

# 17

## The Expected Value and Standard Error

*If you believe in miracles, head for the Keno lounge.*

—JIMMY THE GREEK

### 1. THE EXPECTED VALUE

A chance process is running. It delivers a number. Then another. And another. You are about to drown in random output. But mathematicians have found a little order in this chaos. The numbers delivered by the process vary around the *expected value*, the amounts off being similar in size to the *standard error*. To be more specific, imagine generating a number through the following chance process: count the number of heads in 100 tosses of a coin. You might get 57 heads. This is 7 above the expected value of 50, so the chance error is +7. If you made another 100 tosses, you would get a different number of heads, perhaps 46. The chance error would be -4. A third repetition might generate still another number, say 47; and the chance error would be -3. Your numbers will be off 50, by chance amounts similar in size to the standard error, which is 5 (section 5 below).

The formulas for the expected value and standard error depend on the chance process which generates the number. This chapter deals with the sum of draws from a box, and the formula for the expected value will be introduced with an example: the sum of 100 draws made at random with replacement from the box

1	1	1	5
---	---	---	---

About how large should this sum be? To answer this question, think how the draws should turn out. There are four tickets in the box, so  $\boxed{5}$  should come up on around one-fourth of the draws, and  $\boxed{1}$  on three-fourths. With 100 draws, you can expect to get around twenty-five  $\boxed{5}$ 's and seventy-five  $\boxed{1}$ 's. The sum of the draws should be around

$$25 \times 5 + 75 \times 1 = 200.$$

That is the expected value.

The formula for the expected value is a short-cut. It has two ingredients:

- the number of draws;
- the average of the numbers in the box, abbreviated to "average of box."

The expected value for the sum of draws made at random with replacement from a box equals

$$(\text{number of draws}) \times (\text{average of box}).$$

To see the logic behind the formula, go back to the example. The average of the box is

$$\frac{1 + 1 + 1 + 5}{4} = 2.$$

On the average, each draw adds around 2 to the sum. With 100 draws, the sum must be around  $100 \times 2 = 200$ .

*Example 1.* Suppose you are going to Las Vegas to play Keno. Your favorite bet is a dollar on a single number. When you win, they give you the dollar back and two dollars more. When you lose, they keep the dollar. There is 1 chance in 4 to win.<sup>1</sup> About how much should you expect to win (or lose) in 100 plays, if you make this bet on each play?

*Solution.* The first step is to write down a box model. On each play, your net gain either goes up by \$2 or goes down by \$1. There is 1 chance in 4 to go up; there are 3 chances in 4 to go down. So your net gain after 100 plays is like the sum of 100 draws at random with replacement from the box

\$2	-\$1	-\$1	-\$1
-----	------	------	------

The average of this box is

$$\frac{\$2 - \$1 - \$1 - \$1}{4} = -\$0.25$$

On the average, each play costs you a quarter. In 100 plays, you can expect to lose around \$25. This is the answer. If you continued on, in 1,000 plays you should expect to lose around \$250. The more you play, the more you lose. Perhaps you should look for another game.

3 GREEK

r. And  
is have  
ss vary  
andard  
lowing  
might  
error is  
heads,  
enerate  
mbers  
which

chance  
draws  
with an  
he box

## Exercise Set A

- Find the expected value for the sum of 100 draws at random with replacement from the box—
  - $\boxed{0} \boxed{1} \boxed{1} \boxed{6}$
  - $\boxed{-2} \boxed{-1} \boxed{0} \boxed{2}$
  - $\boxed{-2} \boxed{-1} \boxed{3}$
  - $\boxed{0} \boxed{1} \boxed{1}$
- Find the expected number of squares moved on the first play in Monopoly (p. 279).
- Someone is going to play roulette 100 times, betting a dollar on the number 17 each time. Find the expected value for the net gain. (See pp. 283–284.)
- You are going to play roulette 100 times, staking \$1 on red-or-black each time. Find the expected value for your net gain. (This bet pays even money, and you have 18 chances in 38 of winning; figure 3 on p. 282.)
- Repeat exercise 4 for 1,000 plays.
- A game is *fair* if the expected value for the net gain equals 0: on the average, players neither win nor lose. A generous casino would offer a bit more than \$1 in winnings if a player staked \$1 on red-and-black in roulette and won. How much should they pay to make it a fair game? (Hint: Let  $x$  stand for what they should pay. The box has 18 tickets  $\boxed{x}$  and 20 tickets  $\boxed{-\$1}$ . Write down the formula for the expected value in terms of  $x$  and set it equal to 0.)
- If an Adventurer at the Game of the Royal Oak staked 1 pound on a point and won, how much should the Master of the Ball have paid him, for the Game to be fair? (The rules are explained in exercise 6 on p. 250.)

The answers to these exercises are on pp. A72–73.

## 2. THE STANDARD ERROR

Suppose 25 draws are made at random with replacement from the box

$$\boxed{0} \boxed{2} \boxed{3} \boxed{4} \boxed{6}$$

(There is nothing special about the numbers in the box; they were chosen to make later calculations come out evenly.) Each of the five tickets should appear on about one-fifth of the draws, that is, 5 times. So the sum should be around

$$5 \times 0 + 5 \times 2 + 5 \times 3 + 5 \times 4 + 5 \times 6 = 75.$$

That is the expected value for the sum. Of course, each ticket won't appear exactly one-fifth of the draws, just as Kerrich didn't get heads on exactly half the tosses. The sum will be off the expected value by a chance error

$$\text{sum} = \text{expected value} + \text{chance error}.$$

The chance error is the amount above (+) or below (–) the expected value. For example, if the sum is 70, the chance error is –5.

How big is the chance error likely to be? The answer is given by the *standard error*, usually abbreviated to SE.

A sum is likely to be around its expected value, but to be off by a chance error similar in size to the standard error.

There is a formula to use in computing the SE for a sum of draws made at random with replacement from a box. It is called a *square root law*, because it involves the square root of the number of draws. The statistical procedures presented in the rest of the book depend on this formula.<sup>2</sup>

*The square root law.* When drawing at random with replacement from a box of numbered tickets, the standard error for the sum of the draws is

$$\sqrt{\text{number of draws}} \times (\text{SD of box}).$$

The formula has two ingredients: the square root of the number of draws, and the SD of the list of numbers in the box (abbreviated to "SD of the box"). The SD measures the spread among the numbers in the box. If there is a lot of spread in the box, the SD is big, and it is hard to predict how the draws will turn out. So the standard error must be big too. Now for the number of draws. The sum of two draws is more variable than a single draw; the sum of 100 draws is still more variable. Each draw adds some extra variability to the sum, because you don't know how it is going to turn out. As the number of draws goes up, the sum gets harder to predict, the chance errors get bigger, and so does the standard error. However, the standard error goes up slowly, by a factor equal to the square root of the number of draws. For instance, the sum of 100 draws is only  $\sqrt{100} = 10$  times as variable as a single draw.

The SD and the SE are different.<sup>3</sup> The SD applies to spread in lists of numbers. It is worked out using the method explained on p. 71. By contrast, the SE applies to chance variability—for instance, in the sum of the draws.

The SD is for a list

1 2 3 4 5 6

The SE is for a chance process



At the beginning of the section, we looked at the sum of 25 draws made at random with replacement from the box

0 2 3 4 6

The expected value for this sum is 75: the sum will be around 75, but will be off by a chance error. How big is the chance error likely to be? To find out, calculate the standard error. The average of the numbers in the box is 3. The deviations

from the average are

$$-3 \quad -1 \quad 0 \quad 1 \quad 3$$

The SD of the box is

$$\begin{aligned} \sqrt{\frac{(-3)^2 + (-1)^2 + 0^2 + 1^2 + 3^2}{5}} &= \sqrt{\frac{9 + 1 + 0 + 1 + 9}{5}} \\ &= \sqrt{\frac{20}{5}} = 2. \end{aligned}$$

This measures the variability in the box. According to the square root law, the sum of 25 draws is more variable, by the factor  $\sqrt{25} = 5$ . The SE for the sum of 25 draws is  $5 \times 2 = 10$ . In other words, the likely size of the chance error is 10. And the sum of the draws should be around 75, give or take 10 or so. In general, the sum is likely to be around its expected value, give or take a standard error or so.

To show what this means empirically, we had the computer programmed to draw 25 times at random with replacement from the box  $\boxed{0} \boxed{2} \boxed{3} \boxed{4} \boxed{6}$ . It got

0 0 4 4 0    4 3 2 6 2    2 0 2 6 2    6 4 2 6 3    0 3 6 4 0

The sum of these 25 draws is 71. This is 4 below the expected value, so the chance error is  $-4$ . The computer drew another 25 times and took the sum, getting 76. The chance error was  $+1$ . The third sum was 86, with a chance error of  $+11$ . In fact, we had the computer generate 100 sums, shown in table 1. These numbers are all around 75, the expected value. They are off by chance errors similar in size to 10, the standard error.

The sum of the draws is likely to be around \_\_\_\_\_, give or take \_\_\_\_\_ or so. The expected value for the sum fills in the first blank. The SE for the sum fills in the second blank.

Some terminology: the number 71 in table 1 is an *observed value* for the sum of the draws; the 76 is another observed value. All told, the table has 100 observed values for the sum. These observed values differ from the expected value of 75. The difference is chance error. For example, the chance error in 71 is  $-4$ , because  $71 - 75 = -4$ . The chance error in 76 is  $+1$ , because  $76 - 75 = 1$ . And so forth.

The observed values in table 1 show remarkably little spread around the expected value. In principle, they could be as small as 0, or as large as  $25 \times 6 = 150$ . However, all but one of them are between 50 and 100, that is, within 2.5 SEs of the expected value.

Observed values are rarely more than 2 or 3 SEs away from the expected value.

Table 1. Computer simulation: the sum of 25 draws made at random with replacement from the box  $\boxed{0} \boxed{2} \boxed{3} \boxed{4} \boxed{6}$ .

Repetition	Sum	Repetition	Sum	Repetition	Sum	Repetition	Sum	Repetition	Sum
1	71	21	80	41	64	61	64	81	60
2	76	22	77	42	65	62	70	82	67
3	86	23	70	43	88	63	65	83	82
4	78	24	71	44	77	64	78	84	85
5	88	25	79	45	82	65	64	85	77
6	67	26	56	46	73	66	77	86	79
7	76	27	79	47	92	67	81	87	82
8	59	28	65	48	75	68	72	88	88
9	59	29	72	49	57	69	66	89	76
10	75	30	73	50	68	70	74	90	75
11	76	31	78	51	80	71	70	91	77
12	66	32	75	52	70	72	76	92	66
13	76	33	89	53	90	73	80	93	69
14	84	34	77	54	76	74	70	94	86
15	58	35	81	55	77	75	56	95	81
16	60	36	68	56	65	76	49	96	90
17	79	37	70	57	67	77	60	97	74
18	78	38	86	58	60	78	98	98	72
19	66	39	70	59	74	79	81	99	57
20	71	40	71	60	83	80	72	100	62

Exercise Set B

1. One hundred draws are made at random with replacement from the box  $\boxed{1} \boxed{2} \boxed{3} \boxed{4} \boxed{5} \boxed{6} \boxed{7}$ .

- (a) Find the expected value and standard error for the sum.
- (b) The sum of the draws will be around \_\_\_\_\_, give or take \_\_\_\_\_ or so.
- (c) Suppose you had to guess what the sum was going to be. What would you guess? Would you expect to be off by around 2, 4, or 20?

2. You gamble 100 times on the toss of a coin. If it lands heads, you win \$1. If it lands tails, you lose \$1. Your net gain will be around \_\_\_\_\_, give or take \_\_\_\_\_ or so. Fill in the blanks, using the options

-\$10    -\$5    \$0    +\$5    +\$10

3. The expected value for a sum is 50, with an SE of 5. The chance process generating the sum is repeated ten times. Which of the following is the sequence of observed values, and why?

- (i) 51, 57, 48, 52, 57, 61, 58, 41, 53, 48
- (ii) 51, 49, 50, 52, 48, 47, 53, 50, 49, 47
- (iii) 45, 50, 55, 45, 50, 55, 45, 50, 55, 45

4. Fifty draws are made at random with replacement from the box  $\boxed{1} \boxed{2} \boxed{3} \boxed{4} \boxed{5}$ ; the sum of the draws turns out to be 157. The expected value for the sum is \_\_\_\_\_, the observed value is \_\_\_\_\_, the chance error is \_\_\_\_\_, and the standard error is \_\_\_\_\_. Fill in the blanks, and explain briefly.

5. Tickets are drawn at random with replacement from a box of numbered tickets. The sum of 25 draws has expected value equal to 50, and the SE is 10. If possible, find the expected value and SE for the sum of 100 draws. Or do you need more information?
6. One hundred draws are made at random with replacement from the box  $\boxed{0} \boxed{2} \boxed{3} \boxed{4} \boxed{6}$ . True or false and explain.
  - (a) The expected value for the sum of the draws is 300.
  - (b) The expected value for the sum of the draws is 300, give or take 20 or so.
  - (c) The sum of the draws will be 300.
  - (d) The sum of the draws will be around 300, give or take 20 or so.
7. In the simulation for table 1 (p. 293), if the computer kept on running, do you think it would eventually generate a sum more than 3 SEs away from the expected value? Explain.

The answers to these exercises are on p. A73.

### 3. USING THE NORMAL CURVE

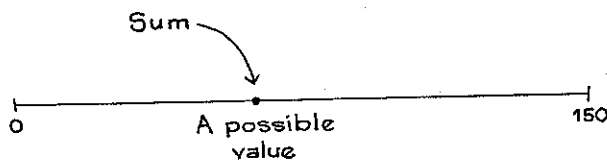
A large number of draws will be made at random with replacement from a box. What is the chance that the sum of the draws will be in a given range? Mathematicians discovered the normal curve while trying to solve problems of this kind. The logic behind the curve will be discussed in the next chapter. The object of this section is only to sketch the method, which applies whenever the number of draws is reasonably large. Basically, it is a matter of converting to standard units (using the expected value and standard error) and then working out areas under the curve, just as in chapter 5.

Now for an example. Suppose the computer is programmed to take the sum of 25 draws made at random with replacement from the magic box

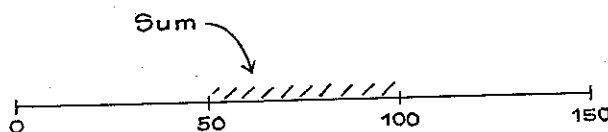
$\boxed{0} \boxed{2} \boxed{3} \boxed{4} \boxed{6}$

It prints out the result, repeating the process over and over again. About what percentage of the observed values should be between 50 and 100?

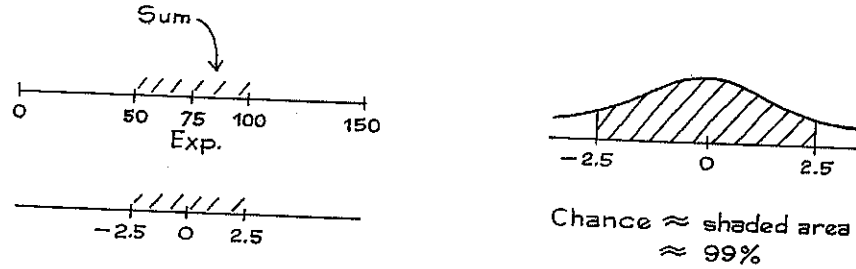
Each sum will be somewhere on the horizontal axis between 0 and  $25 \times 6 = 150$ .



The problem is asking for the chance that the sum will turn out to be between 50 and 100.



To find the chance, convert to standard units and use the normal curve. Standard units say how many SEs a number is away from the expected value.<sup>4</sup> In the example, 100 becomes 2.5 in standard units. The reason: the expected value for the sum is 75 and the SE is 10, so 100 is 2.5 SEs above the expected value. Similarly, 50 becomes -2.5.

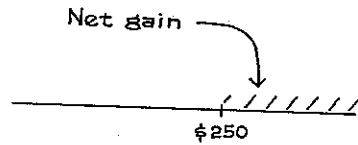


The interval from 50 to 100 is the interval within 2.5 SEs of the expected value, so the sum should be there about 99% of the time.

That finishes the calculation. Now for some data. Table 1 above reported 100 observed values for the sum: about 99 of them should be in the interval from 50 to 100, and in fact 99 of them are. To take some less extreme ranges, about 68% of the observed values should be in the interval from 75 - 10 to 75 + 10. In fact, 73 are. Finally, about 95% of the observed values in table 1 should be in the range 75 ± 20, and 98 of them are. The theory looks pretty good. (Ranges include endpoints; ± is read "plus-or-minus.")

**Example 2.** In a month, there are 10,000 independent plays on a roulette wheel in a certain casino. To keep things simple, suppose the gamblers only stake \$1 on red at each play. Estimate the chance that the house will win more than \$250 from these plays.<sup>5</sup> (Red-or-black pays even money, and the house has 20 chances in 38 to win.)

**Solution.** The problem asks for the chance that the net gain of the house will be more than \$250.



The box model is the first thing. The box is

$$20 \text{ tickets } \boxed{+\$1} \quad 18 \text{ tickets } \boxed{-\$1}$$

The net gain for the house is like the sum of 10,000 draws from this box.

The expected value for the net gain is the number of draws times the average of the numbers in the box. The average is

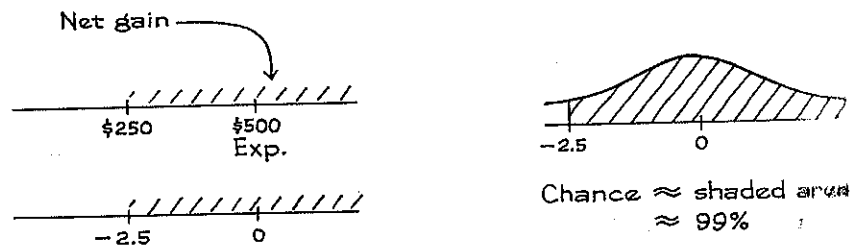
$$\frac{20 \text{ tickets } \overbrace{+\$1 + \dots + \$1} + 18 \text{ tickets } \overbrace{-\$1 - \dots - \$1}}{38} = \frac{\$20 - \$18}{38} = \frac{\$2}{38} \approx \$0.05$$



On the average, each draw adds around \$0.05 to the sum. The sum of 10,000 draws has an expected value of  $10,000 \times \$0.05 = \$500$ . The house averages about a nickel on each play, so in 10,000 plays it can expect to win around \$500. (The gambler and the house are on opposite sides of the box: 20 tickets are good for the house, and 18 are good for the gambler; see pp. 281–283.)

Finding the SE for the net gain comes next. This requires the SD of the numbers in the box. The deviations from average are all just about \$1, because the average is close to \$0. So the SD of the box is about \$1. This \$1 measures the variability in the box. According to the square root law, the sum of 10,000 draws is more variable, by the factor  $\sqrt{10,000} = 100$ . The SE for the sum of 10,000 draws is  $100 \times \$1 = \$100$ . The house can expect to win around \$500, give or take \$100 or so.

Now the normal curve can be used.



This completes the solution. The key idea: the net gain is like the sum of the draws from a box; that provided a logical basis for the square root law.

The house has about a 99% chance to win more than \$250. This may not seem like much, but you have to remember that the house owns many wheels; there often is a crowd of gamblers playing on each spin of each wheel, and a lot of bets are over a dollar. The house can expect to win about 5% of the money that crosses the table, and the square root law virtually eliminates the risk. For instance, suppose the house runs 25 wheels. To be very conservative, suppose each wheel operates under the conditions of example 2. With these assumptions the casino's expected winnings go up by a full factor of 25, to  $25 \times \$500 = \$12,500$ . But their standard error only goes up by the factor  $\sqrt{25} = 5$ , to \$500. Now the casino can be virtually certain—99%—of winning at least \$11,000. For the casino, roulette is a volume business, just like groceries are for Safeway.

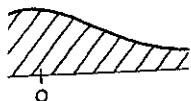
### Exercise Set C

1. One hundred draws will be made at random with replacement from the box

$\boxed{1} \boxed{1} \boxed{2} \boxed{2} \boxed{2} \boxed{4}$ .

- (a) The smallest the sum can be is \_\_\_\_\_, the largest is \_\_\_\_\_.
- (b) The sum of the draws will be around \_\_\_\_\_, give or take \_\_\_\_\_.
- (c) The chance that the sum will be bigger than 250 is almost \_\_\_\_\_.

sum of 10,000  
house averages  
in around \$500.  
tickets are good  
3.)  
as the SD of the  
about \$1, because  
this \$1 measures  
the sum of 10,000  
E for the sum of  
win around \$500,



≡ shaded area  
≡ 99%

ke the sum of the  
root law.

250. This may not  
owns many wheels,  
ch wheel, and a lot  
t 5% of the money  
inates the risk. For  
onservative, suppose  
these assumptions,  
5, to  $25 \times \$500 =$   
 $\sqrt{25} = 5$ , to \$500.  
at least \$11,000. For  
are for Safeway.

ement from the bin

r take \_\_\_\_\_ or m  
almost \_\_\_\_\_%

2. One hundred draws will be made at random with replacement from the box  $\boxed{1} \boxed{3} \boxed{3} \boxed{9}$ .

- (a) How large can the sum be? How small?
- (b) How likely is the sum to be in the range from 370 to 430?

3. You can draw either 10 times or 100 times at random with replacement from the box  $\boxed{-1} \boxed{1}$ . How many times should you draw—

- (a) To win \$1 when the sum is 5 or more, and nothing otherwise?
- (b) To win \$1 when the sum is  $-5$  or less, and nothing otherwise?
- (c) To win \$1 when the sum is between  $-5$  and  $5$ , and nothing otherwise?

No calculations are needed, but explain your reasoning.

4. There are two options:

(i) One hundred draws will be made at random with replacement from the box  $\boxed{1} \boxed{1} \boxed{5} \boxed{7} \boxed{8} \boxed{8}$ .

(ii) Twenty-five draws will be made at random with replacement from the box  $\boxed{14} \boxed{17} \boxed{21} \boxed{23} \boxed{25}$ .

Which is better, if the payoff is—

- (a) \$1 when the sum is 550 or more, and nothing otherwise?
- (b) \$1 when the sum is 450 or less, and nothing otherwise?
- (c) \$1 when the sum is between 450 and 550, and nothing otherwise?

5. Suppose that in one week at a certain casino, there are 25,000 independent plays at roulette. On each play, the gamblers stake \$1 on red. Is the chance that the casino will win more than \$1,000 from these 25,000 plays closest to 2%, 50%, or 98%? Explain briefly.

6. Suppose that one person stakes \$25,000 on one play at red-or-black in roulette. Is the chance that the casino will win more than \$1,000 from this play closest to 2%, 50%, or 98%? Explain briefly.

7. A gambler plays once at roulette, staking \$1,000 on each number (including 0 and 00). So this person has staked \$38,000 in all. What will happen? Explain briefly.

8. A box contains 10 tickets. Each ticket is marked with a whole number between  $-5$  and  $5$ . The numbers are not all the same; their average equals 0. There are two choices:

- (A) 100 draws are made from the box, and you win \$1 if the sum is between  $-15$  and  $15$ .
- (B) 200 draws are made from the box, and you win \$1 if the sum is between  $-30$  and  $30$ .

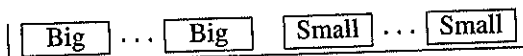
Choose one of the four options below; explain your answer.<sup>6</sup>

- (i) A gives a better chance of winning.
- (ii) B gives a better chance of winning.
- (iii) A and B give the same chance of winning.
- (iv) Can't tell without the SD of the box.

<sup>6</sup>Answers to these exercises are on p. A74.

#### 4. A SHORT-CUT

Finding SDs can be painful, but there is a short-cut for boxes with only two kinds of tickets.<sup>7</sup>



When the tickets in the box show only two different numbers ("big" and "small"), the SD of the box equals

$$\left( \begin{array}{c} \text{big} \\ \text{number} \end{array} - \begin{array}{c} \text{small} \\ \text{number} \end{array} \right) \times \sqrt{\begin{array}{c} \text{fraction with} \\ \text{big number} \end{array} \times \begin{array}{c} \text{fraction with} \\ \text{small number} \end{array}}$$

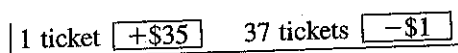
For example, take the box 1 1 1 5. The short-cut can be used because the tickets show only two different numbers, 1 and 5. So the SD is

$$(5 - 1) \times \sqrt{\frac{1}{4} \times \frac{3}{4}} \approx 1.73$$

The short-cut involves much less arithmetic than finding the root-mean-square of the deviations from average (p. 71), and gives exactly the same answer.

*Example 3.* A gambler plays roulette 100 times, staking \$1 on the number 10 each time. The bet pays 35 to 1, and the gambler has 1 chance in 38 to win. Fill in the blanks: The gambler will win \$\_\_\_\_\_, give or take \$\_\_\_\_\_ or 80.

*Solution.* The first thing to do is to make a box model for the net gain. (See example 1 on pp. 283–284.) The gambler's net gain is like the sum of 100 draws made at random with replacement from



What is the expected net gain? This is 100 times the average of the box. The average of the numbers in the box is their total, divided by 38. The winning ticket contributes \$35 to the total, while the 37 losing tickets take away \$37 in all. So the average is

$$\frac{\$35 - \$37}{38} = \frac{-\$2}{38} \approx -\$0.05$$

In 100 plays, the expected net gain is

$$100 \times (-\$0.05) = -\$5$$

In other words, the gambler expects to lose about \$5 in 100 plays.

The next step is to find the SE for the sum of the draws: this is  $\sqrt{100}$  times the SD of the box. The short-cut can be used, and the SD of the box equals

with only two

$$[\$35 - (-\$1)] \times \sqrt{\frac{1}{38} \times \frac{37}{38}} \approx \$36 \times 0.16 \approx \$5.76$$

The SE for the sum of the draws is  $\sqrt{100} \times \$5.76 \approx \$58$ .

The gambler will lose about \$5, give or take \$58 or so. This completes the solution. The big SE gives the gambler a reasonable chance of winning a lot, and that is the attraction. Of course, on average the gambler loses; and the big SE also means that the gambler can lose big.

umbers

with  
umber

it can be used  
the SD is

not-mean-square  
ae answer.

1 on the number  
ce in 38 to win.  
\$\_\_\_\_\_ or so.

for the net gain.  
the sum of 100

verage of the box.  
38. The winning  
take away \$37 in

lays.  
this is  $\sqrt{100}$  times  
the box equals

### Exercise Set D

1. Does the formula give the SD of the list? Explain.

List	Formula
(a) 7, 7, 7, -2, -2	$5 \times \sqrt{3/5 \times 2/5}$
(b) 0, 0, 0, 0, 5	$5 \times \sqrt{1/5 \times 4/5}$
(c) 0, 0, 1	$\sqrt{2/3 \times 1/3}$
(d) 2, 2, 3, 4, 4, 4	$2 \times \sqrt{1/6 \times 2/6 \times 3/6}$

- Suppose a gambler bets a dollar on a single number at Keno (example 1 on p. 289). In 100 plays, the gambler's net gain will be \$\_\_\_\_\_, give or take \$\_\_\_\_\_ or so.
- At Nevada roulette tables, the "house special" is a bet on the numbers 0, 00, 1, 2, 3. The bet pays 6 to 1, and there are 5 chances in 38 to win.
  - For all other bets at Nevada roulette tables, the house expects to make about 5 cents out of every dollar put on the table. How much does it expect to make per dollar on the house special?
  - Someone plays roulette 100 times, betting a dollar on the house special each time. Estimate the chance that this person comes out ahead.
- A gambler plays roulette 100 times. There are two possibilities:
  - Betting \$1 on a section each time (see figure 3 on p. 282).
  - Betting \$1 on red each time.

A section bet pays 2 to 1, and there are 12 chances in 38 to win. Red pays even money, and there are 18 chances in 38 to win. True or false, and explain:

  - The chance of coming out ahead is the same with (i) and (ii).
  - The chance of winning more than \$10 is bigger with (i).
  - The chance of losing more than \$10 is bigger with (i).

The answers to these exercises are on pp. A74-75.

### CLASSIFYING AND COUNTING

Some chance processes involve counting. The square root law can be used to find the standard error for a count, but the box model has to be set up correctly. The next example will show how to do this.

*Example 4.* A die is rolled 60 times.

(a) The total number of spots should be around \_\_\_\_\_, give or take \_\_\_\_\_ or so.

(b) The number of 6's should be around \_\_\_\_\_, give or take \_\_\_\_\_ or so.

By way of illustration, table 2 shows the results of throwing a die 60 times: the first throw was a 4, the second was a 5, and so on.

Table 2. Sixty throws of a die.

4	5	5	2	4	5	3	2	6	3	5	4	6	2	6	4	4	2	5	6
1	5	3	1	2	2	1	2	5	3	3	6	6	1	1	5	1	6	1	2
4	4	2	1	4	4	5	2	6	3	2	4	6	1	6	4	6	1	5	2

*Solution.* Part (a) is familiar. It involves adding. Each throw contributes some number of spots, and we add these numbers up. The total number of spots in 60 throws of the die is like the sum of 60 draws from the box

1	2	3	4	5	6
---	---	---	---	---	---

The average of this box is 3.5 and the SD is 1.71. The expected value for the sum is  $60 \times 3.5 = 210$ ; the SE for the sum is  $\sqrt{60} \times 1.71 \approx 13$ . The total number of spots will be around 210, give or take 13 or so. In fact, the sum of the numbers in table 2 is 212. The sum was off its expected value by around one-sixth of an SE.

Part (b). Filling in the first blank is easy. Each of the six faces should come up on about one-sixth of the throws, so the expected value for the number of 6's is  $60 \times 1/6 = 10$ . The second blank is harder. We need a new kind of box because the sum of the draws from 

1	2	3	4	5	6
---	---	---	---	---	---

 is no longer relevant. Instead of being added, each throw of the die is classified: is it a 6, or not? (There are only two classes here: 6's on one hand, everything else on the other.) Then, the number of 6's is counted up.

The point to notice is that on each throw, the number of 6's either goes up by 1, or stays the same:

- 1 is added to the count if the throw is 6;
- 0 is added to the count if the throw is anything else.

The count has 1 chance in 6 to go up by one, and 5 chances in 6 to stay the same. Therefore, on each draw, the sum must have 1 chance in 6 to go up by one, and 5 chances in 6 to stay the same. The right box to use is

X <sup>0</sup>	Z <sup>0</sup>	3 <sup>0</sup>	4 <sup>0</sup>	5 <sup>0</sup>	6 <sup>1</sup>
----------------	----------------	----------------	----------------	----------------	----------------

As far as the chances are concerned, the number of 6's in 60 throws of the die is just like the sum of 60 draws from the new box. This puts us in a position to use the square root law.

The new box has five  $\boxed{0}$ 's and a  $\boxed{1}$ . The SD is  $\sqrt{1/6 \times 5/6} \approx 0.37$ , by the short-cut method. And the SE for the sum of the draws is  $\sqrt{60} \times 0.37 \approx 3$ . In 60 throws of a die, the number of 6's will be around 10, give or take 3 or so. In fact, in table 2 there were eleven 6's. The observed number of 6's was off its expected value by a third of an SE. This completes the example. It's the old story, for a new box.

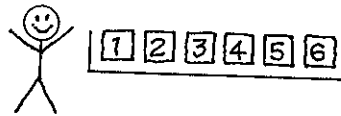
This example makes a general point. Although they may look quite different, many problems about chance processes can be solved in the same way. In these problems, some tickets are drawn at random from a box. An operation is performed on the draws, and the problem asks for the chance that the result will be in a given interval. In this chapter, there are two possible operations on the draws:

- adding,
- classifying and counting.

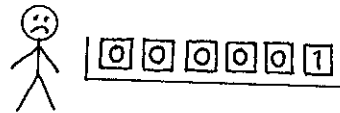
The message is that both operations can be treated the same way—provided you change the box.

If you have to classify and count the draws, put 0's and 1's on the tickets. Mark 1 on the tickets that count for you, 0 on the others.

For adding up the draws,  
the box is



For counting 6's,  
the box is



Remember to change the tickets!

**Example 5.** A coin will be tossed 100 times. Find the expected value and standard error for the number of heads. Estimate the chance of getting between 40 and 60 heads.

**Solution.** The first thing is to make a box model. The problem involves classifying the tosses as heads or tails, and then counting the number of heads. So there should be only 0's and 1's in the box. The chances are 50–50 for heads, so the box should be  $\boxed{0} \boxed{1}$ . The number of heads in 100 tosses of a coin is like the sum of 100 draws made at random with replacement from the box  $\boxed{0} \boxed{1}$ . (The coin is even simpler than the die in example 4: each toss either pushes the number of heads up by 1 or leaves it alone, with a 50–50 chance; likewise, each draw from the box either pushes the sum up by 1 or leaves it alone, with the same 50–50 chance.) This completes the model.

Since the number of heads is like the sum of the draws, the square root law will be used. The SD of the box is  $1/2$ . So the SE for the sum of 100 draws is  $\sqrt{100} \times 1/2 = 5$ . The number of heads will be around 50, give or take 5 or

→ give or take

e \_\_\_\_\_ or so.

ie 60 times: the

5 6

1 2

5 2

now contributes  
number of spots

x

value for the sum  
e total number of  
of the numbers in  
re-sixth of an SE.

aces should come  
or the number of  
a new kind of box  
o longer relevant.  
is it a 6, or not?  
else on the other.)

6's either goes up

es in 6 to stay the  
e in 6 to go up by  
e is

0 throws of the die  
s us in a position to

so. The range from 40 to 60 heads represents the expected value, give or take 2 SEs. And the chance is around 95%. This completes the solution.

To interpret this 95% chance, imagine counting the number of heads in 100 tosses of a coin. You might get 44 heads. Toss again: you might get 54 heads. A third time, the number would change once more, perhaps to 48 heads. And so on. In the long run, about 95% of these counts would come out in the range from 40 to 60. John Kerrich actually did this experiment. Table 3 shows the results, with Kerrich's 10,000 tosses broken down into successive groups of one hundred. In fact, 95 out of 100 groups had 40 to 60 heads (inclusive). The theory looks good.

Table 3. Kerrich's coin tossing experiment, showing the number of heads he got in each successive group of 100 tosses.

Group of 100 tosses	No. of heads	Group of 100 tosses	No. of heads	Group of 100 tosses	No. of heads	Group of 100 tosses	No. of heads
1-100	44	2,501-2,600	44	5,001-5,100	42	7,501-7,600	48
101-200	54	2,601-2,700	34	5,101-5,200	68	7,601-7,700	44
201-300	48	2,701-2,800	59	5,201-5,300	45	7,701-7,800	58
301-400	53	2,801-2,900	50	5,301-5,400	37	7,801-7,900	57
401-500	56	2,901-3,000	51	5,401-5,500	47	7,901-8,000	48
501-600	57	3,001-3,100	51	5,501-5,600	52	8,001-8,100	45
601-700	56	3,101-3,200	48	5,601-5,700	51	8,101-8,200	50
701-800	45	3,201-3,300	56	5,701-5,800	49	8,201-8,300	54
801-900	45	3,301-3,400	57	5,801-5,900	48	8,301-8,400	46
901-1,000	44	3,401-3,500	50	5,901-6,000	37	8,401-8,500	56
1,001-1,100	40	3,501-3,600	54	6,001-6,100	47	8,501-8,600	58
1,101-1,200	54	3,601-3,700	47	6,101-6,200	52	8,601-8,700	54
1,201-1,300	53	3,701-3,800	53	6,201-6,300	45	8,701-8,800	49
1,301-1,400	55	3,801-3,900	50	6,301-6,400	48	8,801-8,900	48
1,401-1,500	52	3,901-4,000	53	6,401-6,500	44	8,901-9,000	47
1,501-1,600	54	4,001-4,100	52	6,501-6,600	51	9,001-9,100	59
1,601-1,700	58	4,101-4,200	54	6,601-6,700	55	9,101-9,200	51
1,701-1,800	50	4,201-4,300	55	6,701-6,800	53	9,201-9,300	51
1,801-1,900	53	4,301-4,400	52	6,801-7,000	52	9,301-9,400	50
1,901-2,000	42	4,401-4,500	51	6,901-7,000	60	9,401-9,500	50
2,001-2,100	56	4,501-4,600	53	7,001-7,100	50	9,501-9,600	50
2,101-2,200	53	4,601-4,700	54	7,101-7,200	57	9,601-9,700	50
2,201-2,300	53	4,701-4,800	47	7,201-7,300	49	9,701-9,800	50
2,301-2,400	45	4,801-4,900	42	7,301-7,400	46	9,801-9,900	50
2,401-2,500	52	4,901-5,000	44	7,401-7,500	62	9,901-10,000	50

It is time to connect the square root law and the law of averages. If a coin is tossed a large number of times. Then heads will come up on about the tosses:

$$\text{number of heads} = \text{half the number of tosses} + \text{chance error}$$

How big is the chance error likely to be? At first, Kerrich's assistant the

would be very small. The record showed him to be wrong. As Kerrich kept tossing the coin, the chance error grew in absolute terms but shrank relative to the number of tosses, just as the mathematics predicts. (See figures 1 and 2, pp. 275-276.)

According to the square root law, the likely size of the chance error is  $\sqrt{\text{number of tosses}} \times 1/2$ . For instance, with 10,000 tosses the standard error is  $\sqrt{10,000} \times 1/2 = 50$ . When the number of tosses goes up to 1,000,000, the standard error goes up too, but only to 500—because of the square root. As the number of tosses goes up, the SE for the number of heads gets bigger and bigger in absolute terms, but smaller and smaller relative to the number of tosses. That is why the percentage of heads gets closer and closer to 50%. The square root law is the mathematical explanation for the law of averages.

Exercise Set E

1. A coin is tossed 16 times.
  - (a) The number of heads is like the sum of 16 draws made at random with replacement from one of the following boxes. Which one and why?
 

(i)  head  tail    (ii)  0  1    (iii)  0  1  1
  - (b) The number of heads will be around \_\_\_\_\_, give or take \_\_\_\_\_ or so.
2. One hundred draws are made at random with replacement from the box  1  2  3  4  5. What is the chance of getting between 8 and 32 tickets marked "5"?
3. According to the simplest genetic model, the sex of a child is determined at random, as if by drawing a ticket at random from the box

male  female

What is the chance that of the next 2,500 births (not counting twins or other multiple births), more than 1,275 will be females?

4. This exercise and the next are based on Kerrich's coin-tossing experiment (table 3, p. 302). For example, in tosses 1-100, the observed number of heads was 44, the expected number was 50, so the chance error was  $44 - 50 = -6$ . Fill in the blanks.

Group of 100 tosses	Observed value	Expected value	Chance error	Standard error
1-100	44	50	-6	—
101-200	54	50	—	—
201-300	48	—	—	—
301-400	—	—	—	—

How many of the counts in table 3 on p. 302 should be in the range 45 to 55? How many are? (Endpoints included.)

2. give or take 2 n.

of heads in 100 get 54 heads. A eads. And so on. ie range from 40 the results, with one hundred. In eory looks good.

ber of heads

Group of 100 tosses	No. of heads
501-7,600	48
601-7,700	43
701-7,800	58
801-7,900	57
901-8,000	48
001-8,100	45
101-8,200	50
201-8,300	53
301-8,400	46
401-8,500	56
501-8,600	58
601-8,700	54
701-8,800	49
801-8,900	48
901-9,000	45
001-9,100	55
101-9,200	51
201-9,300	48
301-9,400	56
401-9,500	55
501-9,600	55
601-9,700	50
701-9,800	48
801-9,900	59
901-10,000	52

averages. Support ne up about half

nce error.

assistant thought