

## Lecture 14

Nancy Pfenning Stats 1000

### Chapter 7: Probability

Last time we established some basic definitions and rules of probability:

#### Rule 1:

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

#### Rule 2:

In general, the probability of one event or another occurring is

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

If the events are mutually exclusive, then  $P(A \text{ and } B) = 0$  and so  $P(A \text{ or } B) = P(A) + P(B)$ .

#### Rule 3:

In general, the probability of one event and another is

$$P(A \text{ and } B) = P(A)P(B|A)$$

which we can re-express as

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

If the events are independent,  $P(B|A) = P(B)$  and so  $P(A \text{ and } B) = P(A)P(B)$ .

This time we utilize these rules to solve some more complicated problems, and discuss why at times probabilities can be counter-intuitive.

**Tree diagrams** are often helpful in understanding conditional probability problems.

#### Example

(Game Show) Suppose a prize is hidden behind one of three doors A, B, or C. after the contestant picks one of the three, the host reveals one of the remaining two doors, showing no prize. He gives the guest a chance to switch to the door he or she did not select originally. Is the probability of winning the prize higher with the “keep” or “switch” strategy? Use a probability tree to calculate the probability of winning using each strategy.

When events occur in stages, we normally are interested in the probability of the second, given that the first has occurred. At times, however, we may want to know the probability of an earlier event having occurred, given that a later event ultimately occurred.

#### Example

Suppose that the proportion of people infected with AIDS in a large population is .01. If AIDS is present, a certain medical test is positive with probability .997 (called the **sensitivity** of the test), and negative with probability .003. If AIDS is not present, the test is positive with probability .015, negative with probability .985 (called the **specificity** fo the test). Does a positive test mean a person almost certainly has the disease? Let’s figure out the following: if a person tests positive, what is the probability of having AIDS?

A tree diagram is very helpful for this type of problem. We'll let  $A$  and not  $A$  denote the events of having AIDS or not,  $T$  and not  $T$  denote the events of testing positive or not.

First we'll find the overall probability of testing positive. Either a person has AIDS and tests positive *or* he/she does not have AIDS and tests positive:

$$P(T) = P(A \text{ and } T) + P(\text{not } A \text{ and } T)$$

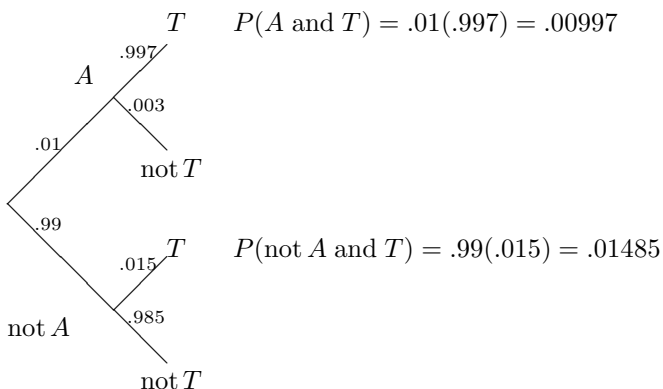
According to the multiplication rule, it follows that

$$\begin{aligned} P(T) &= P(A)P(T|A) + P(\text{not } A)P(T|\text{not } A) \\ &= .01(.997) + .99(.015) = .00997 + .01485 = .02482 \end{aligned}$$

Now, we can apply the definition of conditional probability to find the probability we seek (probability of having AIDS, given that a person tested positive):

$$P(A|T) = \frac{P(A \text{ and } T)}{P(T)} = \frac{.00997}{.02482} = .40$$

Thus, even if a person tests positive, he/she is more likely *not* to have the disease!



Most students—and even most physicians—would have expected the probability to be much higher. Conditional probabilities are often misunderstood, and people often are misled by **confusion of the inverse**: confusing the probability of having the disease, given that you test positive ( $P(A|T) = .40$ ), with the probability of testing positive, given that you have the disease ( $P(T|A) = .997$ ).

### Example

What is the chance that at least two people in a class of 50 share the same birthday? In a survey, the average of 52 responses was 23%. Is this intuitive guess close to the actual probability?

### Example

Here is an easier example to solve before tackling the previous, more difficult one: What is the chance that at least two people in a group of 3 share the same birthday? [Assume all days to be equally likely, and disregard leapyear.] If we call the students A, B, and C, then at least two can share the same birthday in any of these ways: AB or AC or BC or ABC. They are all mutually exclusive, so by the addition rule, the probability of any one *or* the other happening is the sum of the four probabilities. Look first at the probability of A and B having the same birthday, and C different. Whatever A's birthday is, the probability of B having the same birthday is  $\frac{1}{365}$ , and the probability of C's birthday being different is  $\frac{364}{365}$ . So, by the multiplication rule, the probability of B being the same *and* C being different is  $\frac{1}{365} * \frac{364}{365}$ . Similarly, the probabilities of A and C or B and C being the same are each  $\frac{1}{365} * \frac{364}{365}$ . The probability that A and B and C are all the same is  $\frac{1}{365} * \frac{1}{365}$ . Altogether, the probability is  $\frac{1}{365} * \frac{364}{365} + \frac{1}{365} * \frac{364}{365} + \frac{1}{365} * \frac{364}{365} + \frac{1}{365} * \frac{1}{365} = .0082$ .

### Example

Consider this problem: What is the probability of at least two out of 10 people sharing the same birthday? If we call the people A, B, C, D, E, F, G, H, I, J, at least two can share the same birthday in more than 1000 ways! AB AC AD ... AJ BC BD ... ABC...ABCDEFGHJI. Imagine how much more complicated it would be for the original problem, with 50 students instead of 10!

The solutions are much easier if we employ an alternate strategy, taking advantage of probability Rule 1, which tells us the probability of something happening must equal 1 minus the probability of *not* happening! [This is because the probabilities of all possibilities together must sum to 1.] First we'll apply this strategy to re-do the easiest problem, the chance of at least 2 out of 3 sharing a birthday.

The probability of at least 2 out of 3 sharing the same birthday must equal 1 minus the probability of all 3 having different birthdays. The probability of all 3 different is the probability of B different from A [ $\frac{364}{365}$ ] times the probability of C different from both [ $\frac{363}{365}$ ]. Thus, the probability of at least 2 the same is  $1 - \frac{364}{365} * \frac{363}{365} = .0082$  [the same answer we got originally].

Now, we will use this strategy on the probability of at least 2 out of 50 sharing a birthday  
=1 minus probability of all 50 birthdays being different

$$= 1 - \frac{364}{365} * \frac{363}{365} * \frac{362}{365} \cdots * \frac{316}{365} = 1 - \frac{364 * 363 * \cdots * 316}{365^{49}} = .97$$

It is almost certain that at least 2 people in a class of 50 share the same birthday! The fact that students' personal probabilities for this event averaged only about 23% demonstrates that intuition is often an inadequate substitute for systematic application of the laws of probability. [Note: in a class of 80, the probability of at least two birthdays the same is .999915.]

Another way to understand why shared birthdays in a large class are not so unlikely is to realize that if there are many unlikely events possible, it is not so unlikely that at least one of them occurs. This brings us to a discussion of coincidences.

A **coincidence** is a surprising concurrence of events, perceived as meaningfully related, with no apparent causal connection. Should we really be surprised by coincidences?

### Example

On my trip to Denver in 1997, checking into the Sheraton with my husband and three children, I was dismayed to find they only had a single room reserved for us, even though I'd asked for two queens and a cot. After I got a call in our room the next morning from a woman I'd never met, claiming to be a friend of Nancy Pfenning, we gradually pieced together the truth: Nancy Pfenning from Bismarck, North Dakota was supposed to stay at the Sheraton that night, too! In fact, we had usurped her reservation—we were supposed to be at the other Sheraton down the road!

Class members may have had similarly "surprising" experiences. There are so many possible improbable events that may occur in our lives, in the long run some of them are bound to happen. Thus, coincidences, rather than defying the laws of probability, can actually be explained by the laws of probability.

Note: We will not cover computer probability simulation in this course.

**Exercise:** Write up and email me (directly, not as an attachment) a personal coincidence story that happened to you. Were the occurrences really so unlikely?

## Lecture 15

### Chapter 8: Random Variables

A **random variable** (R.V.) is one whose values are quantitative outcomes of a random phenomenon. If it has a finite or countably infinite number of possible values, like the counting numbers 1, 2, 3,..., then it is called **discrete**. Probability distributions of discrete R.V.'s can often be specified in a list.

### Example

Consider the distribution of the R.V.  $X$ , the number of girls in a randomly chosen family with 3 children. This distribution can be found by first examining the sample space of all possible outcomes, with their associated probabilities, then using Rule 2 to specify the probabilities of the events of having 0, 1, 2, or 3 girls:

| Sample Space | Probability   | X |
|--------------|---------------|---|
| BBB          | $\frac{1}{8}$ | 0 |
| BBG          | $\frac{1}{8}$ | 1 |
| BGB          | $\frac{1}{8}$ | 1 |
| GBB          | $\frac{1}{8}$ | 1 |
| BGG          | $\frac{1}{8}$ | 2 |
| GBG          | $\frac{1}{8}$ | 2 |
| GGB          | $\frac{1}{8}$ | 2 |
| GGG          | $\frac{1}{8}$ | 3 |

| Value of $X$ | 0             | 1             | 2             | 3             |
|--------------|---------------|---------------|---------------|---------------|
| Probability  | $\frac{1}{8}$ | $\frac{3}{8}$ | $\frac{3}{8}$ | $\frac{1}{8}$ |

Note that each probability is between 0 and 1, and together they sum to 1.

1. What is the probability that a randomly chosen family of 3 children has 2 girls?  $P(X = 2) = \frac{3}{8}$ .
2. What is the probability of having at least 2 girls?  $P(X \geq 2) = \frac{3}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$ .
3. What is the probability of having more than 2 girls?  $P(X > 2) = \frac{1}{8}$ .

Note that for a discrete R.V. like this, whether or not we have strict inequality makes a difference.

The probability distribution of a discrete R.V. can be displayed in a **probability histogram**, which represents *all* the possible values of a R.V. and their probabilities. Thus, it displays behavior for an entire (possibly abstract) population. The frequency histograms of Chapter 1 displayed behavior of (concrete) sample data values. In a probability histogram, possible values of the R.V.  $X$  are marked along the horizontal axis. As long as the possible values of  $X$  are in increments of 1, the height of each rectangle will be the same as its area, which equals the probability that the R.V.  $X$  takes the value at the rectangle's base.

## Means (Expected Values) and Variances of Random Variables

Sample mean and sample proportion are random variables as long as the sample has been chosen at random. Their distributions are of particular interest, because they will allow us to establish how good an estimate our sample statistics are for the unknown parameters of interest. As we learned in Chapter 2, a distribution may be summarized by telling its center and spread. Now we will focus on the mean as center and variance as spread of a random variable. The mean is simply the average of all the possible values of  $X$ , where more probable values are given more weight. We sometimes call it the **expected value** of  $X$ , written  $E(X)$ .

If  $X$  is a discrete random variable with possible values  $x_1, x_2, x_3, \dots$  occurring with probabilities  $p_1, p_2, p_3, \dots$ , then the **mean** of  $X$ , or equivalently the **expected value**, is

$$\mu = E(X) = x_1p_1 + x_2p_2 + x_3p_3 + \dots = \sum x_i p_i$$

### Example

What is the mean (expected) number of girls in a family of three children?

$$\mu = E(X) = 0\left(\frac{1}{8}\right) + 1\left(\frac{3}{8}\right) + 2\left(\frac{3}{8}\right) + 3\left(\frac{1}{8}\right) = \frac{12}{8} = 1.5$$

The average number of girls for a family with 3 children is 1.5.

### Example

Use the probability distribution below to find the mean  $\mu$  of all dice rolls:

|              |               |               |               |               |               |               |
|--------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Value of $X$ | 1             | 2             | 3             | 4             | 5             | 6             |
| Probability  | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

$$\mu = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{21}{6} = 3.5.$$

### Example

Use the probability distribution below to find the mean  $\mu$  of all household sizes  $X$  in the U.S.:

|              |     |     |     |     |     |     |     |
|--------------|-----|-----|-----|-----|-----|-----|-----|
| Value of $X$ | 1   | 2   | 3   | 4   | 5   | 6   | 7   |
| Probability  | .25 | .32 | .17 | .15 | .07 | .03 | .01 |

$$\mu = 1(.25) + 2(.32) + 3(.17) + 4(.15) + 5(.07) + 6(.03) + 7(.01) = 2.6.$$

Notice that the mean equalled the median (midpoint) in the first two examples because the distributions were perfectly symmetric. Median household size is 2 because 50% of households had 2 people or fewer; mean household size is higher than median because this distribution is right-skewed.

To describe spread of a random variable  $X$ , we focus first on the **variance**  $\sigma^2 = V(X)$ , then take its square root to find the standard deviation:

$$\sigma^2 = V(X) = (x_1 - \mu)^2 p_1 + (x_2 - \mu)^2 p_2 + (x_3 - \mu)^2 p_3 + \cdots = \Sigma(x_i - \mu)^2 p_i$$

The **standard deviation**  $\sigma$  of  $X$  is the square root of the variance.

### Example

Find standard deviation of  $X$  = number of girls in a family of 3 children. [Recall that  $\mu = 1.5$ .]

$$Var(X) = (0 - 1.5)^2\left(\frac{1}{8}\right) + (1 - 1.5)^2\left(\frac{3}{8}\right) + (2 - 1.5)^2\left(\frac{3}{8}\right) + (3 - 1.5)^2\left(\frac{1}{8}\right) = .75$$

$$\sigma = \sqrt{Var(X)} = \sqrt{.75} = .87$$

You should know how to calculate  $\mu$  for a discrete random variable, and you will be required to calculate  $\sigma$  in a homework exercise, but not on quizzes or exams.

## Binomial Distributions

Our primary goal in this course involves the use of statistics *sample mean*  $\bar{x}$  or *sample proportion*  $\hat{p}$  to estimate parameters *population mean*  $\mu$  or *population proportion*  $p$ . If we take a simple random sample of size  $n$  from a population and observe for each individual the value of some quantitative variable (like height, number of girls, or household size), then we can calculate its sample mean value  $\bar{x}$ , and use it to estimate the unknown population mean  $\mu$ . On the other hand, if we observe the value of some *categorical* variable (like gender) to see whether or not each individual has a particular characteristic, we can calculate *sample count*  $X$ , then *sample proportion*  $\hat{p} = \frac{X}{n}$  of units falling into that category, and use  $\hat{p}$  to estimate the unknown  $p$ . In this section we will focus on sample count for categorical data.

### Example

A recent survey of 1012 Americans found that 516 of them opposed gay marriage. We can set up a R.V.  $X$  for the *sample count* opposing, and say  $X$  takes the value 516 for this sample. Or, we can set up a R.V.  $\hat{p} = \frac{X}{1012}$  for the *sample proportion* opposing, and say  $\hat{p}$  takes the value  $\frac{516}{1012} = .51$ .

Just as many quantitative variables fall into a particular pattern known as the normal distribution, the counts for many categorical variables fall into a particular pattern called *binomial*. In this section, we study the distribution of binomial counts  $X$ ; in future chapters we will shift our attention to the distribution of proportions  $\hat{p}$ . The two are directly related because  $\hat{p} = \frac{X}{n}$ .

The distribution of the count  $X$  of “successes” in the “binomial setting” is called the **binomial distribution** with parameters  $n$  and  $p$ . The “binomial setting” has the following requirements:

1. There is a fixed number  $n$  of observations.
2. Each of the  $n$  observations is independent of the others.
3. There are two possible categories, “success” and “failure”, for each observation.
4. The probability  $p$  of success is the same for each observation.

### Example

The following R.V.’s do *not* have a binomial distribution:

1. Pick a card from a deck of 52, replace it, pick another, etc. Let  $X$  be the number of tries until you get an ace. [ $n$  not fixed]
2. Choose 16 cards without replacement from a deck of 52. Let  $X$  be the number of red cards chosen. [observations not independent]
3. Pick a card from a deck of 52, replace it, pick another, etc. Do this 16 times. We are interested in the number of cards in each suit (hearts, diamonds, clubs, spades) picked. [more than 2 possible categories for each observation]
4. Pick a card from a deck of 52, replace it, pick a card from a deck of 32, replace it, back to 52, etc. After 16 tries, let  $X$  be the number of aces picked. [different probabilities for success  $\frac{1}{13}$  or  $\frac{1}{8}$ ]

### Example

The following R.V. does have a binomial distribution:

Pick a card from a deck of 52, replace it, pick another. Do this 16 times. Let  $X$  be the R.V. for the number of red cards picked. Then  $X$  is binomial with  $n = 16$ ,  $p = \frac{1}{2}$ .

Requirement 2 may be fudged slightly: If only a very small degree of dependence is present, we may still treat a R.V. as binomial. In general, sampling with replacement is associated with independence, and sampling without replacement is associated with dependence. Almost all of our examples in this course will assume data arises from a simple random sample, that is, sampling *without* replacement, in which selections are dependent. Under what circumstances can violation of Requirement 2 be overlooked?

### Example

Pick 2 people at random without replacement from a class where 25 out of 75 are male. Let  $X$  be the number of males picked. The probability of “success” for the first person picked is exactly  $\frac{25}{75} = \frac{1}{3} = .333$ . The probability of “success” for the second person is  $\frac{24}{74} = .324$  if the first was male,  $\frac{25}{74} = .337$  if the first was female— pretty close! Since population size (75) is much larger than sample size (2),  $X$  is approximately binomial with  $n = 2$ ,  $p = \frac{1}{3}$ .

However, if 2 people were picked from a group of only 3 where one third are male, the probability of the second being male is either 0 or  $\frac{1}{2}$ — very different! Likewise, there would be a higher degree of dependence if we picked 25 from 75 where one third are male.

**Rule of Thumb:** If the population is at least 10 times the sample size, replacement has little effect. In such cases, when taking a simple random sample of size  $n$  from a population with proportion  $p$  in a certain category, the sample count  $X$  in the category of interest is approximately binomial with the same  $n$  and  $p$ .

Most problems for binomial R.V.'s take the form of a question like this: "If  $X$  is binomial with a certain  $n$  and  $p$  given, what is the probability that  $X$  equals some value, or lies within some range of values?" To answer such questions, we can

1. Use the formula  $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$  for  $k = 0, \dots, n$  [not done in this course]; or
2. use binomial tables [not done in this course]; or
3. use a calculator or MINITAB [not done in this course]; or
4. use a normal approximation to find  $P(X \leq \text{_____}.5in)$  [done later in this chapter] or  $P(\hat{p} \leq \text{_____}.5in)$ , where  $\hat{p} = \frac{X}{n}$  [done in Chapter 9].

Since our normal approximation will be based on a variable having the same mean and standard deviation as the binomial variable of interest  $X$ , we first discuss the binomial mean and standard deviation.

Earlier we found the mean value of  $X$ , the binomial count of girls in a family with three children, by using our formula for means of discrete random variables:

$$\mu = E(X) = 0\left(\frac{1}{8}\right) + 1\left(\frac{3}{8}\right) + 2\left(\frac{3}{8}\right) + 3\left(\frac{1}{8}\right) = \frac{12}{8} = 1.5$$

In fact, there is a much simpler way to find a binomial mean. If sample count  $X$  of successes is a binomial R.V. for  $n$  observations with probability  $p$  of success on each observation, then  $X$  has mean

$$\mu = np$$

### Example

The number of girls  $X$  in a family of 3 children is binomial with  $n = 3$  and  $p = .5$ . The mean number of girls is  $\mu = np = 3(.5) = 1.5$ .

The mean of a binomial R.V. is easy to grasp intuitively: Say the probability of success for each observation is .2 and we make 10 observations. Then on the average we should have  $10 * .2 = 2$  successes. The spread of a binomial distribution is not so intuitive, so we will not justify our formula for standard deviation:

$$\sigma = \sqrt{np(1-p)}$$

### Example

Standard deviation for number of girls  $X$  in a family of 3 children is  $\sigma = \sqrt{3(.5)(1-.5)} = \sqrt{.75} = .87$ , a much easier solution method than the one we used previously.

### Example

Pick a card from a deck of 52, replace it, pick another. Do this 16 times. Let  $X$  be the R.V. for the number of red cards picked. Then  $X$  is binomial with  $n = 16$ ,  $p = .5$ . Find the mean and standard deviation of  $X$ .  $\mu = np = 16(.5) = 8$ ;  $\sigma = \sqrt{np(1-p)} = \sqrt{16(.5)(1-.5)} = 2$ . In this situation, we'd expect on average to get 8 red cards, give or take about 2.

### Example

The population of whole numbers from 1 to 20 have a mean of 10.5 and a standard deviation of 5.77. I am interested in the mean and standard deviation of all the numbers chosen “randomly” from 1 to 20 by students. I suspect the mean may be higher than 10.5, since people tend to think larger numbers are more “random” than smaller numbers. I suspect the standard deviation may be less than 5.77, since people may avoid extreme numbers like 1 and 20. First I record what proportion of students chose each number:

$$\begin{aligned} \#1 : \frac{7}{446} = 1.57\% \quad \#2 : \frac{13}{446} = 2.91\% \quad \#3 : \frac{20}{446} = 4.48\% \quad \#4 : \frac{16}{446} = 3.59\% \quad \#5 : \frac{20}{446} = 4.48\% \\ \#6 : \frac{20}{446} = 4.48\% \quad \#7 : \frac{34}{446} = 7.62\% \quad \#8 : \frac{17}{446} = 3.81\% \quad \#9 : \frac{11}{446} = 2.47\% \quad \#10 : \frac{10}{446} = 2.24\% \\ \#11 : \frac{21}{446} = 4.71\% \quad \#12 : \frac{26}{446} = 5.83\% \quad \#13 : \frac{51}{446} = 11.43\% \quad \#14 : \frac{23}{446} = 5.16\% \quad \#15 : \frac{22}{446} = 4.93\% \\ \#16 : \frac{18}{446} = 4.04\% \quad \#17 : \frac{66}{446} = 14.80\% \quad \#18 : \frac{21}{446} = 4.71\% \quad \#19 : \frac{16}{446} = 3.59\% \quad \#20 : \frac{14}{446} = 3.14\% \end{aligned}$$

Next I calculate mean and standard deviation as follows:

$$\mu = E(X) = 1(.0157) + 2(.0291) + \dots + 20(.0314) = 11.614$$

$$\text{Var}(X) = (1 - 11.614)^2(.0157) + (2 - 11.614)^2(.0291) + \dots + (20 - 11.614)^2(.0314) = 27.85$$

$$\sigma = \sqrt{\text{Var}(X)} = \sqrt{27.85} = 5.28$$

As I suspected, the mean of 11.614 is higher than what it would be if students truly picked at random, and the standard deviation is lower. We can say students’ selections averaged 11.614, and they typically deviated from this average by about 5.28.

**Exercise:** Use the survey data to report the probability distribution of year for the undergraduates in the class (years 1, 2, 3, and 4). [You will need to tally the years and adjust the total to exclude “other” students.] Find the mean, variance, and standard deviation. Use mean and standard deviation in a sentence about the distribution of year in order to tell what is typical for students in the class.

## Lecture 16

Last time we talked about a distribution that often applies when we are interested in a single categorical variable: the binomial count  $X$  for number of successes in a situation that allows for two possible categories, “success” and “failure”. Next we consider a distribution that arises when we are interested in a single quantitative variable whose possible values constitute a continuum.

Unlike discrete R.V.’s, **continuous** R.V.’s can take all values in an interval of real numbers. The probability distribution of a continuous R.V.  $X$  is represented by a **density curve**. The probability of an event is the area under the curve over the values of  $X$  that make up the event:  $P(a \leq X \leq b)$  equals the area under the curve from  $a$  to  $b$ . The total area under the curve is 1, so the probability of any event must be between 0 and 1. The **mean**  $\mu$  of a continuous R.V. tells its center or “balance point”; the **standard deviation**  $\sigma$  tells its spread.

The simplest continuous random variable is the **uniform** R.V., which takes any value over an interval with equal probability.



### Example

A student's friend can call her cell phone any time between 9 a.m. and 5 p.m. (an 8 hour interval) with equal probability. What is the probability that she calls during statistics class, between 10:00 and 11:00?  $\frac{1}{8}$

**Normal probability distributions** are for a particular type of continuous R.V. that we have already encountered in Chapter 2.

### Example

The distribution of verbal SAT scores in a certain population is normal with mean  $\mu = 500$ , standard deviation  $\sigma = 100$ . The Empirical Rule told us, for example, that "95% of the scores fell within 2 standard deviations of the mean, that is, between 300 and 700". Now, we will imagine choosing a score at random and consider its value to be a random variable  $X$ . Using probability notation, we now write " $P(300 < X < 700) = .95$ ".

Recall: The relative size of a normal value  $x$  is expressed by standardizing to find the z-score:

$$z = \frac{\text{observed value} - \text{mean}}{\text{standard deviation}} = \frac{x - \mu}{\sigma}$$

### Example

The z-score for a verbal SAT of 300 is  $\frac{300-500}{100} = -2$ . The z-score for a verbal SAT of 700 is  $\frac{700-500}{100} = +2$ .

By construction, the standard normal ( $z$ ) distribution has mean  $\mu = 0$  and standard deviation  $\sigma = 1$ . The normal table (A.1 in the appendix) gives us the *area* under the standard normal curve to the *left* of any value  $z^*$ , which is also the *proportion* of standard normal observations which are *less* than the value  $z^*$ . These are given for any  $z^*$  between -3.49 and +3.49, with a few extreme cases beyond these values to specify exactly how unlikely such values are. But we often simply state  $P(Z < z^*)$  is approximately zero for any  $z^*$  less than -3.5, which means  $P(Z > z^*)$  is approximately one for  $z^*$  less than -3.5. Similarly,  $P(Z < z^*)$  is approximately one for any  $z^*$  greater than +3.5, which means  $P(Z > z^*)$  is approximately zero for  $z^*$  greater than +3.5.

### Example

Find the following standard normal probabilities:

1.  $P(Z < -1.92) = .0274$ .
2.  $P(Z < 0.68) = .7517$ .
3.  $P(Z > -2.5) = P(Z < +2.5)$  (by the symmetry of the standard normal curve about zero) = .9938. Alternatively, because the total area under the curve is 1, we can write  $P(Z > -2.5) = 1 - P(Z < -2.5) = 1 - .0062 = .9938$ .
4.  $P(Z > 1.22) = P(Z < -1.22) = .1112$  or  $P(Z > 1.22) = 1 - P(Z < 1.22) = 1 - .8888 = .1112$ .
5.  $P(-1.00 < Z < +1.00) = P(Z < +1.00) - P(Z < -1.00) = .8413 - .1587 = .6826$  since the area between -1 and +1 is the area left of +1 minus the area left of -1. [Recall: we said 68% of values are within  $\sigma$  of  $\mu$ . For a standard normal  $z$ , this means 68% are within 1 of 0, or from -1 to +1. The tables have given us two more decimal places of accuracy for our Empirical Rule.]

Note that since Table A.1 only shows us the probability of values  $z$  being *less* than a given value  $z^*$ , we must either use symmetry or the fact that total area is one to find proportion of values *greater* than a given value. The above examples give a certain value  $z^*$  and ask for a probability. The next examples will give us a probability or **percentile** and ask for the corresponding value  $z^*$ . Keep in mind that standard normal values  $z$  are of the form \_\_\_\_\_1em.\_\_\_\_\_1em \_\_\_\_\_1em and are shown along the *margins*. The row is for ones and tenths, the column fine-tunes for the hundredths place. *Probabilities* below given  $z$  values are of the form .\_\_\_\_\_1em \_\_\_\_\_1em \_\_\_\_\_1em \_\_\_\_\_1em and are shown *inside* the table.

### Example

1. What is the z-score for the 33rd percentile? In other words, the probability is .33 that a standard normal variable  $Z$  falls below what value? According to Table A.1,  $.3300 = P(Z < -1.44)$  so -1.44 is the z-score for the 33rd percentile.
2. What is the 90th percentile of  $z$ ?  $.8997 = P(Z < +1.28)$  so 1.28 is the 90th percentile of  $Z$ .
3. What is the 10th percentile of  $z$ ?  $.1003 = P(Z < -1.28)$  so -1.28 is the 10th percentile of  $Z$ .
4. 99% is the probability of  $Z$  being *greater* than what value? The same value for which 1% of the time,  $Z$  is *less* than that value:  $.0100 \approx .0099 = P(Z < -2.33)$  so  $.99 = P(Z > -2.33)$  and -2.33 has 99% of z-scores above it. Alternatively, we can say  $.99 = P(Z < +2.33)$  means  $.99 = P(Z > -2.33)$ .
5. .48 is the probability of  $Z$  being greater than what value?  $.48 = P(Z < -0.05)$  means  $.48 = P(Z > +0.05)$ .
6. .95 is the probability of  $Z$  being how far from zero?  $\frac{1-.95}{2} = .0250 = P(Z < -1.96) = P(Z > +1.96)$  so .95 is the probability of being within 1.96 of 0 (i.e., approximately within  $2\sigma$  of  $\mu$  where  $\sigma$  is 1 and  $\mu$  is 0. The Empirical Rule is only roughly accurate.)

Note: by a “percentile”, we mean a value on a scale of 100 that indicates the percent of a distribution that is equal to or below it. Keep this definition in mind when considering the following example.

### Example

Researchers recently reported that more than 20 percent of black and Hispanic children, and more than 12 percent of white children, were at or above the 95th percentile for body mass (and thus classified as overweight). Is this possible? Strictly speaking, no; the inconsistency results because doctors referred to the 95th percentile on charts made between 1960 and 1980. There are more overweight children today than there were at that time.

## Using Table A.1 for Non-Standard Normal Values

### Example

The distribution of heights of young women in the U.S. is normal with mean 65, standard deviation 2.7. The distribution of heights of young men in the U.S. is normal with mean 69, standard deviation 3. What percentiles are Jane and Joe in, if Jane is 71 inches and Joe is 75 inches tall?

For women,  $P(X < 71) = P\left(\frac{X-65}{2.7} < \frac{71-65}{2.7}\right) = P(Z < 2.22) = .9868$ . For men,  $P(X < 75) = P\left(\frac{X-69}{3} < \frac{75-69}{3}\right) = P(Z < 2) = .9772$ . Jane is in the 98.68th percentile and Joe is in the 97.72nd percentile.