other, may be regarded simply as an abbreviation for convenience. Thus, for example, to denote "the probability of one girl" we can write:

$$\Pr(X = 1) \text{ or just } p(1)$$

Finally, we emphasize the difference between *discrete* and *continuous* variables. A random variable is called *discrete* if it has just a finite (or "countably infinite") set of values. For example, when X = the number of heads in 3 tosses of a coin, then its values are 0, 1, 2, 3—a finite set. Or if X = the number of tosses required to get the first head, then its values are 1, 2, 3, 4, 5, . . .—a "countably infinite" set.

By contrast, a *continuous* random variable takes on a continuum of values. For example, if X = the number of gallons that pass hourly through a meter with a capacity of 50 gallons per hour, then its value could be *any* number between 0 and 50—for example, 17.2 or 39.826. The mathematics required for continuous random variables is more advanced, requiring integration (the analog of summation in calculus).

PROBLEMS

4-1 In planning a family of 3 children, assume boys and girls are equally likely on each birth. Find the probability distribution of:

- **a.** X = the number of girls
- **b.** Y = the number of runs (where a "run" is a run or string of children of the same sex. For example Y = 2 for the outcome BGG).

4-2 Simulate the experiment in Problem 4-1 by using the random numbers in Appendix Table I (even number = heads, odd = tails). Repeat this experiment 50 times, tabulating the frequency of X. Then calculate:

- **a.** The relative frequency distribution.
- **b.** The mean $\overline{X}$.
- **c.** The variance s^2.

4-3 If the experiment in Problem 4-2 were repeated millions of times (rather than 50 times), to what value would the calculated quantities tend?

4-4 A salesperson for a large pharmaceutical company makes 3 calls per year on a drugstore, with the chance of a sale each time being 80%. Let X denote the total number of sales in a year (0, 1, 2, or 3).

- **a.** Tabulate and graph the probability distribution p(x).
- **b.** What in the chance of at least two sales?

4-5 Repeat Problem 4-4 under the different assumption that "nothing succeeds like success." Specifically, while the chance of a sale on the first call is still 80%, the chance of a sale on later calls depends on what happened on the previous call, being 90% if the previous call was a sale, or 40% if the previous call was no sale.

4-2 MEAN AND VARIANCE

In Chapter 2 we calculated the mean $\overline{X}$ and variance s^2 of a sample of observations from its relative frequency distribution (f/n). In the same way, it is natural to calculate the mean and variance of a random variable from its probability distribution $p(x)$:

$$\boxed{\begin{aligned}\text{Mean } \mu &\equiv \Sigma x p(x) \\ \text{Variance } \sigma^2 &\equiv \Sigma(x-\mu)^2 p(x)\end{aligned}}$$

(4-4) like (2-20)

(4-5) like (2-21) and (2-23)

Here we are following the usual custom of reserving Greek letters for theoretical values (μ is the Greek letter mu, equivalent to m for mean, and σ is the Greek letter sigma, equivalent to s for standard deviation). Of course, probabilities can be viewed as just the long-run relative frequencies from the population of all possible repetitions of the experiment. Thus we often call μ and σ^2 the *population* moments, to distinguish them from the *sample* moments $\overline{X}$ and s^2 that are calculated for a mere sample. Table 4-1 makes this clear.

TABLE 4-1 A Comparison of Sample and Population Moments

Sample Moments *Use Relative Frequencies f/n*	*Population Moments (in Greek)* *Use Probabilities p(x)* *(long-run f/n)*
Sample Mean	Population Mean
$\overline{X} = \Sigma x\left(\frac{f}{n}\right)$	$\mu \equiv \Sigma x p(x)$
Sample Variance	Population Variance
$s^2 = \text{MSD} = \Sigma(x-\overline{X})^2\left(\frac{f}{n}\right)$	$\sigma^2 \equiv \Sigma(x-\mu)^2 p(x)$

ing point—a weighted average using probability weights rather than relative frequency weights. And the standard deviation σ is the typical deviation.

We emphasize that the distinction between sample and population moments must not be forgotten: μ is called the population mean since it is based on the population of all possible repetitions of the experiment; on the other hand, we call $\overline{X}$ the sample mean since it is based on a mere sample drawn from the parent population.

PROBLEMS

4-6 Compute μ and σ for each of the following distributions (from Problems 4-4 and 4-5). Graph each distribution, showing the mean as the balancing point, and the standard deviation as a typical deviation (like Figure 2-6).

a.

x	$p(x)$
0	.01
1	.10
2	.38
3	.51

b.

x	$p(x)$
0	.07
1	.10
2	.18
3	.65

4-7 On the basis of past experience, the buyer for a large sports store estimates that the number of 10-speed bicycles sold next year will be somewhere between 40 and 90—with the following distribution:

Number of Bicycles Sold x	*Probability* $p(x)$
40	.05
50	.15
60	.41
70	.34
80	.04
90	.01

a. What is the mean number sold? What is the standard deviation?

b. If 60 are ordered, what is the chance they will all be sold? What is the chance some will be left over (undesired inventory)?

c. To be almost sure (95%) of having enough bicycles, how many should be ordered?

4-8 In planning a huge outdoor concert for June 16, the producer estimates the attendance will depend on the weather according to the following table. He also finds out from the local weather office what the weather has been like, for June days in the past 10 years.

Weather	*Attendance*	*Relative Frequency*
wet, cold	5,000	.20
wet, warm	20,000	.20
dry, cold	30,000	.10
dry, warm	50,000	.50

a. What is the expected (mean) attendance?

b. The tickets will sell for \$9 each. The costs will be \$2 per person for cleaning and crowd control, plus \$150,000 for the band, plus \$60,000 for administration (including the facilities). Would you advise the producer to go ahead with the concert, or not? Why?

4-9 In Problem 4-8, suppose the producer has gone ahead with his plans, and on June 10 has obtained some rather gloomy long-run weather forecasts: The 4 weather conditions now have probabilities .30, .20, .20, and .30, respectively.

If he cancels the concert, he will still have to pay half of the \$60,000 administration cost, plus a \$15,000 cancellation penalty to the band.

Would you advise him to cancel or not?

4-3 THE BINOMIAL DISTRIBUTION

A—THE BINOMIAL ASSUMPTIONS AND FORMULA

There are many types of discrete random variables, and the commonest is called the *binomial*. The classical example of a binomial variable is:

$$S = \text{number of heads in several tosses of a coin.}$$

There are many random variables of this binomial type, a few of which are listed in Table 4-3. (We have already encountered not only the coin-tossing example, but also another: the number of girls in a family of three children.) To handle all such cases, it will be helpful to state the basic assumptions in general notation:

1. We suppose there are n *trials* (tosses of the coin).
2. In each trial, a certain event of interest can occur, or fail to occur; then we say a *success* (head) or *failure* (tail) has occurred. Their respective

easily. In Table IIIc we simply look up $n = 5$, $\pi = .60$, and $s_0 = 3$. This immediately gives the answer .683, which confirms (4-12).

PROBLEMS

4-10 In families with 6 children, let X = the number of boys. For simplicity, assume that births are independent and boys and girls are equally likely.

a. Graph the probability distribution of X.

b. Calculate the mean and standard deviation, and show them on the graph.

c. Of all families with 6 children, what proportion have:

i. Exactly an even split between the sexes (3-3)?

ii. Nearly an even split (3-3, or 4-2)?

iii. 3 or more boys?

4-11 In a desperate gamble, Wildcat Oil Exploration has committed all its remaining funds to finance a sequence of 12 drillings. Each drilling in this region has a 20% chance of successfully producing oil, independent of the other drillings.

To avoid bankruptcy, three or more drillings must produce oil. What is the chance of this?

4-12 **a.** One hundred coins are spilled at random on the table, and the total number of heads S is counted. The distribution of S is binomial, with n = ____ and π = ____. Although this distribution would be tedious to tabulate, the *average* (mean) of S is easily guessed to be ____.

b. Repeat part **a** for S = the number of aces when 30 dice are spilled at random.

c. Repeat part **a** for S = the number of correct answers when 20 true–false questions are answered by pure guessing.

d. Guess what the mean is for *any* binomial variable, in terms of n and π.

4-13 In Problem 4-12, you guessed the mean of a binomial variable. There is also a formula for the standard deviation (both formulas will be proved later, in Problem 6-23):

$$\boxed{\begin{array}{l}\textbf{Binomial mean: } \mu = n\pi \\ \textbf{Standard deviation: } \sigma = \sqrt{n\pi(1-\pi)}\end{array}} \tag{4-13}$$

Now that you understand these formulas, you can use them whenever it is convenient. For example,

a. Verify μ and σ found earlier in Problem 4-10.

b. Calculate μ and σ for Problem 4-11.

4-14 A multiple choice exam consists of 12 questions, each having 5 possible answers. To pass, you must answer at least 8 out of 12 questions correctly. What is the chance of this, if:

a. You go into the exam without knowing a thing, and have to resort to pure guessing?

b. You have studied enough so that on each question, 3 choices can be eliminated. But then you have to make a pure guess between the remaining 2 choices.

c. You have studied enough so that you know *for sure* the correct answer on 2 questions. For the remaining 10 questions you have to resort to pure guessing.

4-15 **a.** Suppose a warship takes 10 shots at a target, and it takes at least 4 hits to sink it. If the warship has a record of hitting with 20% of its shots in the long run, what is the chance of sinking the target?

b. What crucial assumptions did you make in part **a**? Why might they be questionable?

c. To appreciate these crucial assumptions, put yourself in the position of the captain of the British battleship *Prince of Wales* in World War II. The gunners on the German *Bismark* have just homed in on the British *Hood*, and sunk it after several shots. They now turn their fire on you, and after an initial miss they make a direct hit. Do you leave the probability they will hit you on the next shot unchanged, or do you revise it? (In that actual situation, the captain of the *Prince of Wales* broke off the action, and the *Bismark* was sunk by the British a few days later.)

***4-16** **a.** The governing board of a small corporation consists of the chairman and six other members, with a majority vote among these seven deciding any given issue. Suppose that the chairman wants to pass a certain motion, but is not sure of its support. Suppose the other six members vote independently, each with probability 40% of voting for the motion. What is the chance that it will pass?

b. If the chairman had two firm allies who were certain to vote for the motion, how would that improve the chances of its passing? [Assume that the other four members are the same as in part **a**.]

sum of all the *areas* of the bars) equal to 1. We accomplish this by defining:

$$\boxed{\textbf{Relative frequency density} \equiv \frac{\textbf{relative frequency}}{\textbf{cell width}}} \tag{4-14}$$

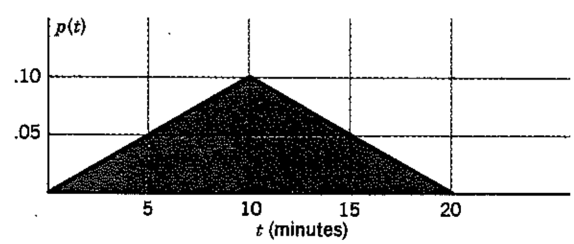
$$= \frac{\text{relative frequency}}{1/4}$$
$$= 4 \text{ (relative frequency)} \tag{4-15}$$

Thus in Figure 4-3, panel (*b*) is 4 times as high as panel (*a*) and now has an area equal to 1.

In Figure 4-4 we show what happens to the relative frequency density of a continuous random variable as sample size increases. With a small sample, chance fluctuations influence the picture. But as sample size increases, chance is averaged out, and relative frequencies settle down to probabilities. At the same time, the increase in sample size allows a finer definition of cells. While the area remains fixed at 1, the relative frequency density becomes approximately a curve, the *probability density function*, $p(x)$, which we informally call the *density*, or the *probability distribution*. (To calculate probabilities and moments of continuous distributions, see Appendix 4-4.)

PROBLEMS

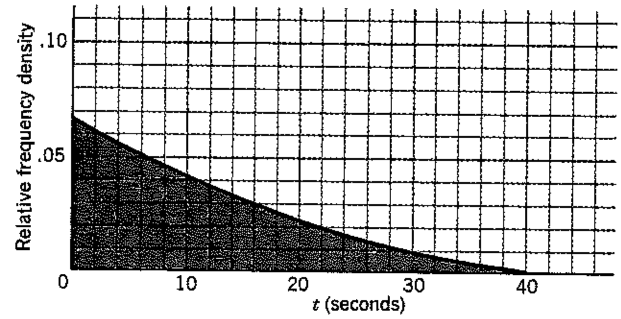
4-17 The total time T that I wait for buses, on a long trip that includes a transfer, has the following probability distribution: Note that the area of a triangle = base × height/2, so that the total area or probability is $20 \times .10/2 = 1.00$.



a. If I wait more than 15 minutes, I will be late for my appointment. What is the chance of this?

b. What is the mean waiting time?

4-18 At a busy switchboard, the waiting time T between one incoming call and the next was recorded thousands of times, as follows:



a. Estimate the probability that T exceeds 30 seconds.

b. Find the modal time and the median time. In relation to these two, where would the mean be?

4-5 THE NORMAL DISTRIBUTION

For many random variables, the probability distribution is a specific bell-shaped curve, called the *normal* curve, or *Gaussian* curve (in honor of the great German scientist Karl Friedrick Gauss, 1777–1855). This is the most common and useful distribution in statistics. For example, errors made in measuring physical and economic phenomena often are normally distributed. In addition, many other probability distributions (such as the binomial) often can be approximated by the normal curve.

A—STANDARD NORMAL DISTRIBUTION

The simplest of the normal distributions is the *standard normal* distribution shown in Figure 4-5, and discussed in detail in Section C. Called the Z distribution, it is distributed around a mean $\mu = 0$ with a standard deviation $\sigma = 1$. Thus, for example, the value $Z = 1.5$ is one-and-a-half standard deviations above the mean, and in general:

> **Each Z value is *the number of standard deviations* away from the mean.** (4-16)

We often want to calculate the probability (i.e., the area under the curve) beyond a given value of Z, like the value $Z = 1.5$ in Figure 4-5.

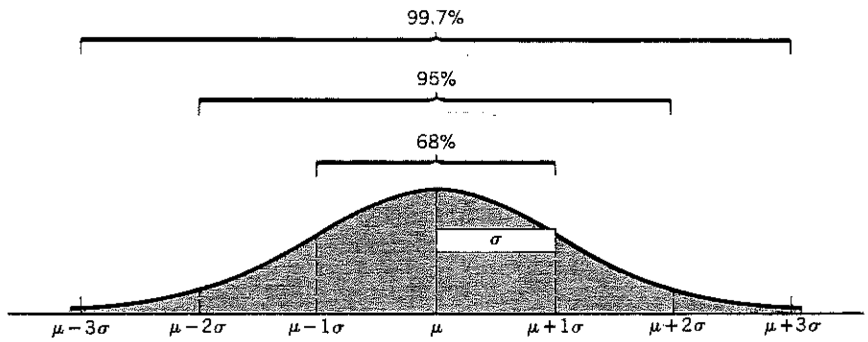


FIGURE 4-10
To graph a normal distribution, it helps to know that nearly all (99.7%) of its probability is within three standard deviations of the mean.

In the exponent we find the familiar $z = (x - \mu)/\sigma$ as in (4-23). The shape of this distribution is therefore the same bell shape as the standard normal—being centered at μ, and spread out by the factor σ.

To graph the normal distribution, it helps to know that nearly all of the area or probability lies within 3 standard deviations of the mean. (By looking up $z = 3$ in the standard normal table, we find the probability is 99.7%.) This is shown, along with several other areas, in Figure 4-10.

PROBLEMS

Recall that Table IV is repeated inside the front cover where it is easy to find.

4-19 If Z is a standard normal variable, calculate:

a. $\Pr(Z > 1.60)$

b. $\Pr(1.60 < Z < 2.30)$

c. $\Pr(Z < 1.64)$

d. $\Pr(-1.64 < Z < -1.02)$

e. $\Pr(0 < Z < 1.96)$

f. $\Pr(-1.96 < Z < 1.96)$

g. $\Pr(-1.50 < Z < .67)$

h. $\Pr(Z < -2.50)$

4-20 If X is normally distributed around a mean of 16 with a standard deviation of 5, find

a. $\Pr(X > 20)$

b. $\Pr(20 < X < 25)$

c. $\Pr(X < 10)$

d. $\Pr(12 < X < 24)$

4-21 Phil and Kim Bell don't know whether to buy a house now or wait a year, in which case a price increase may put a house beyond their reach. Their best guess is that, if they wait a year, the price increase will be approximately normal, with a mean of 8% and, reflecting the uncertainty of the market, a standard deviation of 10%.

a. If the price increase exceeds 25% they feel they will be unable to afford a house. What is the chance of this?

b. On the other hand, if the price drops, they will have won their gamble handsomely. What is the chance of this?

4-22 The time required to complete a college achievement test was found to be normally distributed, with a mean of 110 minutes and a standard deviation of 20 minutes.

a. What proportion of the students will finish in 2 hours (120 minutes)?

b. When should the test be terminated to allow just enough time for 90% of the students to complete the test?

4-23 Wearever tires have a track record of lasting 56,000 miles on average, with a standard deviation of 8,000 miles, and a normal distribution.

a. What is the chance that a given tire will last 50,000 miles?

b. What is the chance that all four Wearever tires on my car will last 50,000 miles?

c. State the assumptions made in **b**, and how more realistic assumptions would change your answer.

4-6 A FUNCTION OF A RANDOM VARIABLE

A—FUNCTIONS IN GENERAL

Consider again the planning of a family of three children. Suppose the annual cost of clothing (R) is a function of the number of girls (X) in the family, that is:

$$R = g(X) \tag{4-27}$$

Specifically, suppose $R = -100X^2 + 300X + 500$. Then, for each X, the

TABLE 4-4 Annual Clothing Cost *R* as a Function of *X*.

Number of Girls x	*Probability* $p(x)$	*Clothing Cost* $r = g(x)$
0	.14	$500
1	.39	$700
2	.36	$700
3	.11	$500

Why is clothing cost R lower if $X = 0$ or 3, that is, if all children are the same sex? The answer is that they will be better able to re-use the same clothes.

That is, noting (4-5):

$$\boxed{E(X-\mu)^2 = \sigma^2} \tag{4-34}$$

This emphasizes that σ^2 *may be regarded as just a kind of expected value—the expected squared deviation* [like the mean squared deviation in (2-10)]. In this new E notation, we also can rewrite (4-6) as:

$$\boxed{\sigma^2 = E(X^2) - \mu^2} \tag{4-35}$$

Sometimes we find it useful to solve this for $E(X^2)$:

$$E(X^2) = \mu^2 + \sigma^2 \tag{4-36}$$

The E notation is so useful that we shall continue to use it throughout the book.

PROBLEMS

4-24 **a.** If X is a random variable, then a function of X, say $g(X)$, will be a random variable too. What is the easiest way to find its mean

- **i.** If $g(X)$ is linear?
- **ii.** If $g(X)$ is nonlinear?

b. Classify as linear or nonlinear:

- **i.** $g(X) = 2X + 4$
- **ii.** $g(X) = (X+2)(X-4)$
- **iii.** $g(X) = 5(X-32)/9$
- **iv.** $g(X) = (X+7)/3$
- **v.** $g(X) = 3/(X+7)$
- **vi.** $g(X) = \sqrt{X^2+4}$

4-25 In her new job of selling computers, Dawn Elliot faces uncertain prospects next year. She guesses that her taxable income X might be anywhere from 20 to 50 thousand dollars according to her schedule of personal probabilities $p(x)$ given below. The corresponding tax is given in the final column.

Income x (\$000)	*Probability* p(x)	*Tax t(x)* (\$000)
20	.10	4
30	.30	6
40	.40	9
50	.20	13

a. Calculate her expected income.

b. Calculate her expected tax.

c. Calculate her expected disposable income (after tax) in two ways:

i. Calculate first the table of disposable incomes, and then take their expected value.

ii. As an easier way, just use the answers in **a** and **b**.

4-26 To fill up the last week of the season, Larry's Paving bids on 6 similar but statistically independent small jobs. Larry has a 40% chance of getting each contract, and will net \$200 on each contract he gets. Then he has to defray a \$300 expense preparing the bids.

a. What is the expected number of contracts he will get?

b. What is the expected profit $E(P)$?

c. What is the chance he will make a (positive) profit? And the chance he will lose?

4-27 Now let's change Problem 4-26. Suppose the more contracts Larry wins (X), the more overtime he must pay, so that his profit is now:

$$P = \$200X - \$300 - \$20X^2$$

with this last quadratic term being the downward adjustment in his profit for overtime pay. Now answer the same questions as before.

4-28 Larry's accountant in Problem 4-27 pointed out that Larry was perhaps unwise to submit that sixth bid. After all, if he won all his bids he would be so busy paying overtime that he would *lose* money on that sixth contract.

Larry has asked you, as a consultant, to check out whether his accountant is right. (Assume that it still costs \$300 to prepare the bids, whether 6 or fewer.)

4-29 Suppose that X is a very simple discrete random variable, distributed as follows:

x	p(x)
2	.50
4	.50

For each of the following, state True or False, and back up your answer either by calculating both sides of the equation, or by appealing to a general theorem.

a. $E(X + 10) = E(X) + 10$

b. $E(X/10) = E(X)/10$

c. $E(10/X) = 10/E(X)$

d. $E(X^2) = [E(X)]^2$
e. $E(5X + 2)/10 = (5E(X) + 2)/10$

*4-7 EXPECTED VALUE IN BIDDING[3]

Bidding and negotiating are always chancy. You never know at the beginning what package you will get at the end. So this is one of the most interesting applications of probability and expected value.

Negotiating skills are essential for business and labor leaders in their collective bargaining over wages and benefits; for government officials in negotiating arms limitations or trade agreements; and for individuals negotiating a salary or the price of a car. One of the simplest negotiating problems arises in bidding, as an example will show.

EXAMPLE 4-8
HOW HIGH SHOULD YOU BID?

You are thinking of buying a summer cottage listed for $40,000. Since it has been listed unsold for six months, the owner is offering it for public auction during the regatta weekend—with the cottage going to the highest sealed bid.

You feel you might possibly get it for as little as $30,000 or $32,000, and you are almost certain to get it for $38,000 or $40,000. In fact, you estimate that the chance of winning increases as the bid increases, roughly as shown in Figure 4-11.

What should you bid?

SOLUTION

You don't have sufficient information yet. In negotiating any price, two things are vital to know:

1. What do your opponents—in this case, the competitive bidders—think it is worth? This is the information that has already been captured in Figure 4-11.

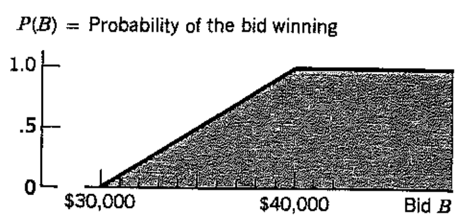


FIGURE 4-11
Higher bids give a higher probability of winning.

[3] The star means this is an optional section that is not needed in later chapters.

i. $H = \$44{,}000$ **iii.** $H = \$36{,}000$

ii. $H = \$60{,}000$ **iv.** $H = \$33{,}000$

b. Answer True or False; if false, correct it:

i. To determine the optimal bid B_0 the most important figure is V, the value of the house to you.

ii. If V is somewhere between the low and high bids L and H, then an equally important figure is the high bid H.

4-31 In looking for a used car, I find what I'm looking for—a three-year old Toyota Tercel for \$5000. A brief talk with the seller plus some market information suggests that any bid less than \$4200 would have no chance whatsoever of striking a deal.

a. What advice would you give me about what to bid?

b. Suppose I tell you that any bid over \$4800 is certain to be accepted by the seller. And the value of the car to me is \$5600—it's exactly what I want, at a price I didn't expect to find. Now what would you advise me to bid?

c. On the other hand, suppose the value to me is not \$5600, but only \$4600. Now what should I bid?

4-32 Bidding not to buy, but to sell (tendering).

Suppose you are bidding to *sell* something, instead of to buy. For example, as vice-president of a construction firm, suppose you are bidding to sell the city of Seattle a new overpass. Now it is a *low* bid that has the best chance of succeeding. For example, the probability of your bid succeeding may be 1 if you bid 4.0 million dollars, and gradually fall (linearly) to 0 if you bid 6.0 million dollars.

Suppose finally that V, your cost of building this project, is 5.2 million. Thus, for example, if you successfully bid 5.5 million, you would make a profit of 0.3 million.

a. The formula for the optimal bid is now the mirror image of (4-39). Write it down.

b. Calculate the optimal bid in this case.

c. If I bid .2 million higher than the optimal figure in **b**, would it be almost as good?

***d.** Can you think of any circumstances where you would be willing to bid slightly less than the answer in **b**?

4-33 **a.** Graph the expected gain $E(G)$ in the last column of Table 4-7, as a function of the possible bid B.

b. Is $E(G)$ a parabola? Why?

c. True or False: If false, correct it: Since $E(G)$ is a symmetric parabola, its maximum value must occur halfway between its zero values (at the low bid $L = 30$ and at the buyer's value $V = 38$). This rigorously proves that (4-38) is true.

REVIEW PROBLEMS

4-35 In the 1984 U.S. presidential election, approximately 60% voted Republican and 40% voted Democratic. Calculate the probability that a random sample would correctly forecast the election winner—that is, that a majority of the sample would be Republicans, if the sample size were:

a. $n = 1$

b. $n = 3$

c. $n = 9$

Note how the larger sample increases the probability of a correct forecast.

4-36 A manufacturer needs washers between .1180 and .1220 inches thick; any thickness outside this range is unusable. One machine shop will sell washers at $3.00 per 1000. Their thickness is normally distributed with a mean of .1200 inch and a standard deviation of .0010 inch.

A second machine shop will sell washers at $2.60 per 1000. Their thickness is normally distributed with a mean of .1200 inch and a standard deviation of .0015 inch.

Which shop offers the better deal? (Use the price per usable washer as the criterion. The costs for sorting the washers can be assumed about the same in both cases.)

4-37 American women born in 1950 have completed their families by now, as follows:

Number of Children	*Relative Frequency*
0	17%
1	19%
2	36%
3	17%
4	7%
5 (or more)	4%
Total	100%

Estimates based on *Stat. Abst. of U.S.*, 1987, pp. 59, 64.

a. Find approximately the average number of children per woman. Also the median, mode, and standard deviation.

To determine population growth, which of these measures is most appropriate?

b. Graph the relative frequency distribution. Mark the average as the balancing point. Show the standard deviation as the unit of scale as in Figure 2-6.

4-38 How many Americans would there be a generation later if American women continued to reproduce at the rate shown in Problem 4-37? Let us see how the average produces the key to the answer.

a. To start, the average number of children per woman was found to be about 1.90. But to get to the next generation of women, we must allow that only 48.8% of these children are girls (slightly less than half) and only 97% will survive to adulthood (projecting the present 3% mortality rate). These reductions leave an average of how many females surviving to the next generation?

This figure is called the "net reproduction rate," NRR. Is it consistent with your answer to Problem 2-31?

b. Suppose the NRR doesn't change. And suppose we ignore the relatively minor effects of immigration, change in life expectancy, and other changes in age distribution. Then we can get a simple *projection* of America's future population.

Starting from 250 million in 1980, roughly graph America's population as far as four generations (about 100 years). How useful do you think this projection is? Why might the NRR change?

4-39 Hawaii contains 1,100,000 people, 60% of whom are Asian, 39% white, and 1% black. If a random sample of 7 persons is drawn:

a. What is the chance that a majority will be Asians?

b. What is the chance that none will be black?

c. For the number of Asians in the sample, what is the expected value? And the standard deviation?

4-40 Mercury Mufflers guarantees its mufflers for 3 years. On each muffler, they make a profit of $15. However, they must pay $50 for any replacement under the guarantee. (If this replacement also burns out within 3 years, they are *not* committed to a second replacement.)

Their muffler life is approximately normally distributed around a mean of 4.2 years, with a standard deviation of 1.8 years.

a. What is the average profit per sale (net, after paying for a possible replacement under the guarantee)?

b. If they want this average profit per sale to be $5, by how much should they reduce the time period of the guarantee?

4-41 A family has four smoke alarms in their home, all battery operated and working independently of each other. Each has a reliability of 90%—that is, has a 90% chance of working. If fire breaks out that engulfs all four in smoke, what is the chance that at least one of them will sound the alarm?

***4-42** In a large New England college, there are 20 classes of introductory statistics with the following distribution of class size.

Class Size	*Relative Frequency*
10	.50
20	.30
90	.20
Total	1.00

The student newspaper reported that the average statistics student faced a class size "over 50." Alarmed, the Dean asked the 20 professors to calculate their average class size, and they reported, "under 30." Who's telling the truth? Or are there two truths? Specifically, calculate:

a. What class size does the average professor have?

b. What class size does the average student have? (Hint: Problem 2-41).

4-43 (The Sign Test) Eight volunteers are to have their breathing capacity measured before and after a new treatment for asthma, with the data recorded in a layout like the following:

	Breathing Capacity		
Person	*Before*	*After*	*Improvement*
H.J.	750	850	+100
K.L.	860	880	+20
M.M.	950	930	−20
.	.	.	.
.	.	.	.
.	.	.	.

a. Ignore the actual numbers in the table, and instead suppose that the treatment has *no effect whatever*. Then the "improvements" will be mere random fluctuations (resulting from minor variations in people's performance, and as likely to be negative as positive. Also assume that measurement is so precise that an improvement of exactly zero is never observed.)

What is the probability that seven or more signs will be +?

b. If it actually turned out that seven of the eight signs were +, would you question the hypothesis in part **a** that the treatment has no effect whatever?

4-44 In the 1984 presidential election, 60% voted for Reagan (Republican). A small random sample of 5 voters was simulated, and the number who voted Republican turned out to be $R = 4$. This simulation was repeated by each of 50 students in an econometrics course. Usually R turned out to be 2, 3, or 4, but occasionally as extreme as 5 or 1 or even 0. The results were arrayed in the following table:

r	Frequency
0	1
1	3
2	9
3	15
4	16
5	6
	50

a. Graph the relative frequency distribution of R, and calculate the mean and standard deviation.

b. If the sampling simulation were repeated millions of times (not merely 50), what would be your answers to part **a**?

c. Do this simulation yourself 10 more times, using the random digits in Appendix Table I.

***4-45** A small plumbing contractor specializing in home repairs employs five plumbers on a full-time basis—that is, 40 hours a week each. Business is so good that he has to turn away customers occasionally, and so he wonders whether adding a sixth plumber would be profitable.

A new full-time plumber would cost the contractor \$800 per week (\$20/hour). Yet the additional revenue generated would be uncertain. Past records indicate that the total weekly demand (X) in hours would vary as follows:

x	Relative Frequency
180–190	.03
190–200	.09
200–210	.12
210–220	.15
220–230	.22
230–240	.21
240–250	.13
250–260	.05

a. If the revenue from this work is \$30 per hour, will it pay to hire the extra plumber,

i. assuming inflexible hours (i.e., no overtime is possible).

ii. assuming each plumber could work overtime up to 5 hours per week, at a cost of $25 per hour to the contractor.

b. How much is it worth to the contractor to be able to hire plumbers overtime?

4-46 ***A Final Challenge: Estimating the Chances of "No Shows"***

Transamerican Airlines has opened a new daily flight from Chicago to Boise, Idaho. It is so popular that its 75 seats have all been reserved in its first 20 flights. Unfortunately, in each flight some of the passengers failed to show up, so that the plane left with some empty seats and lost revenue.

The number of "no shows" varied from 2 on the best flight to 11 on the worst flight. Here are the details:

Number of "No Shows," x	*Frequency*
2	1
3	4
4	0
5	4
6	2
7	5
8	1
9	1
10	0
11	2
	20 Flights

To reduce the number of empty seats, the airline can sell more than 75 reservations, of course. But that introduces a new risk of overbooking, if more than 75 people show up with reservations. In order to properly balance its risks, the airline needs to estimate the chances of various "no shows." We therefore imagine collecting data for many many flights (under similar conditions) so that the relative frequencies would settle down to probabilities $p(x)$. Let us estimate one of these probabilities—for example, the chance of exactly 3 "no shows," $p(3)$.

a. To estimate this long-run relative frequency $p(3)$, the short-run relative frequency in the table above is $4/20 = .20$. Do you think this estimate is too high, or too low? Why?

b. Graph the relative frequency distribution of X from the table above. Sketch a smooth curve through it—a crude probability model. What is its height at $X = 3$? Do you think this is a better estimate of the probability of $p(3)$?