2 Repeated Trials and Sampling

This chapter studies a mathematical model for repeated trials, each of which may result in some event either happening or not happening. Occurrence of the event is called *success*, and non-occurrence called *failure*. For instance:

Nature of trial	Meaning of success	Meaning of failure	Probabilities p and q
Tossing a fair coin	head	tail	$1/2$ and $1/2$
Rolling a die	six	not six	$1/6$ and $5/6$
Rolling a pair of dice	double six	not double six	$1/36$ and $35/36$
Birth of a child	girl	boy	0.487 and 0.513

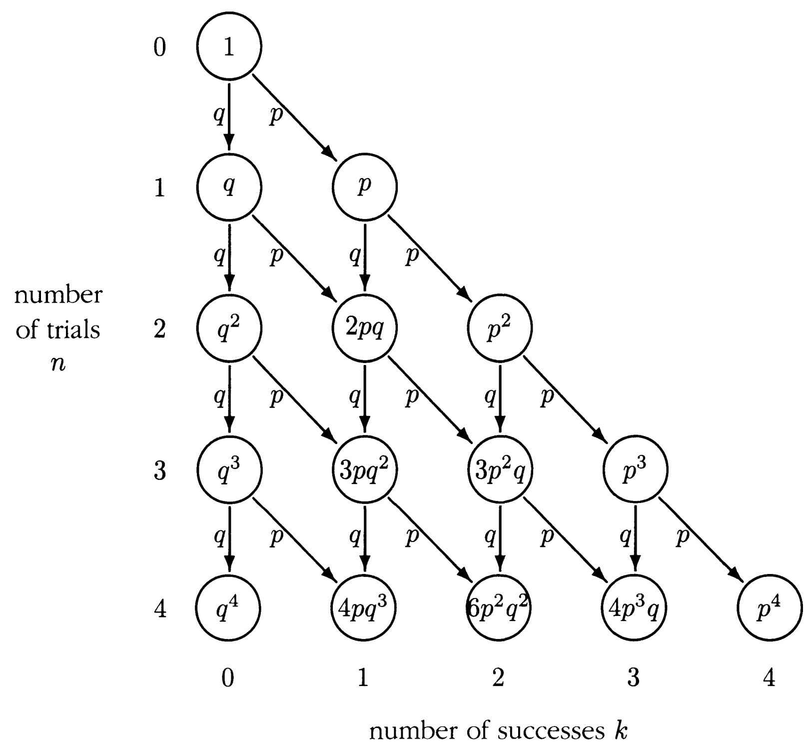
Suppose that on each trial there is success with probability p, failure with probability $q = 1-p$, and assume the trials are independent. Such trials are called *Bernoulli trials* or *Bernoulli* (p) *trials* to indicate the success probability p. The number of successes in n trials then cannot be predicted exactly. But if n is large we expect the number of successes to be about np, so the relative frequency of successes will, most likely, be close to p. The important questions treated in this chapter are: how likely? and how close? The answers to these questions, first discovered by the mathematicians James Bernoulli and Abraham De Moivre, around 1700, are the mathematical basis of the long-run frequency interpretation of probabilities.

The first step in Section 2.1 is to find a formula for the probability of getting k successes in n trials. This formula defines the *binomial probability distribution* over the possible numbers of successes from 0 to n. For large values of n, the histogram of the distribution turns out to follow a smooth curve quite closely.

2.1 The Binomial Distribution

The problem is to find a formula for the probability of getting k successes in n independent trials. This is solved by analysis of a tree diagram representing all possible results of the n trials, shown in Figure 1 for $n = 4$.

FIGURE 1. Tree diagram for derivation of the binomial distribution.



Each path down n steps through the tree diagram represents a possible outcome of the first n trials. The kth *node* in the nth row represents the event of k successes in n trials. The expression inside each node is its probability in terms of p and $1 - p = q$ (the probabilities of success and failure on each trial). This expression is the sum of the probabilities of all paths leading to this node. For example, in row 3 the probabilities of $k = 0, 1, 2, 3$ successes in $n = 3$ trials are the terms in the expansion

$$(p + q)^3 = q^3 + 3pq^2 + 3p^2q + p^3$$

For $k = 0$ or 3 there is only one path leading to k successes, hence the probability of q^3 or p^3 by the multiplication rule. For $k = 1$ the factor of 3 arises because there are three ways to get just one success in three trials, FFS, FSF, SFF, represented by the three paths through the diagram leading to the first node in row 3. The probabilities of these events are the terms qqp, qpq, and pqq in the expansion of $(q+p)^3$. These terms add to give the probability $3pq^2$ of $k=1$ success in 3 trials. Similarly, the probability of $k=2$ successes in 3 trials is $3p^2q$.

The tree diagram can be imagined drawn down to any number of trials n. To achieve k successes in n trials, the path must move down to the right k times, corresponding to the k successes, and straight down $n-k$ times, corresponding to the $n-k$ failures. The probability of every such path is the product of k factors of p, and $n-k$ factors of q, which is $p^k q^{n-k}$, regardless of the order of the factors. Therefore, the probability of k successes in the n trials is the sum of as many equal contributions of $p^k q^{n-k}$ as there are paths down through the diagram leading to the kth node of row n, or this number of paths times $p^k q^{n-k}$. This number of paths is denoted $\binom{n}{k}$ and called *n choose k*. So the probability of k successes in n trials is $\binom{n}{k}p^k q^{n-k}$. This conclusion and a formula for $\binom{n}{k}$ are summarized in the next box.

Binomial Distribution

For n independent trials, with probability p of success and probability $q=1-p$ of failure on each trial, the probability of k successes is given by the *binomial probability formula*:

$$P(k \text{ successes in } n \text{ trials}) = \binom{n}{k} p^k q^{n-k}$$

where $\binom{n}{k}$, called *n choose k*, is the number of different possible patterns of k successes and $n-k$ failures in n trials, given by the formula

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1} = \frac{n!}{k!(n-k)!}$$

Here the $k!$ is *k factorial*, the product of the first k integers for $k \geq 1$, and $0! = 1$. For fixed n and p, as k varies, these binomial probabilities define a probability distribution over the set of $n+1$ integers $\{0, 1, \ldots, n\}$, called the *binomial (n,p) distribution*. This is the distribution of the number of successes in n independent trials, with probability p of success in each trial. The binomial (n,p) probabilities are the terms in the *binomial expansion*:

$$(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$$

Appendix 1 gives the background on counting and a derivation of the formula for $\binom{n}{k}$ in the box. The first expression for $\binom{n}{k}$ in the box is the simplest to use for numerical evaluations if $k < \frac{1}{2}n$. For example,

$$\binom{8}{3} = \frac{8 \times 7 \times 6}{3 \times 2 \times 1} = 8 \times 7 = 56$$

In this expression for $\binom{n}{k}$ there are always k factors in both the numerator and denominator. If $k > \frac{1}{2}n$, needless cancellation is avoided by first using symmetry:

$$\binom{n}{k} = \binom{n}{n-k}$$

as you can easily check. For instance, $\binom{9}{7} = \binom{9}{2} = \frac{9 \times 8}{2 \times 1} = 9 \times 4 = 36.$

To illustrate the binomial probability formula, the chance of getting 2 sixes and 7 non-sixes in 9 rolls of a die is therefore

$$\binom{9}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^7 = \frac{36 \times 5^7}{6^9} = 0.279$$

The convention $0! = 1$ makes the factorial formula for $\binom{n}{k}$ work even if k or n is 0. This formula is sometimes useful for algebraic manipulations. Because $n!$ increases so rapidly as a function of n, the factorial formula is awkward for numerical calculations of $\binom{n}{k}$. But for large values of n and k there are simple approximations to be described in the following sections.

The binomial expansion. Often called the *binomial theorem*, this is the expansion of $(p+q)^n$ as a sum of coefficients times powers of p and q. The coefficient $\binom{n}{k}$ of $p^k q^{n-k}$ is often called a *binomial coefficient*. For $p+q=1$ the binomial expansion of $(p+q)^n$ amounts to the fact that the probabilities in the binomial (n, p) distribution sum up to 1 over $k = 0$ to n:

$$\sum_{k=0}^{n} P(k \text{ successes in } n \text{ trials}) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = 1$$

This illustrates the addition rule for probabilities: as k varies from 0 to n, the $n+1$ events of getting, respectively,

$$0 \text{ successes, } 1 \text{ success, } 2 \text{ successes, } \ldots, n \text{ successes,}$$

in n trials, form a partition of all possible outcomes. For example, you can't get both 2 successes and 3 successes in 10 trials. And in n trials, you must get some number of successes between 0 and n.

The case of fair coin tossing. Then $p = q = 1/2$, so

$$p^k q^{n-k} = (1/2)^k (1/2)^{n-k} = (1/2)^n \qquad \text{and}$$

$$P(k \text{ heads in } n \text{ fair coin tosses}) = \binom{n}{k}/2^n \qquad (0 \le k \le n)$$

All possible patterns of heads and tails of length n are equally likely in this case. So the above probability of k heads in n tosses is just the number of such patterns with k heads, namely $\binom{n}{k}$, relative to the total number of such patterns, namely 2^n. A consequence is that

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = \sum_{k=0}^{n} \binom{n}{k} = 2^n$$

This is the binomial expansion of $(x+y)^n$ for $x = y = 1$.

Example 1. Coin tossing and sex of children.

Problem 1. Find the probability of getting four or more heads in six tosses of a fair coin.

Solution. $P(4 \text{ or more heads in 6 tosses}) = P(4) + P(5) + P(6)$, where

$$P(k) = P(k \text{ heads in 6 tosses}) = \binom{6}{k}/2^6 \qquad \text{so}$$

$$P(4 \text{ or more heads in 6 tosses}) = (15 + 6 + 1)/2^6 = 11/32$$

Problem 2. What is the probability that among five families, each with six children, at least three of the families have four or more girls?

Solution. Assume that each child in each family is equally likely to be a boy or a girl, independently of all other children. Then the chance that any particular family has four or more girls is $p = 11/32$, by the solution of the previous problem. Call this event a success in the present problem. Then the probability that at least 3 of the families have 4 or more girls is the probability of at least 3 successes in $n = 5$ trials, with probability $p = 11/32$ of success on each trial. So the required probability is

$$P(3 \text{ successes}) + P(4 \text{ successes}) + P(5 \text{ successes})$$

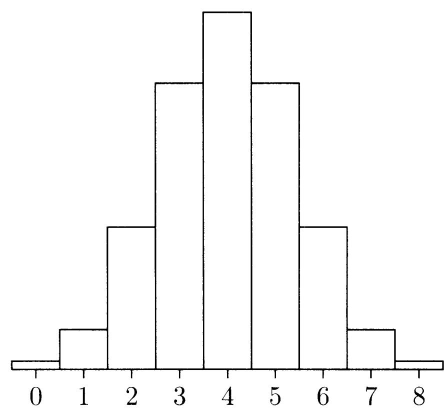
$$= \binom{5}{3}\left(\frac{11}{32}\right)^3\left(\frac{21}{32}\right)^2 + \binom{5}{4}\left(\frac{11}{32}\right)^4\left(\frac{21}{32}\right) + \binom{5}{5}\left(\frac{11}{32}\right)^5 = 0.226$$

Consecutive Odds Ratios

The binomial (n, p) distribution is most easily analyzed in terms of the chance of k successes relative to $k-1$ successes. These odds ratios are much simpler than the probabilities $P(k) = P(k \text{ successes})$. But the ratios determine the probabilities, so the whole distribution can be understood in terms of the consecutive odds ratios.

Consider first the case when $p = 1/2$. The nth row of Pascal's triangle displays the binomial $(n, 1/2)$ distribution as multiples of 2^{-n}. The numbers in this nth row first increase rapidly, then less rapidly. Then they level off, and start decreasing just as they have increased. This gives rise to the characteristic bell shape of the histogram of a symmetric binomial distribution.

FIGURE 2. **The binomial $(8, 1/2)$ distribution.** This is the distribution of the number of heads in eight fair coin tosses.



The aim now is to understand the shape of such a binomial distribution in terms of the ratio of the heights of consecutive bars. The numbers from the eighth row of Pascal's triangle are:

$$1 \quad 8 \quad 28 \quad 56 \quad 70 \quad 56 \quad 28 \quad 8 \quad 1$$

So the consecutive odds ratios are

$$\frac{8}{1} \quad \frac{28}{8} \quad \frac{56}{28} \quad \frac{70}{56} \quad \frac{56}{70} \quad \frac{28}{56} \quad \frac{8}{28} \quad \frac{1}{8}$$

which simplify to

$$\frac{8}{1} \quad \frac{7}{2} \quad \frac{6}{3} \quad \frac{5}{4} \quad \frac{4}{5} \quad \frac{3}{6} \quad \frac{2}{7} \quad \frac{1}{8}$$

So the ratios start big, and steadily decrease, crossing 1 in the middle. In the nth row of Pascal's triangle,

$$\binom{n}{0} \quad \binom{n}{1} \quad \binom{n}{2} \quad \binom{n}{3} \quad \cdots \quad \binom{n}{n-3} \quad \binom{n}{n-2} \quad \binom{n}{n-1} \quad \binom{n}{n}$$

the consecutive ratios decrease steadily as follows:

$$\frac{n}{1} \quad \frac{n-1}{2} \quad \frac{n-2}{3} \quad \cdots \quad \cdots \quad \frac{3}{n-2} \quad \frac{2}{n-1} \quad \frac{1}{n}$$

This simple pattern displays the special case $p = q = 1/2$ of the result stated in the following box:

Consecutive Odds for the Binomial Distribution

For independent trials with success probability p, the odds of k successes relative to $k-1$ successes are $R(k)$ to 1, where

$$R(k) = \frac{P(k \text{ successes in } n \text{ trials})}{P(k-1 \text{ successes in } n \text{ trials})} = \left[\frac{n-k+1}{k}\right]\frac{p}{q}$$

This follows from the binomial probability formula and the formula for $\binom{n}{k}$ by cancelling common factors. This simple formula for ratios makes it easy to calculate all the probabilities in a binomial distribution recursively.

Example 2. Computing all probabilities in a binomial distribution.

Problem 1. A pair of fair coins is tossed 8 times. Find the probability of getting both heads on k of these double tosses, for $k = 0$ to 8.

Solution. The chance of getting both heads on each double toss is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. So the required probabilities form the binomial $(8, 1/4)$ distribution. The following table shows how simply these probabilities can be found, starting with $P(0)$ and then using the consecutive odds formula with $p/q = (\frac{1}{4})/(\frac{3}{4}) = \frac{1}{3}$.

Value of k	0	1	2	3	4	5	6	7	8
How $P(k)$ found	$(\frac{3}{4})^8$	$\frac{8}{1}\frac{1}{3}P(0)$	$\frac{7}{2}\frac{1}{3}P(1)$	$\frac{6}{3}\frac{1}{3}P(2)$	$\frac{5}{4}\frac{1}{3}P(3)$	$\frac{4}{5}\frac{1}{3}P(4)$	$\frac{3}{6}\frac{1}{3}P(5)$	$\frac{2}{7}\frac{1}{3}P(6)$	$\frac{1}{8}\frac{1}{3}P(7)$
Value of $P(k)$	.100	.267	.311	.208	.087	.023	.004	.0004	.00001

Notice how the ratios from Pascal's triangle first dominate the odds against a success ratio of 3 in the denominator, as the probabilities $P(k)$ increase for $k \leq 2$. Then for $k \geq 3$ the ratios from Pascal's triangle are smaller than the odds against success, and the probabilities $P(k)$ steadily decrease. Something similar happens, no matter what the values of n and p. See Figure 3 where this binomial $(8, 1/4)$ distribution is displayed along with other binomial (n, p) distributions for $n = 1$ to 8 and selected values of p.

What is the most likely number of successes in n independent trials with probability of success p on each trial? Intuitively, we expect about proportion p of the trials to be successes . In n trials, we therefore expect around np successes. So it is reasonable to guess that the most likely number of successes m, called the *mode* of the distribution, is an integer close to np. According to the following formula, the mode differs by at most 1 from np:

Most Likely Number of Successes (Mode of Binomial Distribution)

For $0 < p < 1$, the most likely number of successes in n independent trials with probability p of success on each trial is m, the greatest integer less than or equal to $np + p$:

$$m = \text{int}\,(np + p) \qquad \text{where int denotes the integer part function.}$$

If $np + p$ is an integer, as in the case $p = 1/2$, n odd, then there are two most likely numbers, m and $m - 1$. Otherwise, there is a unique most likely number. In either case, the probabilities in the binomial (n, p) distribution are strictly increasing before they reach the maximum, and strictly decreasing after the maximum.

These features of the binomial distribution can be seen in Figure 3. Note the double maxima for $n = 3$, p a multiple of $1/4$, and $n = 7$, p a multiple of $1/8$. Check the formula in a few of these cases to see how it works.

Proof of the formula for the mode. Fix n and p, and consider the following statements about an integer k between 1 and n. Each statement may be true for some k and false for others. By manipulating inequalities and using the formula for consecutive odds, these statements (1) to (5) are logically equivalent:

$$P(k-1) \leq P(k) \tag{1}$$

$$1 \leq P(k)/P(k-1) \tag{2}$$

$$1 \leq \frac{(n-k+1)}{k}\frac{p}{1-p} \tag{3}$$

$$k(1-p) \leq (n-k+1)p \tag{4}$$

$$k \leq np + p \tag{5}$$

FIGURE 3. **Histograms of some binomial distributions.** The histogram in row n, column p shows the binomial (n, p) distribution for the number of successes in n independent trials, each with success probability p. In row n, the range of values shown is 0 to n. The horizontal scale changes from one row to the next, but equal probabilities are represented by equal areas, even in different histograms.

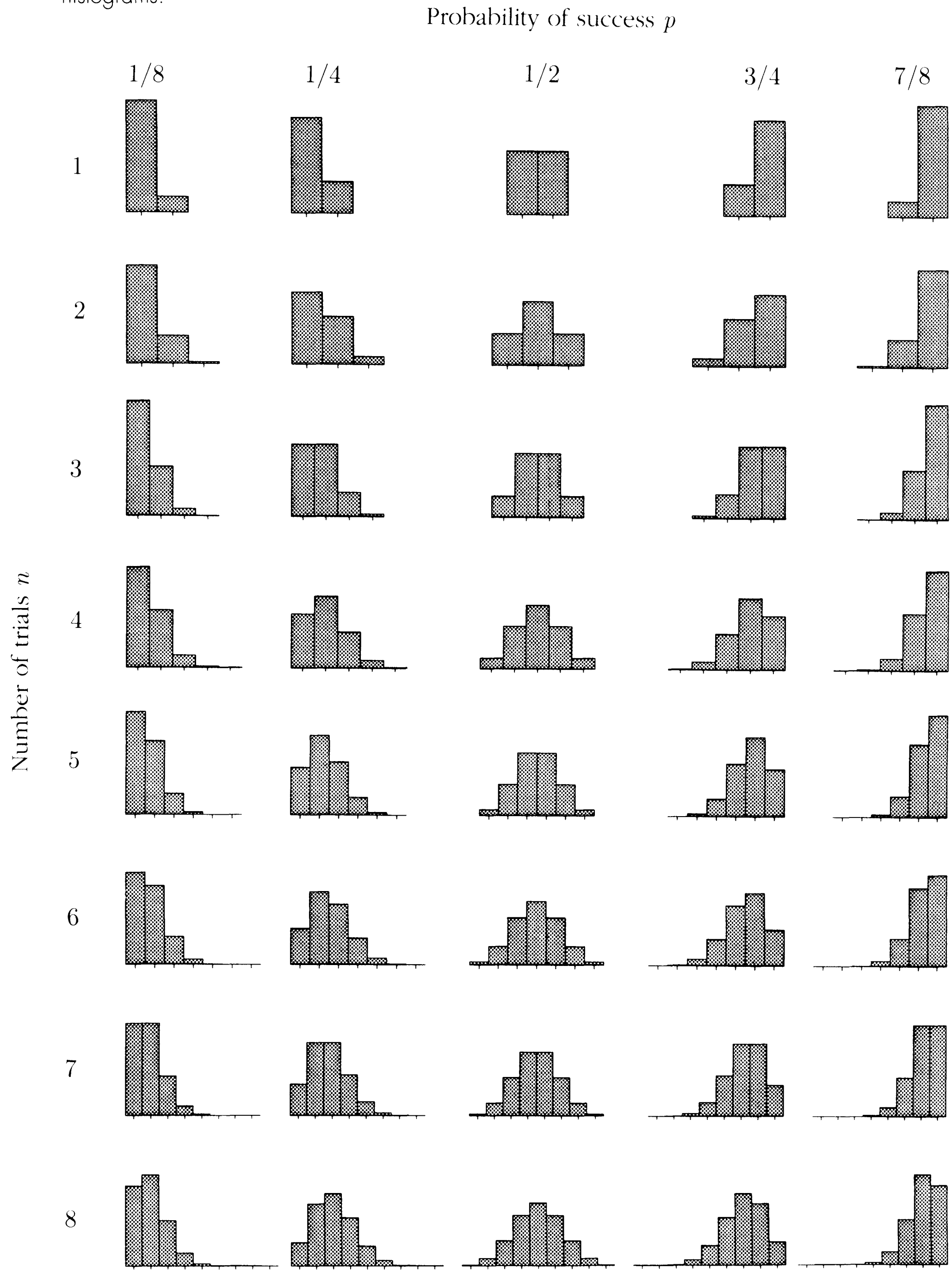


FIGURE 4. **Distribution of the number of heads in n coin tosses.** Histograms of the binomial (n, 1/2) distribution are shown for $n = 10$ to 100 by steps of 10. Each histogram is a bar graph of the probability of k successes $P(k)$ as a function of k, plotted with the same horizontal and vertical scale. Notice the following features: as n increases the distribution shifts steadily to the right, so as always to be centered on the expected number $n/2$; each distribution is symmetric about $n/2$; as n increases the distribution gradually spreads out, covering a wider range of values; still, the range of values on which the probability is concentrated becomes a smaller and smaller fraction of the whole range of possible values from 0 to n; and apart from these variations in height and width, the histograms all appear to follow the same bell-shaped curve.

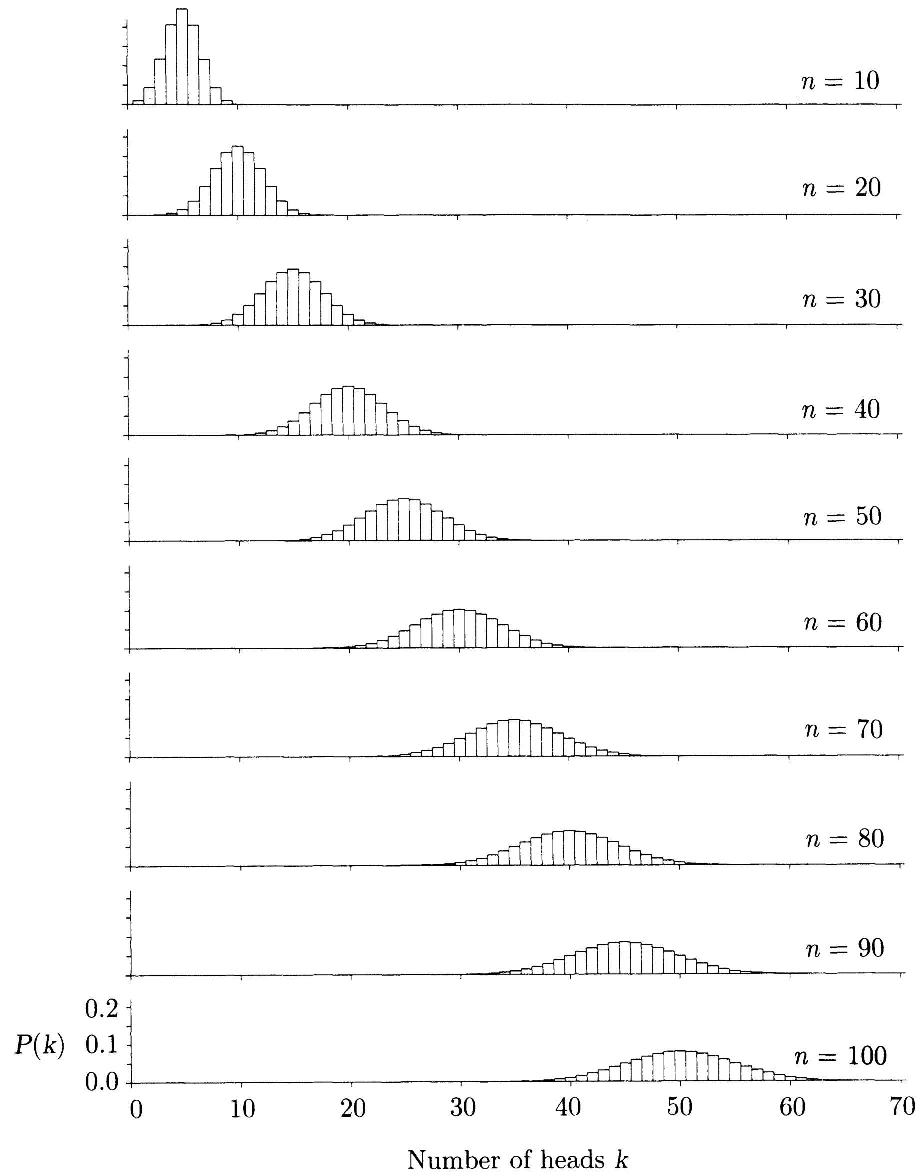
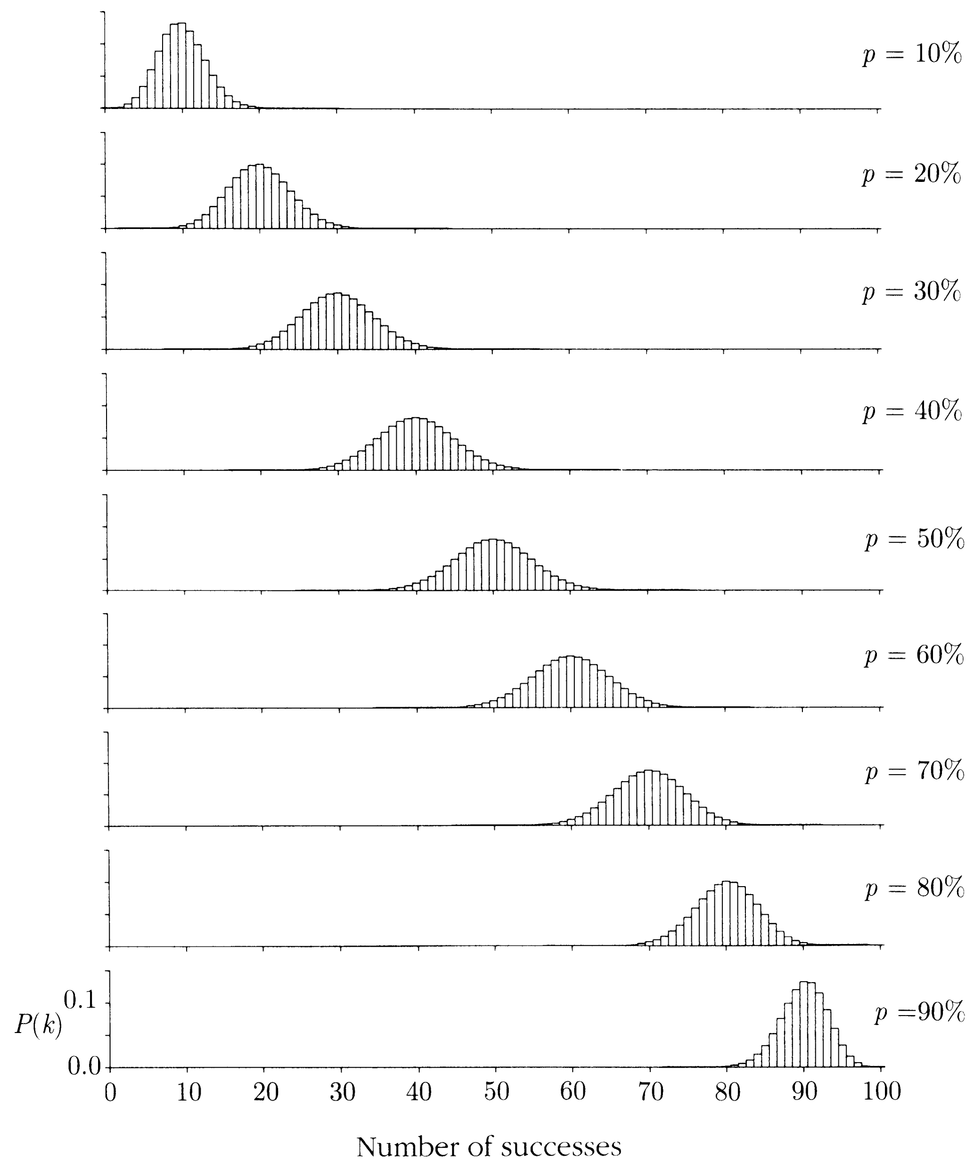


FIGURE 5. **Distribution of the number of successes in 100 trials.** Histograms of the binomial $(100, p)$ distribution are shown for $p = 10\%$ to 90% by steps of 10%. Each histogram is a bar graph of the probability of k successes $P(k)$ as a function of k, plotted with the same horizontal and vertical scale. Notice the following features: as p increases the distribution shifts steadily to the right, so as always to be centered around the expected number $100p$; the distribution is most spread out for $p = 50$; for all values of p the distribution concentrates on a range of numbers that is small in comparison to $n = 100$; and apart from these variations in height and width, and slight skewness toward the edges, the histograms all follow a symmetric bell-shaped curve quite closely.



Let m be the largest k attaining the maximum value of $P(k)$ over all $0 \le k \le n$. By definition of m, $P(m-1) \le P(m) > P(m+1)$. That is,

$$m \le np + p < m + 1$$

by the equivalence of (1) and (5) for $k = m$ and $k = m + 1$. Thus m is the greatest integer less than or equal to $np + p$. (Strictly speaking, the cases $m = 0$ and $m = n$ should be considered separately, but the conclusion is the same.) □

The mean. The number np, which is always close to the mode of the binomial distribution, is called the *expected number of successes*, or the *mean* of the binomial (n, p) distribution, usually denoted μ (Greek letter mu). In case the mean μ is an integer, it turns out that μ is the most likely number of successes. But if μ is not an integer, μ is not even a possible number of successes.

Expected Number of Successes (Mean of Binomial Distribution)

$$\mu = np$$

Remark. For the time being this formula is taken as the definition of the mean of a binomial distribution. Chapter 3 gives a more general, consistent definition.

Behavior of the binomial distribution for large n. This is displayed in the last two figures. As a general rule, for large values of n, the binomial distribution concentrates on a range of values around the expected value np which, while becoming larger on an absolute numerical scale, becomes smaller on a relative scale in comparison with n. Put another way, as n increases, it becomes harder to predict the number of successes exactly, but easier to predict the proportion of successes, which will most likely be close to p. This is made more precise by the *square root law* and the *law of large numbers,* discussed in the following sections. Apart from slight variations in height and width, and some slight skewness toward the edges, all the histograms follow a bell-shaped curve of roughly the same form. This is the famous *normal curve,* first discovered by De Moivre, around 1730, as an approximation to binomial distribution for large values of n.

Exercises 2.1

1. a) How many sequences of zeros and ones of length 7 contain exactly 4 ones and 3 zeros?

b) If you roll 7 dice, what is the chance of getting exactly 4 sixes?

2. Suppose that in 4-child families, each child is equally likely to be a boy or a girl, independently of the others. Which would then be more common, 4-child families with 2 boys and 2 girls, or 4-child families with different numbers of boys and girls? What would be the relative frequencies?

3. Suppose 5 dice are rolled. Assume they are fair and the rolls are independent. Calculate the probability of the following events:
$A =$ (exactly two sixes); $B =$ (at least two sixes); $C =$ (at most two sixes);
$D =$ (exactly three dice show 4 or greater); $E =$ (at least 3 dice show 4 or greater).

4. A die is rolled 8 times. Given that there were 3 sixes in the 8 rolls, what is the probability that there were 2 sixes in the first five rolls?

5. Given that there were 12 heads in 20 independent coin tosses, calculate

a) the chance that the first toss landed heads;

b) the chance that the first two tosses landed heads;

c) the chance that at least two of the first five tosses landed heads.

6. A man fires 8 shots at a target. Assume that the shots are independent, and each shot hits the bull's eye with probability 0.7.

a) What is the chance that he hits the bull's eye exactly 4 times?

b) Given that he hit the bull's eye at least twice, what is the chance that he hit the bull's eye exactly 4 times?

c) Given that the first two shots hit the bull's eye, what is the chance that he hits the bull's eye exactly 4 times in the 8 shots?

7. You roll a die, and I roll a die. You win if the number showing on your die is strictly greater than the one on mine. If we play this game five times, what is the chance that you win at least four times?

8. For each positive integer n, what is the largest value of p such that zero is the most likely number of successes in n independent trials with success probability p?

9. The chance of winning a bet on 00 at roulette is $1/38 = 0.026315$. In 325 bets on 00 at roulette, the chance of six wins is 0.104840. Use this fact, and consideration of odds ratios, to answer the following questions without long calculations.

a) What is the most likely number of wins in 325 bets on 00, and what is its probability?

b) Find the chance of ten wins in 325 bets on 00.

c) Find the chance of ten wins in 326 bets on 00.

10. Suppose a fair coin is tossed n times. Find simple formulae in terms of n and k for

a) $P(k-1 \text{ heads} \mid k-1 \text{ or } k \text{ heads})$;

b) $P(k \text{ heads} \mid k-1 \text{ or } k \text{ heads})$.

11. 70% of the people in a certain population are adults. A random sample of size 15 will be drawn, with replacement, from this population.

a) What is the most likely number of adults in the sample?

b) What is the chance of getting exactly this many adults?

12. A gambler decides to keep betting on red at roulette, and stop as soon as she has won a total of five bets.

a) What is the probability that she has to make exactly 8 bets before stopping?

b) What is the probability that she has to make at least 9 bets?

13. **Genetics.** Hereditary characteristics are determined by pairs of *genes*. A gene pair for a particular characteristic is transmitted from parents to offspring by choosing one gene at random from the mother's pair, and, independently, one at random from the father's. Each gene may have several forms, or *alleles*. For example, human beings have an allele (B) for brown eyes, and an allele (b) for blue eyes. A person with allele pair BB has brown eyes, and a person with allele pair bb has blue eyes. A person with allele pair Bb or bB will have brown eyes—the allele B is called *dominant* and b *recessive*. So to have blue eyes, one must have the allele pair bb. The alleles don't "mix" or "blend".

a) A brown-eyed (BB) woman and a blue-eyed man plan to have a child. Can the child have blue eyes?

b) A brown-eyed (Bb) woman and a blue-eyed man plan to have a child. Find the chance that the child has brown eyes.

c) A brown-eyed (Bb) woman and a brown-eyed (Bb) man plan to have a child. Find the chance that the child has brown eyes.

d) A brown-eyed woman has brown-eyed parents, both Bb. She and a blue-eyed man have a child. Given that the child has brown eyes, what is the chance that the woman carries the allele b?

14. **Genetics.** In certain pea plants, the allele for tallness (T) dominates over the allele for shortness (s), and the allele for purple flowers (P) dominates over the allele for white flowers (w) (see Exercise 13). According to the *principle of independent assortment*, alleles for the two characteristics (flower color and height) are chosen independently of each other.

a) A (TT, PP) plant is crossed with a (ss, ww) plant. What will the offspring look like?

b) The offspring in part a) is self-fertilized, that is, crossed with itself. Write down the possible genetic combination (of flower color and height) that the offspring of this fertilization can have, and find the chance with which each such combination occurs.

c) Ten (Ts, Pw) plants are self-fertilized, each producing a new plant. Find the chance that at least 2 of the new plants are tall with purple flowers.

15. Consider the mode m of the binomial (n, p) distribution. Use the formula $m = int\,(np + p)$ to show the following:

a) If np happens to be an integer, then $m = np$.

b) If np is not an integer, then the most likely number of successes m is one of the two integers to either side of np.

c) Show by examples that m is not necessarily the closest integer to np. Neither is m always the integer above np, nor the integer below it.

2.2 Normal Approximation: Method

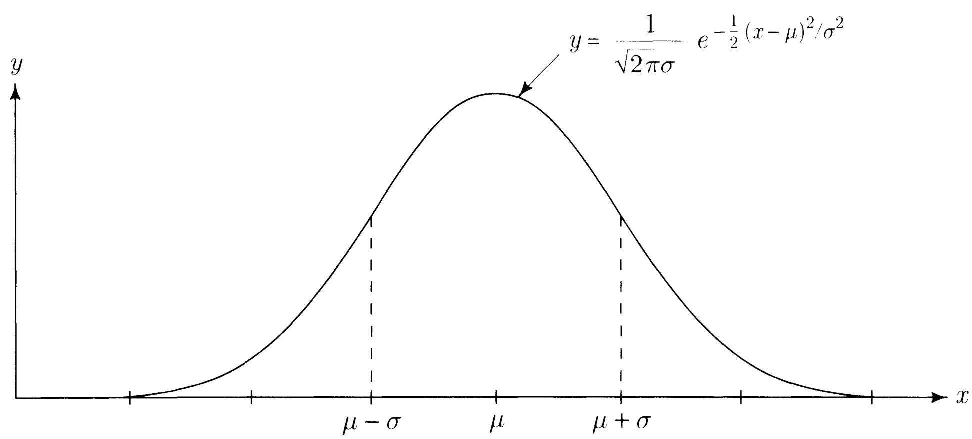
The figures of the previous section illustrate the general fact that no matter what the value of p, provided n is large enough, binomial (n, p) histograms have roughly the same bell shape. As n and p vary, the binomial (n, p) distributions differ in where they are centered, and in how spread out they are. But when the histograms are suitably scaled they all follow the same curve provided n is large enough. This section concerns the practical technique of using areas under the curve to approximate binomial probabilities. This can be understood without following the derivation of the curve in the next section.

The *normal curve* has equation

$$y = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} \qquad (-\infty < x < \infty)$$

The equation involves the two fundamental constants $\pi = 3.14159265358\ldots$, and $e = 2.7182818285\ldots$, the base of natural logarithms. The curve has two ***parameters***, the ***mean*** μ, and the ***standard deviation*** σ. Here μ can be any real number positive or negative, while σ can be any strictly positive number. The mean μ indicates where the curve is located, while the standard deviation σ marks a horizontal scale. You can check by calculus that the curve is symmetric about the point marked μ, concave on either side of μ, out to the points of inflection $\mu - \sigma$ and $\mu + \sigma$, where it switches to become convex (Exercise 15).

FIGURE 1. The normal curve.



Think of the normal curve as a continuous histogram, defining a probability distribution over the line by relative areas under the curve. Then μ indicates the general location of the distribution, while σ measures how spread out the distribution is. The constant $1/\sqrt{2\pi}\sigma$ is put in the definition of the curve by convention, so that the total area under the curve is 1. This is shown by calculus in Section 5.3. See also Chapter 4 for a general treatment of continuous probability distributions like the normal.

The Normal Distribution

The normal distribution with mean μ and standard deviation σ is the distribution over the x-axis defined by areas under the normal curve with these parameters.

The equation of the normal curve with parameters μ and σ, can be written as

$$y = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}z^2}$$

where $z = (x - \mu)/\sigma$ measures the number of standard deviations from the mean μ to the number x, as shown in Figure 2. We say that z is *x in standard units*. The *standard normal distribution* is the normal distribution with mean 0 and standard deviation 1. This is the distribution defined by areas under the *standard normal curve* $y = \phi(z)$ where

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

is called the *standard normal density function*. The standard normal distribution is the distribution on the standard unit or z-scale derived from a normal distribution with arbitrary parameters μ and σ on the x-scale. As shown in Figure 2, the probability to the left of x in the normal distribution with mean μ and standard deviation σ is the probability to the left of $z = (x - \mu)/\sigma$ in the standard normal distribution. This probability is denoted $\Phi(z)$. This function of z is called the *standard normal cumulative distribution function*, or standard normal c.d.f. for short.

Standard Normal Cumulative Distribution Function

The standard normal c.d.f $\Phi(z)$ gives the area to the left of z under the standard normal curve:

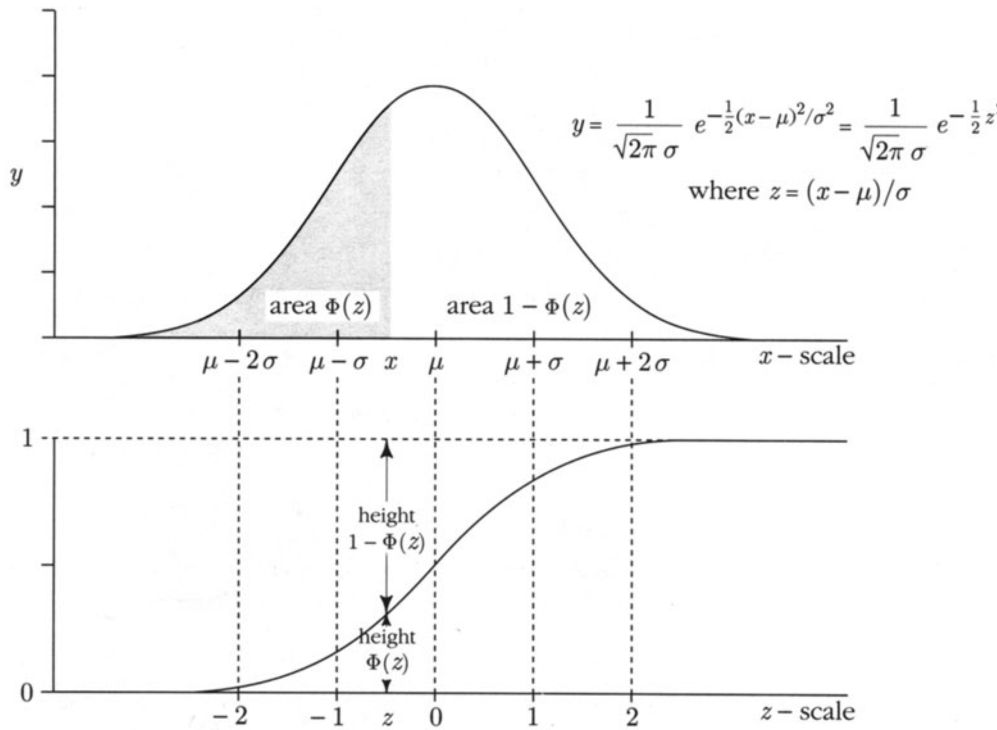
$$\Phi(z) = \int_{-\infty}^{z} \phi(y)\, dy$$

For the normal distribution with mean μ and standard deviation σ, the probability between a and b is

$$\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Because the function $e^{-\frac{1}{2}z^2}$ does not have a simple indefinite integral, there is no simple exact formula for $\Phi(z)$. But $\Phi(z)$ has been calculated numerically. Values of $\Phi(z)$ are tabulated in Appendix 5 for $z \geq 0$.

FIGURE 2. **A normal distribution and the standard normal c.d.f.** The top graph shows the curve that defines the normal distribution with mean μ and standard deviation σ. The lower graph shows the standard normal c.d.f. $\Phi(z)$, the probability in the normal distribution to the left of z on the standard unit scale. The area shaded under the normal curve is $\Phi(z)$ for a particular value z between -1 and 0. This area appears as a height in the graph of the normal c.d.f. $\Phi(z)$.



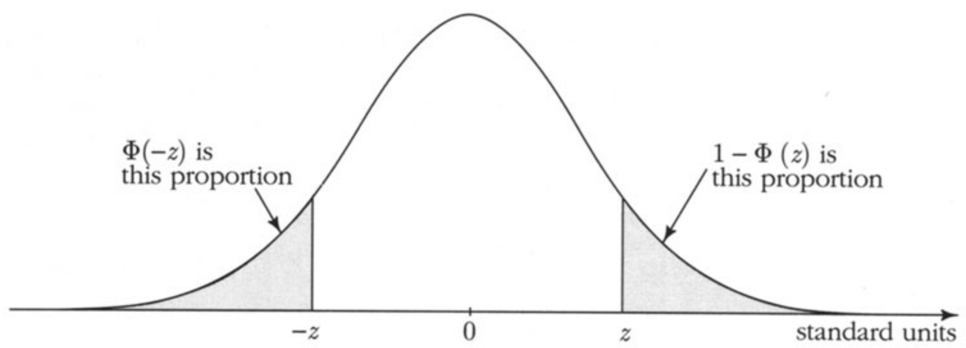
Remark. Instead of using the normal table, you may prefer to program an approximate formula for $\Phi(z)$ on a calculator. A formula, good enough for most purposes, is

$$\Phi(z) \approx 1 - \tfrac{1}{2}\left(1 + c_1 z + c_2 z^2 + c_3 z^3 + c_4 z^4\right)^{-4} \quad (z \geq 0)$$

$$\text{where} \quad \begin{array}{ll} c_1 = 0.196854 & c_2 = 0.115194 \\ c_3 = 0.000344 & c_4 = 0.019527 \end{array}$$

For every value of $z \geq 0$, the absolute error of this approximation is less than 2.5×10^{-4} [Abramowitz and Stegun, *Handbook of Mathematical Functions*].

FIGURE 3. Symmetry of the normal curve.



By the symmetry of the normal curve (see Figure 3),

$$\Phi(-z) = 1 - \Phi(z) \qquad (-\infty < z < \infty)$$

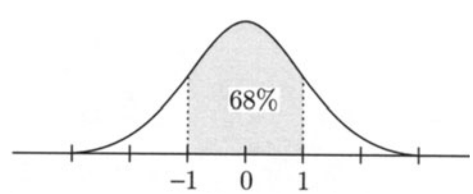
In particular, this implies $\Phi(0) = 1/2$. The probability of the interval (a, b) for the standard normal distribution, denoted $\Phi(a, b)$, is

$$\Phi(a, b) = \Phi(b) - \Phi(a)$$

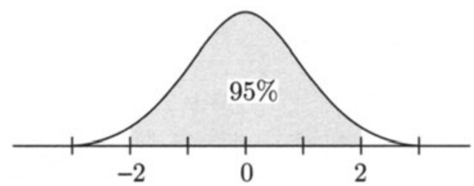
by the difference rule for probabilities. From Figure 3 and the rule of complements, it is clear that

$$\begin{aligned} \Phi(-z, z) &= \Phi(z) - \Phi(-z) \\ &= \Phi(z) - (1 - \Phi(z)) \\ &= 2\Phi(z) - 1 \end{aligned}$$

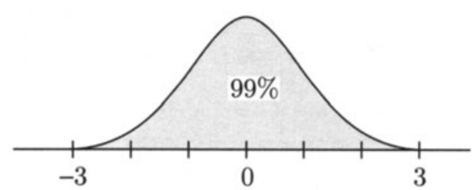
These formulae are used constantly when working with the normal distribution. But, to avoid mistakes, it is best not to memorize them. Rather sketch the standard normal curve each time. Remember the symmetry of the curve, and the definition of $\Phi(z)$, as the proportion of area under the curve to the left of z. Then the formulae are obvious from the diagram. There are three standard normal probabilities which are worth remembering:



$\Phi(-1, 1) \approx 68\%$, the probability within one standard deviation of the mean,



$\Phi(-2,2) \approx 95\%$, the probability within two standard deviations of the mean,



$\Phi(-3,3) \approx 99.7\%$, the probability within three standard deviations of the mean.

From these probabilities you can easily find $\Phi(a,b)$ for several other intervals. For example,

$$\Phi(0,1) = \tfrac{1}{2}\Phi(-1,1) \approx \tfrac{1}{2}68\% = 34\%$$

$$\Phi(2,\infty) = \tfrac{1}{2}(1 - \Phi(-2,2)) \approx \tfrac{1}{2}(100\% - 95\%) = 2.5\%$$

The probability $\Phi(-z,z)^c$ beyond z standard deviations from the mean in a normal distribution is

$$\Phi(-z,z)^c = 1 - \Phi(-z,z) = 2(1 - \Phi(z)) < 2\phi(z)/z$$

as shown in Table 1 for $z = 1$ to 6. The factor $\exp\left(-\frac{1}{2}z^2\right)$ in the definition of $\phi(z)$ makes $\phi(z)$ extremely small for large z. The above inequality, left as an exercise, shows that $\Phi(-z,z)^c$ is even smaller for $z \geq 2$.

Not too much significance should be placed on the extremely small probabilities $\Phi(-z,z)^c$ for z larger than about 3. The point is that the normal distribution is mostly applied as an approximation to some other distribution. Typically the errors involved in such an approximation, though small, are orders of magnitude larger than $\Phi(-z,z)^c$ for $z > 3$.

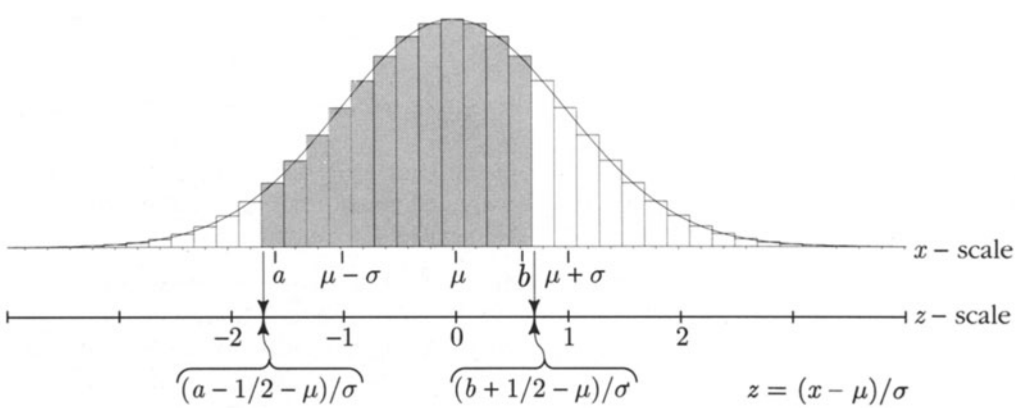
TABLE 1. **Standard normal probability outside $(-z, z)$.** The probability $\Phi(-z,z)^c$ is tabulated along with $2\phi(z)/z$, which is larger than $\Phi(-z,z)^c$ for all z, and a very good approximation to it for large z.

z	1	2	3	4	5	6
$\Phi(-z,z)^c$	0.317	0.046	2.7×10^{-3}	6.3×10^{-5}	5.7×10^{-7}	1.97×10^{-9}
$2\phi(z)/z$	0.484	0.054	2.9×10^{-3}	6.7×10^{-5}	5.9×10^{-7}	2.03×10^{-9}

The Normal Approximation to the Binomial Distribution

In fitting a normal curve to the binomial (n, p) distribution the main question is how the mean μ and standard deviation σ are determined by n and p. As noted in Section 2.1, the number $\mu = np$, called the mean of the binomial (n, p) distribution, is always within ± 1 of the most likely value, $m = \text{int}\,(np + p)$. So $\mu = np$ is a convenient place to locate the center. How to find the right value of σ is less obvious. As explained in the next section, provided $\sqrt{npq}$ is sufficiently large, good approximations to binomial probabilities are obtained by areas under the normal curve with mean $\mu = np$ and $\sigma = \sqrt{npq}$. Later, in Section 3.3, it will be explained how this formula for σ is consistent with the right general definition of the standard deviation of a probability distribution.

FIGURE 4. **A binomial histogram, with the normal curve superimposed.** Both the x scale (number of successes) and the z scale (standard units) are shown.



Let $P(a \text{ to } b)$ be the probability of getting between a and b successes (inclusive) in n independent trials with success probability p. Then, from Figure 4, we see that:

$$\begin{aligned} P(a \text{ to } b) &= \text{proportion of area under the binomial } (n,p) \text{ histogram} \\ &\quad \text{between } a - \tfrac{1}{2} \text{ and } b + \tfrac{1}{2} \\ &\approx \text{proportion of area under the normal curve } \quad \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} \\ &\quad \text{between } x = a - \tfrac{1}{2} \text{ and } b + \tfrac{1}{2} \\ &= \text{proportion of area under the normal curve } \quad \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \\ &\quad \text{between } z = (a - \tfrac{1}{2} - \mu)/\sigma \text{ and } z = (b + \tfrac{1}{2} - \mu)/\sigma. \end{aligned}$$

In terms of the standard normal c.d.f. Φ, this gives the following:

Normal Approximation to the Binomial Distribution

For n independent trials with success probability p

$$P(a \text{ to } b \text{ successes}) \approx \Phi\left(\frac{b + \frac{1}{2} - \mu}{\sigma}\right) - \Phi\left(\frac{a - \frac{1}{2} - \mu}{\sigma}\right)$$

where $\mu = np$ is the *mean*, and $\sigma = \sqrt{npq}$ is the *standard deviation*.

Use of $a - \frac{1}{2}$ and $b + \frac{1}{2}$ in the normal approximation rather than a and b is called the *continuity correction*. This correction is essential to obtain good approximations for small values of $\sqrt{npq}$. For large $\sqrt{npq}$ it makes little difference unless a and b are very close.

Example 1. 100 fair coin tosses.

Problem. Find, approximately, the chance of getting 50 heads in 100 tosses of a fair coin.

Solution. Here $n = 100$, $p = 1/2$, so $\mu = 50$, $\sigma = 5$. The normal approximation above with $a = b = 50$ gives

$$\begin{aligned} P(50) &\approx \Phi\left((50 + \tfrac{1}{2} - 50)/5\right) - \Phi\left((50 - \tfrac{1}{2} - 50)/5\right) \\ &= \Phi(0.1) - \Phi(-0.1) \\ &= 2\Phi(0.1) - 1 = 2 \times 0.5398 - 1 = 0.0796 \quad \text{(exact value 0.0795892)} \end{aligned}$$

Continuation. Other probabilities can be computed in the same way—for example

$$\begin{aligned}
P(45 \text{ to } 55) &\approx \Phi\left((55\tfrac{1}{2} - 50)/5\right) - \Phi\left((44\tfrac{1}{2} - 50)/5\right) \\
&= \Phi(1.1) - \Phi(-1.1) \\
&= 2\Phi(1.1) - 1 = 2 \times 0.8643 - 1 \\
&= 0.7286 \qquad \text{(exact value 0.728747)} \\
P(40 \text{ to } 60) &\approx 2\Phi(2.1) - 1 = 2 \times 0.9821 - 1 \\
&= 0.9642 \qquad \text{(exact value 0.9648)} \\
P(35 \text{ to } 65) &\approx 2\Phi(3.1) - 1 = 2 \times 0.9990 - 1 \\
&= 0.9980 \qquad \text{(exact value 0.99821)}
\end{aligned}$$

Fluctuations in the number of successes. For any fixed p, the normal approximation to the binomial (n, p) distribution gets better and better as n gets larger. So, in a large number of independent trials with success probability p, the typical size of the random fluctuations in the number of successes is of the order of $\sigma = \sqrt{npq}$. For example,

$$\begin{aligned}
P(\mu - \sigma \text{ to } \mu + \sigma \text{ successes in } n \text{ trials}) &\approx 68\% \\
P(\mu - 2\sigma \text{ to } \mu + 2\sigma \text{ successes in } n \text{ trials}) &\approx 95\% \\
P(\mu - 3\sigma \text{ to } \mu + 3\sigma \text{ successes in } n \text{ trials}) &\approx 99.7\%
\end{aligned}$$

It can be shown that for fixed p, as $n \to \infty$, each probability on the left approaches the exact value of the corresponding proportion of area under the normal curve.

Fluctuations in the proportion of successes. While the typical size of random fluctuations of the *number* of successes in n trials away from the expected number np is a moderate multiple of $\sqrt{npq}$, the typical size of random fluctuations in the *relative frequency* of successes about the expected proportion p is correspondingly of order $\sqrt{npq}/n = \sqrt{pq/n}$. Since $\sqrt{pq} \le \frac{1}{2}$ for all $0 < p < 1$, and $1/\sqrt{n} \to 0$ as $n \to \infty$, this makes precise the rate at which we can expect relative frequencies to stabilize under ideal conditions.

Square Root Law

For large n, in n independent trials with probability p of success on each trial:

- the *number* of successes will, with high probability, lie in a relatively small interval of numbers, centered on np, with width a moderate multiple of $\sqrt{n}$ on the numerical scale;
- the *proportion* of successes will, with high probability, lie in a small interval centered on p, with width a moderate multiple of $1/\sqrt{n}$.

Numerical computations show that the square root law also holds for small values of n, but its most important implications are for large n. In particular, it implies the following mathematical confirmation of our intuitive idea of probability as a limit of long-run frequencies:

Law of Large Numbers

If n is large, the proportion of successes in n independent trials will, with overwhelming probability, be very close to p, the probability of success on each trial. More formally:

- for independent trials, with probability p of success on each trial, for each $\epsilon > 0$, no matter how small, as $n \to \infty$,

$$P(\text{proportion of successes in } n \text{ trials differs from } p \text{ by less than } \epsilon) \to 1$$

Confidence Intervals

The normal approximation is the basis of the statistical method of *confidence intervals*. Suppose you think that you are observing the results of a sequence of independent trials with success probability p, but you don't know the value of p. For example, you might be observing whether or not a biased die rolled a six (success) or not six (failure). Suppose in n trials you observe that the relative frequency of successes is $\hat{p}$. If n is large, it is natural to expect that the unknown probability p is most likely fairly close to $\hat{p}$. For example, since

$$\Phi(-4, 4) \approx 99.99\%$$

the above results state that if n is large enough, no matter what p is, it is 99.99% certain that the observed number of successes, $n\hat{p}$, differs from np by less than $4\sqrt{npq}$, so the relative frequency $\hat{p}$ will differ from p by less than $4\sqrt{pq/n}$, which is at most $2/\sqrt{n}$. Having observed the value of $\hat{p}$, it is natural to suppose that this overwhelmingly likely event has occurred, which implies that p is within $2/\sqrt{n}$ of $\hat{p}$. The interval $\hat{p} \pm 2/\sqrt{n}$, within which p can reasonably be expected to lie, is called a 99.99% *confidence interval* for p.

Example 2. **Estimating the bias on a die.**

Problem. In a million rolls of a biased die, the number 6 shows 180,000 times. Find a 99.99% confidence interval for the probability that the die rolls six.

Solution. The observed relative frequency of sixes is $\hat{p} = 0.18$. So a 99.99% confidence interval for the probability that the die rolls six is

$$0.18 \pm 2/\sqrt{1,000,000} \quad \text{or} \quad (0.178, 0.182)$$

Remark. This procedure of going $\pm 2/\sqrt{n}$ from the observed $\hat{p}$ to make the confidence interval is somewhat conservative, meaning the coverage probability will be even higher than 99.99% for large n. This is due to neglecting the factor $\sqrt{pq} \leq 0.5$ and so overestimating the standard deviation $\sigma = \sqrt{npq}$ in case p is not 0.5, as the above $\hat{p}$ would strongly suggest. The usual statistical procedure is to estimate $\sqrt{pq}$ by $\sqrt{\hat{p}(1-\hat{p})}$, which is $\sqrt{0.18 \times 0.82} = 0.384$ in the above example. This reduces the length of the interval by a factor of $0.384/0.5 = 77\%$ in this case.

The most important thing to note in this kind of calculation is how the length of the confidence interval depends on n through the square root law. Suppose the confidence interval is $\hat{p} \pm c/\sqrt{n}$, for some constant c. No matter what c is, to reduce the length of the confidence interval by a factor of f requires an increase of n by a factor of f^2. So to halve the length of a confidence interval, you must quadruple the number of trials.

Example 3. Random sampling.

Problem. Two survey organizations make 99% confidence intervals for the proportion of women in a certain population. Both organizations take random samples with replacement from the population; the first uses a sample of size 350 while the second uses a sample of size 1000. Which confidence interval will be shorter, and by how much?

Solution. The interval based on the larger sample size will be shorter. The size of the second sample is $1000/350 = 2.86$ times the size of the first, so the length of the second interval is $1/\sqrt{2.86}$ times the length of the first, that is, 0.59 times the length of the first.

Example 4. How many trials?

Suppose you estimate the probability p that a biased coin lands heads by tossing it n times and estimating p by the proportion $\hat{p}$ of the times the coin lands heads in the n tosses.

Problem. How many times n must you toss the coin to be at least 99% sure that $\hat{p}$ will be: a) within 0.1 of p? b) within .01 of p?

Solution. First find z such that $\Phi(-z, z) = 99\%$,

$$\text{i.e.,} \quad 2\Phi(z) - 1 = 0.99 \quad \text{i.e.,} \quad \Phi(z) = 0.995$$

Inspection of the table gives $z \approx 2.575$. For large n, $\hat{p}$ will with probability at least 99% lie in the interval $p \pm 2.575\sqrt{pq}/\sqrt{n}$. Since $\sqrt{pq} \leq 0.5$, the difference between

$\hat{p}$ and p will then be less than

$$2.575 \times 0.5/\sqrt{n}$$

For a), set this equal to 0.1 and solve for n:

$$2.575 \times 0.5/\sqrt{n} = 0.1$$

$$n = \left(\frac{2.575 \times 0.5}{0.1}\right)^2 = 165.77$$

So 166 trials suffice for at least 99% probability of accuracy to within 0.1.

b) By the square root law, to increase precision by a factor of 10, requires an increase in the number of trials by $10^2 = 100$. So about $16,577$ trials would be required for 99% probability of accuracy to within .01.

How good is the normal approximation? As a general rule, the larger the standard deviation $\sigma = \sqrt{npq}$, and the closer p is to $1/2$, the better the normal approximation to the binomial (n, p) distribution. The approximation works best for $p = 1/2$ due to the symmetry of the binomial distribution in this case. For $p \neq 1/2$ the approximation is not quite as good, but as the graphs at the end of Section 2.1 show, as n increases the binomial distribution becomes more and more symmetric about its mean. It is shown in the next section that the shape of the binomial distribution approaches the shape of the normal curve as $n \to \infty$ for every fixed p with $0 < p < 1$.

How good the normal approximation is for particular n and p can be measured as follows. Let $N(a \text{ to } b)$ denote the normal approximation with continuity correction to a binomial probability $P(a \text{ to } b)$. Define $W(n, p)$, the *worst error* in the normal approximation to the binomial(n, p) distribution, to be the biggest absolute difference between $P(a \text{ to } b)$ and $N(a \text{ to } b)$, over all integers a and b with $0 \leq a \leq b \leq n$:

$$W(n, p) = \max_{0 \leq a \leq b \leq n} |P(a \text{ to } b) - N(a \text{ to } b)|$$

Numerical calculations show that $W(n, 1/2)$ is less than 0.01 for all $n \geq 10$, and less than 0.005 for all $n \geq 20$. Such a small error of approximation is negligible for most practical purposes. For $p \neq 1/2$ there is a systematic error in the normal approximation because an asymmetric distribution is approximated by a symmetric one. A refinement of the normal approximation described in the next paragraph shows that

$$W(n, p) \approx \frac{1}{10} \frac{|1 - 2p|}{\sqrt{npq}} \tag{1}$$

where the error of the approximation is negligible for all practical purposes provided $\sigma = \sqrt{npq}$ is at least about 3. This formula shows clearly how the larger σ, and the closer p is to $1/2$, the smaller $W(n,p)$ tends to be. Because $|1-2p| \leq 1$ for all $0 \leq p \leq 1$, even if p is close to 0 or 1, the worst error is small provided σ is large enough. For $\sigma \geq 3$ the worst error is about $1/10\sigma$ for p close to 0 or 1 and large n. Numerical calculations confirm the following consequences of (1): the worst error $W(n,p)$ is

- less than 0.01 for $n \geq 20$ and p between 0.4 and 0.6
- less than 0.02 for $n \geq 20$ and p between 0.3 and 0.7
- less than 0.03 for $n \geq 25$ and p between 0.2 and 0.8
- less than 0.05 for $n \geq 30$ and p between 0.1 and 0.9

The systematic error in the normal approximation of magnitude about $1/10\sigma$ can be reduced to an error that is negligible in comparison by the *skewness correction* explained in the next paragraph. This method gives satisfactory approximations to binomial probabilities for arbitrary n and p with $\sigma \geq 3$. For p close to 0 or 1, and $\sigma \leq 3$, a better approximation to the binomial distribution is provided by using the Poisson distribution described in the next section.

The skew-normal approximation. Figures 5 and 7 show how the histogram of the binomial$(100, 1/10)$ distribution is slightly skewed relative to its approximating normal curve. The histogram is better approximated by adding to the standard normal curve $\phi(z)$ a small multiple of the curve $y = \phi'''(z)$, where

$$\phi'''(z) = (3z - z^3)\phi(z)$$

is the third derivative of $\phi(z)$ (Exercise 16), as graphed in Figure 6. By careful analysis of the histogram of a binomial(n,p) distribution plotted on a standard units scale, it can be shown that for $p \neq 1/2$ adding the right small multiple of the anti-symmetric function $\phi'''(z)$ to the symmetric function $\phi(z)$ gives a curve which respects the slight asymmetry of the binomial histogram, and so follows it much more closely than the plain normal curve $\phi(z)$. The resulting *skew-normal curve* has equation

$$y = \phi(z) - \frac{1}{6}\text{Skewness}\,(n,p)\,\phi'''(z) \quad \text{where} \tag{2}$$

$$\text{Skewness}\,(n,p) = (1-2p)/\sqrt{npq} = (1-2p)/\sigma$$

is a number called the *skewness* of the binomial(n,p) distribution, which measures its degree of asymmetry. The skewness is 0 if $p = 1/2$, when the distribution is perfectly symmetric about $n/2$. The skewness positive for $p < 1/2$ when the distribution is called *skewed to the right*, and negative for $p > 1/2$ when the distribution is *skewed to the left*. The meaning of these terms is made precise by the way the binomial histogram follows the skew-normal curve (2) more closely than it does the

FIGURE 5. **Normal curve approximating the binomial $(100, 1/10)$ histogram.** Notice how the bars are slightly above the normal curve just to the left of the mean, and slightly below the curve just to the right of the mean. Further away from the mean, the bars lie below the curve in the left tail, and above the curve in the right tail.

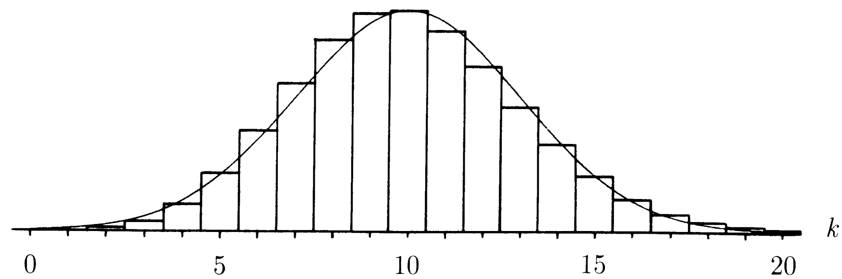


FIGURE 6. **Graph of $\phi'''(z) = (3z - z^3)\phi(z)$.** Note how the function is positive in the intervals $(-\infty, -\sqrt{3})$ and $(0, \sqrt{3})$, and negative in the intervals $(-\sqrt{3}, 0)$ and $(\sqrt{3}, \infty)$. The zeros are at 0 and $\pm\sqrt{3}$. The z-scale is the standard unit scale derived from the histogram in Figures 5 and 7.

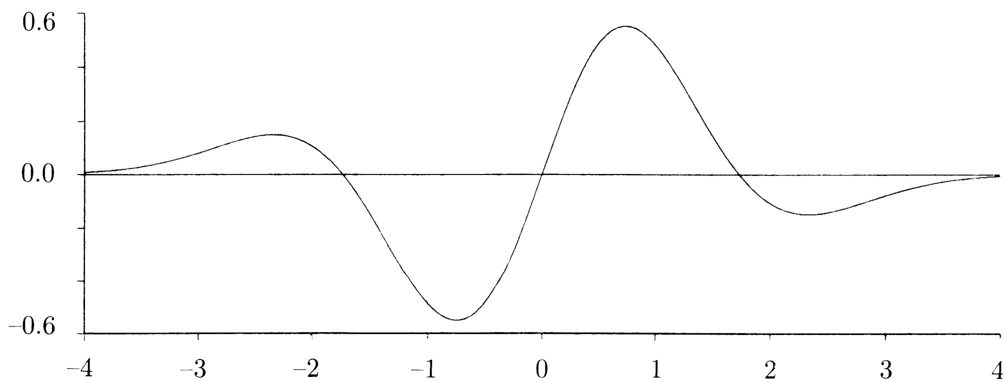
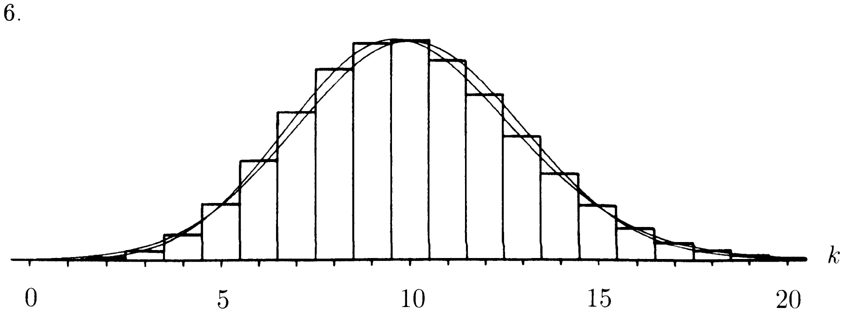


FIGURE 7. **Skew–normal curve approximating the binomial $(100, 1/10)$ histogram.** Refer to Example 5. Both the normal curve $y = \phi(z)$ and the skew–normal curve $y = \phi(z) - (2/45)\phi'''(z)$ are shown. The skew–normal curve follows the binomial histogram much more closely. The difference between the normal and skew–normal curves is $2/45$ times the curve $\phi'''(z)$ graphed in Figure 6.



plain normal curve. Figure 7 illustrates how in the case $p < 1/2$ when the binomial histogram is skewed to the right, there are numbers $z_- < z_0 < z_+$ on the standard units scale, with $z_0 \approx 0$ and $z_\pm \approx \pm\sqrt{3}$, (the three zeros of $\phi'''(z)$) such that

- the histogram is lower than the normal curve on the intervals $(-\infty, z_-)$ and (z_0, z_+)
- the histogram is higher than the normal curve on the intervals (z_-, z_0) and (z_+, ∞)

For $1/2 < p < 1$, the same thing happens, except that the words "higher" and "lower" must be switched in the above description. The distribution is then skewed to the left. Integrating the skew-normal curve (2) from $-\infty$ to the point z on the standard unit scale (Exercise 16) gives the following:

Skew-Normal Approximation to the Binomial Distribution

For n independent trials with success probability p,

$$P(0 \text{ to } b \text{ successes}) \approx \Phi(z) - \frac{1}{6}\text{Skewness}\,(n,p)(z^2-1)\phi(z)$$

where $z = (b + \frac{1}{2} - \mu)/\sigma$ for $\mu = np$ and $\sigma = \sqrt{npq}$, $\Phi(z)$ is the standard normal c.d.f., $\phi(z) = (1/\sqrt{2\pi})\exp(-\frac{1}{2}z^2)$ is the standard normal curve, and

$$\text{Skewness}(n,p) = (1-2p)/\sqrt{npq}$$

The term involving the skewness in the skew-normal approximation is called the *skewness correction*. The skew-normal approximation to an interval probability

$$P(a \text{ to } b) = P(0 \text{ to } b) - P(0 \text{ to } a-1)$$

is found by using the above approximation twice and taking the difference. The resulting normal approximation with skewness correction to $P(a \text{ to } b)$ differs from the plain normal approximation $N(a \text{ to } b)$ by $1/6$ of the skewness times the area under the curve $\phi'''(z)$ between points corresponding to a and b on the standard units scale. You can show (Exercise 16) that this area is always between ± 0.577, and that these extremes are attained over the intervals from $z = -\sqrt{3}$ to $z = 0$, and from $z = 0$ to $z = \sqrt{3}$. It follows that for $p \neq 1/2$, the worst error $W(n,p)$ in the normal approximation without skewness correction occurs for $a \approx \mu - \sqrt{3}\sigma$ and $b \approx \mu$, or for $a \approx \mu$ and $b \approx \mu + \sqrt{3}\sigma$. The errors of the normal approximation for these two intervals will be of opposite signs with approximately equal magnitudes of

$$W(n,p) \approx \frac{1}{6} \times |1-2p|/\sigma \times 0.577 \approx |1-2p|/10\sigma$$

Thus the skew-normal approximation implies this simple estimate for the worst error in the plain normal approximation, and shows the intervals on which such an error is to be expected. This formula shows the plain normal approximation is rather rough for σ in the range from 3 to 10 and p close to 0 or 1. Numerical calculations show that provided $\sigma \geq 3$ (no matter what p) the skew-normal approximation gives interval probabilities correct to two decimal places (error at most 0.005) which is adequate for most practical purposes. For fixed p, as $n \to \infty$, the skewness of the binomial distribution converges to 0, so in the limit of large n the skewness correction can be ignored, just like the continuity correction, which is of the same order of magnitude $1/\sigma$.

Example 5. Distribution of the number of 0's in 100 random digits.

Consider the distribution of the random number of times a particular digit, say 0, appears among 100 random digits picked independently and uniformly at random from the set of 10 digits $\{0, 1, \ldots, 9\}$. This is the binomial$(100, 1/10)$ distribution which is displayed in Figure 7, along with the approximating normal and skew-normal curves. The mean is $\mu = 100 \times 1/10 = 10$, the standard deviation is $\sigma = \sqrt{npq} = \sqrt{100 \times (1/10) \times (9/10)} = 3$, and the skewness is $(1-2p)/\sqrt{npq} = (1 - (2/10))/3 = 4/15$. From (2), the skew-normal curve approximating the shape of the binomial histogram has equation $y = \phi(z) - \frac{2}{45}(3z - z^3)\phi(z)$, as graphed in Figure 7. The probability of 4 or fewer 0's is

$$P(0 \text{ to } 4) = \sum_{k=0}^{4} \binom{100}{k} \left(\frac{1}{10}\right)^k \left(\frac{9}{10}\right)^{100-k} = 0.024$$

by exact calculation, correct to three decimal places. The normal approximation to this probability is $\Phi(z)$ for $z = (4\frac{1}{2} - 10)/3 = -11/6$, i.e., $\Phi(-11/6) = 0.033$, which is not a very good approximation. The skew-normal approximation, which is not much harder to compute, is

$$\begin{aligned}
&\Phi(z) - \frac{1}{6}\text{Skewness}(100, 1/10)(z^2 - 1)\phi(z) \\
&= 0.033 - \frac{1}{6}\frac{4}{15}\left(\left(\frac{-11}{6}\right)^2 - 1\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{-11}{6}\right)^2\right) \\
&= 0.026
\end{aligned}$$

which differs from the exact value by only 0.002. Similar calculations yield the numbers displayed in Table 2. The numbers are correct to three decimal places. The ranges selected, 0 to 4, 5 to 9, 10 to 15, and 16 to 100, are the ranges over which the normal approximation is first too high, then too low, too high, and too low again. The normal approximation is very rough in this example, but the skew-normal approximation is excellent.

TABLE 2. **Approximations to the binomial $(100, 1/10)$ distribution.** The probability $P(a \text{ to } b)$ of from a and b successes (inclusive) in 100 independent trials, with probability $1/10$ of success on each trial, is shown along with approximations using the normal and skew–normal curves.

value range	exact probability	skew–normal approximation	normal approximation
0 – 4	0.024	0.026	0.033
5 – 9	0.428	0.425	0.400
10 – 15	0.509	0.508	0.533
16 – 100	0.040	0.041	0.033

Exercises 2.2

1. Let H be the number of heads in 400 tosses of a fair coin. Find normal approximations to: a) $P(190 \leq H \leq 210)$; b) $P(210 \leq H \leq 220)$; c) $P(H = 200)$; d) $P(H = 210)$.

2. Recalculate the approximations above for a biased coin with $P(\text{heads}) = 0.51$.

3. A fair coin is tossed repeatedly. Consider the following two possible outcomes:

55 or more heads in the first 100 tosses
220 or more heads in the first 400 tosses

a) Without calculation, say which of these outcomes is more likely. Why?

b) Confirm your answer to a) by a calculation.

4. Suppose that each of 300 patients has a probability of $1/3$ of being helped by a treatment independent of its effect on the other patients. Find approximately the probability that more than 120 patients are helped by the treatment.

5. Suppose you bet a dollar on red, 25 times in a row, at roulette. Each time you win a dollar with probability $18/38$, lose with probability $20/38$. Find, approximately, the chance that after 25 bets you have at least as much money as you started with.

6. To estimate the percent of district voters who oppose a certain ballot measure, a survey organization takes a random sample of 200 voters from a district. If 45% of the voters in the district oppose the measure, estimate the chance that:

a) exactly 90 voters in the sample oppose the measure;

b) more than half the voters in the sample oppose the measure.

[Assume that all voters in the district are equally likely to be in the sample, independent of each other.]

7. City A has a population of 4 million, and city B has 6 million. Both cities have the same proportion of women. A random sample (with replacement) will be taken from each city, to estimate this proportion. In each of the following cases, say whether the two samples give equally good estimates; and if you think one estimate is better than the other, say how much better it is.

a) A 0.01% sample from each city.

b) A sample of size 400 from each city.

c) A 0.1% sample from city A, and a 0.075% sample from city B.

8. Find, approximately, the chance of getting 100 sixes in 600 rolls of a die.

9. **Airline overbooking.** An airline knows that over the long run, 90% of passengers who reserve seats show up for their flight. On a particular flight with 300 seats, the airline accepts 324 reservations.

a) Assuming that passengers show up independently of each other, what is the chance that the flight will be overbooked?

b) Suppose that people tend to travel in groups. Would that increase or decrease the probability of overbooking? Explain your answer.

c) Redo the calculation a) assuming that passengers always travel in pairs. Check that your answers to a), b), and c) are consistent.

10. A probability class has 30 students. As part of an assignment, each student tosses a coin 200 times and records the number of heads. Approximately what is the chance that no student gets exactly 100 heads?

11. **Batting averages.** Suppose that a baseball player's long-run batting average (number of hits per time at bat) is .300. Assuming that each time at bat yields a hit with a consistent probability, independently of other times, what is the chance that the player's average over the next 100 times at bat will be

a) .310 or better? b) .330 or better? c) .270 or worse?

d) Suppose the player tends to have periods of good form and periods of bad form. Would different times at bat then be independent? Would that tend to increase or decrease the above chances?

e) Suppose the player actually hits .330 over the 100 times at bat. Would you be convinced that his form had improved significantly? or could the improvement just as well be due to chance?

12. A fair coin is tossed $10,000$ times. Find a number m such that the chance of the number of heads being between $5000 - m$ and $5000 + m$ is approximately $2/3$.

13. A pollster wishes to know the percentage p of people in a population who intend to vote for a particular candidate. How large must a random sample with replacement be in order to be at least 95% sure that the sample percentage is within one percentage point of p?

14. Wonderful Widgets Inc. has developed electronic devices which work properly with probability 0.95, independently of each other. The new devices are shipped out in boxes containing 400 each.

a) What percentage of boxes contains 390 or more working devices?

b) The company wants to guarantee, say, that k or more devices per box work. What is the largest k such that at least 95% of the boxes meet the warranty?

15. First two derivatives of the normal curve. Let $\phi'(z)$, $\phi''(z)$ be the first and second derivatives of the standard normal curve $\phi(z) = (1/\sqrt{2\pi})\exp(-\frac{1}{2}z^2)$. Show that:

a) $\phi'(z) = -z\phi(z)$

b) $\phi''(z) = (z^2 - 1)\phi(z)$

c) Sketch the graphs of $\phi(z)$, $\phi'(z)$, $\phi''(z)$ on the same scale for z between -4 and 4. What are the graphs like outside of this range?

d) Use b) and the chain rule of calculus to find the second derivative at x of the normal curve with parameters μ and σ^2.

e) Use the result of d) to verify the assertions in the sentence above Figure 1 on page 93.

16. Third derivative of the normal curve.

a) Show that $\phi(z)$ has third derivative $\phi'''(z) = (-z^3 + 3z)\phi(z)$

b) Show that $\int_{-\infty}^{x} \phi'''(z)dz = \phi''(x)$, and hence

$$\int_{-\infty}^{-\sqrt{3}} \phi'''(z)dz = -\int_{\sqrt{3}}^{\infty} \phi'''(z)dz = 2\phi(\sqrt{3}) \approx 0.178$$

and

$$-\int_{-\sqrt{3}}^{0} \phi'''(z)dz = \int_{0}^{\sqrt{3}} \phi'''(z)dz = \phi(0) + 2\phi(\sqrt{3}) \approx 0.577$$

c) Show that $\int_a^b \phi'''(z)dz$ lies between $\pm[\phi(0) + 2\phi(\sqrt{3})]$ for every $a < b$. [*Hint:* No more calculation required. Consider the graph of $\phi'''(z)$ and the interpretation of the integral in terms of areas.]

17. Standard normal tail bound. Show that $1 - \Phi(z) < \phi(z)/z$ for positive z by the following steps.

a) Show that

$$1 - \Phi(z) = \int_z^{\infty} \phi(x)dx.$$

(This integral cannot be evaluated by calculus.)

b) Show that multiplying the integrand by x/z gives a new integral whose value is strictly larger.

c) Evaluate the new integral.

2.3 Normal Approximation: Derivation (Optional)

This section is more mathematical than the previous and following ones and can be skipped at first reading. Its main aim is to derive the formula for the normal curve by study of binomial probabilities for large n. The basic idea is that for any p with $0 < p < 1$, as n increases the binomial (n, p) distribution becomes better and better approximated by a normal distribution with parameters $\mu = np$ and $\sigma = \sqrt{npq}$. Why this happens is the subject of this section.

Recall first the calculus definition of e, the base of natural logarithms, as the unique number such that the function $y = \log_e x$ has derivative

$$\frac{d}{dx}\log_e x = \frac{1}{x}$$

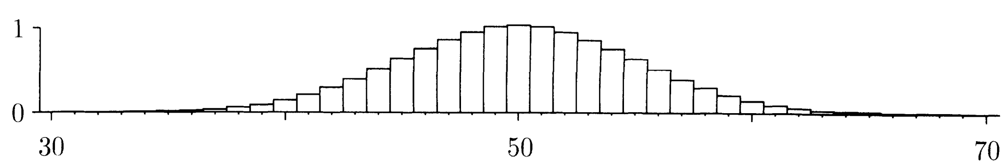
Here $y = \log_e x$ means $x = e^y$. In the following, all logarithms are to the base e: log means $\log_e$. See Appendix 4 for further background on exponentials and logarithms. Since $\log(1) = 0$ and the derivative of $\log x$ at $x = 1$ is $1/1 = 1$,

$$\log(1+\delta) \approx \delta \quad \text{for small } \delta$$

with an error of approximation which becomes negligible in comparison to δ as $\delta \to 0$. This simple approximation makes e the preferred or *natural* base of logarithms, and makes e turn up in almost any limit of a product of an increasing number of factors. The emergence of the normal curve from the binomial probability formula is a case in point.

Let $H(k) = P(k)/P(m)$ be the height at k of a binomial histogram scaled to have maximum height 1 at $k = m$, where $m = \text{int}(np+p)$ is the mode. Note that $H(m) = 1$. The normal approximation will now be derived by a sequence of steps, starting with an approximation for $H(k)$. Consider for illustration the distribution of the number of heads in 100 fair coin tosses:

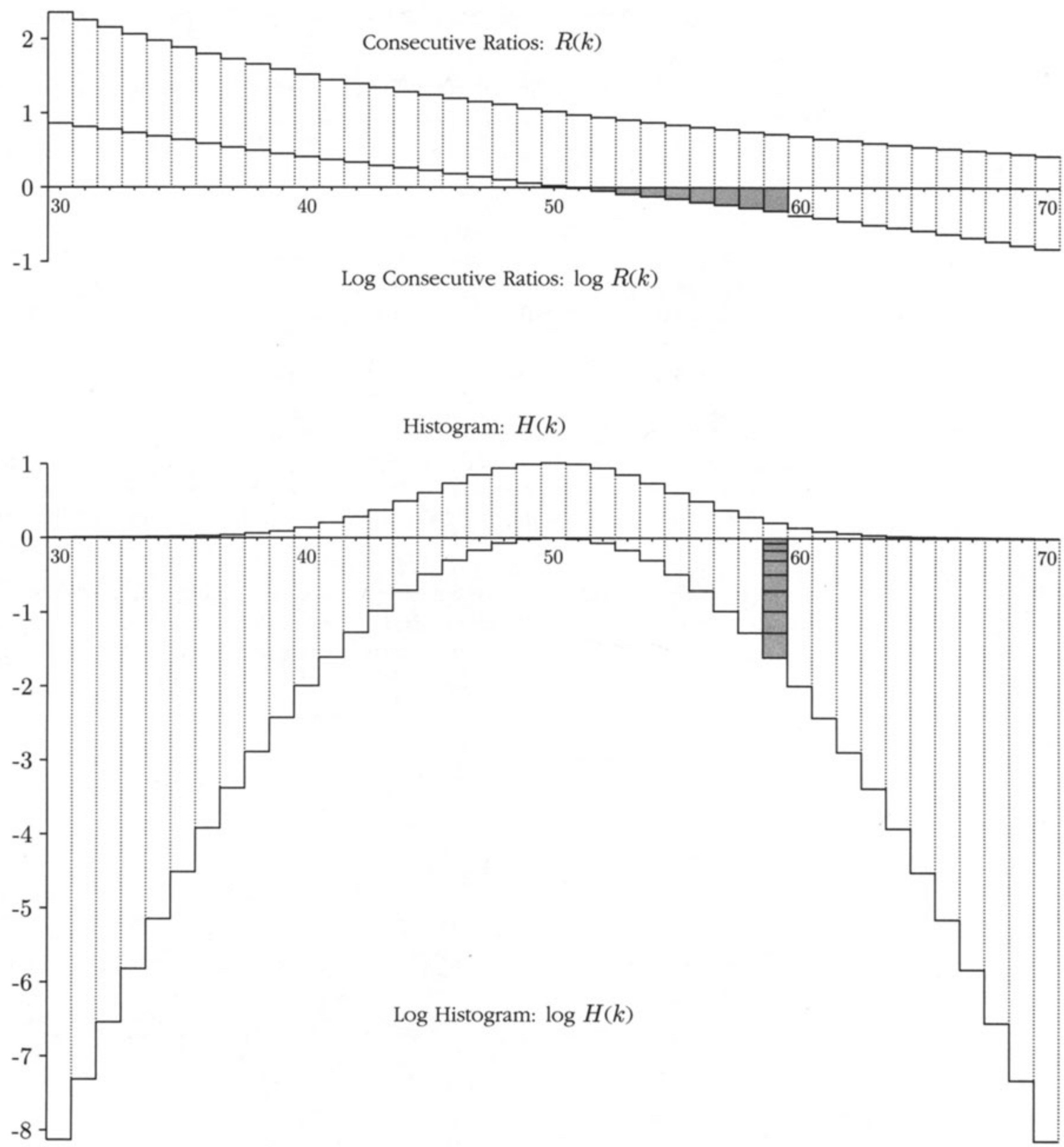
FIGURE 1. **Binomial $(100, 0.5)$ histogram.** Bar graph of $H(k) = P(k)/P(m)$.



The histogram heights $H(k)$ can be found by multiplying the consecutive odds ratios

$$R(k) = H(k)/H(k-1) = P(k)/P(k-1) = \frac{n-k+1}{k}\frac{p}{q}$$

FIGURE 2. **Binomial $(100, 0.5)$ consecutive odds, histogram, and their logarithms.** These graphs are drawn to scale. You can see how $\log R(k)$ is nearly linear with a gentle *slope* of about $-1/25$. Because $\log H(k)$ is a sum of increments of this nearly linear function (see equal shaded areas for $k = 59$), its graph is nearly parabolic. By approximation of the area in the top graph with a right-angled triangle with sides $(k - 50)$ and slope $\times$ $(k - 50)$, the area is $\log H(k) \approx \frac{1}{2}$ slope $\times$ $(k - 50)^2 \approx -\frac{1}{2}(k - \mu)^2/\sigma^2$ for $\mu = 50, \sigma = 5 = \sqrt{25}$. This is formula (1).



For $k > m$, $H(k)$ is the product of $(m - k)$ consecutive ratios

$$H(k) = H(m)\frac{P(m+1)}{P(m)}\frac{P(m+2)}{P(m+1)}\cdots\frac{P(k)}{P(k-1)} = R(m+1)R(m+2)\cdots R(k)$$

and there is a similar expression for $k < m$. The key to the normal approximation is that as the ratios $R(k)$ decrease for values of k near m, crossing near m from more than 1 to less than 1, they do so *very slowly*, and due to the formula for $R(k)$, *almost linearly*.

This is shown in a particular case in Figure 2, and is true no matter what the value of p, provided n is large enough. As n gets larger, the consecutive odds ratios $R(k)$ decrease more and more slowly near $k = m$. Consequently, as n increases, $R(k)$ stays close to 1 over a wider and wider range of numbers k. This means that for large n, for a wide range of k near $m \approx np$, $H(k)$ is the product of factors that are all very close to 1. The way to handle this product is to take logs to the base e:

$$\log H(k) = \log R(m+1) + \cdots + \log R(k) \qquad \text{as graphed in Figure 2.}$$

Now write $k = m + x \approx np + x$, $k + 1 \approx k$, assume x is small in comparison to npq, and use $\log(1+\delta) \approx \delta$ for small δ to justify the following approximation:

$$\begin{aligned}\log R(k) = \log\left(\frac{n-k+1}{k}\cdot\frac{p}{q}\right) &\approx \log\left(\frac{(n-np-x)p}{(np+x)q}\right)\\ &= \log\left(1-\frac{px}{npq}\right) - \log\left(1+\frac{qx}{npq}\right)\\ &\approx -\frac{px}{npq} - \frac{qx}{npq} = \frac{-x}{npq} = -\frac{(k-m)}{npq}\end{aligned}$$

This shows that if $x = k - m$ is kept small in comparison to n, then $\log R(k)$ is an approximately linear function of k, as in Figure 2, with slope approximately $-1/npq$. Adding up these approximations, using $1 + 2 + \cdots + x = \frac{1}{2}x(x+1) \approx \frac{1}{2}x^2$, gives

$$\log H(k) \approx -\frac{1}{npq} - \frac{2}{npq} - \cdots - \frac{(k-m)}{npq} \approx -\frac{1}{2}\frac{(k-m)^2}{npq} \approx -\frac{1}{2}\frac{(k-np)^2}{npq}$$

This is illustrated by the roughly triangular area shaded in Figure 2. A similar argument works for $k < m$. So for the heights $H(k) = P(k)/P(m)$ of the binomial (n,p) histogram there is a preliminary form of the normal approximation:

$$H(k) \approx e^{-\frac{1}{2}(k-\mu)^2/\sigma^2} \tag{1}$$

where $\mu = np$ is the *mean* and $\sigma = \sqrt{npq}$ is the *standard deviation*.

The argument shows this approximation will be good provided $|k - m|$ is small in comparison with npq. A more careful argument shows that this range of k is really all that matters. Now approximate $P(k)$ instead of $H(k)$:

$$P(k)=H(k)P(m)=H(k)/H(0 \text{ to } n) \quad \text{where } H(0 \text{ to } n)=H(0)+\cdots+H(n) \qquad (2)$$

Here $H(0 \text{ to } n)$, the total area under the binomial (n, p) histogram with maximum height 1, can be approximated by the total area under the approximating normal curve (1), which is an integral:

$$\begin{aligned} H(0 \text{ to } n) &\sim \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}\,dx \\ &= \sigma\left[\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2}\,dz\right] \quad \begin{array}{l}\text{by the calculus change of variable}\\ (x-\mu)/\sigma = z, \quad dx = \sigma dz\end{array} \\ &= \sigma\sqrt{2\pi} \qquad \text{as shown by calculus in Section 5.3} \end{aligned}$$

It can be shown that the relative error of approximation can be made arbitrarily small, no matter what the values of n and p, provided that $\sigma = \sqrt{npq}$ is sufficiently large. Now combine this with (1) and (2):

$$P(k) \approx \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(k-\mu)^2/\sigma^2} \qquad \text{where} \qquad \mu = np, \quad \sigma = \sqrt{npq} \qquad (3)$$

The precise meaning of the $\approx$ involved here is somewhat technical. As $\sigma \to \infty$, both sides tend to zero. But the *relative* error of approximation tends to 0 provided $(k-\mu)/\sigma$ remains bounded. See Feller's book *An Introduction to Probability Theory and its Applications*, Vol. I, for more details.

The equation of the normal curve appears in formula (3) as a function of k. The probability of an interval of numbers is now approximated by replacing relative areas under the histogram by relative areas under the approximating curve.

What makes the normal curve a better and better approximation as $n \to \infty$, is that for large n, as k moves away from m, the histogram heights $H(k)$ approach zero before the consecutive ratios $R(k)$ differ significantly from 1. In the expression

$$\log H(k) = \log R(m+1) + \cdots + \log R(k)$$

a large number of terms on the right, each nearly zero, add up to a total $\log H(k)$ which is significantly different from 0.

Probability of the Most Likely Number of Successes

A consequence of the normal approximation (3) for $k = m$, closely related to the square root law discussed in the previous section, is that the most likely value $m = \text{int}(np + p)$ in the binomial (n, p) distribution has probability

$$P(m) \sim \frac{1}{\sqrt{2\pi}\sigma} = \frac{1}{\sqrt{2\pi npq}} \quad \text{as} \quad n \to \infty \tag{5}$$

For fixed p, as $n \to \infty$, the relative error in this approximation tends to 0. In particular, no matter what the success probability p, the probability of the most likely number of successes in n independent trials tends to zero as $n \to \infty$, like a constant divided by $\sqrt{n}$. For fixed n, the approximation is always best for p near $\frac{1}{2}$, and worst for p close to 0 or 1 when the binomial distribution is skewed and the normal approximation not so accurate. In particular, if $p = \frac{1}{2}$, so $m = \frac{n}{2}$ if n is even, $\frac{n}{2} \pm \frac{1}{2}$ if n is odd,

$$P(m \text{ heads in } n \text{ fair coin tosses}) = \binom{n}{m} 2^{-n} \sim \sqrt{\frac{2}{n\pi}} \quad \text{as} \quad n \to \infty \tag{6}$$

As you can check on a pocket calculator, the asymptotic formula gives excellent results even for quite small values of n, and the relative error of the approximation decreases as n increases. According to the asymptotic formula, this relative error tends to 0 as $n \to \infty$. As $n \to \infty, 1/\sqrt{n} \to 0$, so the chance of getting exactly as many heads as tails tends to zero as the number of tosses tends to ∞.

To understand why this is so, recall the basis of the normal approximation. For large n the binomial (n, p) probabilities are distributed almost uniformly if you look close to the center of the distribution. The consecutive odds ratios are very close to one over an interval containing nearly all the probability. Still, these ratios conspire over larger distances to produce the gradual decreasing trend of the histogram away from its maximum, following the normal curve. By a distance of $4\sigma = 2\sqrt{n}$ or so from the center the histogram has almost vanished. And nearly all the probability must lie in this interval. Because a total probability of nearly 1 is distributed smoothly over an interval of length about $4\sqrt{n}$, the probabilities of even the most likely numbers in the middle cannot be much greater than $1/\sqrt{n}$. Thus even the most likely value m has a probability $P(m)$ which tends to zero as $n \to \infty$ like a constant over $\sqrt{n}$. See the exercises for another derivation of this, and a different evaluation of the constant, which leads to a remarkable infinite product formula for π.

Exercises 2.3

1. Suppose you knew the consecutive odds ratios $R(k) = P(k)/P(k-1)$ of a distribution $P(0), \ldots, P(n)$. Find a formula for $P(k)$ in terms of $R(1), \ldots, R(n)$. Thus the consecutive odds ratios determine a distribution.

2. A fair coin is tossed $10,000$ times. The probability of getting exactly 5000 heads is closest to:

0.001, 0.01, 0.1, 0.2, 0.5, 0.7, 0.9, 0.99, 0.999.

Pick the correct number and justify your choice.

3. Equalizations in coin tossing. Let $P(k \text{ in } n)$ be the probability of exactly k heads in n independent fair coin tosses. Let $n = 2m$ be even, and consider $P(m \text{ in } 2m)$, the chance of getting m heads and m tails in $2m$ tosses. Derive the following formulae:

a) $P(m-1 \text{ in } 2m) = P(m+1 \text{ in } 2m) = P(m \text{ in } 2m)\left(1 - \frac{1}{m+1}\right)$

b) $P(m+1 \text{ in } 2m+2) = \frac{1}{4}P(m-1 \text{ in } 2m) + \frac{1}{2}P(m \text{ in } 2m) + \frac{1}{4}P(m+1 \text{ in } 2m)$

c) By a) and b)

$$\frac{P(m+1 \text{ in } 2m+2)}{P(m \text{ in } 2m)} = 1 - \frac{1}{2(m+1)}$$

Check this also by cancelling factorials in the binomial formula.

d) By repeated application of c),

$$P(m \text{ in } 2m) = \left(1 - \frac{1}{2\times 1}\right)\left(1 - \frac{1}{2\times 2}\right)\cdots\left(1 - \frac{1}{2\times m}\right)$$

e) $0 < P(m \text{ in } 2m) < e^{-\frac{1}{2}\left(\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{m}\right)} < \dfrac{1}{\sqrt{m}}$

f) $P(m \text{ in } 2m) \to 0$ as $m \to \infty$. The bound of $1/\sqrt{m}$ is of the right order of magnitude, as shown by both the following calculations and the normal approximation. Let $\alpha_m = P(m \text{ in } 2m)$. Then verify the following:

$$\frac{(m+1/2)\alpha_m^2}{(m-1+1/2)\alpha_{m-1}^2} = 1 - \frac{1}{4m^2}$$

g)

$$2(m+1/2)\alpha_m^2 = \left(1 - \frac{1}{2^2}\right)\left(1 - \frac{1}{4^2}\right)\cdots\left(1 - \frac{1}{(2m)^2}\right)$$

$$= \frac{1}{2}\cdot\frac{3}{2}\cdot\frac{3}{4}\cdot\frac{5}{4}\cdot\frac{5}{6}\cdot\frac{7}{6}\cdots\frac{(2m-1)}{2m}\cdot\frac{(2m+1)}{2m}$$

h) $\alpha_m \sim K/\sqrt{m}$ as $m \to \infty$, where

$$2K^2 = 2\lim_{m\to\infty}\left(m + \frac{1}{2}\right)\alpha_m^2 = \frac{1}{2}\cdot\frac{3}{2}\cdot\frac{3}{4}\cdot\frac{5}{4}\cdot\frac{5}{6}\cdot\frac{7}{6}\cdots$$

Deduce by comparison with the normal approximation that the value of the infinite product is $2/\pi$.

2.4 Poisson Approximation

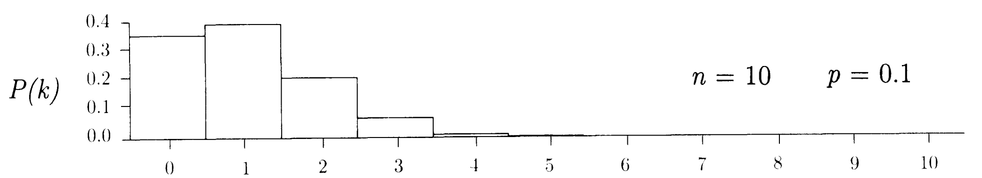
Even if n is very large, if p is close enough to 0 or 1 the standard deviation $\sigma = \sqrt{npq}$ is small. The binomial (n, p) distribution then does not follow the normal curve at all closely. By switching consideration from successes to failures, if necessary, we need only consider the case when p is nearly 0 and q is nearly 1. Then the standard deviation $\sigma = \sqrt{npq}$ is

$$\sigma = \sqrt{\mu q} \approx \sqrt{\mu} \qquad \text{where } \mu = np \text{ is the mean.}$$

If, for example, $\mu = 1$, so we are considering n trials with probability $p = 1/n$ of success on each trial, then $\sigma \approx 1$. The normal approximation will be very bad no matter how large n is. This is because the normal curve is symmetric, while the binomial distribution is not even approximately symmetric, due to the impossibility of negative values.

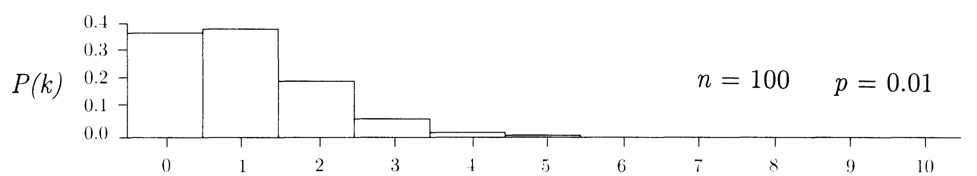
Example 1. The binomial (10, 1/10) distribution.

This is the distribution of the number of black balls obtained in 10 random draws with replacement from a box containing 1 black ball and 9 white ones.



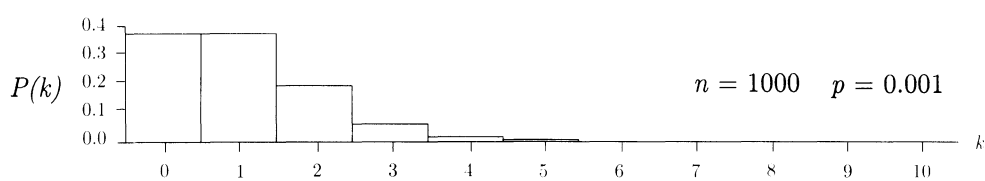
Example 2. The binomial (100, 1/100) distribution.

This is the distribution of the number of black balls obtained in 100 random draws with replacement from a box containing 1 black ball and 99 white ones.



Example 3. The binomial (1000, 1/1000) distribution.

Now take 1000 random draws with replacement from a box with 1 black ball and 999 white ones. This is the distribution of the number of black balls drawn:



As these examples show, binomial distributions with parameters n and $1/n$ are always concentrated on a small number of values near the mean value $\mu = 1$, with a shape which approaches a limit as $n \to \infty$ and $p = 1/n \to 0$. This limit corresponds to sampling more and more times with replacement from a box containing a smaller and smaller proportion of black balls. If μ, the expected number of black balls in the sample, is kept constant, the binomial (n, p) distribution with mean $\mu = np$ approaches a limit as $n \to \infty$ and $p \to 0$. This limit distribution, called the *Poisson distribution* with *parameter* μ, provides useful approximations to binomial probabilities in case n is large and p is so small that the normal approximation is bad.

The limit involved here is essentially the same as for the gambler's rule of Section 1.6. As in that example, the chance of getting zero successes in n trials with probability p of success on each trial is

$$P(0) = (1-p)^n \approx (e^{-p})^n = e^{-np} = e^{-\mu}$$

by the exponential approximation

$$1 - p \approx e^{-p} \qquad \text{if} \quad p \approx 0$$

It can be shown that no matter what the value of n, the error in this approximation to $P(0)$ is of the same order of magnitude as p. Consequently, this error tends to 0 as $p \to 0$, regardless of the value of n, and

$$P(0) \to e^{-\mu} \qquad \text{as } n \to \infty \text{ and } p \to 0 \text{ with } np \to \mu$$

To see what happens to the probability of k successes under the same conditions, look at the consecutive odds ratio:

$$R(k) = \frac{P(k)}{P(k-1)} = \frac{n-k+1}{k}\frac{p}{1-p} = \frac{np}{k}\frac{(1-(k-1)/n)}{1-p} \approx \frac{\mu}{k}$$

if n is large and p is small. In particular, if $\mu = 1$ as in the examples above, the first two odds ratios are

$$R(1) \approx 1/1 \qquad R(2) \approx 1/2$$

as apparent in the histograms. In the limit as $n \to \infty$ the binomial $(n, 1/n)$ distribution approaches a distribution with

$$P(0) = e^{-1}$$

and odds ratios $R(1) = 1$, $R(2) = 1/2$. This is the Poisson (μ) distribution defined below in case $\mu = 1$. More generally, for any fixed value of $\mu = np$, as $n \to \infty$ and $p \to 0$, the consecutive odds ratio $R(k)$ tends to μ/k, and

$$P(k) = P(0)R(1)R(2)\cdots R(k) \to e^{-\mu}\frac{\mu}{1}\cdot\frac{\mu}{2}\cdots\frac{\mu}{k} = e^{-\mu}\frac{\mu^k}{k!}$$

To summarize, we have the following:

Poisson Approximation to the Binomial Distribution

If n is large and p is small, the distribution of the number of successes in n independent trials is largely determined by the value of the mean $\mu = np$, according to the *Poisson approximation*

$$P(k \text{ successes}) \approx e^{-\mu}\frac{\mu^k}{k!}$$

Remark. It can be shown that the accuracy of the approximation depends largely on the value of p, and hardly at all on the value of n. Roughly speaking, absolute errors in using this approximation will be of the same order of magnitude as p.

Example 4. Defectives in a sample.

Problem 1. Suppose that over the long run a manufacturing process produces 1% defective items. What is the chance of getting two or more defective items in a sample of 200 items produced by the process?

Solution. Assume each item is defective with probability p, independently of other items. The long-run percentage of defectives would then be $100p\%$, so we can estimate $p = 1/100$. The number of defectives in a sample size of 200 then has binomial $(200, 1/100)$ distribution, with mean $\mu = 200 \times 1/100 = 2$. Using the Poisson approximation

$$\begin{aligned} P(2 \text{ or more defectives}) &= 1 - P(0) - P(1) \\ &\approx 1 - e^{-2}\frac{2^0}{0!} - e^{-2}\frac{2^1}{1!} \\ &= 1 - 3e^{-2} = 0.594 \end{aligned}$$

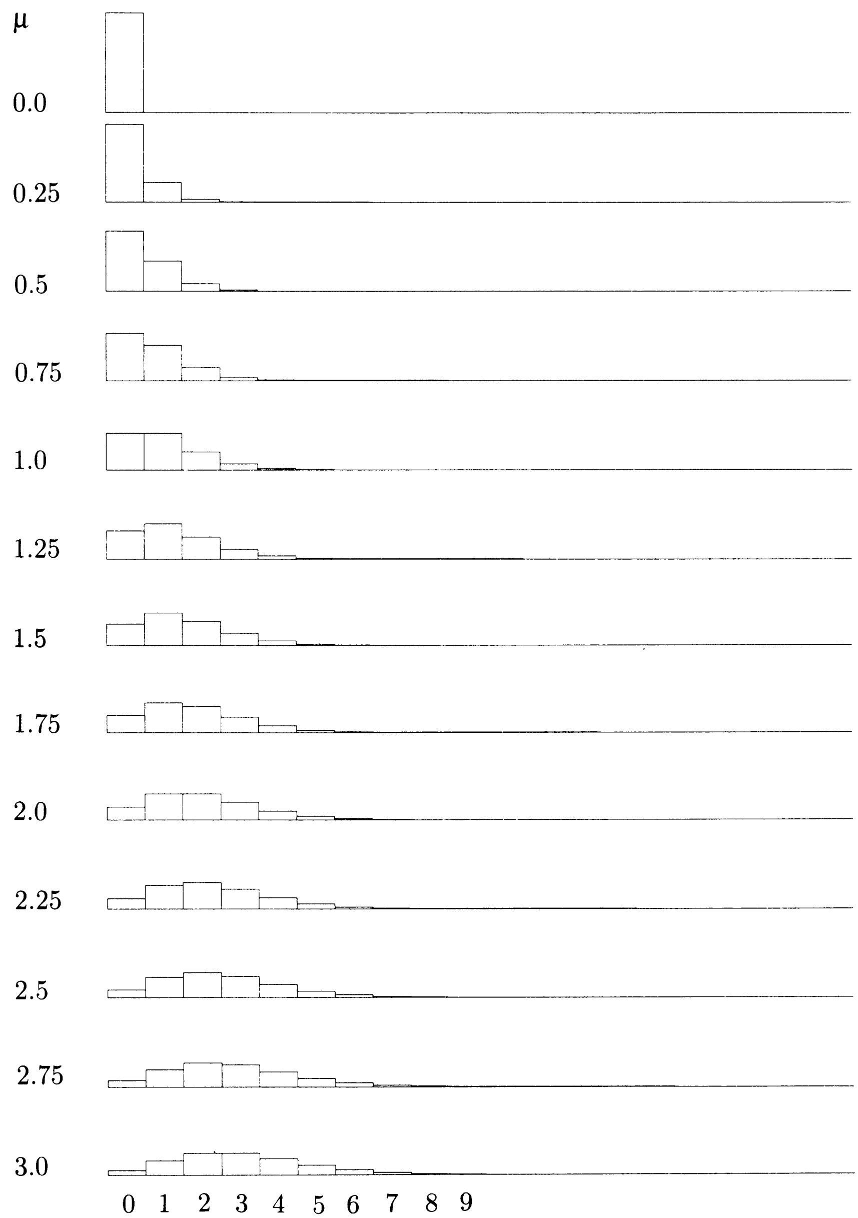
A check on the Poisson approximation. As a check on the approximation

$$P(k \text{ successes in } n \text{ trials}) \approx e^{-\mu}\mu^k/k! \qquad \text{where } \mu = np,$$

sum both sides from $k = 0$ to n to obtain

$$1 \approx e^{-\mu}\sum_{k=0}^{n}\frac{\mu^k}{k!}$$

FIGURE 1. **Poisson distributions.** Notice how when μ is small the distribution is piled up on values near zero. As μ increases, the distribution shifts to the right and spreads out, gradually approaching the normal distribution in shape as $\mu \to \infty$. This can be shown by a variation of the argument in Section 2.3.



For fixed μ, as $n \to \infty$ and $p = \mu/n \to 0$, this approximation becomes better and better. The limit of the sum is a well-known formula for e^{μ}:

$$\lim_{n\to\infty} \sum_{k=0}^{n} \frac{\mu^k}{k!} = \sum_{k=0}^{\infty} \frac{\mu^k}{k!} = e^{\mu}$$

and

$$e^{-\mu}e^{\mu} = e^{-\mu+\mu} = e^0 = 1$$

See Appendix 4 for further details.

This calculation does show that the limiting probabilities $P_\mu(k) = e^{-\mu}\mu^k/k!$ form a probability distribution on $\{0, 1, 2, \ldots\}$, meaning that

$$P_\mu(k) \geq 0 \text{ and } \sum_{k=0}^{\infty} P_\mu(k) = 1.$$

This kind of distribution over an infinite set of possible values is discussed more generally in Section 3.4. More about the Poisson distribution can be found in Section 3.5.

The Poisson (μ) Distribution

The *Poisson distribution with parameter* μ or *Poisson* (μ) *distribution* is the distribution of probabilities $P_\mu(k)$ over $\{0, 1, 2, \ldots\}$ defined by

$$P_\mu(k) = e^{-\mu}\mu^k/k! \qquad (k = 0, 1, 2, \ldots)$$

Exercises 2.4

1. Sketch the histograms of binomial distributions with the following parameters (n, p):

a) $(10^6, 10^{-6})$; b) $(10^6, 2 \times 10^{-6})$; c) $(3284, 10^{-4})$; d) $(1000, 0.998)$.

2. Find Poisson approximations to the probabilities of the following events in 500 independent trials with probability 0.02 of success on each trial:

a) 1 success; b) 2 or fewer successes; c) more than 3 successes.

3. The chance of getting 25 or more sixes in 100 rolls of a die is 0.022. If you rolled 100 dice once every day for a year, find the chance that you would see 25 or more sixes:

a) at least once; b) at least twice.

4. Repeat the previous problem for the event of getting 30 or more sixes in 100 die rolls, which has probability 0.00068.

5. Suppose that each week you buy a ticket in a lottery which gives you a chance of $1/100$ of a win. You do this each week for a year. What is the chance that you get k wins during the year, approximately? Calculate as a decimal for $k = 0, 1, 2$.

6. A box contains 1000 balls, of which 2 are black and the rest are white.

a) Which of the following is most likely to occur in 1000 draws with replacement from the box?

fewer than 2 black balls, exactly 2 black balls, more than 2 black balls

b) If two series of 1000 draws are made at random from this box, what, approximately, is the chance that they produce the same number of black balls?

7. Let X be the number of successes in 25 independent trials with probability $1/10$ of success on each trial. Let m be the most likely value of S.

a) Find m.

b) Find $P(S = m)$ correct to 3 decimal places.

c) What is the value of the normal approximation to $P(S = m)$?

d) What is the value of the Poisson approximation to $P(S = m)$?

e) Repeat a) for $n = 2500$ trials instead of 25. Which would now give the better approximation to $P(S = m)$, the normal or the Poisson approximation? Find $P(S = m)$ approximately using the best approximation.

f) Repeat e) for 2500 trials and $p = 1/1000$ instead of $p = 1/10$.

8. **Mode of the Poisson distribution.** Use consecutive odds ratios to find the largest k that maximizes the Poisson (μ) probability $P_\mu(k)$. For what values of μ is there a double maximum? What are the two values of k in that case? Is there ever a triple maximum?

9. A cereal company advertises a prize in every box of its cereal. In fact, only about 95% of their boxes have prizes in them. If a family buys one box of this cereal every week for a year, estimate the chance that they will collect more than 45 prizes. What assumptions are you making?

10. Let N be a fixed large integer. Consider n independent trials, each of which is a success with probability $1/N$. Recall that the gambler's rule (see Example 1.6.3) says that if $n \approx \frac{2}{3}N$, the chance of at least one success in n trials is about $1/2$. Show that if $n \approx \frac{5}{3}N$, then the chance of at least two successes is about $1/2$.

2.5 Random Sampling

Random sampling is a statistical technique for gaining information about the composition of a large population from the composition of a random sample from the population. Suppose that each element of the population can be classified into one of two categories, say "good" and "bad". Of course, the designation of which elements are good is quite arbitrary, and will depend on the problem at hand. In practical problems the fraction of good elements in the population will be unknown. The problem is to estimate this fraction based on the composition of the sample, and to know how accurate this estimate is likely to be. The natural estimate of the fraction of good elements in the population is the fraction of good elements in the sample. That is to say, population percentages are estimated by sample percentages. The accuracy of this estimate depends on exactly what procedure was used to obtain the sample. The ideal is to obtain a sample that is as representative as possible of the whole population. This ideal is approached by picking the sample at random. Provided the sample size is large enough, the proportion in the sample will most likely be close to the proportion in the population.

Sampling with Replacement

Suppose n individuals are drawn one by one at random from a population of size N, with replacement between draws. On each draw it is assumed that each of the N individuals has the same chance of being chosen, and the successive draws are assumed independent. So all N^n possible sequences of choices are equally likely. This might be done, for example, by drawing tickets from a box, with replacement of the tickets and mixing between draws. There is no restriction on the sample size n. In principle, the procedure can be repeated indefinitely.

Consider now the distribution of the number of good elements in a sample of size n with replacement from a population of G good and B bad elements, with $G + B = N$. This is the distribution of the number of successes in n independent trials, with probability $p = G/N$ of success in each trial, that is to say the binomial (n, p) distribution for $p = G/N$. Provided the sample size n is large enough, this binomial distribution with parameters n and $p = G/N$ is well approximated by the normal curve with parameters $\mu = np$ and $\sigma = \sqrt{npq}$. According to the law of large numbers, if n is sufficiently large, the proportion of good elements in the sample is likely to be close to the proportion $p = G/N$ of good elements in the population. By the normal approximation, if n is sufficiently large, the number of good elements in the sample will lie in the range $np \pm 2\sqrt{npq}$ with probability about 95%. So if n is sufficiently large, the proportion of good elements in the sample will lie in the range $p \pm 2\sqrt{pq/n}$ with probability about 95%. Since $\sqrt{pq} \leq 1/2$, this means that

$$P(p - 1/\sqrt{n} \leq \text{sample proportion} \leq p + 1/\sqrt{n}) \geq 95\%$$

If the proportion of good elements in a population is not known, the result above can be used to estimate the unknown proportion by the method of confidence intervals. If the sample size is large, then with probability greater than 95% the sample proportion of good elements will lie within $1/\sqrt{n}$ of the population proportion. So if the observed proportion of good elements in a large sample is $\hat{p}$, guess that the population proportion lies in the range $\hat{p} \pm 1/\sqrt{n}$. The interval $\hat{p} \pm 1/\sqrt{n}$ is an *approximate* 95% *confidence interval* for the unknown population proportion.

Sampling Without Replacement

In this procedure, elements in a population of size N are drawn one by one at random as before, but without replacement between draws. The sample size n is now restricted to $n \leq N$. At each stage it is assumed that no matter what elements have been drawn so far, all remaining elements are equally likely on the next draw. Equivalently, all possible orderings of n of the N elements are assumed equally likely.

The number of different possible orderings of n out of N elements is denoted by $(N)_n$, a symbol which can be read "N *order* n". As explained in Appendix 1, the product rule for counting gives the formula

$$(N)_n = N(N-1)\cdots(N-n+1)$$

where there are n factors in the product. Compare with N to the power n:

$$N^n = N \cdot N \cdots N \qquad (n \text{ factors})$$

which is the larger number of possible samples with replacement, and N choose n:

$$\binom{N}{n} = (N)_n/n!$$

which is the smaller number of different unordered samples or subsets of size n. This is just the formula for $\binom{n}{k}$ of Section 2.1 with N instead of n and n instead of k. When rewritten in the form

$$(N)_n = \binom{N}{n} n!$$

this formula can be understood as follows: Each of the $\binom{N}{n}$ possible unordered samples of size n can be ordered in $n!$ different ways to obtain $n!$ different ordered samples of size n. Thus $(N)_n$, the number of ordered samples of size n, is $\binom{N}{n}$ times $n!$ by the product rule of counting.

Consider now the distribution of the number of good elements in a sample of size n without replacement from a population of G good and B bad elements with $G+B=N$. The problem is to find the chance of getting g good and b bad elements in the sample, for $0 \leq g \leq n$ and $g+b=n$. Thinking in terms of an ordered random sample, one way to get g good and b bad in the sample is if the first g elements in the sample are good and the last b are bad. Either by the product rule for conditional probabilities, or by the product rule for counting, the chance of this event is

$$\frac{G}{N}\cdot\frac{G-1}{N-1}\cdots\frac{G-g+1}{N-g+1}\cdot\frac{B}{N-g}\cdot\frac{B-1}{N-g-1}\cdots\frac{B-b+1}{N-g-b+1}=\frac{(G)_g(B)_b}{(N)_n}$$

This is the chance of just one of $\binom{n}{g}$ different possible patterns of g good and b bad elements in an ordered sample of size n. But the chance of any other pattern of g good and b bad, for example, the first b elements bad and the next g elements good, is just the same, because the same factors then appear in a different order. Thus, multiplying the above expression by $\binom{n}{g}$ gives the chance of g good and b bad elements appearing in an unspecified pattern, as in the second formula of the following box:

Sampling With and Without Replacement

Suppose a population of size N contains G good and B bad elements, with $N = G + B$. For a sample of size $n = g + b$, where $0 \le g \le n$, the probability of getting g good elements and b bad elements is

- **for sampling with replacement**

$$P(g \text{ good and } b \text{ bad}) = \binom{n}{g}\frac{G^g B^b}{N^n}$$

- **for sampling without replacement**

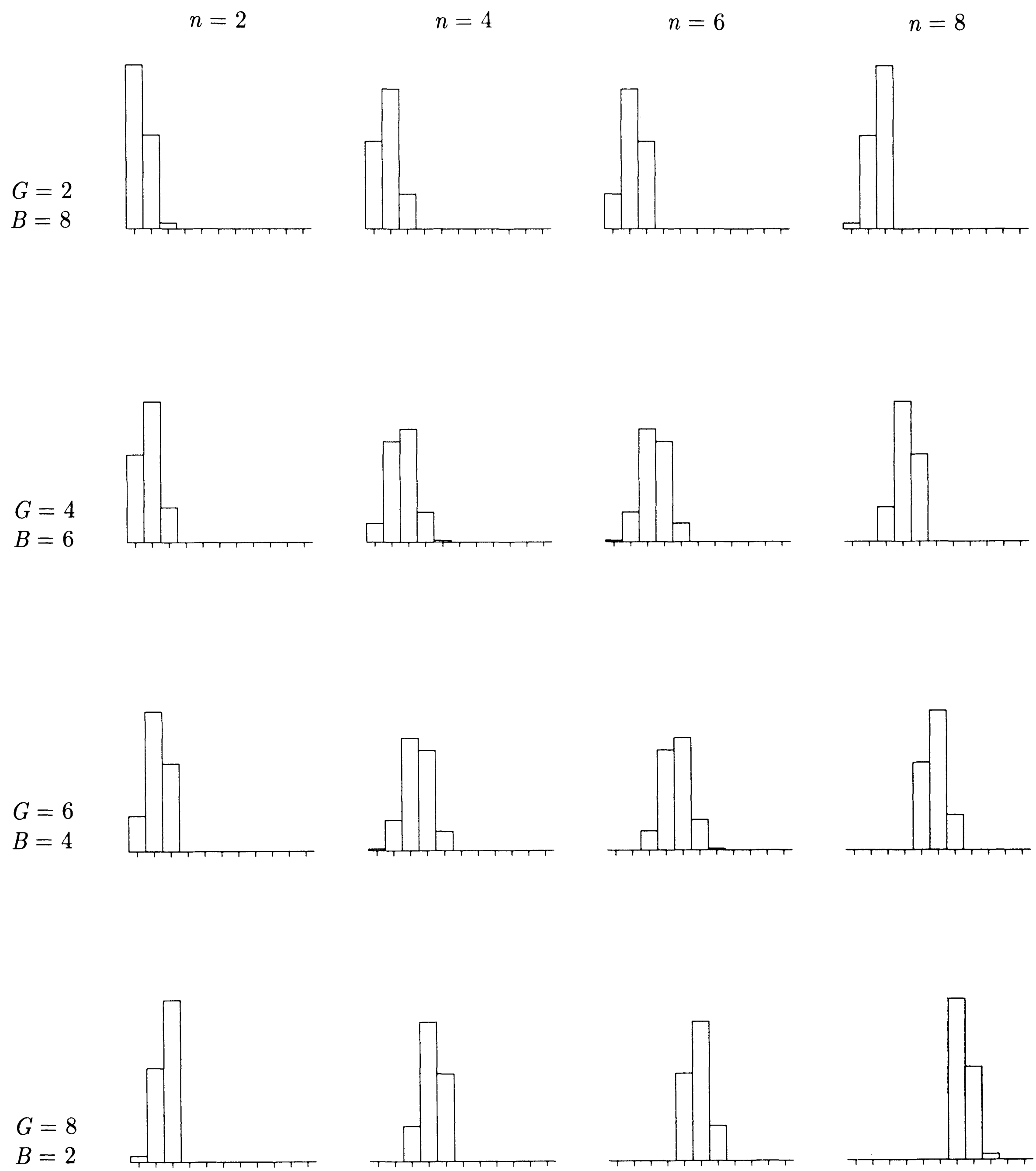
$$P(g \text{ good and } b \text{ bad}) = \binom{n}{g}\frac{(G)_g(B)_b}{(N)_n} = \frac{\binom{G}{g}\binom{B}{b}}{\binom{N}{n}}$$

The formula for sampling with replacement is just the usual binomial formula written in a way that parallels with the first formula for sampling without replacement. The second formula for sampling without replacement follows from the first by cancellation after using the formula $\binom{M}{m} = (M)_m/m!$ three times. This expression can also be derived another way, by working in the outcome space of all $\binom{N}{n}$ possible unordered samples. Since there are $n!$ ordered samples corresponding to each unordered sample, each possible unordered sample has the same chance

$$\frac{n!}{(N)_n} = 1\Big/\binom{N}{n}$$

And $\binom{G}{g}\binom{B}{b}$ is the number of possible unordered samples with g good and b bad elements, by yet another application of the product rule of counting. The good

FIGURE 1. **Some hypergeometric distributions.** The histograms display the distribution of the number of good elements in a sample of size n without replacement from a population of $N = 10$ elements, containing G good elements and $B = 10 - G$ bad ones, for $n = 2, 4, 6, 8$ (different columns) and $G = 2, 4, 6, 8$ (different rows). Each horizontal scale is marked by ticks at $0, 1, \ldots, 10$



elements can be chosen in $\binom{G}{g}$ ways, and no matter how these are chosen, the bad ones may be chosen in $\binom{B}{b}$ ways. This method of counting unordered samples is what is used to calculate the probabilities of various poker hands. See Exercise 12.

The hypergeometric distribution. This is the name of the distribution of the number of good elements in a sample of size n without replacement from a population of G good and $N-G$ bad elements. The distribution has three parameters, n, N and G. The probability that this distribution assigns to $g \in \{0, 1, \ldots, n\}$ is the probability $P(g \text{ good and } b \text{ bad})$ for sampling without replacement, as in the box, for $b = n - g$ and $B = N - G$. Note that this probability may be zero for some g between 0 and n. (See Exercise 11). The fact that these probabilities add up to 1, and so define a distribution on $\{0, 1, \ldots, n\}$, is not obvious from the formula, but it follows at once from the rules of probability: as g varies from 0 to n the events of getting g good elements and b bad elements in sampling without replacement form a partition of the whole outcome space.

Binomial approximation to the hypergeometric distribution. If N, G, and B are large in comparison to n, g, and b, the formulae for sampling with and without replacement give nearly identical probabilities. More precisely, for fixed n, b, and g, and $N \to \infty$, $G \to \infty$, and $B \to \infty$, the ratio of the two probabilities tends to 1. This follows from the fact that for any fixed n,

$$(N)_n/N^n \to 1 \quad \text{as} \quad N \to \infty$$

In practice, this makes the binomial distribution a useful approximation to the more complicated hypergeometric distribution. The approximation is quite intuitive, because if the sample size is small in comparison to the population size there is very little chance of a duplicate in sampling with replacement. The chance of getting a duplicate in sampling with replacement is just $1 - (N)_n/N^n \approx 0$ if $n \ll \sqrt{N}$ (see the birthday problem of Section 1.6). And given that there are no duplicates, the sample with replacement is just like a sample without replacement, in the sense that all orderings are equally likely.

Normal approximation to the hypergeometric distribution. This is discussed in Section 3.6.

Exercises 2.5

1. Suppose you take a random sample of 10 tickets without replacement from a box containing 20 red tickets and 30 blue tickets.

a) What is the chance of getting exactly 4 red tickets?

b) Repeat a) for sampling with replacement.

2. Three cards are dealt from a standard deck of 52 cards, containing 26 red cards and 26 black cards. Write down the probability that:

a) the first card is red and the second two black;

b) exactly one of the cards dealt is red;

c) at least one of the cards dealt is red.

3. A deck of cards is shuffled and dealt to four players, with each receiving 13 cards. Find:

a) the probability that the first player holds all the aces;

b) the probability that the first player holds all the aces given that she holds the ace of hearts;

c) the probability that the first player holds all the aces given that she holds at least one;

d) the probability that the second player holds all the aces given that he holds all the hearts.

4. A population of $100,000$ people consists of 40% men and 60% women. A random sample of size 100 is drawn from this population without replacement. Write down an expression for the probability that there are at least 45 men in the sample. Approximately what is the value of this probability?

5. Suppose 55% of a large population of voters actually favor candidate A. How large a random sample must be taken for there to be a 99% chance that the majority of voters in the sample will favor candidate A?

6. In a hand of 13 cards drawn randomly from a pack of 52, find the chance of:

a) no court cards (J, Q, K, A);

b) at least one ace but no other court cards;

c) at most one kind of court card.

7. A box contains 50 black balls and 30 red balls. Four balls are drawn at random from the box, one after the other, without replacement. Find the chance that:

a) all four balls are black;

b) exactly three balls are black;

c) the first red ball appears on the last draw.

8. In a raffle with 100 tickets, 10 people buy 10 tickets each. If there are 3 winning tickets drawn at random find the probability that:

a) one person gets all 3 winning tickets;

b) there are 3 different winners;

c) some person gets two winners and someone else gets just one.

9. A lot of 50 items is inspected by the following two-stage plan.

(i) A first sample of 5 items is drawn without replacement. If all are good the lot is passed; if two or more are bad the lot is rejected.

(ii) If the first sample contains just one bad item, a second sample of 10 more items is drawn without replacement (from the remaining 45 items) and the lot is rejected if two or more of these are bad. Otherwise it is accepted.

Suppose there are 10 bad items in the lot.

a) What is the probability that the second sample is drawn and contains more than one bad item?

b) Write down an expression for the probability that the lot is accepted.

10. Suppose a population of N elements consists of G good, B bad, and I indifferent elements, with $B + G + I = N$. If a random sample of size n is drawn with replacement from this population, explain why the chance that the sample contains k_1 good elements, k_2 bad elements, and k_3 indifferent elements, where $k_1 + k_2 + k_3 = n$, is

$$\frac{n!}{k_1!k_2!k_3!}(G/N)^{k_1}(B/N)^{k_2}(I/N)^{k_3}$$

11. Range of the hypergeometric distribution. For $1 \le n \le N$ and $0 \le G \le N$, describe the set of g with $0 \le g \le n$ such that there is strictly positive probability of getting g good elements in a random sample of size n without replacement from a population of G good and $N - G$ bad elements. Explain why the formula for the probability in question gives the correct value, (possibly 0) for all $0 \le g \le n$.

12. Poker hands. Assume all $\binom{52}{5}$ hands equally likely. Find the probability of being dealt:

a) a straight flush (5 consecutive cards of the same suit);

b) four of a kind (ranks a, a, a, a, b);

c) a full house (ranks a, a, a, b, b);

d) a flush (5 of the same suit, not a straight flush);

e) a straight (5 consecutive ranks, not a flush);

f) three of a kind (ranks a, a, a, b, c);

g) two pairs (ranks a, a, b, b, c);

h) a pair (ranks a, a, b, c, d);

i) none of the above.

13. A factory which produces chips in lots of ten thousand uses the following scheme to check the quality of its product. From each lot of chips produced, a random sample of size 500 is taken. If the sample contains 10 or less defectives, the lot is passed. If the sample contains more than 10 defectives, another random sample of size 500 is chosen from the lot. If this sample contains 10 or less defectives, the lot is passed. Otherwise, the lot is rejected. If a lot actually contains 5% defectives, find the chance that it will pass. [Approximate by sampling with replacement, and use the normal curve.]

Repeated Trials and Sampling: Summary

Binomial Probability Formula

$$P(k \text{ successes in } n \text{ trials}) = \binom{n}{k} p^k q^{n-k} \qquad \text{for independent trials with}$$

$p =$ probability of success on each trial,

$q = 1 - p =$ probability of failure on each trial.

For fixed n, as k varies from 0 to n, these probabilities define the *binomial* (n, p) *distribution* on $\{0, 1, \ldots, n\}$. That the probabilities add to 1 amounts to the

Binomial Theorem: $(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$

Here, $$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}$$

$=$ binomial coefficient called n *choose* k

$=$ number of ways to pick k places out of n

$=$ number of subsets of k of a set of n

$=$ number in row n, column k of Pascal's triangle

Note: $\binom{n}{n} = \binom{n}{0} = 1$

Recursion Formula for Pascal's Triangle

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \qquad (\text{for } 0 < k < n, \quad n = 1, 2, \ldots)$$

Symmetry of Pascal's Triangle

$$\binom{n}{k} = \binom{n}{n-k}$$

Consecutive Ratios in Pascal's Triangle

$$\binom{n}{k} \Big/ \binom{n}{k-1} = \frac{n-k+1}{k}$$

Consecutive Ratios in the Binomial (n, p) Distribution

$$R(k) = \frac{P(k)}{P(k-1)} = \frac{(n-k+1)}{k}\frac{p}{q}$$

Mode of Binomial (n, p) Distribution: m = most likely value = $\text{int}(np + p)$

Normal Approximation to the Binomial Distribution

$$P(k) \approx \frac{1}{\sigma}\phi\left(\frac{k-\mu}{\sigma}\right)$$

where $\mu = np$ is the *mean*,

$\sigma = \sqrt{npq}$ is the *standard deviation*,

$z = (k-\mu)/\sigma$ is k in *standard units*,

$\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$ is the *standard normal density function.*

$$P(a \text{ to } b) \approx \Phi\left(\frac{b+\frac{1}{2}-\mu}{\sigma}\right) - \Phi\left(\frac{a-\frac{1}{2}-\mu}{\sigma}\right)$$

where $\Phi(z) = \int_{-\infty}^{z} \phi(x)dx$ is the *standard normal c.d.f.*

This approximation should be used only if $\sigma \geq 3$. The larger σ, the better.

$$\Phi(-z) = 1 - \Phi(z)$$
$$\Phi(a, b) = \Phi(b) - \Phi(a)$$
$$\Phi(-b, b) = 2\Phi(b) - 1$$
$$P(\mu - \sigma \text{ to } \mu + \sigma \text{ success in } n \text{ trials}) \approx \Phi(-1, 1) \approx 68\%$$
$$P(\mu - 2\sigma \text{ to } \mu + 2\sigma \text{ success in } n \text{ trials}) \approx \Phi(-2, 2) \approx 95\%$$
$$P(\mu - 3\sigma \text{ to } \mu + 3\sigma \text{ success in } n \text{ trials}) \approx \Phi(-3, 3) \approx 99.7\%$$

Square Root Law for Independent Trials: The deviation from the expected number of successes np will most likely be a small multiple of $\sigma = \sqrt{npq} \leq \frac{1}{2}\sqrt{n}$.

$$P(p - \tfrac{1}{\sqrt{n}} \leq \text{sample proportion} \leq p + \tfrac{1}{\sqrt{n}}) \geq 95\% \qquad \text{for large } n.$$

Poisson Approximation to the Binomial Distribution

If p is close to zero

$$P(k) \approx e^{-\mu}\mu^k/k! \qquad \text{where } \mu = np$$

Random Sampling: See box on page 125.

Review Exercises

1. Ten dice are rolled. Write down numerical expressions for

a) the probability that exactly 4 dice are sixes.

b) the probability that exactly 4 dice are sixes given that none of the dice is a five.

c) the probability of 4 sixes, 3 fives, 2 fours, and a three.

d) the probability that none of the first three dice is a six given 4 sixes among the ten dice.

2. A fair die is rolled 36 times. Approximate the probability that 12 or more sixes appear.

3. Suppose I roll a fair die, then toss as many coins as there are spots on the die.

a) What is the probability that exactly three heads appear among the coins?

b) Given three heads appear, what is the probability that the die showed 4?

4. A fair coin is tossed 10 times. Given that at least 9 of the tosses resulted in tails, what is the probability that exactly 9 of the tosses resulted in tails?

5. A thumb tack was tossed 100 times, and landed point up on 40 tosses and point down on 60 tosses. Given this information, what is the probability that the first three tosses landed point down?

6. Four numbers are drawn at random from a box of ten numbers $0, 1, \ldots, 9$. Find the probability that the largest number drawn is a six:

a) if the draws are made with replacement;

b) if the draws are made without replacement.

7. 10^6 fair coins are tossed. Find a number k such that the chance that the number of heads is between $500,000 - k$ and $500,000 + k$ is approximately 0.96.

8. Suppose you and I each roll ten dice. What is the probability that we each roll the same number of sixes?

9. In a certain town, 10% of the families have no children, 10% have one child, 40% have two children, 30% have three children, and 10% have four children. Assume that births are independent of each other, and equally likely to produce male or female.

a) One family is picked at random from all of the families in this town. What is the probability that there are at least two children in the family?

b) One family is picked at random from all of the families in this town. Guess the size of the family, given that it has at least two girls. Give reasons for your guess.

c) A family is picked at random from among the families with four children. Then a child is picked at random from the selected family. What is the chance that the child picked is a girl with at least one brother?

10. **Lie detectors.** According to a newspaper report, in 2 million lie detector tests, $300,000$ were estimated to have produced erroneous results. Assuming these figures to be correct, answer the following:

a) If ten tests were picked at random from these 2 million tests, what would be the chance that at least one of them produced an erroneous result? Sketch the histogram of the distribution of the number of erroneous results among these ten tests.

b) Suppose these 2 million tests were done on a variety of machines. If a machine were picked at random, then ten tests picked at random from these tests performed on that machine, would it be reasonable to suppose that the chance that at least one of them produced an erroneous result would be the same as in a)? Explain.

11. Consider two machines, A and B, each producing the same items. Each machine produces a large number of these items every day. However, production per day from machine B, being newer, is twice that of A. Further the rate of defectives is 1% for B and 2% for A. The daily output of the machines is combined and then a random sample of size 12 taken. Find the probability that the sample contains 2 defective items. What assumptions are you making?

12. In poker, a hand containing face values of the form (x, x, y, z, w) is called one pair.

a) If I deal a poker hand, what is the probability that I get one pair?

b) I keep dealing independent poker hands. Write an expression for the probability that I get my 150th 'one pair' on or after the 400th deal.

c) Approximately what is the value of the probability in b)?

13. A seed manufacturer sells seeds in packets of 50. Assume that each seed germinates with a chance of 99%, independently of all others. The manufacturer promises to replace, at no cost to the buyer, any packet that has 3 or more seeds that do not germinate. What is the chance that the manufacturer has to replace more than 40 of the next 4000 packets sold?

14. a) If Ted and Jim are among 10 people arranged randomly in a line, what is the chance that they stand next to each other?

b) What if the ten people are arranged at random in a circle?

c) Generalize to find the chance of k particular people ending up all together if n people are arranged at random in a line or a circle.

15. Draws are made at random with replacement from a box of colored balls with the following composition:

color	red	blue	green	yellow
proportion	0.1	0.2	0.3	0.4

Write down and justify unsimplified expressions for the probabilities of the following events:

a) exactly 5 yellow balls appear in 20 draws;

b) exactly 2 red, 4 blue, 6 green and 8 yellow balls appear in 20 draws;

c) the number of draws required to produce 3 red balls is 25.

16. Eight cards are drawn from a well-shuffled deck of 52 cards. What is the probability that the 8 cards contain: a) 4 aces; b) 4 aces and 4 kings;

c) 4 of a kind (any kind, including the possibility of 4 of two kinds).

17. If four dice are rolled, what is the probability of:

a) four of a kind; b) three of a kind; c) two pairs?

18. Seven dice are rolled. Write down unsimplified expressions for the probabilities of each of the following events:

a) exactly three sixes;

b) three of one kind and four of another;

c) two fours, two fives, and three sixes; d) each number appears;

e) the sum of the dice is 9 or more.

19. In a World Series, teams A and B play until one team wins four games. Suppose all games are independent, and that on each game, the probability that team A beats team B is $2/3$.

a) What is the probability that team A wins the series in four games?

b) What is the probability that team A wins the series, given team B won games 1 and 2?

20. A computer communication channel transmits words of n bits using an error-correcting code which is capable of correcting errors in up to k bits. Here each bit is either a 0 or a 1. Assume each bit is transmitted correctly with probability p and incorrectly with probability q independently of all other bits.

a) Find a formula for the probability that a word is correctly transmitted.

b) Calculate the probability of correct transmission for $n = 8$, $k = 2$, and $q = 0.01$.

21. Suppose a single bit is transmitted by repeating it n times and the message is interpreted by majority decoding. For example, for $n = 5$, if the message received is 10010, it is concluded that a 0 was sent. Assuming n is odd and each bit in the message is transmitted correctly with probability p, independently of the other bits, find a formula for the probability that the message is correctly received.

22. Suppose that, on average, 3% of the purchasers of airline tickets do not appear for the departure of their flight. Determine how many tickets should be sold for a flight on an airplane which has 400 seats, such that with probability 0.95 everybody who appears for the departure of the flight will have a seat. What assumptions are you making?

23. Ten percent of the families in a town have no children, twenty percent have one child, forty percent have two children, twenty percent have three, and ten percent have four. Assume each child in a family is equally likely to be a boy or a girl, independently of all the others. A family is picked at random from this town. Given that there is at least one boy in the family, what is the chance that there is also at least one girl?

24. In a large population, the distribution of the number of children per family is as follows:

Number of children n	0	1	2	3	4	5
Proportion families with n children	0.15	0.2	0.3	0.2	0.1	0.05

Assume that each child in a family is a boy or a girl with probability $1/2$, independently.

a) If a family is picked at random, what is the chance that it contains exactly two girls?

b) If a child is picked at random from the children of this population, what is the chance that the child comes from a family with exactly two girls?

25. At Wimbledon, men's singles matches are played on a "best of five sets" basis, that is, players A and B play until one of them has won 3 sets. Suppose each set is won by A with probability p, independently of all previous sets.

a) For each $i = 3, 4, 5$, find a formula in terms of p and $q = 1 - p$ that player A wins in exactly i sets.

b) In terms of p and q, what is the probability that player A wins the match?

c) Given that player A won the match, what is the probability (in terms of p and q) that the match lasted only three sets?

d) Compute the probability in c) for the case $p = 2/3$.

e) Do you think the assumption of independence made above is reasonable?

26. Suppose 3 points are picked at random from 10 points equally spaced around the circumference of a circle.

a) What is the probability that two particular adjacent points, say A and B, are both among the 3 points picked at random?

b) What is the probability that among the 3 points picked at random there is least one pair of adjacent points?

27. A university schedules its final examinations in 18 "examination groups", so that courses held at different times are in different examination groups. The examination times are spread over 6 days, with 3 examinations each day. Suppose all students take 4 examinations. About what proportion of students will have their 4 examinations on different days? [You need to make some assumptions—state what the assumptions are.]

28 **The matching problem.** There are n letters addressed to n people at n different addresses. The n addresses are typed on n envelopes. A disgruntled secretary shuffles the letters and puts them in the envelopes in random order, one letter per envelope.

a) Find the probability that at least one letter is put in a correctly addressed envelope. [*Hint*: Use the inclusion-exclusion formula of Exercise 1.3.12]

b) What is this probability approximately, for large n?

29. **Cosmic wimpout.** In this game five dice are rolled. Four of the dice have the same set of symbols and numbers on their faces. The numbers are 5 and 10, and let us call the symbols A, B, C, and D. The fifth die is the same, except symbol D is replaced by a different symbol W, indicating a wild roll. In one version of the game, the following kinds of rolls count for a score:

- any roll that shows one or more numbers;
- any roll that shows a triple of symbols, where the wild symbol W can count as any symbol you like, e.g., WAABC scores a triple, the W counting as A;
- a roll that shows W together with one of each of the other symbols A, B, C, and D.

Any other combination fails to score, and is called a wimpout. Calculate the probability of a wimpout.

30. Stirling's formula. Use logarithms and calculus to derive an approximation of the form

$$n! \sim C\left(\frac{n}{e}\right)^n \sqrt{n}$$

for some constant C. Now compare with the normal approximation to the probability of m heads and m tails in $2m$ fair coin tosses to deduce that $C = \sqrt{2\pi}$.

31. The normal approximation works reasonably well whenever the area under the normal curve over the range of the binomial distribution is close to one. Show that if $\sqrt{npq} \geq 3$, then at least 99% of the area under the normal curve is between 0 and n by showing:

a) $np - 3\sqrt{npq} \geq 0$; b) $nq - 3\sqrt{npq} \geq 0$; c) $0 \leq np \pm 3\sqrt{npq} \leq n$.

32. Call a card hand of h cards a *straight* if the denominations can be arranged as $d, d+1, \ldots, d+h$, for some $1 \leq d \leq 13 - h$, where $d = 1$ represents ace, $d = 11$ for jack, 12 for queen, and 13 for king (so aces only count low). Call the hand a *flush* if all h cards are of the same suit. Assume for simplicity that a straight flush counts both as a straight and as a flush. For which h is a straight more likely than a flush?

33. a) How could you simulate a biased coin landing heads with probability $p = 1/3$ if you only had available a fair coin?

b) How could you simulate fair coin tossing if you only had available a coin with unknown bias p strictly between 0 and 1?

34. a) Explain why if you and I each toss m fair coins, the chance that we both get the same (unspecified) number of heads equals the chance that we get exactly m heads and m tails between us.

b) If I toss m fair coins, and you toss $m+1$ fair coins, what is the chance that you get strictly more heads than I do?

35. At roulette, the chance of winning a bet on a single number is $1/38$.

a) Write down a numerical expression for the chance of winning between 20 and 35 bets (inclusive) out of 1000 bets on a single number. Do not evaluate this expression.

b) Should the normal curve be used to approximate the chance in a)? (Give a reason.) If yes, find the normal approximation. If no, use some better method of approximation.

36. An efficient way of computing probabilities in the binomial (n, p) distribution to any desired degree of accuracy is to use the following method. Let $m = \text{int}(np + p)$, and fix some small number $\epsilon > 0$. Starting from $H(m) = 1$, find the histogram heights

$H(k) = P(k)/P(m)$ for $k = m+1, m+2, \ldots$ by repeatedly multiplying consecutive odds ratios, until $k = b$ say, the least $k > m$ such that $H(k) < \epsilon$. Find $H(k)$ for $a \leq k < m$ similarly, where a is the greatest $k < m$ such that $H(k) < \epsilon$.

a) Show that the binomial (n, p) probability $P(a \text{ to } b)$ is at least $1 - \epsilon$.

[*Hint*: Use the fact that the consecutive odds ratios are decreasing to show $P(b + j) \leq \epsilon P(m + j)$ for $j = 1, 2, \ldots$, hence $P(b + 1 \text{ to } n) \leq \epsilon P(m + 1 \text{ to } n)$. Bound the left tail similarly. This argument was discovered by N. Bernoulli around 1700.]

b) For $a \leq k \leq b$ let $P(k \,|\, a \text{ to } b) = H(k)/\Sigma$ where Σ is the sum of the $H(j)$ over $a < j < b$. Deduce from a) that for $a \leq k \leq b$

$$(1 - \epsilon)P(k \,|\, a \text{ to } b) \leq P(k) \leq P(k \,|\, a \text{ to } b)$$

So $P(k \mid a \text{ to } b)$ computed as above is an approximation to $P(k)$ with a relative error of at most ϵ. The computer run time to compute $P(k \mid a \text{ to } b)$ for every $a \leq k \leq b$ is approximately $K(b - a)$ for some constant K depending on the speed of the computer.

c) Use the normal approximation to find an approximate formula for the run time in terms of n, p, K, and ϵ which will be asymptotically correct as $n \to \infty$.

d) If it takes my computer 2 seconds to compute this approximation to the distribution of the number of reds in 100 spins of a roulette wheel with $\epsilon = 0.001$, approximately how long should it take my computer to approximate the distribution for 1000 spins with the same ϵ?

37. Integrals related to equalizations in coin tossing. Let $I_n = \int_{-\pi/2}^{\pi/2} \cos^n(x)dx$.

a) Show that for $n = 2, 3, \ldots$

$$I_n = \frac{n-1}{n} I_{n-2} \quad \text{and} \quad \sqrt{\frac{2\pi}{n+1}} < I_n < \sqrt{\frac{2\pi}{n}}.$$

b) Referring to Exercise 2.3.3, show that these formulae yield much sharper bounds on $\alpha_m = P(m \text{ in } 2m)$, the probability of exactly m heads in $2m$ fair coin tosses, as well as the value of $K = \lim_{m\to\infty} \sqrt{m}\alpha_m$.

c) Use $\cos(x) \approx 1 - \frac{1}{2}x^2$ for $x \approx 0$ and an exponential approximation to deduce that $I_n \approx C/\sqrt{n}$ for large n where

$$C = \int_{-\infty}^{\infty} \phi(z)dz$$

Compare with the estimates of I_n in a) to conclude that $C = \sqrt{2\pi}$.