The law of large numbers

In this section we shall study some further properties of the independent trials process with two outcomes. In Section 4.8 we saw that the probability for x successes in n trials is given by

$$f(n, x; p) = \binom{n}{x} p^x q^{n-x}.$$

In Figure 4.16 we show these probabilities graphically for $n = 8$ and $p = \frac{3}{4}$. In Figure 4.17 we have done similarly for the case of $n = 7$ and $p = \frac{3}{4}$.

We see in the first case that the values increase up to a maximum value at $x = 6$ and then decrease. In the second case the values increase up to a maximum value at $x = 5$, have the same value for $x = 6$, and

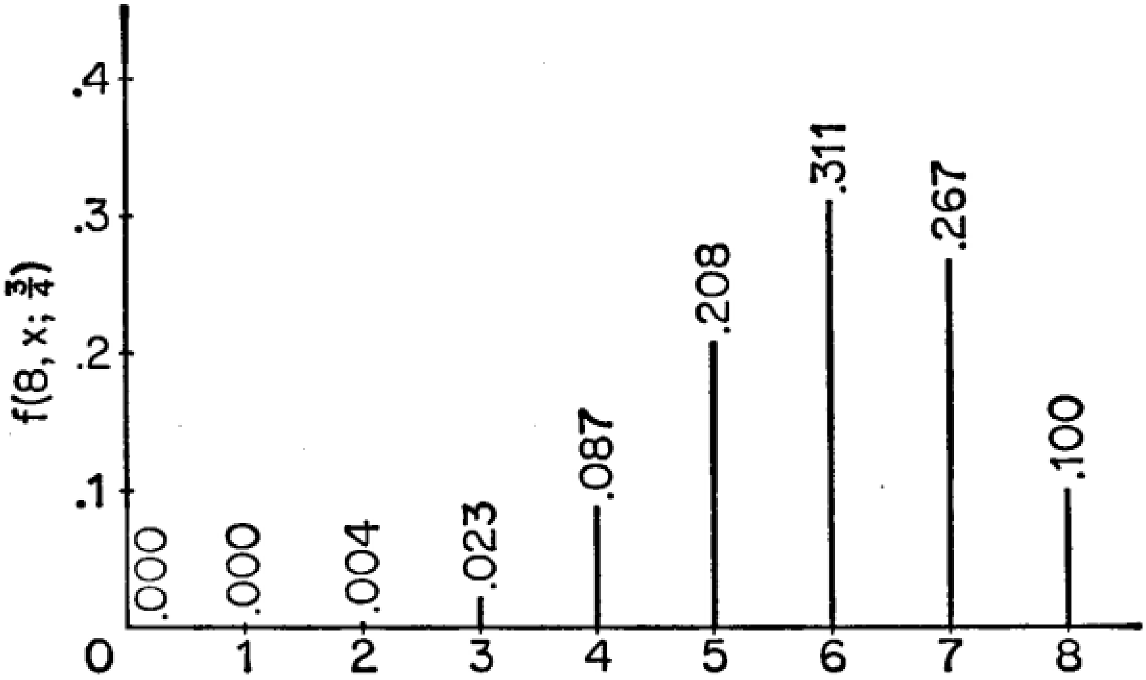


Figure 4.16: ♢

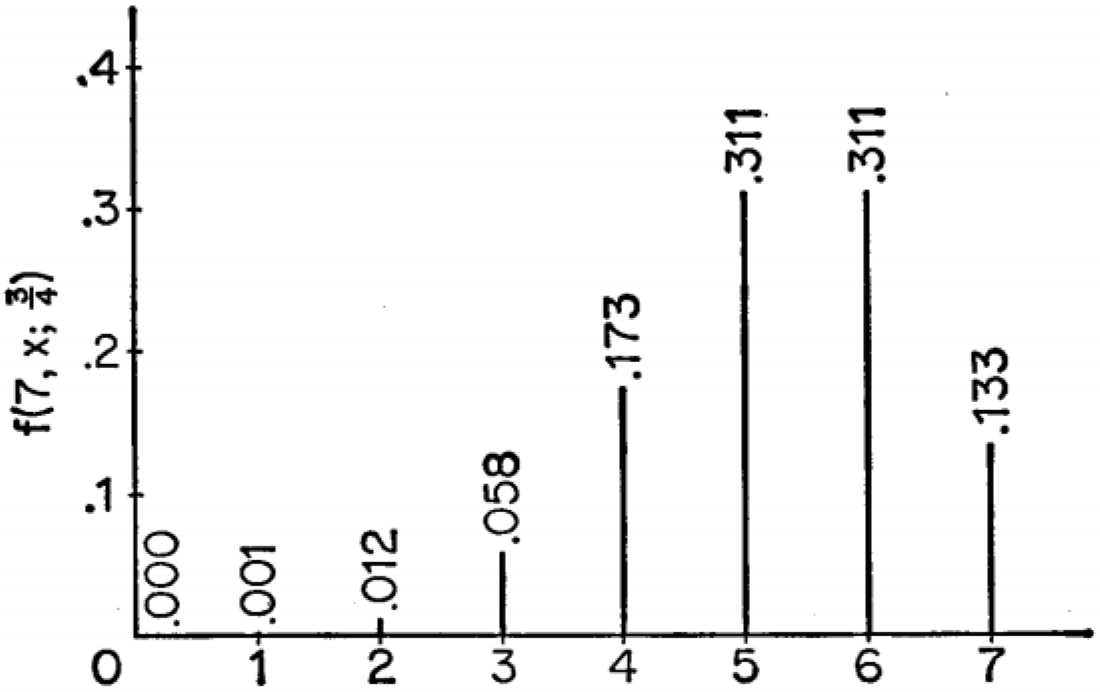


Figure 4.17: $\diamondsuit$

then decrease. These two cases are typical of what can happen in general.

Consider the ratio of the probability of $x+1$ successes in n trials to the probability of x successes in n trials, which is

$$\frac{\binom{n}{x+1}p^{x+1}q^{n-x-1}}{\binom{n}{x}p^x q^{n-x}} = \frac{n-x}{x+1}\cdot\frac{p}{q}.$$

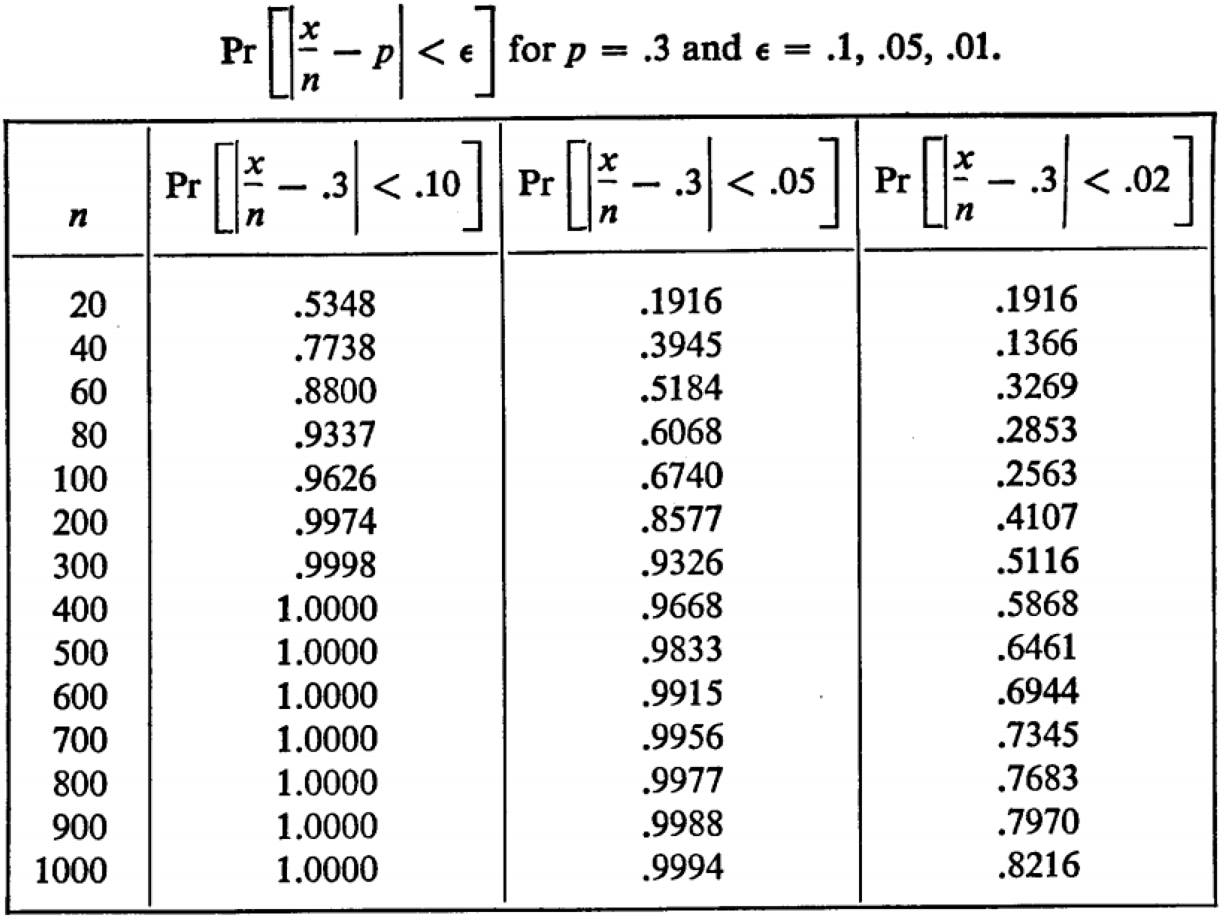
This ratio will be greater than one as long as $(n-x)p > (x+1)q$ or as long as $x < np - q$. If $np - q$ is not an integer, the values $\binom{n}{x}p^x q^{n-x}$ increase up to a maximum value, which occurs at the first integer greater than $np - q$, and then decrease. In case $np - q$ is an integer, the values $\binom{n}{x}p^x q^{n-x}$ increase up to $x = np - q$, are the same for $x = np - q$ and $x = np - q + 1$, and then decrease.

Thus we see that, in general, values near np will occur with the largest probability. It is not true that one particular value near np is highly likely to occur, but only that it is relatively more likely than a value further from np. For example, in 100 throws of a coin, $np = 100 \cdot \frac{1}{2} = 50$. The probability of exactly 50 heads is approximately .08. The probability of exactly 30 is approximately .00002.

More information is obtained by studying the probability of a given deviation of the proportion of successes x/n from the number p; that is, by studying for $\epsilon > 0$,

$$\Pr[|\frac{x}{n} - p| < \epsilon].$$

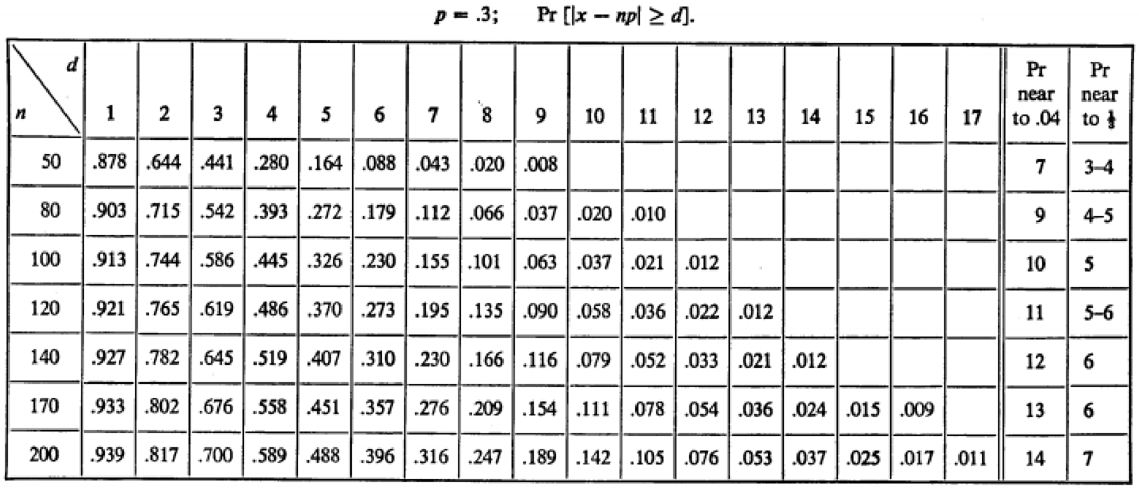
For any fixed n, p, and ϵ, the latter probability can be found by adding all the values of $f(n, x; p)$ for values of x for which the inequality $p - \epsilon < x/n < p + \epsilon$ is true. In Figure 4.18 we have given these probabilities for the case $p = .3$ with various values for ϵ and n. In the first column we have the case $\epsilon = .1$. We observe that as n increases, the probability that the fraction of successes deviates from .3 by less than .1 tends to the value 1. In fact to four decimal places the answer is 1 after $n = 400$. In column two we have the same probabilities for the smaller value of $\epsilon = .05$. Again the probabilities are tending to 1 but not so fast. In the third column we have given these probabilities for the case $\epsilon = .02$. We see now that even after 1000 trials there is still a reasonable chance that the fraction x/n is not within .02 of the value of $p = .3$. It is natural to ask if we can expect these probabilities also to tend to 1 if we increase n sufficiently. The answer is yes and this is



$\Pr\left[\left|\frac{x}{n} - p\right| < \epsilon\right]$ **for** $p = .3$ **and** $\epsilon = .1, .05, .01.$

n	$\Pr\left[\left\|\frac{x}{n} - .3\right\| < .10\right]$	$\Pr\left[\left\|\frac{x}{n} - .3\right\| < .05\right]$	$\Pr\left[\left\|\frac{x}{n} - .3\right\| < .02\right]$
20	.5348	.1916	.1916
40	.7738	.3945	.1366
60	.8800	.5184	.3269
80	.9337	.6068	.2853
100	.9626	.6740	.2563
200	.9974	.8577	.4107
300	.9998	.9326	.5116
400	1.0000	.9668	.5868
500	1.0000	.9833	.6461
600	1.0000	.9915	.6944
700	1.0000	.9956	.7345
800	1.0000	.9977	.7683
900	1.0000	.9988	.7970
1000	1.0000	.9994	.8216

Figure 4.18: ◊



$p = .3$; $\Pr[|x - np| \geq d]$.

n \ d	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Pr near to .04	Pr near to $\frac{1}{3}$
50	.878	.644	.441	.280	.164	.088	.043	.020	.008									7	3–4
80	.903	.715	.542	.393	.272	.179	.112	.066	.037	.020	.010							9	4–5
100	.913	.744	.586	.445	.326	.230	.155	.101	.063	.037	.021	.012						10	5
120	.921	.765	.619	.486	.370	.273	.195	.135	.090	.058	.036	.022	.012					11	5–6
140	.927	.782	.645	.519	.407	.310	.230	.166	.116	.079	.052	.033	.021	.012				12	6
170	.933	.802	.676	.558	.451	.357	.276	.209	.154	.111	.078	.054	.036	.024	.015	.009		13	6
200	.939	.817	.700	.589	.488	.396	.316	.247	.189	.142	.105	.076	.053	.037	.025	.017	.011	14	7

Figure 4.19: ◊

assured by one of the fundamental theorems of probability called the *law of large numbers.* This theorem asserts that, for any $\epsilon > 0$,

$$\Pr[|\frac{x}{n} - p| < \epsilon]$$

tends to 1 as n increases indefinitely.

It is important to understand what this theorem says and what it does not say. Let us illustrate its meaning in the case of coin tossing. We are going to toss a coin n times and we want the probability to be very high, say greater than .99, that the fraction of heads which turn up will be very close, say within .001 of the value .5. The law of large numbers assures us that we can have this if we simply choose n large enough. The theorem itself gives us no information about how large n must be. Let us however consider this question.

To say that the fraction of the times success results is near p is the same as saying that the actual number of successes x does not deviate too much from the expected number np. To see the kind of deviations which might be expected we can study the value of $\Pr[|x - np| \geq d]$. A table of these values for $p = .3$ and various values of n and d are given in Figure 4.19. Let us ask how large d must be before a deviation as large as d could be considered surprising. For example, let us see for each n the value of d which makes $\Pr[|x - np| \geq d]$ about .04. From the table, we see that d should be 7 for $n = 50$, 9 for $n = 80$, 10 for $n = 100$, etc. To see deviations which might be considered more typical we look for the values of d which make $\Pr[|x - np| \geq d]$ approximately $\frac{1}{3}$. Again from the table, we see that d should be 3 or 4 for $n = 50$, 4 or 5 for $n = 80$, 5 for $n = 100$, etc. The answers to these two questions are given in the last two columns of the table. An examination of these numbers shows us that deviations which we would consider surprising are approximately $\sqrt{n}$ while those which are more typical are about one-half as large or $\sqrt{n}/2$.

This suggests that $\sqrt{n}$, or a suitable multiple of it, might be taken as a unit of measurement for deviations. Of course, we would also have to study how $\Pr[|x - np| \geq d]$ depends on p. When this is done, one finds that $\sqrt{npq}$ is a natural unit; it is called a *standard deviation*. It can be shown that for large n the following approximations hold.

$$\Pr[|x - np| \geq \sqrt{npq}] \approx .3174$$

$$\Pr[|x - np| \geq 2\sqrt{npq}] \approx .0455$$

$$\Pr[|x - np| \geq 3\sqrt{npq}] \approx .0027$$

That is, a deviation from the expected value of one standard deviation is rather typical, while a deviation of as much as two standard deviations is quite surprising and three very surprising. For values of p not too near 0 or 1, the value of $\sqrt{pq}$ is approximately $\frac{1}{2}$. Thus these approximations are consistent with the results we observed from our table.

For large n, $\Pr[x - np \geq k\sqrt{npq}]$ or $\Pr[x - np \leq -k\sqrt{npq}]$ can be shown to be approximately the same. Hence these probabilities can be estimated for $k = 1, 2, 3$ by taking $\frac{1}{2}$ the values given above.

Example 4.19 In throwing an ordinary coin 10,000 times, the expected number of heads is 5000, and the standard deviation for the number of heads is $\sqrt{10,000(\frac{1}{2})(\frac{1}{2})} = 50$. Thus the probability that the number of heads which turn up deviates from 5000 by as much as one standard deviation, or 50, is approximately .317. The probability of a deviation of as much as two standard deviations, or 100, is approximately .046. The probability of a deviation of as much as three standard deviations, or 150, is approximately .003. ◇

Example 4.20 Assume that in a certain large city, 900 people are chosen at random and asked if they favor a certain proposal. Of the 900 asked, 550 say they favor the proposal and 350 are opposed. If, in fact, the people in the city are equally divided on the issue, would it be unlikely that such a large majority would be obtained in a sample of 900 of the citizens? If the people were equally divided, we would assume that the 900 people asked would form an independent trials process with probability $\frac{1}{2}$ for a “yes” answer and $\frac{1}{2}$ for a “no” answer. Then the standard deviation for the number of “yes” answers in 900 trials is $\sqrt{900(\frac{1}{2})(\frac{1}{2})} = 15$. Then it would be very unlikely that we would obtain a deviation of more than 45 from the expected number of 450. The fact that the deviation in the sample from the expected number was 100, then, is evidence that the hypothesis that the voters were equally divided is incorrect. The assumption that the true proportion is any value less than $\frac{1}{2}$ would also lead to the fact that a number as large as 550 favoring in a sample of 900 is very unlikely. Thus we are led to suspect that the true proportion is greater than $\frac{1}{2}$. On the other hand, if the number who favored the proposal in the sample of 900 were 465, we would have only a deviation of one standard deviation, under the assumption of an equal division of opinion. Since such a deviation is not unlikely, we could not rule out this possibility on the evidence of the sample. $\diamond$

Example 4.21 A certain Ivy League college would like to admit 800 students in their freshman class. Experience has shown that if they admit 1250 students they will have acceptances from approximately 800. If they admit as many as 50 too many students they will have to provide additional dormitory space. Let us find the probability that this will happen assuming that the acceptances of the students can be considered to be an independent trials process. We take as our estimate for the probability of an acceptance $p = \frac{800}{1250}$. Then the expected number of acceptances is 800 and the standard deviation for the number of acceptances is $\sqrt{1250 \cdot .64 \cdot .36} \approx 17$. The probability that the number accepted is three standard deviations or 51 from the mean is approximately .0027. This probability takes into account a deviation above the mean or below the mean. Since in this case we are only interested in a deviation above the mean, the probability we desire is half of this or approximately .0013. Thus we see that it is highly unlikely that the college will have to have new dormitory space under the assumptions we have made. $\diamond$

We finish this discussion of the law of large numbers with some final remarks about the interpretation of this important theorem.

Of course no matter how large n is we cannot prevent the coin from coming up heads every time. If this were the case we would observe a fraction of heads equal to 1. However, this is not inconsistent with the theorem, since the probability of this happening is $(\frac{1}{2})^n$ which tends to

0 as n increases. Thus a fraction of 1 is always possible, but becomes increasingly unlikely.

The law of large numbers is often misinterpreted in the following manner. Suppose that we plan to toss the coin 1000 times and after 500 tosses we have already obtained 400 heads. Then we must obtain less than one-half heads in the remaining 500 tosses to have the fraction come out near $\frac{1}{2}$. It is tempting to argue that the coin therefore owes us some tails and it is more likely that tails will occur in the last 500 tosses. Of course this is nonsense, since the coin has no memory. The point is that something very unlikely has already happened in the first 500 tosses. The final result can therefore also be expected to be a result not predicted before the tossing began.

We could also argue that perhaps the coin is a biased coin but this would make us predict more heads than tails in the future. Thus the law of averages, or the law of large numbers, should not give you great comfort if you have had a series of very bad hands dealt you in your last 100 poker hands. If the dealing is fair, you have the same chance as ever of getting a good hand.

Early attempts to define the probability p that success occurs on a single experiment sounded like this. If the experiment is repeated indefinitely, the fraction of successes obtained will tend to a number p, and this number p is called the probability of success on a single experiment. While this fails to be satisfactory as a definition of probability, the law of large numbers captures the spirit of this frequency concept of probability.

Exercises

1. If an ordinary die is thrown 20 times, what is the expected number of times that a six will turn up? What is the standard deviation for the number of sixes that turn up?

   [Ans. $\frac{10}{3}$; $\frac{5}{3}$.]

2. Suppose that an ordinary die is thrown 450 times. What is the expected number of throws that result in either a three or a four? What is the standard deviation for the number of such throws?

3. In 16 tosses of an ordinary coin, what is the expected number of heads that turn up? What is the standard deviation for the number of heads that occur?

4. In 16 tosses of a coin, find the exact probability that the number of heads that turn up differs from the expected number by (a) as much as one standard deviation, and (b) by more than one standard deviation. Do the same for the case of two standard deviations, and for the case of three standard deviations. Show that the approximations given for large n lie between the values obtained, but are not very accurate for so small an n.

   [Ans. .454; .210; .077; .021; .004; .001.]

5. Consider n independent trials with probability p for success. Let r and s be numbers such that $p < r < s$. What does the law of large numbers say about $\Pr[r < \frac{x}{n} < s]$ as we increase n indefinitely? Answer the same question in the case that $r < p < s$.

6. A drug is known to be effective in 20 per cent of the cases where it is used. A new agent is introduced, and in the next 900 times the drug is used it is effective 250 times. What can be said about the effectiveness of the drug?

7. In a large number of independent trials with probability p for success, what is the approximate probability that the number of successes will deviate from the expected number by more than one standard deviation but less than two standard deviations?

   [Ans. .272.]

8. What is the approximate probability that, in 10,000 throws of an ordinary coin, the number of heads which turn up lies between 4850 and 5150? What is the probability that the number of heads lies in the same interval, given that in the first 1900 throws there were 1600 heads?

9. Suppose that it is desired that the probability be approximately .95 that the fraction of sixes that turn up when a die is thrown n times does not deviate by more than .01 from the value $\frac{1}{6}$. How large should n be?

   [Ans. Approximately 5555.]

10. Suppose that for each roll of a fair die you lose \$1 when an odd number comes up and win \$1 when an even number comes up. Then after 10,000 rolls you can, with approximately 84 per cent confidence, expect to have lost not more than \$(how much?).

11. Assume that 10 per cent of the people in a certain city have cancer. If 900 people are selected at random from the city, what is the expected number which will have cancer? What is the standard deviation? What is the approximate probability that more than 108 of the 900 chosen have cancer?

[Ans. 90;9;.023.]

12. Suppose that in Exercise 11, the 900 people are chosen at random from those people in the city who smoke. Under the hypothesis that smoking has no effect on the incidence of cancer, what is the expected number in the 900 chosen that have cancer? Suppose that more than 120 of the 900 chosen have cancer, what might be said concerning the hypothesis that smoking has no effect on the incidence of cancer?

13. In Example 4.20, we made the assumption in our calculations that, if the true proportion of voters in favor of the proposal were p, then the 900 people chosen at random represented an independent trials process with probability p for a "yes" answer, and $1-p$ for a "no" answer. Give a method for choosing the 900 people which would make this a reasonable assumption. Criticize the following methods.

    (a) Choose the first 900 people in the list of registered Republicans.

    (b) Choose 900 names at random from the telephone book.

    (c) Choose 900 houses at random and ask one person from each house, the houses being visited in the mid-morning.

14. For n throws of an ordinary coin, let t_n be such that

$$\Pr[-t_n < \frac{x}{n} - \frac{1}{2} < t_n] = .997,$$

where x is the number of heads that turn up. Find t_n for $n = 10^4$, $n = 10^6$, and $n = 10^{20}$.

[Ans. .015; .0015; .000,000,000,15.]

15. Assume that a calculating machine carries out a million operations to solve a certain problem. In each operation the machine gives the answer 10^{-5} too small, with probability $\frac{1}{2}$, and 10^{-5} too large, with probability $\frac{1}{2}$. Assume that the errors are independent of one another. What is a reasonable accuracy to attach to the answer? What if the machine carries out 10^{10} operations?

[Ans. $\pm .01; \pm 1$.]

16. A computer tosses a coin 1 million times, and obtains 499,588 heads. Is this number reasonable?

4.11 Independent trials with more than two outcomes

By extending the results of Section 4.8, we shall study the case of independent trials in which we allow more than two outcomes. We assume that we have an independent trials process where the possible outcomes are $a_1, a_2, \ldots, a_k$, occurring with probabilities $p_1, p_2, \ldots, p_k$, respectively. We denote by

$$f(r_1, r_2, \ldots, r_k; p_1, p_2, \ldots, p_k)$$

the probability that, in $n = r_1 + r_2 + \ldots + r_k$ such trials, there will be r_1 occurrences of a_1, r_2 occurrences of a_2, etc. In the case of two outcomes this notation would be $f(r_1, r_2; p_1, p_2)$. In Section 4.8 we wrote this as $f(n, r+1; p)$ since r_2 and p_2 are determined from n, r_1, and p_1. We shall indicate how this probability is found in general, but carry out the details only for a special case. We choose $k = 3$, and $n = 5$ for purposes of illustration. We shall find $f(1, 2, 2; p_1, p_2, p_3)$.

We show in Figure 4.20 enough of the tree for this process to indicate the branch probabilities for a path (heavy lined) corresponding to the outcomes a_2, a_3, a_1, a_2, a_3. The tree measure assigns weight $p_2 \cdot p_3 \cdot p_1 \cdot p_2 \cdot p_3 = p_1 \cdot p_2^2 \cdot p_3^2$ to this path.

There are, of course, other paths through the tree corresponding to one occurrence of a_1, two of a_2, and two of a_3. However, they would all be assigned the same weight $p_1 \cdot p_2^2 \cdot p_3^2$, by the tree measure. Hence to

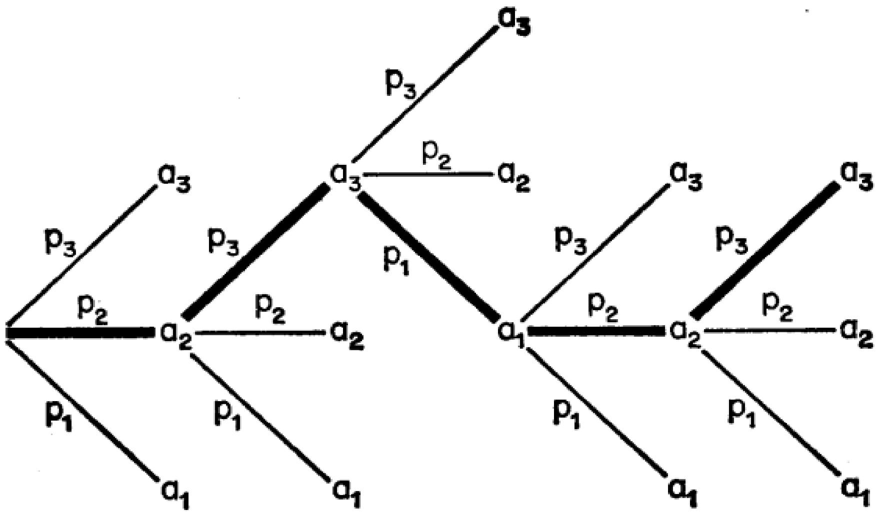


Figure 4.20: ◇

find $f(l, 2, 2; p_1, p_2, p_3)$ we must multiply this weight by the number of paths having the specified number of occurrences of each outcome.

We note that the path a_2, a_3, a_1, a_2, a_3 can be specified by the three-cell partition $[\{3\}, \{1, 4\}, \{2, 5\}]$ of the numbers from 1 to 5. Here the first cell shows the experiment which resulted in a_1, the second cell shows the two that resulted in a_2, and the third shows the two that resulted in a_3. Conversely, any such partition of the numbers from 1 to 5 with one element in the first cell, two in the second, and two in the third corresponds to a unique path of the desired kind. Hence the number of paths is the number of such partitions. But this is

$$\binom{5}{1,2,2} = \frac{5!}{1!2!2!}$$

(see 3.4), so that the probability of one occurrence of a_1, two of a_2, and two of a_3 is

$$\binom{5}{1,2,2} \cdot p_1 \cdot p_2^2 \cdot p_3^2 .$$

The above argument carried out in general leads, for the case of independent trials with outcomes $a_1, a_2, \ldots, a_k$ occurring with probabilities $p_1, p_2, \ldots, p_k$, to the following.

The probability for r_1 occurrences of a_1, r_2 occurrences of a_2, etc., is given by

$$f(r_1, r_2, \ldots, r_k; p_1, p_2, \ldots, p_k) = \binom{n}{r_1, r_2, \ldots, r_k} p_1^{r_1} \cdot p_2^{r_2} \cdot \ldots \cdot p_k^{r_k} .$$

Example 4.22 A die is thrown 12 times. What is the probability that each number will come up twice? Here there are six outcomes, 1, 2, 3, 4, 5, 6 corresponding to the six sides of the die. We assign each outcome probability $\frac{1}{6}$. We are then asked for

$$f(2,2,2,2,2,2;\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6})$$

which is

$$\binom{12}{2,2,2,2,2,2}(\frac{1}{6})^2(\frac{1}{6})^2(\frac{1}{6})^2(\frac{1}{6})^2(\frac{1}{6})^2(\frac{1}{6})^2 = .0034.$$

◇

Example 4.23 Suppose that we have an independent trials process with four outcomes a_1, a_2, a_3, a_4 occurring with probability p_1, p_2. p_3, p_4, respectively. It might be that we are interested only in the probability that r_1 occurrences of a_1 and r_2 occurrences of a_2 will take place with no specification about the number of each of the other possible outcomes. To answer this question we simply consider a new experiment where the outcomes are $a_1, a_2, \bar{a_3}$. Here $\bar{a_3}$ corresponds to an occurrence of either a_3 or a_4 in our original experiment. The corresponding probabilities would be p_1, p_2 and $\bar{p_3}$ with $\bar{p_3} = p_3 + p_4$. Let $\bar{r_3} = n - (r_1 + r_2)$ Then our question is answered by finding the probability in our new experiment for r_1 occurrences of a_1, r_2 of a_2, and $\bar{r_3}$ of $\bar{a_3}$, which is

$$\binom{n}{r_1, r_2, \bar{r_3}} p_1^{r_1} \cdot p_2^{r_2} \cdot \bar{p_3}^{\bar{r_3}}.$$

◇

The same procedure can be carried out for experiments with any number of outcomes where we specify the number of occurrences of such particular outcomes. For example, if a die is thrown ten times the probability that a one will occur exactly twice and a three exactly three times is given by

$$\binom{10}{2,3,5}(\frac{1}{6})^2(\frac{1}{6})^2(\frac{4}{6})^5 = .043.$$

Exercises

1. Suppose that in a city 60 per cent of the population are Democrats, 30 per cent are Republicans, and 10 per cent are Independents. What is the probability that if three people are chosen at random there will be one Republican, one Democrat, and one Independent voter?

   [Ans. .108.]

2. Three horses, A, B, and C, compete in four races. Assuming that each horse has an equal chance in each race, what is the probability that A wins two races and B and C win one each? What is the probability that the same horse wins all four races?

   [Ans. $\frac{4}{27}$; $\frac{1}{27}$.]

3. Assume that in a certain large college 40 per cent of the students are freshmen, 30 per cent are sophomores, 20 per cent are juniors, and 10 per cent are seniors. A committee of eight is chosen at random from the student body. What is the probability that there are equal numbers from each class on the committee?

4. Let us assume that when a batter comes to bat, he or she has probability .6 of being put out, .1 of getting a walk, .2 of getting a single, .1 of getting an extra base hit. If he or she comes to bat five times in a game, what is the probability that

   (a) He gets two walks and three singles?

   [Ans. .0008.]

   (b) He gets a walk, a single, an extra base hit (and is out twice)?

   [Ans. .043.]

   (c) He has a perfect day (i.e., never out)?

   [Ans. .010.]

5. Assume that a single torpedo has a probability $\frac{1}{2}$ of sinking a ship, probability $\frac{1}{4}$ of damaging it, and probability $\frac{1}{4}$ of missing. Assume further that two damaging shots sink the ship. What is the probability that four torpedos will succeed in sinking the ship?

[Ans. $\frac{251}{256}$.]

6. Jones, Smith, and Green live in the same house. The mailman has observed that Jones and Smith receive the same amount of mail on the average, but that Green receives twice as much as Jones (and hence also twice as much as Smith). If he or she has four letters for this house, what is the probability that each resident receives at least one letter?

7. If three dice are thrown, find the probability that there is one six and two fives, given that all the outcomes are greater than three.

   [Ans. $\frac{1}{9}$.]

8. An athlete plays a tournament consisting of three games. In each game he or she has probability $\frac{1}{2}$ for a win, $\frac{1}{4}$ for a loss, and $\frac{1}{4}$ for a draw, independently of the outcomes of other games. To win the tournament he or she must win more games than he or she loses. What is the probability that he or she wins the tournament?

9. Assume that in a certain course the probability that a student chosen at random will get an A is .1, that he or she will get a B is .2, that he or she will get a C is .4, that he or she will get a D is .2, and that he or she will get an F is .1. What distribution of grades is most likely in the case of four students?

   [Ans. One B, two C's, one D.]

10. Let us assume that in a World Series game a batter has probability $\frac{1}{4}$ of getting no hits, $\frac{1}{2}$ for getting one hit, $\frac{1}{4}$ for getting two hits, assuming that the probability of getting more than two hits is negligible. In a four-game World Series, find the probability that the batter gets

    (a) Exactly two hits.

    [Ans. $\frac{7}{64}$.]

    (b) Exactly three hits.

    [Ans. $\frac{7}{32}$.]

    (c) Exactly four hits.

[Ans. $\frac{35}{128}$.]

(d) Exactly five hits.

[Ans. $\frac{7}{32}$.]

(e) Fewer than two hits or more than five.

[Ans. $\frac{23}{128}$.]

11. Gypsies sometimes toss a thick coin for which heads and tails are equally likely, but which also has probability $\frac{1}{5}$ of standing on edge (i.e., neither heads nor tails). What is the probability of exactly one head and four tails in five tosses of a gypsy coin?

12. A family car is driven by the father, two sons, and the mother. The fenders have been dented four times, three times while the mother was driving. Is it fair to say that the mother is a worse driver than the men?