CHAPTER 9

Probability and Random Variables

IN THIS CHAPTER

Summary: We've completed the basics of data analysis and we now begin the transition to inference. In order to do inference, we need to use the language of probability. In order to use the language of probability, we need an understanding of random variables and probabilities. The next two chapters lay the probability foundation for inference. In this chapter, we'll learn about the basic rules of probability, what it means for events to be independent, and about discrete and continuous random variables, simulation, and rules for combining random variables.

KEY IDEA

Key Ideas
- Probability
- Random Variables
- Discrete Random Variables
- Continuous Random Variables
- Probability Distributions
- Normal Probability
- Simulation
- Transforming and Combining Random Variables

Probability

The second major part of a course in statistics involves making *inferences* about populations based on sample data (the first was *exploratory data analysis*). The ability to do this is based on being able to make statements such as, "The probability of getting a finding as different, or more different, from expected as we got by chance alone, under the assumption that the null hypothesis is true, is 0.6." To make sense of this statement, you need to have a understanding of what is meant by the term "probability" as well as an understanding of some of the basics of probability theory.

An **experiment or chance experiment (random phenomenon):** An activity whose outcome we can observe or measure but we do not know how it will turn out on any single trial. Note that this is a somewhat different meaning of the word "experiment" than we developed in the last chapter.

example: if we roll a die, we know that we will get a 1, 2, 3, 4, 5, or 6, but we don't know *which* one of these we will get on the next trial. Assuming a fair die, however, we *do* have a good idea of approximately what proportion of each possible outcome we will get over a large number of trials.

Outcome: One of the possible results of an experiment (random phenomenon).

example: the possible outcomes for the roll of a single die are 1, 2, 3, 4, 5, 6. Individual outcomes are sometimes called **simple events**.

Sample Spaces and Events

Sample space: The set of all possible outcomes, or simple events, of an experiment.

example: For the roll of a single die, S = {1, 2, 3, 4, 5, 6}.

Event: A collection of outcomes or simple events. That is, an event is a subset of the sample space.

example: For the roll of a single die, the sample space (all outcomes or simple events) is S = {1, 2, 3, 4, 5, 6}. Let event A = "the value of the die is 6." Then A = {6}. Let B = "the face value is less than 4." Then B = {1, 2, 3}. Events A and B are subsets of the sample space.

example: Consider the experiment of flipping two coins and noting whether each coin lands heads or tails. The sample space is S = {HH, HT, TH, TT}. Let event B = "at least one coin shows a head." Then B = {HH, HT, TH}. Event B is a subset of the sample space S.

Probability of an event: the relative frequency of the outcome. That is, it is the fraction of time that the outcome would occur if the experiment were repeated indefinitely. If we let E = the event in question, s = the number of ways an outcome can succeed, and f = the number of ways an outcome can fail, then

$$P(\mathrm{E}) = \frac{s}{s+f}.$$

Note that $s + f$ equals the number of outcomes in the sample space. Another way to think of this is that the probability of an event is the sum of the probabilities of all outcomes that make up the event.

For any event A, P(A) ranges from 0 to 1, inclusive. That is, $0 \le P(\mathrm{A}) \le 1$. This is an algebraic result from the definition of probability when success is guaranteed ($f = 0$, $s = 1$) or failure is guaranteed ($f = 1$, $s = 0$).

The sum of the probabilities of all possible outcomes in a sample space is one. That is, if the sample space is composed of n possible outcomes,

$$\sum_{i=1}^{n} p_i = 1.$$

example: In the experiment of flipping two coins, let the event A = obtain at least one head. The sample space contains four elements ({HH, HT, TH, TT}). $s = 3$ because there are three ways for our outcome to be considered a success ({HH, HT, TH}) and $f = 1$.

Thus

$$P(\text{A}) = \frac{3}{3+1} = \frac{3}{4}.$$

example: Consider rolling two fair dice and noting their sum. A sample space for this event can be given in table form as follows:

Face	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Let B = "the sum of the two dice is greater than 4." There are 36 outcomes in the samples space, 30 of which are greater than 4. Thus,

$$P(\text{B}) = \frac{30}{36} = \frac{5}{6}.$$

Furthermore,

$$\sum p_i = P(2) + P(3) + \cdots + P(12) = \frac{1}{36} + \frac{2}{36} + \cdots + \frac{1}{36} = 1.$$

Probabilities of Combined Events

***P*(A or B):** The probability that **either** event A **or** event B occurs. (They can both occur, but only one needs to occur.) Using set notation, *P*(A or B) can be written $P(\text{A} \cup \text{B})$. $\text{A} \cup \text{B}$ is spoken as, "A union B."

***P*(A and B):** The probability that **both** event A **and** event B occur. Using set notation, *P*(A and B) can be written $P(\text{A} \cap \text{B})$. $\text{A} \cap \text{B}$ is spoken as, "A intersection B."

example: Roll two dice and consider the sum (see table). Let A = "one die shows a 3," B = "the sum is greater than 4." Then *P*(A or B) is the probability that *either* one die shows a 3 *or* the sum is greater than 4. Of the 36 possible outcomes in the sample space, there are 32 possible outcomes that are successes [30 outcomes greater than 4 as well as (1,3) and (3,1)], so

$$P(\text{A or B}) = \frac{32}{36}.$$

There are nine ways in which a sum has one die showing a 3 and has a sum greater than 4: [(3,2), (3,3), (3,4), (3,5), (3,6), (2,3), (4,3), (5,3), (6,3)], so

$$P(\text{A and B}) = \frac{9}{36}.$$

Complement of an event A: events in the sample space that are not in event A. The complement of an event A is symbolized by $\bar{A}$, or A^c. Furthermore, $P(\bar{A}) = 1 - P(A)$.

Mutually Exclusive Events

Mutually exclusive (disjoint) events: Two events are said to be *mutually exclusive* (some texts refer to mutually exclusive events as *disjoint*) if and only if they have no outcomes in common. That is, $A \cap B = \varnothing$. If A and B are mutually exclusive, then $P(\text{A and B}) = P(A \cap B) = 0$.

example: in the two-dice rolling experiment, A = "face shows a 1" and B = "sum of the two dice is 8" are mutually exclusive because there is no way to get a sum of 8 if one die shows a 1. That is, events A and B cannot both occur.

Conditional Probability

Conditional Probability: "The probability of A given B" assumes we have knowledge of an event B having occurred before we compute the probability of event A. This is symbolized by $P(A|B)$. Also,

$$P(A \mid B) = \frac{P(\text{A and B})}{P(B)}.$$

Although this formula will work, it's often easier to think of a condition as reducing, in some fashion, the original sample space. The following example illustrates this "shrinking sample space."

example: Once again consider the possible sums on the roll of two dice. Let A = "the sum is 7," B = "one die shows a 5." We note, by counting outcomes in the table, that $P(A) = 6/36$. Now, consider a slightly different question: what is $P(A|B)$ (that is, what is the probability of the sum being 7 *given that* one die shows a 5)?

solution: Look again at the table:

Face	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

The condition has effectively reduced the sample space from 36 outcomes to only 11 (you do not count the "10" twice). Of those, two are 7s. Thus, the *P*(*the sum is* 7 | *one die shows a* 5) = 2/11.

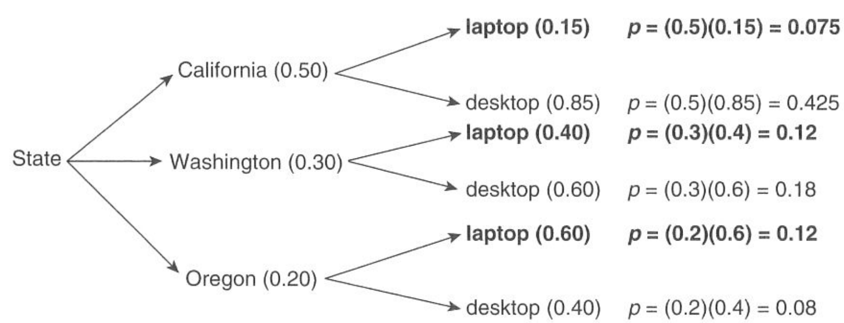
alternate solution: If you insist on using the formula for conditional probability, we note that *P*(A and B) = *P*(*the sum is* 7 and *one die shows a* 5) = 2/36, and *P*(B) = *P*(*one die shows a* 5) = 11/36. By formula

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{2/36}{11/36} = \frac{2}{11}.$$

Some conditional probability problems can be solved by using a **tree diagram.** A tree diagram is a schematic way of looking at all possible outcomes.

example: Suppose a computer company has manufacturing plants in three states. 50% of its computers are manufactured in California, and 85% of these are desktops; 30% of computers are manufactured in Washington, and 40% of these are laptops; and 20% of computers are manufactured in Oregon, and 40% of these are desktops. All computers are first shipped to a distribution site in Nebraska before being sent out to stores. If you picked a computer at random from the Nebraska distribution center, what is the probability that it is a laptop?

solution:



Note that the final probabilities add to 1 so we know we have considered all possible outcomes. Now, *P*(*laptop*) = **0.075 + 0.12 + 0.12 = 0.315.**

Independent Events

Independent Events: Events A and B are said to be *independent* if and only if

(i) Are A and B independent?

solution: $P(A|B) = P$(the card drawn is an ace | the card is a 10, J, Q, K, or A) = 4/20 = 1/5 (there are 20 cards to consider, 4 of which are aces). Since $P(A) = 1/13$, knowledge of B has changed what we know about A. That is, in this case, $P(A) \neq P(A|B)$, so events A and B are *not* independent.

(ii) Are A and C independent?

solution: $P(A|C) = P$(the card drawn is an ace | the card drawn is a diamond) = 1/13 (there are 13 diamonds, one of which is an ace). So, in this case, $P(A) = P(A|C)$, so that the events "the card drawn is an ace" and "the card drawn is a diamond" are independent.

Probability of A and B or A or B

The Addition Rule: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.

Special case of *The Addition Rule*: If A and B are *mutually exclusive,*

$P(A \text{ and } B) = 0$, so $P(A \text{ or } B) = P(A) + P(B)$.

The Multiplication Rule: $P(A \text{ and } B) = P(A) \bullet P(B|A)$.

Special case of *The Multiplication Rule:* If A and B are *independent,*

$P(B|A) = P(B)$, so $P(A \text{ and } B) = P(A) \bullet P(B)$.

example: If A and B are two mutually exclusive events for which $P(A) = 0.3$, $P(B) = 0.25$. Find $P(A \text{ or } B)$.

solution: $P(A \text{ or } B) = 0.3 + 0.25 = 0.55$.

example: A basketball player has a 0.6 probability of making a free throw. What is his probability of making two consecutive free throws if

(a) he gets very nervous after making the first shot and his probability of making the second shot drops to 0.4.

solution: P(making the first shot) = 0.6, P(making the second shot | he made the first) = 0.4. So, P(making both shots) = (0.6)(0.4) = 0.24.

(b) the events "he makes his first shot" and "he makes the succeeding shot" are independent.

solution: Since the events are independent, his probability of making each shot is the same. Thus, P(he makes both shots) = (0.6)(0.6) = 0.36.

Random Variables

Recall our earlier definition of an **experiment (random phenomenon):** An activity whose outcome we can observe and measure, but for which we can't predict the result of any single trial. A **random variable,** X**,** is a numerical value assigned to an outcome of a random phenomenon. Particular values of the random variable X are often given small case names, such as x. It is common to see expressions of the form $P(X = x)$, which refers to the probability that the random variable X takes on the particular value x.

example: If we roll a fair die, the random variable X could be the face-up value of the die. The possible values of X are $\{1, 2, 3, 4, 5, 6\}$. $P(X = 2) = 1/6$.

example: The score a college-hopeful student gets on her SAT test can take on values from 200 to 800. These are the possible values of the random variable *X*, the score a randomly selected student gets on his/her test.

There are two types of random variables: **discrete random variables** and **continuous random variables.**

Discrete Random Variables

A **discrete random variable (DRV)** is a random variable with a countable number of outcomes. Although most discrete random variables have a finite number of outcomes, note that "countable" is not the same as "finite." A discrete random variable can have an infinite number of outcomes. For example, consider $f(n) = (0.5)^n$. Then $f(1) = 0.5$, $f(2) = (0.5)^2 = 0.25$, $f(0.5)^3 = 0.125$,. . . There are an infinite number of outcomes, but they are countable in that you can identify $f(n)$ for any n.

example: the number of votes earned by different candidates in an election.

example: the number of successes in 25 trials of an event whose probability of success on any one trial is known to be 0.3.

Continuous Random Variables

A **continuous random variable (CRV)** is a random variable that assumes values associated with one or more intervals on the number line. The continuous random variable *X* has an infinite number of outcomes.

example: Consider the *uniform* distribution $y = 3$ defined on the interval $1 \le x \le 5$. The area under $y = 3$ and above the x axis for any interval corresponds to a continuous random variable. For example, if $2 \le x \le 3$, then $X = 3$. If $2 \le x \le 4.5$, then $X = (4.5 - 2)(3) = 7.5$. Note that there are an infinite number of possible outcomes for *X*.

Probability Distribution of a Random Variable

A **probability distribution for a random variable** is the possible values of the random variable *X* together with the probabilities corresponding to those values.

A **probability distribution for a discrete random variable** is a list of the possible values of the DRV together with their respective probabilities.

example: Let *X* be the number of boys in a three-child family. Assuming that the probability of a boy on any one birth is 0.5, the probability distribution for *X* is

X	0	1	2	3
P(X)	1/8	3/8	3/8	1/8

The probabilities P_i of a DRV satisfy two conditions:

(1) $0 \le P_i \le 1$ (that is, every probability is between 0 and 1).

(2) $\Sigma P_i = 1$ (that is, the sum of all probabilities is 1).

(Are these conditions satisfied in the above example?)

The **mean** of a discrete random variable, also called the **expected value,** is given by

$$\mu_X = \sum x \cdot P(x).$$

The **variance of a discrete random variable** is given by

$$\sigma_X^2 = \sum (x - \mu_X)^2 \cdot P(x).$$

The **standard deviation of a discrete random variable** is given by

$$\sigma_X = \sqrt{\sum (x - \mu_X)^2 \cdot P(x)}.$$

example: Given that the following is the probability distribution for a DRV, find $P(X = 3)$.

X	2	3	4	5	6
$P(X)$	0.15		0.2	0.2	0.35

solution: Since $\sum P_i = 1$, $P(3) = 1 - (0.15 + 0.2 + 0.2 + 0.35) = 0.1$.

example: For the probability distribution given above, find μ_x and σ_x.

solution:

$$\mu_X = 2(0.15) + 3(0.1) + 4(0.2) + 5(0.2) + 6(0.35) = 4.5.$$

$$\sigma_X = \sqrt{(2-4.5)^2(0.15) + (3-4.5)^2(0.1) + \ldots + (6-4.5)^2(0.35)} = 1.432.$$

Calculator Tip: While it's important to know the formulas given above, in practice it's easier to use your calculator to do the computations. The TI-83/84 can do this easily by putting the *x*-values in, say, `L1`, and the values of *P*(*X*) in, say, `L2`. Then, entering `1-Var Stats L1,L2` and pressing `ENTER` will return the desired mean and standard deviation. Note that the only standard deviation given is σx—the *Sx* is blank. Your calculator, in its infinite wisdom, recognizes that the entries in `L2` are relative frequencies and assumes you are dealing with a probability distribution (if you are taking measurements on a *distribution*, there is no such thing as a *sample* standard deviation).

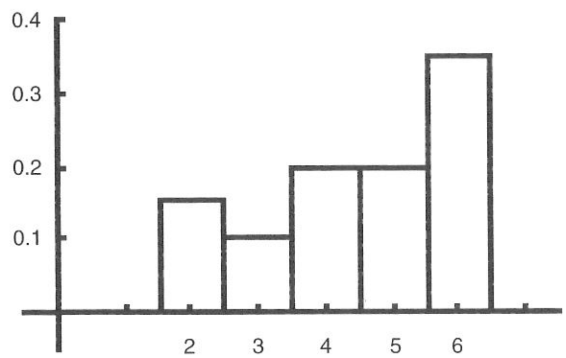
example: Redo the previous example using the TI-83/84, or equivalent, calculator.

solution: Enter the *x* values in a list (say, `L1`) and the probabilities in another list (say, `L2`). Then enter "`1-Var Stats L1,L2`" and press `ENTER`. The calculator will read the probabilities in L2 as relative frequencies and return 4.5 for the mean and 1.432 for the standard deviation.

Probability Histogram

A **probability histogram** of a DRV is a way to picture the probability distribution. The following is a TI-83/84 histogram of the probability distribution we used in a couple of the examples above.

x	2	3	4	5	6
p(x)	0.15	0.1	0.2	0.2	0.35



Probability Distribution for a Continuous Random Variable (CRV). The probability distribution of a continuous random variable has several properties.

- There is a smooth curve, called a **density curve** (defined by a **density function**), that describes the probability distribution of a CRV (sometimes called a probability distribution function). A density curve is always on or above the horizontal axis (that is, it is always non-negative) and has a total area of 1 underneath the curve and above the axis.
- The probability of any individual event is 0. That is, if a is a point on the horizontal axis, $P(X = a) = 0$.
- The probability of a given event is the probability that x will fall in some given interval on the horizontal axis and equals the area under the curve and above the interval. That is, $P(a < X < b)$ equals the area under the graph of the curve and above the horizontal axis between $X = a$ and $X = b$.
- The previous two bulleted items imply that $P(a < X < b) = P(a \leq X \leq b)$.

In this course, there are several CRVs for which we know the *probability density functions* (a probability distribution defined in terms of some density curve). The *normal distribution* (introduced in Chapter 4) is one whose probability density function is the **normal probability distribution.** Remember that the normal curve is "bell-shaped" and is symmetric about the mean μ of the population. The tails of the curve extend to infinity, although there is very little area under the curve when we get more than, say, three standard deviations away from the mean (the empirical rule stated that about 99.7% of the terms in a normal distribution are within three standard deviations of the mean. Thus, only about 0.3% lie beyond three standard deviations of the mean).

Areas between two values on the number line and under the normal probability distribution correspond to probabilities. In Chapter 4, we found the proportion of terms falling within certain intervals. Because the total area under the curve is 1, in this chapter we will consider those proportions to be probabilities.

Remember that we *standardized* the normal distribution by converting the data to z-scores

$$\left(z = \frac{x - \bar{x}}{s_x}\right).$$

We learned in Chapter 4 that a standardized distribution has a mean of 0 and a standard deviation of 1.