Example 1.18 Cindy shuffles a deck of playing cards. Is it likely that she is the first person in history to achieve this particular ordering of the cards?

This is another of those problems that defies intuition. Of all of the people in history, almost surely *someone* must have attained the same shuffle as Cindy. By the multiplication rule, there are

$$52! = 80658175170943878571660636856403766975289505440883277824000000000000$$

different shufflings. Yikes! Perhaps Cindy's shuffle is likely unique after all. To address the likelihood of her shuffle being unique, some back-of-the-envelope calculations are required. The world population is about seven billion people. Approximately half of the people that have ever lived are currently alive, so assume that 14 billion people have lived through the ages. Now assume that everyone lives 100 years on average (dubious), and shuffles a deck of cards ten times a day on average (even more dubious), then there have been a total of a mere

$$14000000000 \cdot 100 \cdot 365 \cdot 10 = 5110000000000000$$

total shuffles. Hence Cindy's shuffle is almost certainly unique. Every shuffle of a deck of cards is almost always making playing-card history.

Although simple to state and use, the multiplication rule is a surprisingly versatile tool for addressing counting (combinatorics) problems. There is a special case of the multiplication rule that arises so often that it gets special treatment here. The object of interest is known as a *permutation*.

Permutations

The notion of whether a sample is taken with or without replacement is a critical notion in combinatorics and probability. When a sample of size r, for example, is selected at random and *with replacement* from a set of n distinct objects, there are n^r different ordered samples that can be taken. On the other hand, when the items are selected *without replacement*, the ordered items that are selected are a *permutation*.

Definition 1.1 A *permutation* is an ordered arrangement of r objects selected from a set of n objects without replacement.

One key question to be addressed in a counting problem is whether the ordering of the objects is relevant. If the ordering is relevant, then using permutations might be appropriate.

Example 1.19 List the permutations from the set $\{a, b, c\}$ selected 2 at a time.

Applying Definition 1.1 with $n = 3$ and $r = 2$ yields the 6 ordered pairs:

$$\begin{array}{cc} (a, b) & (b, a) \\ (a, c) & (c, a) \\ (b, c) & (c, b). \end{array}$$

The second column of permutations is the same as the first column in reverse order.

Theorem 1.2 The number of permutations of n distinct objects selected r at a time without replacement is

$$n \cdot (n-1) \cdot (n-2) \cdot \ \ldots \ \cdot (n-r+1) = \frac{n!}{(n-r)!}$$

for $r = 0, 1, 2, \ldots, n$ and n is a positive integer, and $0! = 1$.

We close this section with one final unifying example that stresses the importance of the following two questions associated with a counting problem. (a) Is the sampling performed *with replacement* or *without replacement*? (b) Is the sample considered *ordered* or *unordered*?

Example 1.42 How many ways are there to select 4 billiard balls from a bag containing the 15 balls numbered $1, 2, \ldots, 15$?

The question as stated is (deliberately) vague. It has not been specified whether

- the billiard balls are replaced (that is, returned to the bag) after being sampled, and
- the order that the balls are being drawn from the bag is important.

So there are really $2 \times 2 = 4$ different questions being asked here. The answers to these questions are given in the 2×2 matrix below.

	Without replacement	With replacement
Ordered sample	$15 \cdot 14 \cdot 13 \cdot 12$	$15 \cdot 15 \cdot 15 \cdot 15$
Unordered sample	$\binom{15}{4}$	$\binom{18}{4}$

These simplify to

	Without replacement	With replacement
Ordered sample	32,760	50,625
Unordered sample	1365	3060

There are several observations that can be made on the numbers in this 2×2 matrix. First of all, the entries in column 2 are always greater than the corresponding entries in column 1. This is because sampling with replacement allows for more possible draws due to the fact that the size of the population from which a draw is made remains constant rather than diminishing. Secondly, the entries in row 1 are always greater than the corresponding entries in row 2. This is because the count of ordered draws (permutations) will always exceed the corresponding number of unordered draws (combinations).

A further explanation of the lower-right entry of the matrix might be needed. Consider 15 bins and 4 balls, where ◯ denotes a billiard ball. One draw of 4 balls is depicted below.

| ◯ ◯ | | ◯ | | ⋯ | | ◯
1 2 3 4 5 ⋯ 14 15

This arrangement of bins and markers corresponds to the unordered draw 2, 2, 4, 15 taken with replacement from the bag. We need to count the number of arrangements of 14 dividers plus 4 balls, or a total of 18 objects. Since the ◯'s are indistinguishable, there are

$$\binom{18}{4}$$

different orderings (the outer walls are ignored).

The previous example has highlighted two important issues that arise in combinatorial problems: order and replacement. These concerns lead to a generic class of problems known as "urn models" in which objects are drawn sequentially from an urn.

This has been an unusually long section, so it ends with an outline of the topics considered, and their associated formulas.

coins, the rolling of dice, and the random sampling of items from an urn or bag. These settings are used to allow for exact results in the case of the analytic solution of the simple problems presented here.

Example 2.12 Three men and two women sit in a row of chairs in a random order. Let the event A be that men and women alternate (that is, MWMWM). Find $P(A)$.

Using the multiplication rule, there are $5! = 120$ equally likely outcomes to the random ordering, and, again by the multiplication rule, there are 12 orderings that correspond to men and women alternating (see Examples 1.14 and 1.15). Therefore,

$$P(A) = \frac{3 \cdot 2 \cdot 2 \cdot 1 \cdot 1}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = \frac{12}{120} = \frac{1}{10}.$$

This analytic solution is exact and correct. It can be checked by Monte Carlo simulation. The R function `sample` can be used to generate a random ordering of the five people, who will be numbered 1, 2, 3, 4, 5 (the women are even and the men are odd, an agreeable convention for the women). The event A associated with men and women alternating is equivalent to chairs 1, 3, and 5 being occupied by the men, so the product of their indices will be 15. Increasing the number of replications from 1000 (in the coin flipping experiment in Example 2.7) to 100,000, the R code for the Monte Carlo simulation experiment is shown below.

```
nrep = 100000
count = 0
for (i in 1:nrep) {
  x = sample(5)
  if (x[1] * x[3] * x[5] == 15) count = count + 1
}
print(count / nrep)
```

Using indices 1, 3, and 5 to denote both the chairs and the men is coincidental. We could have used 1, 2, and 3, for example, to denote the men. After a call to `set.seed(3)` to initialize the random number seed, the code segment is run five times yielding

0.09948 0.10076 0.10100 0.10105 0.09859.

The fact that the five probability estimates are closer to the analytic value than in the previous Monte Carlo simulation experiment is due to the larger number of replications. Two of these estimates of the probability that men and women alternate are less than the true value $(1/10)$ and three of these estimates are greater than the true value, so the analytic solution is considered "verified." Although we will use the terms like *verified* and *confirmed* when Monte Carlo simulation results hover around the analytic value, these terms are a bit misleading. The Monte Carlo simulation results provide supporting evidence, but they do not provide a mathematical verification of an analytic solution. Monte Carlo can be helpful to see when an analytic solution is incorrect, however, because the estimates will not hover around the analytic solution.

Example 2.13 A *hatcheck girl* collects n hats and returns them at random. Let the event A be the proper return of the hats to their owners. Find $P(A)$.

By the multiplication rule, the hats can be returned in $n!$ different orders. Of these orders, only one of the orders is correct. Thus the probability of returning all hats to the correct owners is

$$P(A) = \frac{1 \cdot 1 \cdot \ldots \cdot 1}{n \cdot (n-1) \cdot \ldots \cdot 1} = \frac{1}{n!}.$$

Example 3.40 Let n be a positive integer. A cube is comprised of n^3 smaller cubes, as illustrated in Figure 3.28 for $n = 4$. If one of the n^3 smaller cubes is selected at random, give an expression for the expected number of exposed faces. (*Hint*: an interior smaller cube has no exposed faces; a corner smaller cube has three exposed faces, etc.)

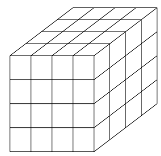

Figure 3.28: A $4 \times 4 \times 4$ array of cubes.

Solution 1 (brute force). Table 3.2 classifies the number of smaller cubes of various types for $n > 1$. The sum of the elements in the second column of this table is n^3 as expected. If the random variable X models the number of exposed faces in a cube selected at random, then the probability mass function of X is

$$f(x) = \begin{cases} \dfrac{(n-2)^3}{n^3} & x = 0 \\ \dfrac{6(n-2)^2}{n^3} & x = 1 \\ \dfrac{12(n-2)}{n^3} & x = 2 \\ \dfrac{8}{n^3} & x = 3. \end{cases}$$

Thus an expression for the expected number of exposed faces is:

$$E[X] = 0 \cdot \frac{(n-2)^3}{n^3} + 1 \cdot \frac{6(n-2)^2}{n^3} + 2 \cdot \frac{12(n-2)}{n^3} + 3 \cdot \frac{8}{n^3} = \frac{6}{n}$$

for $n = 1, 2, \ldots$. As n becomes large, $\lim_{n \to \infty} E[X] = 0$ because of the overwhelming number of interior smaller cubes with no exposed faces.

Type of smaller cube	Number of smaller cubes	Number of exposed faces
Interior	$(n-2)^3$	0
Face	$6(n-2)^2$	1
Edge	$12(n-2)$	2
Corner	8	3

Table 3.2: Classifying the cubes.

Solution 2 (finesse). The total number of *faces* on the smaller cubes is $6n^3$. The total number of *exposed faces* on the smaller cubes is $6n^2$. If a *face* of a smaller cube is selected at random, the probability that the face is exposed is

$$\frac{6n^2}{6n^3} = \frac{1}{n}$$

which is also the expected number of exposed faces. Since there are six faces on a smaller cube selected at random, the expected number of exposed faces is

$$\frac{6}{n}.$$

Using similar methodology, the population variance of a binomial(n, p) random variable is

$$\sigma^2 = np(1-p)$$

and the population skewness and kurtosis are

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{1-2p}{\sqrt{np(1-p)}} \qquad \text{and} \qquad E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = 3 + \frac{1-6p(1-p)}{np(1-p)}.$$

The population skewness and kurtosis converge to 0 and 3, respectively, in the limit as $n \to \infty$. Finally, the moment generating function for a binomial(n, p) random variable is

$$M(t) = \left(1-p+pe^t\right)^n \qquad -\infty < t < \infty.$$

The shape of the probability mass function for a binomial(n, p) random variable typically follows a bell shape. Consider the following three binomial random variables.

- The number of fours in 60 rolls of a fair die: $X \sim \text{binomial}(60, 1/6)$.
- The number of even numbers in 60 rolls of a fair die: $X \sim \text{binomial}(60, 1/2)$.
- The number of non-fours in 60 rolls of a fair die: $X \sim \text{binomial}(60, 5/6)$.

Plots of the bell-shaped probability mass functions are shown in Figure 4.1, with identical vertical scales on the three probability mass functions. The left-hand probability mass function is centered around $\mu = 60 \cdot \frac{1}{6} = 10$ and is skewed to the right; the middle probability mass function is centered around $\mu = 60 \cdot \frac{1}{2} = 30$ and is symmetric; the right-hand probability mass function is centered around $\mu = 60 \cdot \frac{5}{6} = 50$ and is skewed to the left. The R commands that create these plots are given below.

```
par(mfrow = c(1, 3))
x = 0:60
plot(x, dbinom(x, 60, 1 / 6), type = "h")
plot(x, dbinom(x, 60, 1 / 2), type = "h")
plot(x, dbinom(x, 60, 5 / 6), type = "h")
```

The `mfrow` argument in `par` indicates that a 1×3 array of plots is to be displayed. The `dbinom` function returns the probability mass function for the binomial distribution.

The binomial distribution is one of the pillars in applied probability because it arises so often in applications. Applications of the distribution are now considered in the following sequence of examples.

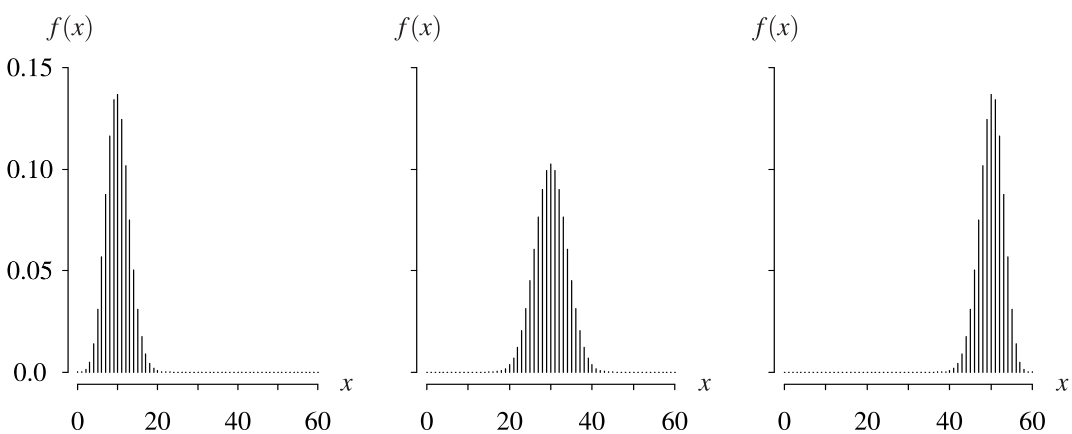

Figure 4.1: Three binomial probability mass functions.

6.4 Bivariate Normal Distribution

Just as the normal distribution plays a central role as a univariate distribution, the bivariate normal distribution is the fundamental bivariate distribution. This distribution has several mathematically and geometrically elegant properties and also proves to be quite useful in applications. Some of the early development associated with a distribution like the bivariate normal was done by Sir Francis Galton in the late 19th century concerning data pairs (X, Y) consisting of the average heights of parents, adjusted for gender differences, X, versus the adult heights of their offspring, adjusted for gender differences, Y.

Definition 6.10 The continuous random variables X and Y with joint probability density function

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]},$$

which is defined on the support $\mathcal{A} = \{(x,y) \mid -\infty < x < \infty, -\infty < y < \infty\}$ with the associated parameter space

$$\Omega = \{(\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho) \mid -\infty < \mu_X < \infty, -\infty < \mu_Y < \infty, \sigma_X > 0, \sigma_Y > 0, -1 < \rho < 1\}$$

are *bivariate normal random variables* with parameters $\mu_X, \mu_Y, \sigma_X, \sigma_Y$, and ρ.

The choice of symbols used for the five parameters will come as no surprise in that they also happen to be the following expected values:

$$E[X] = \mu_X \qquad E[Y] = \mu_Y \qquad V[X] = \sigma_X^2 \qquad V[Y] = \sigma_Y^2$$

and ρ is the population correlation. A plot of the joint probability density function for one particular choice of the five parameters is shown in Figure 6.18. Although the support for X and Y covers all of $\mathcal{R}^2$, only a square region is shown in the figure. For all values of the parameters, the bivariate normal probability density function is unimodal with the mode at (μ_X, μ_Y). The height of the joint probability density function at the mode is

$$f(\mu_X, \mu_Y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}.$$

It is hard to distinguish one bivariate normal distribution from another based on three-dimensional graphs of the joint probability density function like the one in Figure 6.18. They all look like mountains. Level surfaces of the joint probability density function tend to be more visually distinct.

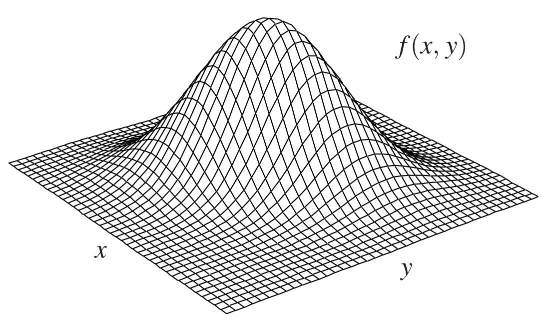

Figure 6.18: The joint probability density function of a bivariate normal distribution.

The rather complicated joint probability density function $f(x, y)$ has level surfaces that are concentric ellipses. Level surfaces are also known as contours and they are the set of points at which $f(x, y)$ assumes a constant value. Consider first the simple case when $\mu_X = \mu_Y = \rho = 0$. In this case, equating the joint probability density function to a constant and performing some algebra gives the usual form of an ellipse

$$\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_Y^2} = c$$

for some constant c. In a more general case where ρ is nonzero, but the population means continue to be 0, the equation for the ellipse becomes a bit more complicated:

$$\frac{x^2}{\sigma_X^2} - \frac{2\rho xy}{\sigma_X \sigma_Y} + \frac{y^2}{\sigma_Y^2} = c$$

for some constant c. Finally, in the most general case, the ellipse has the form

$$\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X \sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} = c$$

for some constant c. One particular ellipse gets its own name. The *population concentration ellipse* is the level surface containing the ordered pairs

$$(\mu_X - \sigma_X, \mu_Y - \sigma_Y), (\mu_X - \sigma_X, \mu_Y + \sigma_Y), (\mu_X + \sigma_X, \mu_Y - \sigma_Y), (\mu_X + \sigma_X, \mu_Y + \sigma_Y).$$

Figure 6.19 displays four population concentration ellipses for four different sets of parameters for the bivariate normal distribution. The ellipse in the upper-left plot is a circle (a special case of an ellipse), so one can conclude that the variances are equal, that is $\sigma_X^2 = \sigma_Y^2$, and $\rho = 0$. The ellipse

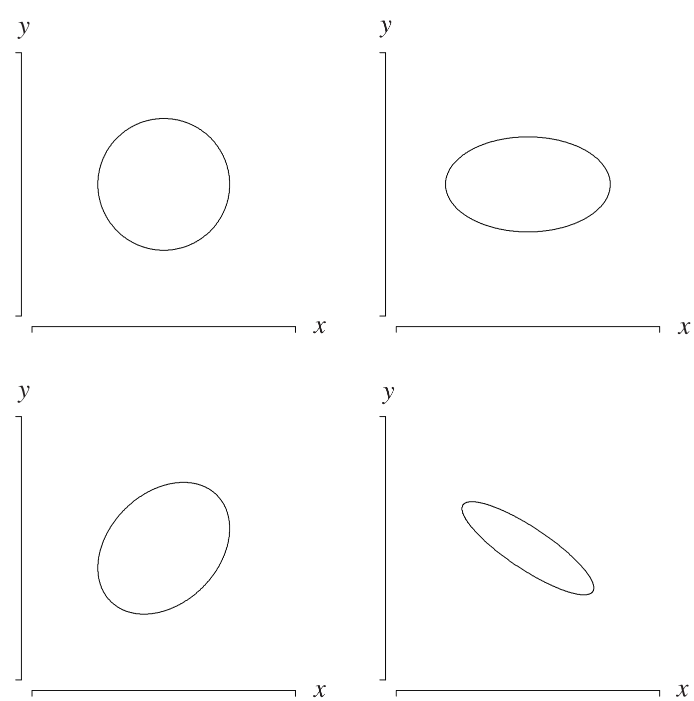

Figure 6.19: Level surfaces of the joint probability density function.