b. Are \( X \) and \( Y \) independent? Explain your answer.
   c. Find \( P(X \leq \frac{1}{2} \) and \( Y \geq \frac{1}{2}) \).
   d. Find \( P(X + Y < \frac{1}{2}) \).

4. Let \( X \) and \( Y \) be independent discrete random variables and assume that the moment-generating functions \( M_X \) and \( M_Y \) both exist. Prove that \( M_{X+Y}(t) = M_X(t) \cdot M_Y(t) \). Clearly indicate where the assumption of independence is used in your proof.
   Hint: To get started, note that by definition, \( M_{X+Y}(t) = E(e^{(x+y)t}) = \sum_x \sum_y e^{(x+y)t} f(x, y) \).

5. Let \( X_1, X_2, X_3, X_4, X_5 \) be a random sample from a population with geometric distribution with \( p = \frac{1}{3} \).
   a. Find the moment generating function of \( Y = X_1 + X_2 + X_3 + X_4 + X_5 \).
   b. What is the distribution of \( Y \)? What are the mean and variance of \( Y \)?

6. Let \( W = X_1 + X_2 + \cdots + X_n \), where the \( X_i \) are independent and identically distributed exponential random variables with parameter \( \lambda \). Show that \( W \) has a gamma distribution. What are the p.d.f., mean, and variance of \( W \)?

7. A soft drink company uses a filling machine to fill cans. Each 12 oz. can is to contain 355 milliliters of beverage. In fact, the amount varies according to a normal distribution with mean \( \mu = 355.2 \) ml and standard deviation \( \sigma = 0.5 \) ml.
   a. What is the probability that an individual can contains less than 355 ml?
   b. What is the probability that the mean content of a six-pack of cans is less than 355 ml?
8. The lifetime of disk brake pads varies according to a normal distribution with mean \( \mu = 50,000 \) miles and standard deviation \( \sigma = 3000 \) miles. Suppose that a sample of nine brake pads is tested.

a. What is the distribution of the sample mean \( \bar{X} \)? Give the mean and standard deviation of this distribution.

b. Suppose that a sample mean less than 47,000 miles is considered good evidence that the true mean lifetime \( \mu \) of the break pads is less than 50,000 miles. What is the probability that this will happen (a sample mean of less than 47,000 miles in a sample of size nine) even when the true mean is in fact 50,000 miles, leading to an incorrect conclusion?