LECTURE 2: PROBABILITY

VICTORIA UNIVERSITY OF WELLINGTON
TE HERENGA WAKA

Section 2.1: Basic Ideas

Definition: An **experiment** is a process that results in an outcome that cannot be predicted in advance with certainty.

Examples:

- rolling a die
- tossing a coin
- weighing the contents of a box of cereal

Sample Space

Definition: The set of all possible outcomes of an experiment is called the **sample space** for the experiment.

Examples:

- For rolling a fair die, the sample space is {1, 2, 3, 4, 5, 6}.
- For a coin toss, the sample space is {heads, tails}.
- For weighing a cereal box, the sample space is $(0, \infty)$, a more reasonable sample space is $(12, 20)$ for a 16 oz. box.

More Terminology

Definition: A subset of a sample space is called an **event**.

- For any sample space, the empty set ϕ is an event, as is the entire sample space.
- A given event is said to have occurred if the outcome of the experiment is one of the outcomes in the event. For example, if a die comes up 2, the events {2, 4, 6} and {1, 2, 3} have both occurred, along with every other event that contains the outcome "2."

Example 1

An electrical engineer has on hand two boxes of resistors, with four resistors in each box. The resistors in the first box are labeled 10 ohms, but in fact their resistances are 9, 10, 11, and 12 ohms. The resistors in the second box are labeled 20 ohms, but in fact their resistances are 18, 19, 20, and 21 ohms. The engineer chooses one resistor from each box and determines the resistance of each.

Example 1 cont.

Let *A* be the event that the first resistor has a resistance greater than 10, let *B* be the event that the second resistor has resistance less than 19, and let C be the event that the sum of the resistances is equal to 28.

1. Find the sample space for this experiment.

2. Specify the subsets corresponding to the events *A*, *B*, and C.

$$\mathcal{S} = \{(9, 18), (9, 19), (9, 20), (9, 21), (10, 18), (10, 19), (10, 20), (10, 21), (11, 18), (11, 19), (11, 20), (11, 21), (12, 18), (12, 19), (12, 20), (12, 21)\}$$

The events A, B, and C are given by

$$A = \{(11, 18), (11, 19), (11, 20), (11, 21), (12, 18), (12, 19), (12, 20), (12, 21)\}$$
$$B = \{(9, 18), (10, 18), (11, 18), (12, 18)\}$$
$$C = \{(9, 19), (10, 18)\}$$

Combining Events

The **union** of two events *A* and *B*, denoted $A \cup B$, is the set of outcomes that belong either to *A*, to *B*, or to both.

In words, $A \cup B$ means "*A* or *B*." So the event "*A* or *B*" occurs whenever either *A* or *B* (or both) occurs.

Example 2

Let $A = \{1, 2, 3\}$ and $B = \{2, 3, 4\}$.

What is $A \cup B$?

Intersections

The **intersection** of two events A and B, denoted by $A \cap B$, is the set of outcomes that belong both to A and to B.

In words, $A \cap B$ means "A and B."

Thus the event "A and B" occurs whenever both A and B occur.

Example 3

Let $A = \{1, 2, 3\}$ and $B = \{2, 3, 4\}$.

What is $A \cap B$?

Complements

The **complement** of an event A, denoted A^c, is the set of outcomes that do not belong to A. In words, A^c means "not A." Thus the event "not A" occurs whenever A does **not** occur.

Example 4

Consider rolling a fair six-sided die. Let A be the event: "rolling a six" = {6}.

What is A^c = "not rolling a six"?

Mutually Exclusive Events

Definition: The events *A* and *B* are said to be **mutually exclusive** if they have no outcomes in common.

More generally, a collection of events $A_1, A_2, \ldots, A_n$ is said to be mutually exclusive if no two of them have any outcomes in common.

Sometimes mutually exclusive events are referred to as disjoint events.

Back to Example 1

- If the experiment with the resistors is performed
  - Is it possible for events A and B both to occur?
  - How about B and C?
  - A and C?
  - Which pair of events is mutually exclusive?

Solution

If the outcome is (11, 18) or (12, 18), then events A and B both occur. If the outcome is (10, 18), then both B and C occur. It is impossible for A and C both to occur, because these events are mutually exclusive, having no outcomes in common.

Probabilities

Definition: Each event in the sample space has a **probability** of occurring. Intuitively, the probability is a quantitative measure of how likely the event is to occur.

Given any experiment and any event A:

- The expression $P(A)$ denotes the probability that the event A occurs.
- $P(A)$ is the proportion of times that the event A would occur in the long run, if the experiment were to be repeated over and over again.

Axioms of Probability

1. Let ***S*** be a sample space. Then *P*(***S***) = 1.

2. For any event *A*, $0 \le P(A) \le 1$.

3. If *A* and *B* are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$. More generally, if $A_1, A_2, \ldots$ are mutually exclusive events, then

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$$

A Few Useful Things

- For any event *A*,

$$P(A^C) = 1 - P(A).$$

- Let ϕ denote the empty set. Then

$$P(\phi) = 0.$$

- If *A* is an event containing outcomes $O_1, \ldots, O_n$, that is, if $A = \{O_1, \ldots, O_n\}$, then

$$P(A) = P(O_1) + \cdots + P(O_n)$$

- Addition Rule (for when *A* and *B* are not mutually exclusive): $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Example 5

A target on a test firing range consists of a bull's-eye with two concentric rings around it. A projectile is fired at the target. The probability that it hits the bull's-eye is 0.10, the probability that it hits the inner ring is 0.25, and the probability that it hits the outer ring is 0.45.

1. What is the probability that the projectile hits the target?
2. What is the probability that it misses the target?

$$\begin{aligned} P(\text{hits target}) &= P(\text{bull's-eye}) + P(\text{inner ring}) + P(\text{outer ring}) \\ &= 0.10 + 0.25 + 0.45 \\ &= 0.80 \end{aligned}$$

We can now compute the probability that the projectile misses the target by using Equation (2.1):

$$\begin{aligned} P(\text{misses target}) &= 1 - P(\text{hits target}) \\ &= 1 - 0.80 \\ &= 0.20 \end{aligned}$$

Example 6

In a process that manufactures aluminum cans, the probability that a can has a flaw on its side is 0.02, the probability that a can has a flaw on the top is 0.03, and the probability that a can has a flaw on both the side and the top is 0.01.

1. What is the probability that a randomly chosen can has a flaw?

2. What is the probability that it has no flaw?

We are given that $P(\text{flaw on side}) = 0.02$, $P(\text{flaw on top}) = 0.03$, and $P(\text{flaw on side and flaw on top}) = 0.01$. Now $P(\text{flaw}) = P(\text{flaw on side or flaw on top})$. Using Equation (2.5),

$$\begin{aligned} P(\text{flaw on side or flaw on top}) &= P(\text{flaw on side}) + P(\text{flaw on top}) \\ &\quad - P(\text{flaw on side and flaw on top}) \\ &= 0.02 + 0.03 - 0.01 \\ &= 0.04 \end{aligned}$$

To find the probability that a can has no flaw, we compute

$$P(\text{no flaw}) = 1 - P(\text{flaw}) = 1 - 0.04 = 0.96$$

Venn diagrams can sometimes be useful in computing probabilities by showing how to express an event as the union of disjoint events.

Section 2.2: Counting Methods

The Fundamental Principle of Counting:

Assume that k operations are to be performed. If there are n_1 ways to perform the first operation, and if for each of these ways there are n_2 ways to perform the second calculation, and so on, then the total number of ways to perform the sequence of k operations is $n_1 n_2 \cdots n_k$.

Example 7

When ordering a certain type of computer, there are 3 choices of hard drive, 4 choices for the amount of memory, 2 choices of video card, and 3 choices of monitor. In how many ways can a computer be ordered?

Permutations

- A **permutation** is an ordering of a collection of objects. The number of permutations of *n* objects is *n*!, where $n! = n(n-1)(n-2)\cdots(3)(2)(1)$.
  - Note: We define 0! = 1.

- The number of permutations of *k* objects chosen from a group of *n* objects is

$$\frac{n!}{(n-k)!}$$

- When order matters, use permutations.

Example 8

1. Five people stand in line at a movie theater. Into how many different orders can they be arranged?

2. Five lifeguards are available for duty one Saturday afternoon. There are three lifeguard stations. In how many ways can three lifeguards be chosen and ordered among the stations?

Combinations

- **Combinations** are an unordered collection of objects.

- The number of combinations of *k* objects chosen from a group of *n* objects is

$$\frac{n!}{k!(n-k)!}$$

- The number of ways to divide a group of *n* objects into groups of $k_1, \ldots, k_r$ objects where $k_1 + \cdots + k_r = n$, is $\frac{n!}{k_1! \cdots k_r!}$

Example 9

1. At a certain event, 30 people attend, and 5 will be chosen at random to receive door prizes. The prizes are all the same, so the order in which people are chosen does not matter. How many different groups of five people can be chosen?

2. A die is rolled 20 times. Given that three of the rolls came up 1, five came up 2, four came up 3, two came up 4, three came up 5, and three came up 6, how many different arrangements of the outcomes are there?

Section 2.3: Conditional Probability and Independence

Definition: A probability that is based on part of the sample space is called a **conditional probability**.

Let A and B be events with $P(B) \neq 0$. The conditional probability of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad .$$

Back to Example 6

What is the probability that a can will have a flaw on the side, given that it has a flaw on the top?

Independence

Definition: Two events A and B are **independent** if the probability of each event remains the same whether or not the other occurs.

If $P(A) \neq 0$ and $P(B) \neq 0$, then A and B are independent if $P(B \mid A) = P(B)$ or, equivalently, $P(A \mid B) = P(A)$.

If either $P(A) = 0$ or $P(B) = 0$, then A and B are independent.

Independence

Events $A_1, A_2, \ldots, A_n$ are independent if the probability of each remains the same no matter which of the others occur.

The Multiplication Rule

- If A and B are two events and $P(B) \neq 0$, then $P(A \cap B) = P(B)P(A \mid B)$.

- If A and B are two events and $P(A) \neq 0$, then $P(A \cap B) = P(A)P(B \mid A)$.

- If $P(A) \neq 0$, and $P(B) \neq 0$, then both of the above hold.

- If A and B are two independent events, then

  $P(A \cap B) = P(A)P(B)$.

Extended Multiplication Rule

- If A_1, A_2,..., A_n are independent results, then for each collection of A_{j1},..., A_{jm} of events

$$P\left(A_{j1} \cap A_{j2} \cap \cdots \cap A_{jm}\right) = P\left(A_{j1}\right)P\left(A_{j2}\right) \cdots P\left(A_{jm}\right)$$

- In particular,

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2) \cdots P(A_n)$$

Example 10

Of the microprocessors manufactured by a certain process, 20% are defective. Five microprocessors are chosen at random. Assume they function independently. What is the probability that they all work?

Law of Total Probability

Law of Total Probability:

If $A_1, \ldots, A_n$ are mutually exclusive and exhaustive events, and B is any event, then

$$P(B) = P(A_1 \cap B) + \cdots + P(A_n \cap B)$$

Equivalently, if $P(A_i) \neq 0$ for each A_i,

$$P(B) = P(B \mid A_1)P(A_1) + \cdots + P(B \mid A_n)P(A_n)$$

Example 11

Customers who purchase a certain make of car can order an engine in any of three sizes. Of all cars sold, 45% have the smallest engine, 35% have the medium-size one, and 20% have the largest. Of cars with the smallest engine, 10% fail an emissions test within two years of purchase, while 12% of the those with the medium size and 15% of those with the largest engine fail. What is the probability that a randomly chosen car will fail an emissions test within two years?

First Steps Toward a Solution

Let B denote the event that a car fails an emissions test within two years. Let A_1 denote the event that a car has a small engine, A_2 the event that a car has a medium size engine, and A_3 the event that a car has a large engine. Then $P(A_1) = 0.45$, $P(A_2) = 0.35$, and $P(A_3) = 0.20$. Also, $P(B|A_1) = 0.10$, $P(B|A_2) = 0.12$, and

$P(B|A_3) = 0.15$. What is the probability that a car fails an emissions test with two years?

Bayes' Rule

Bayes' Rule (Special Case): Let *A* and *B* be events with $P(A) \neq 0$, $P(A^C) \neq 0$, and $P(B) \neq 0$. Then

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)}$$

Bayes' Rule

Bayes' Rule (General Case): Let $A_1,\ldots, A_n$ be mutually exclusive and exhaustive events, with $P(A_i) \neq 0$ for each A_i. Let B be any event with $P(B) \neq 0$. Then

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}$$

Example 12

The proportion of people in a given community who have a certain disease is 0.005. A test is available to diagnose the disease. If a person has the disease, the probability that the test will produce a positive signal is 0.99. If a person does not have the disease, the probability that the test will produce a positive signal is 0.01. If a person tests positive, what is the probability that the person actually has the disease?

Solution

Let D represent the event that a person actually has the disease, and let + represent the event that the test gives a positive signal. We wish to find $P(D|+)$. We know $P(D) = 0.005$, $P(+|D) = 0.99$, and $P(+|D^C) = 0.01$.

Using Bayes' rule:

$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^C)P(D^C)}$$

$$= \frac{0.99(0.005)}{0.99(0.005) + 0.01(0.995)} = 0.332$$

Section 2.4: Random Variables

Definition: A **random variable** assigns a numerical value to each outcome in a sample space.

Definition: A random variable is **discrete** if its possible values form a discrete set.

This means that if the possible values are arranged in order, there is a gap between each value and the next one. The set of possible values may be infinite; for example, the set of all integers is a discrete set.

Example 13

The number of flaws in a 1-inch length of copper wire manufactured by a certain process varies from wire to wire. Overall, 48% of the wires produced have no flaws, 39% have one flaw, 12% have two flaws, and 1% have three flaws. Let X be the number of flaws in a randomly selected piece of wire.

Then $P(X = 0) = 0.48$, $P(X = 1) = 0.39$, $P(X = 2) = 0.12$, and $P(X = 3) = 0.01$. The list of possible values 0, 1, 2, and 3, along with the probabilities of each, provide a complete description of the population from which X was drawn.

Probability Mass Function

- The description of the possible values of X and the probabilities of each has a name: the probability mass function.

Definition: The **probability mass function** (pmf) of a discrete random variable X is the function $p(x) = P(X = x)$. The probability mass function is sometimes called the **probability distribution**.

Cumulative Distribution Function

- The probability mass function specifies the probability that a random variable is equal to a given value.

- A function called the **cumulative distribution function** (cdf) specifies the probability that a random variable is less than or equal to a given value.

- The cumulative distribution function of the random variable X is the function $F(x) = P(X \leq x)$.

Summary for Discrete Random Variables

Let X be a discrete random variable. Then

- The probability mass function of X is the function $p(x) = P(X = x)$.

- The cumulative distribution function of X is the function $F(x) = P(X \leq x)$.

- $F(x) = \sum_{t \leq x} p(t) = \sum_{t \leq x} P(X = t)$.

- $\sum_{x} p(x) = \sum_{x} P(X = x) = 1$, where the sum is over all

The possible values of X.

Example 14

Recall the example of the number of flaws in a randomly chosen piece of wire. The following is the pmf: $P(X = 0) = 0.48$, $P(X = 1) = 0.39$, $P(X = 2) = 0.12$, and $P(X = 3) = 0.01$.

For any value x, we compute $F(x)$ by summing the probabilities of all the possible values of x that are less than or equal to x.

$F(0) = P(X \leq 0) = 0.48$

$F(1) = P(X \leq 1) = 0.48 + 0.39 = 0.87$

$F(2) = P(X \leq 2) = 0.48 + 0.39 + 0.12 = 0.99$

$F(3) = P(X \leq 3) = 0.48 + 0.39 + 0.12 + 0.01 = 1$

Mean for Discrete Random Variables

- Let X be a discrete random variable with probability mass function $p(x) = P(X = x)$.
- The **mean** of X is given by

$$\mu_X = \sum_x xP(X = x) \quad ,$$

where the sum is over all possible values of X.

- The mean of X is sometimes called the expectation, or expected value, of X and may also be denoted by $E(X)$ or by μ.

Variance for Discrete Random Variables

- Let X be a discrete random variable with probability mass function $p(x) = P(X = x)$.
- The **variance** of X is given by

$$\sigma_X^2 = \sum_x (x - \mu_X)^2 P(X = x)$$
$$= \sum_x x^2 P(X = x) - \mu_X^2 .$$

- The variance of X may also be denoted by $V(X)$ or by σ^2.
- The standard deviation is the square root of the variance:

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

Example 15

A certain industrial process is brought down for recalibration whenever the quality of the items produced falls below specifications. Let *X* represent the number of times the process is recalibrated during a week, and assume that *X* has the following probability mass function.

x	0	1	2	3	4
$p(x)$	0.35	0.25	0.20	0.15	0.05

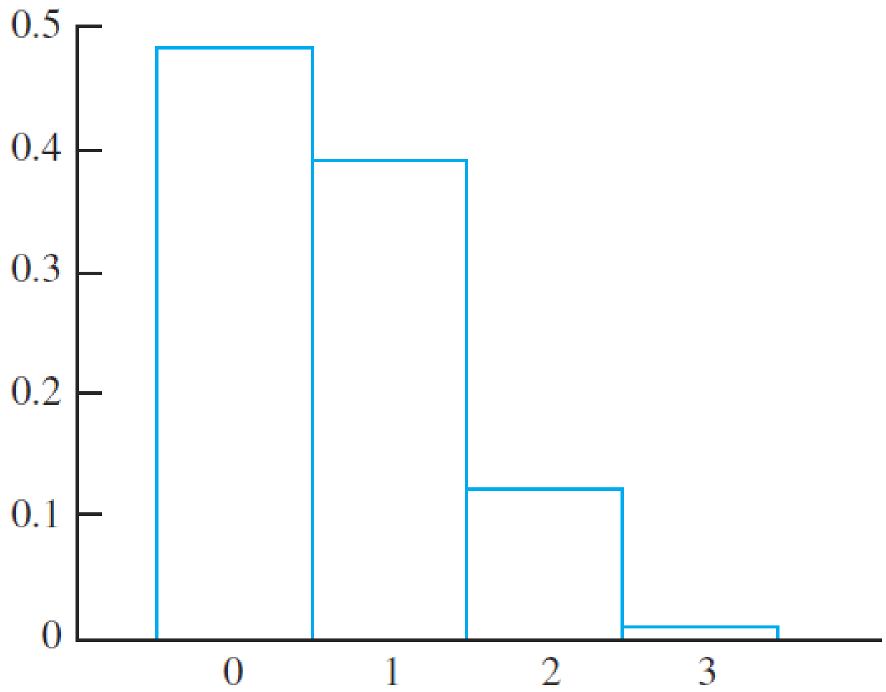
Find the mean and variance of *X*.

The Probability Histogram

- When the possible values of a discrete random variable are evenly spaced, the probability mass function can be represented by a histogram, with rectangles centered at the possible values of the random variable.
- The area of the rectangle centered at a value x is equal to $P(X = x)$.
- Such a histogram is called a **probability histogram**, because the areas represent probabilities.

Probability Histogram for the Number of Flaws in a Wire

The pmf is: $P(X = 0) = 0.48$, $P(X = 1) = 0.39$, $P(X = 2) = 0.12$, and $P(X = 3) = 0.01$.



Example 16

Construct a probability histogram for the example with the number of weekly recalibrations.

Continuous Random Variables

- A random variable is **continuous** if its probabilities are given by areas under a curve.
- The curve is called a **probability density function** (pdf) for the random variable. Sometimes the pdf is called the **probability distribution**.
- The function $f(x)$ is the probability density function of X.
- Let X be a continuous random variable with probability density function $f(x)$. Then

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

Computing Probabilities

Let X be a continuous random variable with probability density function $f(x)$. Let a and b be any two numbers, with $a < b$. Then

$$P(a \le X \le b) = P(a \le X < b) = P(a < X \le b) = \int_a^b f(x)dx.$$

In addition,

$$P(X \le a) = P(X < a) = \int_{-\infty}^a f(x)dx$$

$$P(X \ge a) = P(X > a) = \int_a^\infty f(x)dx.$$

More on Continuous Random Variables

- Let X be a continuous random variable with probability density function *f*(*x*). The **cumulative distribution function** of X is the function

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt.$$

- The mean of X is given by

$$\mu_X = \int_{-\infty}^{\infty} x f(x)\,dx.$$

- The variance of X is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\,dx$$

$$= \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu_X^2.$$

Example 17

A hole is drilled in a sheet-metal component, and then a shaft is inserted through the hole. The shaft clearance is equal to the difference between the radius of the hole and the radius of the shaft. Let the random variable *X* denote the clearance, in millimeters. The probability density function of *X* is

$$f(x) = \begin{cases} 1.25(1-x^4), & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

1. Components with clearances larger than 0.8 mm must be scraped. What proportion of components are scrapped?
2. Find the cumulative distribution function *F*(*x*) .

Median for a Continuous Random Variable

Let X be a continuous random variable with probability mass function $f(x)$ and cumulative distribution function $F(x)$.

- The median of X is the point x_m that solves the equation

$$F(x_m) = P(X \leq x_m) = \int_{-\infty}^{x_m} f(x)dx = 0.5.$$

Percentiles

- If p is any number between 0 and 100, the pth percentile is the point x_p that solves the equation

$$F(x_p) = P(X \leq x_p) = \int_{-\infty}^{x_p} f(x)dx = p/100.$$

- Note: the median is the 50th percentile.

Chebyshev's Inequality

- Let X be a random variable with mean μ_X and standard deviation σ_X. Then

$$P(|X - \mu_X| \geq k\sigma_X) \leq \frac{1}{k^2}$$

- Chebyshev's inequality is valid for any random variable and does not require knowledge of the distribution. The bound tends to overestimate the desired probability.

Section 2.5: Linear Functions of Random Variables

If X is a random variable, and a and b are constants, then

- $\mu_{aX+b} = a\mu_X + b$,
- $\sigma^2_{aX+b} = a^2\sigma^2_X$,
- $\sigma_{aX+b} = |a|\sigma_X$.

More Linear Functions

If X and Y are random variables, and a and b are constants, then

$$\mu_{aX+bY} = \mu_{aX} + \mu_{bY} = a\mu_X + b\mu_Y .$$

More generally, if X_1, …, X_n are random variables and c_1, …, c_n are constants, then the mean of the linear combination $c_1 X_1 + \ldots + c_n X_n$ is given by

$$\mu_{c_1X_1+c_2X_2+\ldots+c_nX_n} = c_1\mu_{X_1} + c_2\mu_{X_2} + \ldots + c_n\mu_{X_n} .$$

Two Independent Random Variables

If X and Y are **independent** random variables, and S and T are sets of numbers, then

$$P(X \in S \text{ and } Y \in T) = P(X \in S)P(Y \in T).$$

More generally, if $X_1, \ldots, X_n$ are independent random variables, and $S_1, \ldots, S_n$ are sets, then

$$P(X_1 \in S_1, X_2 \in S_2, \cdots, X_n \in S_n) = P(X_1 \in S_1)P(X_2 \in S_2) \cdots P(X_n \in S_n)$$

.

Variance Properties

If $X_1, \ldots, X_n$ are *independent* random variables, then the variance of the sum $X_1 + \cdots + X_n$ is given by $\sigma^2_{X_1+X_2+\ldots+X_n} = \sigma^2_{X_1} + \sigma^2_{X_2} + \ldots. + \sigma^2_{X_n}$.

If $X_1, \ldots, X_n$ are *independent* random variables and $c_1, \ldots, c_n$ are constants, then the variance of the linear combination $c_1 X_1 + \cdots + c_n X_n$ is given by $\sigma^2_{c_1X_1+c_2X_2+\ldots+c_nX_n} = c_1^2\sigma^2_{X_1} + c_2^2\sigma^2_{X_2} + \ldots. + c_n^2\sigma^2_{X_n}$.

More Variance Properties

If X and Y are *independent* random variables with variances σ_X^2 and σ_Y^2, then the variance of the sum $X + Y$ is

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2.$$

The variance of the difference $X - Y$ is

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2.$$

Example 18

A piston is placed inside a cylinder. The clearance is the distance between the edge of the piston and the wall of the cylinder and is equal to one-half the difference between the cylinder diameter and the piston diameter. Assume the piston diameter has a mean of 80.85 cm with a standard deviation of 0.02 cm. Assume the cylinder diameter has a mean of 80.95 cm with a standard deviation of 0.03 cm. Find the mean clearance. Assuming that the piston and cylinder are chosen independently, find the standard deviation of the clearance.

Solution

Let X_1 represent the diameter of the cylinder and let X_2 the diameter of the piston. The clearance is given by $C = 0.5X_1 - 0.5X_2$. From above equation, the mean clearance is

$$\begin{aligned}\mu_C &= \mu_{0.5X_1-0.5X_2} \\ &= 0.5\mu_{X_1} - 0.5\mu_{X_2} \\ &= 0.5(80.95) - 0.5(80.85) \\ &= 0.050\end{aligned}$$

Since X_1 and X_2 are independent, we can use this equation to find the standard deviation σ_C:

$$\begin{aligned}\sigma_C &= \sqrt{\sigma^2_{0.5X_1-0.5X_2}} \\ &= \sqrt{(0.5)^2\sigma^2_{X_1} + (-0.5)^2\sigma^2_{X_2}} \\ &= \sqrt{0.25(0.02)^2 + 0.25(0.03)^2} \\ &= 0.018\end{aligned}$$

Independence and Simple Random Samples

Definition: If $X_1, \ldots, X_n$ is a **simple random sample**, then $X_1, \ldots, X_n$ may be treated as independent random variables, all with the same distribution.

When $X_1, \ldots, X_n$ are independent random variables, all with the same distribution, we sometimes say that $X_1, \ldots, X_n$ are **independent and identically distributed (i.i.d)**.

Properties of $\bar{X}$

If $X_1, \ldots, X_n$ is a simple random sample from a population with mean μ and variance σ^2, then the sample mean $\bar{X}$ is a random variable with

$$\mu_{\bar{X}} = \mu$$

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}.$$

The standard deviation of $\bar{X}$ is

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}.$$

Example 19

A process that fills plastic bottles with a beverage has a mean fill volume of 2.013 L and a standard deviation of 0.005 L. A case contains 24 bottles. Assuming that the bottles in a case are a simple random sample of bottles filled by this method, find the mean and standard deviation of the average volume per bottle in a case.

Solution

Let $V_1, \ldots, V_{24}$ represent the volumes in 24 bottles in a case. This is a simple random sample from a population with mean $\mu = 2.013$ and standard deviation $\sigma = 0.005$. The average volume is $\overline{V} = (V_1 + \cdots + V_{24})/24$. We find that

$$\mu_{\overline{V}} = \mu = 2.013$$

And

$$\sigma_{\overline{V}} = \frac{\sigma}{\sqrt{24}} = 0.001$$

Section 2.6: Jointly Distributed Random Variables

If X and Y are jointly discrete random variables:

- The **joint probability mass function** of X and Y is the function

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

- The joint probability mass function has the property that

$$\sum_x \sum_y p(x, y) = 1$$

where the sum is taken over all the possible values of X and Y.

Marginal Probability Mass Functions

- The **marginal probability mass functions** of *X* and *Y* can be obtained from the joint probability mass function as follows:

$$p_X(x) = P(X = x) = \sum_y p(x, y)$$

$$p_Y(y) = P(Y = y) = \sum_x p(x, y)$$

where the sums are taken over all the possible values of *Y* and of *X*, respectively.

Jointly Continuous Random Variables

If X and Y are jointly continuous random variables, with joint probability density function $f(x, y)$, and $a < b$, $c < d$, then

$$P(a \le X \le b \text{ and } c \le Y \le d) = \int_a^b \int_c^d f(x, y)\,dy\,dx.$$

The joint probability density function has the property that

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

Marginal Distributions of *X* and *Y*

If *X* and *Y* are jointly continuous with joint probability density function $f(x,y)$, then the marginal probability density functions of *X* and *Y* are given, respectively, by

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

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

Example 20

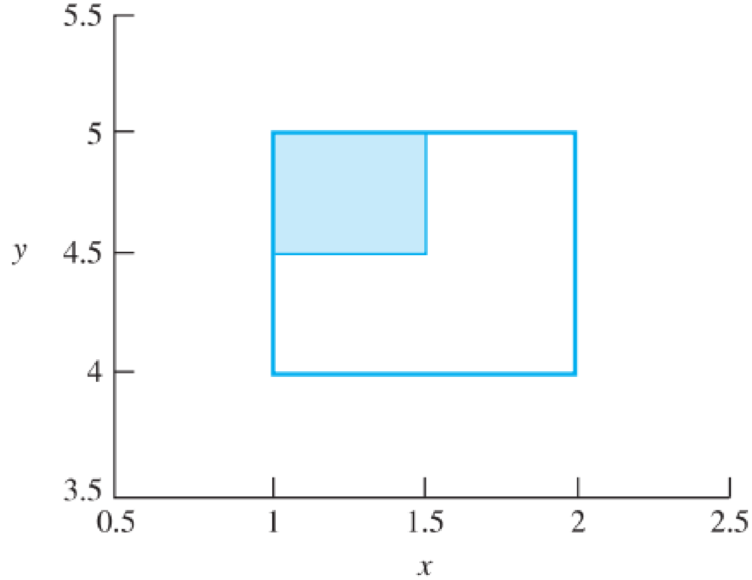
Assume that for a certain type of washer, both the thickness and the hole diameter vary from item to item. Let *X* denote the thickness in millimeters and let *Y* denote the hole diameter in millimeters, for a randomly chosen washer. Assume that the joint probability density of *X* and *Y* is given by

$$f(x,y) = \begin{cases} \frac{1}{6}(x+y), \text{ if } 1 \le x \le 2 \text{ and } 4 \le y \le 5 \\ 0, \text{ otherwise} \end{cases}$$

Questions on Ex. 20

1. Find the probability that a randomly chosen washer has a thickness between 1.0 and 1.5 mm, and a hole diameter between 4.5 and 5 mm.
2. Find the marginal probability density function of the thickness X of a washer.
3. Find the marginal probability density function of the hole diameter Y of a washer.

Solution 1.



We integrate the joint probability density function over the indicated region:

$$
\begin{aligned}
P(1 \le X \le 1.5 \text{ and } 4.5 \le Y \le 5) &= \int_1^{1.5} \int_{4.5}^{5} \frac{1}{6}(x+y)\,dy\,dx \\
&= \int_1^{1.5} \left\{ \frac{xy}{6} + \frac{y^2}{12} \Big|_{y=4.5}^{y=5} \right\} dx \\
&= \int_1^{1.5} \left(\frac{x}{12} + \frac{19}{48} \right) dx \\
&= \frac{1}{4}
\end{aligned}
$$

Solution 2.

Denote the marginal probability density function of X by $f_X(x)$, and the marginal probability density function of Y by $f_Y(y)$. Then

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y)\, dy \\ &= \int_4^5 \frac{1}{6}(x + y)\, dy \\ &= \frac{1}{6}\left(x + \frac{9}{2}\right) \qquad \text{for } 1 \le x \le 2 \end{aligned}$$

Solution 3.

and

$$
\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f(x, y)\, dx \\
&= \int_{1}^{2} \frac{1}{6}(x + y)\, dx \\
&= \frac{1}{6}\left(y + \frac{3}{2}\right) \qquad \text{for } 4 \leq y \leq 5
\end{aligned}
$$

More Than Two Random Variables

If the random variables $X_1, \ldots, X_n$ are jointly discrete, the joint probability mass function is

$$p(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n).$$

If the random variables $X_1, \ldots, X_n$ are jointly continuous, they have a joint probability density function $f(x_1, x_2, \ldots, x_n)$, where

$$P(a_1 \le X_1 \le b_1, \ldots, a_n \le X_n \le b_n)$$

$$= \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f(x_1, \ldots, x_n)\,dx_1 \cdots dx_n.$$

for any constants $a_1 \le b_1, \ldots, a_n \le b_n$.

Means of Functions of Random Variables

Let X be a random variable, and let $h(X)$ be a function of X. Then:

- If X is discrete with probability mass function $p(x)$, the mean of $h(X)$ is given by

$$\mu_{h(X)} = \sum_x h(x)p(x).$$

where the sum is taken over all the possible values of X.

- If X is continuous with probability density function $f(x)$, the mean of $h(X)$ is given by

$$\mu_{h(X)} = \int_{-\infty}^{\infty} h(x)f(x)dx.$$

Functions of Joint Random Variables

If X and Y are jointly distributed random variables, and $h(X, Y)$ is a function of X and Y, then

If X and Y are jointly discrete with joint probability mass function $p(x, y)$,

$$\mu_{h(X,Y)} = \sum_x \sum_y h(x,y)\,p(x,y).$$

where the sum is taken over all possible values of X and Y.

If X and Y are jointly continuous with joint probability mass function $f(x, y)$,

$$\mu_{h(X,Y)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)\,f(x,y)\,dx\,dy.$$

Example 21

An internal combustion engine contains several cylinders bored into an engine block. Let X denote the bore diameter of a cylinder, in millimeters. Assume that the probability distribution of X is

$$f(x) = \begin{cases} 10, & 80.5 < x < 80.6 \\ 0, & \text{otherwise} \end{cases}$$

Let $A = \pi X^2/4$ represent the area of the bore. Find the mean of A.

Solution

$$\mu_A = \int_{-\infty}^{\infty} \frac{\pi x^2}{4} f(x)\, dx$$

$$= \int_{80.5}^{80.6} \frac{\pi x^2}{4} (10)\, dx$$

$$= 5096$$

The mean area is 5096 mm^2.

Conditional Distributions

Let X and Y be jointly discrete random variables, with joint probability density function $p(x, y)$. Let $p_X(x)$ denote the marginal probability mass function of X and let x be any number for which $p_X(x) > 0$.

The **conditional probability mass function of Y given $X = x$** is

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

Note that for any particular values of x and y, the value of $p_{Y|X}(y \mid x)$ is just the conditional probability $P(Y = y \mid X = x)$.

Continuous Conditional Distributions

Let X and Y be jointly continuous random variables, with joint probability density function $f(x, y)$. Let $f_X(x)$ denote the marginal density function of X and let x be any number for which $f_X(x) > 0$.

The **conditional probability density function of Y given $X = x$** is

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

Example 20 cont.

Find the probability that the hole diameter is less than or equal to 4.8 mm given that the thickness is 1.2 mm.

Earlier, we have computed,

$$f_X(x) = \frac{1}{6}(x + 4.5) \quad \text{for } 1 \le x \le 2 \qquad f_Y(y) = \frac{1}{6}(y + 1.5) \quad \text{for } 4 \le y \le 5$$

The conditional probability density function of Y given $X = 1.2$ is

$$f_{Y|X}(y \mid 1.2) = \frac{f(1.2,\ y)}{f_X(1.2)}$$

$$= \begin{cases} \dfrac{(1/6)(1.2 + y)}{(1/6)(1.2 + 4.5)} & \text{if } 4 \le y \le 5 \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \dfrac{1.2 + y}{5.7} & \text{if } 4 \le y \le 5 \\ 0 & \text{otherwise} \end{cases}$$

Therefore,

The probability that the hole diameter is less than or equal to 4.8 mm given that the thickness is 1.2 mm is $P(Y \leq 4.8 \mid X = 1.2)$. This is found by integrating $f_{Y|X}(y|1.2)$ over the region $y \leq 4.8$:

$$P(Y \leq 4.8 \mid X = 1.2) = \int_{-\infty}^{4.8} f_{Y|X}(y \mid 1.2)\, dy$$

$$= \int_{4}^{4.8} \frac{1.2 + y}{5.7}\, dy$$

$$= 0.786$$

Conditional Expectation

- Expectation is another term for mean.
- A **conditional expectation** is an expectation, or mean, calculated using a conditional probability mass function or conditional probability density function.
- The conditional expectation of Y given $X = x$ is denoted by $E(Y|X = x)$ or $\mu_{Y|X=x}$.

Example 20 cont.

Find the conditional expectation of Y (hole diameter) given that the thickness is $X = 1.2$.

Since X and Y are jointly continuous, we use the definition of the mean of a continuous random variable to compute the conditional expectation.

$$E(Y \mid X = 1.2) = \int_{-\infty}^{\infty} y f_{Y|X}(y \mid 1.2)\, dy$$

$$= \int_{4}^{5} y \frac{1.2 + y}{5.7}\, dy$$

$$= 4.5146$$

Independence for Two Random Variables

Two random variables X and Y are independent, provided that:

- If X and Y are jointly discrete, the joint probability mass function is equal to the product of the marginals:

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

- If X and Y are jointly continuous, the joint probability density function is equal to the product of the marginals:

$$f(x_1, y) = f(x)\, f(y).$$

Independence for More Than Two Random Variables

Random variables $X_1, \ldots, X_n$ are independent, provided that:

- If $X_1, \ldots, X_n$ are jointly discrete, the joint probability mass function is equal to the product of the marginals:

$$p(x_1, \ldots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n).$$

- If $X_1, \ldots, X_n$ are jointly continuous, the joint probability density function is equal to the product of the marginals:

$$f(x_1, \ldots, x_n) = f(x_1) \cdots f(x_n).$$

Independence (cont.)

If X and Y are independent random variables, then:

- If X and Y are jointly discrete, and x is a value for which $p_X(x) > 0$, then

$$p_{Y|X}(y \mid x) = p_Y(y).$$

- If X and Y are jointly continuous, and x is a value for which $f_X(x) > 0$, then

$$f_{Y|X}(y \mid x) = f_Y(y).$$

Example 20 cont.

Are the length *X* and thickness *Y* independent?

Earlier, we have computed,

$$f_X(x) = \frac{1}{6}\left(x + \frac{9}{2}\right) \qquad f_Y(y) = \frac{1}{6}\left(y + \frac{3}{2}\right)$$

Clearly $f(x, y) \neq f_X(x) f_Y(y)$. Therefore X and Y are not independent.

Covariance

- Let X and Y be random variables with means μ_X and μ_Y.
- The **covariance** of X and Y is

$$\text{Cov}(X,Y) = \mu_{(X-\mu_X)(Y-\mu_Y)}.$$

- An alternate formula is

$$\text{Cov}(X,Y) = \mu_{XY} - \mu_X \mu_Y.$$

Example 22

A mobile computer is moving in the region A bounded by the x axis, the line $x = 1$, and the line $y = x$. If (X, Y) denotes the position of the computer at a given time, the joint density of X and Y is given by

$$f(x,y) = \begin{cases} 8xy, & (x,y) \in A \\ 0, & (x,y) \notin A \end{cases}$$

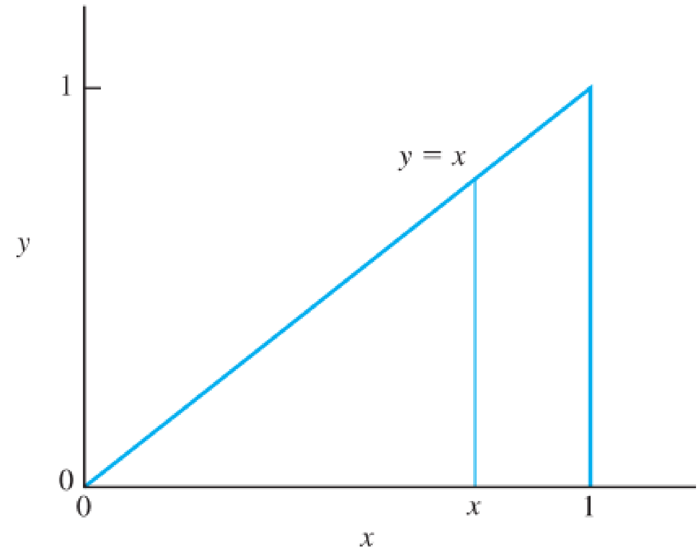
Find Cov(X, Y).

Solution

We will use the formula $\text{Cov}(X,Y) = \mu_{XY} - \mu_X \mu_Y$. First we compute μ_{XY}:

$$\mu_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f(x,y)\, dy\, dx$$

Now the joint density is positive on the triangle shown.



To compute the integral over this region, we fix a value of x, as shown. We compute the inner integral by integrating with respect to y along the vertical line through x. The limits of integration along this line are $y = 0$ to $y = x$. Then we compute the outer integral by integrating with respect to x over all possible values of x, so the limits of integration on the outer integral are $x = 0$ to $x = 1$.

Therefore

$$\begin{aligned}\mu_{XY} &= \int_0^1 \int_0^x xy(8xy)\,dy\,dx \\ &= \int_0^1 \left(\int_0^x 8x^2y^2\,dy \right) dx \\ &= \int_0^1 \frac{8x^5}{3}\,dx \\ &= \frac{4}{9}\end{aligned}$$

To find μ_X and μ_Y, we will use the marginal densities computed in Example 2.57. These are

$$f_X(x) = \begin{cases} 4x^3 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} 4y - 4y^3 & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

We now compute μ_X and μ_Y:

$$\mu_X = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_0^1 4x^4\, dx = \frac{4}{5}$$

$$\mu_Y = \int_{-\infty}^{\infty} y f_Y(y)\, dy = \int_0^1 \left(4y^2 - 4y^4\right)\, dy = \frac{8}{15}$$

Now $\mathrm{Cov}(X,Y) = \frac{4}{9} - \left(\frac{4}{5}\right)\left(\frac{8}{15}\right) = \frac{4}{225} = 0.01778.$

Correlation

- Let X and Y be jointly distributed random variables with standard deviations σ_X and σ_Y.

- The **correlation** between X and Y is denoted $\rho_{X,Y}$ and is given by

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

- For any two random variables X and Y;

$$-1 \leq \rho_{X,Y} \leq 1.$$

Example 22 cont.

Find $\rho_{X,Y}$ in the mobile computer example.

From earlier results, we find that:

$$\sigma_X^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - \mu_X^2$$
$$= \int_0^1 4x^5\,dx - \left(\frac{4}{5}\right)^2$$
$$= 0.02667$$
$$\sigma_Y^2 = \int_{-\infty}^{\infty} y^2 f_Y(y)\,dy - \mu_Y^2$$
$$= \int_0^1 \left(4y^3 - 4y^5\right)\,dy - \left(\frac{8}{15}\right)^2$$
$$= 0.04889$$

It follows that $\rho_{X,Y} = \dfrac{0.01778}{\sqrt{(0.02667)(0.04889)}} = 0.492.$

Covariance, Correlation, and Independence

- If $\mathrm{Cov}(X,Y) = \rho_{X,Y} = 0$, then X and Y are said to be uncorrelated.
- If X and Y are independent, then X and Y are uncorrelated.
- It is mathematically possible for X and Y to be uncorrelated without being independent. This rarely occurs in practice.

Covariance of Random Variables

If $X_1, \ldots, X_n$ are random variables and $c_1, \ldots, c_n$ are constants, then

$$\mu_{c_1X_1+\cdots+c_nX_n} = c_1\mu_{X_1} + \cdots + c_n\mu_{X_n}$$

$$\sigma^2_{c_1X_1+\cdots+c_nX_n} = c_1^2\sigma^2_{X_1} + \cdots + c_n^2\sigma^2_{X_n} + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} c_i c_j \operatorname{Cov}(X_i, X_j).$$

Covariance of Independent Random Variables

- If $X_1, \ldots, X_n$ are *independent* random variables and $c_1, \ldots, c_n$ are constants, then

$$\sigma^2_{c_1X_1+\cdots+c_nX_n} = c_1^2\sigma^2_{X_1} + \cdots + c_n^2\sigma^2_{X_n}.$$

- In particular,

$$\sigma^2_{X_1+\cdots+X_n} = \sigma^2_{X_1} + \cdots + \sigma^2_{X_n}.$$

Example 22 cont.

Assume that the mobile computer moves from a random position (X, Y) vertically to the point $(X, 0)$, and then along the x axis to the origin. Find the mean and variance of the distance traveled.

Summary

- Probability and rules
- Counting techniques
- Conditional probability
- Independence
- Random variables: discrete and continuous
- Probability mass functions

Summary Continued

- Probability density functions
- Cumulative distribution functions
- Means and variances for random variables
- Linear functions of random variables
- Mean and variance of a sample mean
- Jointly distributed random variables