Chapter 6

Limit Theorems

6.1 Sums of Independent Random Variables

If X and Y are independent then

$$\begin{aligned} P(X+Y=z) &= \sum_x P(X=x, Y=z-x) \\ &= \sum_x P(X=x)P(Y=z-x) \end{aligned} \tag{6.1}$$

To see the first equality, note that if the sum is z then X must take on some value x and Y must be $z-x$. The first equality is valid for any random variables. The second holds since we have supposed X and Y are independent.

Example 6.1. Suppose X and Y are independent and have the following distribution

k	1	2	3	4
probability	.1	.2	.3	.4

Find the distribution of $X+Y$.

To compute a single probability is straightforward:

$$\begin{aligned} P(X+Y=4) &= P(X=1)P(Y=3) + P(X=2)P(Y=2) + P(X=3, Y=1) \\ &= 0.1(0.3) + 0.2(0.2) + 0.3(0.1) = 0.10 \end{aligned}$$

To compute the entire distribution, we begin by writing out the joint distribution of X and Y:

X	Y=1	2	3	4
1	.01	.02	.03	.04
2	.02	.04	.06	.08
3	.03	.06	.09	.12
4	.04	.08	.12	.16

The numbers with a given sum are diagonals in the table so

$$
\begin{aligned}
P(X+Y=2) &= 0.01 \\
P(X+Y=3) &= 0.02+0.02 = 0.04 \\
P(X+Y=4) &= 0.03+0.04+0.03 = 0.10 \\
P(X+Y=5) &= 0.04+0.06+0.06+0.04 = 0.20 \\
P(X+Y=6) &= 0.08+0.09+0.08 = 0.25 \\
P(X+Y=7) &= 0.12+0.12 = 0.24 \\
P(X+Y=8) &= 0.16
\end{aligned}
$$

Example 6.2. If $X = \text{binomial}(n, p)$ and $Y = \text{binomial}(m, p)$ are independent then $X + Y = \text{binomial}(n + m, p)$.

Proof by thinking. The easiest way to see the conclusion is to note that if X is the number of successes in the first n trials and Y is the number of successes in the next m trials, then $X + Y$ is the number of successes in $n + m$ trials.

Proof by computation. Using (6.1), noting that $P(X = j) = 0$ when $j < 0$, $P(Y = k - j) = 0$ when $j > k$, and plugging in the definition of the binomial distribution we get

$$
\begin{aligned}
P(X+Y=k) &= \sum_{j=0}^{k} P(X=j)P(Y=k-j) \\
&= \sum_{j=0}^{k} C_{n,j} p^j (1-p)^{n-j} C_{m,k-j} p^{k-j} (1-p)^{m-(k-j)} \\
&= p^k (1-p)^{n+m-k} \sum_{j=0}^{k} C_{n,j} C_{m,k-j} \\
&= p^k (1-p)^{n+m-k} C_{n+m,k}
\end{aligned}
$$

To see the last equality, note that we can pick k students out of a class of n boys and m girls in $C_{n+m,k}$ ways but this can also be done by first deciding on the number j of boys to be chosen and then picking j of the n boys (which can be done in $C_{n,j}$ ways) and $k - j$ of the m girls (which can be done in $C_{m,k-j}$ ways). The multiplication rule implies that for fixed j the number of ways the j boys and $k - j$ girls can be selected is $C_{n,j} C_{m,k-j}$, so summing from $j = 0$ to k gives

$$\sum_{j=0}^{k} C_{n,j} C_{m,k-j} = C_{n+m,k} \qquad \square$$

Since Poissons arise as the limit of Binomials we should guess that

Example 6.3. If $X = \text{Poisson}(\lambda)$ and $Y = \text{Poisson}(\mu)$ are independent then $X + Y = \text{Poisson}(\lambda + \mu)$.

Proof by thinking. Let $[x]$ be the largest integer $\leq x$, i.e., x rounded down to the next integer. By Example 6.2 if $X = \text{Binomial}([n\lambda], 1/n)$ and $Y = \text{Binomial}([n\mu], 1/n)$ are independent $X + Y = \text{Binomial}([n\lambda] + [n\mu], 1/n)$. Letting $n \to \infty$ and using the Poisson approximation to the binomial now gives the result.

Proof by computation. Again we use (6.1), note that $P(X = j) = 0$ when $j < 0$, $P(Y = k - j) = 0$ when $j > k$, and plug in the definition of the Poisson distribution to get

$$\begin{aligned} P(X + Y = k) &= \sum_{j=0}^{k} P(X = j)P(Y = k - j) \\ &= \sum_{j=0}^{k} e^{-\lambda} \frac{\lambda^j}{j!} e^{-\mu} \frac{\mu^{k-j}}{(k-j)!} \\ &= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{j=0}^{k} C_{k,j} \lambda^j \mu^{k-j} \\ &= e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^k}{k!} \end{aligned}$$

where the last equality follows from the binomial theorem.

Example 6.4. Suppose $X_1, \ldots, X_n$ are independent and have a geometric distribution with parameter p and let $T = X_1 + \cdots + X_n$ be the amount of time we have to wait for n successes when each trial is independent and results in success with probability p. Then

$$P(T = m) = C_{m-1,n-1} p^n (1 - p)^{m-n}$$

To see this consider the following possibility with $m = 10$ and $n = 3$:

$$FSFFFFSFFS$$

We know that there must be $m - n = 10 - 3 = 7$ failures and the last trial must be a success. In the first $m-1 = 9$ trials there must be $n-1 = 2$ successes. Once we choose their locations, which can be done in $C_{m-1,n-1}$ ways, the outcome is determined. Each of these outcomes has probability $p^n(1 - p)^{m-n}$.

6.2 Mean and Variance of Sums

In this section we will show that the expected value of a sum of random variables is the sum of the expected values. We begin with the case of two random variables.

Theorem 6.1. *For any random variables X and Y,*

$$E(X+Y) = EX + EY \tag{6.2}$$

Proof. We have not derived a formula for the expected value of a function of two random variables, but the recipe is the same: sum the function times the probability over all possible pairs:

$$\begin{aligned} E(X+Y) &= \sum_{x,y} (x+y)\, P(X=x, Y=y) \\ &= \sum_{x,y} xP(X=x, Y=y) + \sum_{x,y} yP(X=x, Y=y) \end{aligned}$$

By the definition of the marginal distribution $\sum_y P(X=x, Y=y) = P(X=x)$ and $\sum_x P(X=x, Y=y) = P(Y=y)$. This converts the above into

$$= \sum_x xP(X=x) + \sum_y yP(Y=y) = EX + EY$$

which is the desired result. □

From (6.2) and induction it follows that

Theorem 6.2. *For any random variables $X_1, \ldots, X_n$,*

$$E(X_1 + \cdots + X_n) = EX_1 + \cdots + EX_n \tag{6.3}$$

Proof. (6.2) gives the result for $n = 2$. Applying (6.2) to $X = X_1 + \cdots + X_n$ and $Y = X_{n+1}$ we see that if the result is true for n then

$$\begin{aligned} E(X_1 + \cdots + X_{n+1}) &= E(X_1 + \cdots + X_n) + EX_{n+1} \\ &= EX_1 + \cdots + EX_n + EX_{n+1} \end{aligned} \tag{6.4}$$

and the result holds for $n+1$. □

Formula (6.3) is very useful in doing computations.

Example 6.5. Pick two cards out of a deck of 52 and let X be the number of spades. Calculate the expected value of X.

To do this directly from the definition we have to calculate the distribution

$$P(X=2) = \frac{C_{13,2}}{C_{52,2}} = \frac{13 \cdot 12}{52 \cdot 51}$$

$$P(X = 1) \quad = \quad \frac{C_{13,1}C_{39,1}}{C_{52,2}} = \frac{2 \cdot 13 \cdot 39}{52 \cdot 51}$$

Form this it follows that

$$EX = 2 \cdot \frac{13 \cdot 12}{52 \cdot 51} + \frac{2 \cdot 13 \cdot 39}{52 \cdot 51} = 2 \cdot \frac{13 \cdot 51}{52 \cdot 51} = \frac{1}{2}$$

To see this more easily let $X_i = 1$ if the ith card drawn was a spade and 0 otherwise. $X = X_1 + X_2$ so it follows from (6.3) that

$$EX = EX_1 + EX_2 = 1/4 + 1/4 = 1/2$$

since $P(X_i = 1) = 1/4$ and $P(X_i = 0) = 3/4$.

In a similar way we can conclude that if we draw 13 cards out of a deck of 52 the expected number of spades is $13/4 = 3.25$. Doing this directly from the distribution would be extremely tedious. In our next example it would be difficult to calculate the distribution, but it easy to compute the expected value.

Example 6.6. Balls in boxes. Suppose we put n balls randomly into m boxes. What is the expected number of empty boxes?

Let $X_i = 1$ if the ith box is empty. The total number of empty boxes is given by $N = X_1 + \cdots + X_m$, so so it follows from (6.3) that

$$EN = EX_1 + \cdots + EX_m = mEX_1$$

The probability box 1 is empty is $(1 - 1/m)^n$, so the expected number of empty boxes $= m(1 - 1/m)^n$.

For a concrete example suppose $m = 100$, and $n = 500$. In this case

$$(1 - 1/100)^{500} \approx e^{-5} = 0.00673$$

so the expected number of empty boxes is $\approx 2/3$. We will return to this computation in Example 6.14.

The next goal of this section is to show that if $X_1, \ldots, X_n$ are independent then

$$\text{var}\,(X_1 + \cdots + X_n) = \text{var}\,(X_1) + \cdots + \text{var}\,(X_n).$$

The first step is to prove

Theorem 6.3. *If X and Y are independent then*

$$EXY = EX \cdot EY \tag{6.5}$$

Proof. Since X and Y are independent, $P(X = x, Y = y) = P(X = x)P(Y = y)$ and

$$\begin{aligned} EXY &= \sum_{x,y} xyP(X = x)P(Y = y) = \sum_y yP(Y = y) \sum_x xP(X = x) \\ &= \sum_y yP(Y = y)EX = EX \cdot EY \end{aligned}$$

which proves the desired result. □

(6.2) says that $E(X+Y) = EX + EY$ holds for ANY random variables. The next example shows that $EXY = EX \cdot EY$ does not hold in general.

Example 6.7. Suppose X and Y have joint distribution given by

X	$Y=1$	0
1	0	.3
0	.5	.2

We have arranged things so that XY is always 0 so $EXY = 0$. On the other hand, $EX = P(X=1) = 0.3$ and $EY = P(Y=1) = 0.5$ so

$$EXY = 0 < 0.15 = EX\,EY$$

Our next example shows that we may have $EXY = EX\,EY$ without X and Y being independent.

Example 6.8. Suppose X and Y have joint distribution given by

X	$Y=-1$	0	1	
1	0	.25	0	.25
0	.25	.25	.25	.75
	.25	.5	.25	

Again we have arranged things so that XY is always 0, so $EXY = 0$. The symmetry of the marginal distribution for Y (or simple arithmetic) shows $EY = 0$, so we have $EXY = 0 = EX\,EY$. X and Y are not independent since

$$P(X=1, Y=-1) = 0 < P(X=1)P(Y=-1).$$

Our next topic is the variance of sums. To state our first result we need a definition. The **covariance** of X and Y is

$$\text{cov}\,(X,Y) = E\{(X-EX)(Y-EY)\}$$

Now $\text{cov}\,(X,X) = E\{(X-EX)^2\} = \text{var}\,(X)$, so repeating the proof of (1.11) we can rewrite the definition of the covariance in a form that is more convenient for computations:

$$\text{cov}\,(X,Y) = E\{XY - YEX - XEY + EXEY\} = EXY - EXEY \qquad (6.6)$$

Remembering (6.5), we see that if X and Y are independent $\text{cov}\,(X,Y) = 0$.

Example 6.9. Consider the calculus grade joint distribution (Example 3.27)

X	Y= 4	3	2	
5	.10	.05	0	.15
4	.15	.15	0	.30
3	.10	.15	.10	.35
2	0	.05	.10	.15
1	0	0	.05	.05
	.35	.40	.25	

$EX = 5(.15) + 4(.30) + 3(.35) + 2(.15) + 1(.05) = 3.35$ $EY = 4(.35) + 3(.40) + 2(.25) = 3.10$. Patiently adding up all the possibilities we see $E(XY) = 10.9$ so

$$\text{cov}(X, Y) = 10.9 - (3.35)(3.10) = 0.515 > 0$$

Positive covariance means that when one variable is large then the other has a greater tendency to be large. In the opposite direction is

Example 6.10. Consider the urn joint distribution (Example 3.26)

X	Y=0	1	2	
0	6/105	20/105	10/105	36/105
1	24/105	30/105	0	54/104
2	15/105	0	0	15/105
	45/105	50/105	10/105	

$$\begin{aligned} EX &= (54/105) \cdot 1 + (15/105) \cdot 2 = 84/105 = 4/5 \\ EY &= (50/105) \cdot 1 + (10/105) \cdot 2 = 70/105 = 2/3 \\ EXY &= (30/105) \cdot 1 = 2/7 \end{aligned}$$

so we have

$$\text{cov}(X, Y) = \frac{2}{7} - \frac{4}{5} \cdot \frac{2}{3} = -.2476$$

Negative covariance means that when one variable is large then the other has a greater tendency to be small. In this example, this is clear since there is a triangle of 0's in the joint distribution which comes from the restriction $X+Y \leq 2$.

The next result explains our interest in the covariance

$$\text{var}(X + Y) = \text{var}(X) + 2\,\text{cov}(X, Y) + \text{var}(Y) \tag{6.7}$$

Proof. Using $E(X + Y) = EX + EY$ and working out the square

$$\begin{aligned} E(X + Y - E(X + Y))^2 &= E((X - EX) + (Y - EY))^2 \\ &= E(X - EX)^2 + 2E((X - EX)(Y - EY)) + E(Y - EY)^2 \\ &= \text{var}(X) + 2\,\text{cov}(X, Y) + \text{var}(Y) \end{aligned}$$

which proves the desired formula. □

Reasoning similar to the proof of (6.7) with more algebra leads to

$$\text{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j) \tag{6.8}$$

The most important special case of (6.8) is

Theorem 6.4. *If $X_1, \ldots, X_n$ are independent then*

$$var(X_1 + \cdots + X_n) = var(X_1) + \cdots + var(X_n). \tag{6.9}$$

Proof. Our assumption implies that the second sum in (6.8) vanishes. □

(6.9) is useful in computing variances. For example it reduces the computation of the variance of the binomial to the following trivial case.

Example 6.11. Bernoulli distribution. $X = 1$ with probability p and 0 with probability $1-p$. $EX = 1 \cdot p + 0 \cdot (1-p) = p$ and $E(X^2) = 1^2 \cdot p + 0^2 \cdot (1-p) = p$, so $\text{var}(X) = p - p^2 = p(1-p)$.

Example 6.12. Binomial distribution. Consider a sequence of independent trials in which success has probability p. Let $X_i = 1$ if the ith trial results in a success and 0 otherwise. The total number of successes, $S_n = X_1 + \cdots + X_n$, has a binomial distribution with parameters n and p. Compute the variance of X.

(6.9) and Example 6.11 imply $\text{var}(S_n) = n \,\text{var}(X_1) = np(1-p)$. To see what this says consider $n = 400$ and $p = 1/2$, i.e., S_n is the number of heads when we flip 400 coins. (6.3) implies that $ES_{400} = 400(1/2)$ so the expected number of heads is 200. $\text{var}(S_{400}) = 400(1/2)(1/2) = 100$ so the standard deviation $\sigma(S_{400}) = 10$.

What have we learned? The most important facts from this section and the previous one are:

(6.3) *For any random variables $X_1, \ldots, X_n$,*

$$E(X_1 + \cdots + X_n) = EX_1 + \cdots + EX_n$$

(6.9) *If $X_1, \ldots, X_n$ are independent then*

$$\text{var}(X_1 + \cdots + X_n) = \text{var}(X_1) + \cdots + \text{var}(X_n).$$

These two facts imply that if S_n is a sum of n independent random variables with mean μ and variance σ^2 then

$$ES_n = n\mu \qquad \text{var}(S_n) = n\sigma^2$$

and hence the size of the typical deviation from the mean is $\sigma\sqrt{n}$.

To illustrate the use of these results, we will consider two more examples.

Example 6.13. Rolling dice. Suppose we roll two dice 100 times. What is the mean, variance, and standard deviation of the sum of the points?

This is equivalent to rolling one die 200 times. Let S_{200} be the sum. (6.3) implies that $ES_{200} = 200(7/2) = 700$, while (6.9) and Example 1.26 imply $\text{var}(S_{200}) = 200(105/36) = 583.33$ so $\sigma(S_{200}) = 24.15$

In our final example the variables being summed do not all have the same distribution.

Example 6.14. Coupon Collector's Problem. Suppose that we record the birthday of every person we meet. Let N be the number of people we have to meet until we have seen someone with every birthday. Ignoring February 29, find the mean and variance of N.

Let T_k be the time at which we see our kth different birthday, so $N = T_{365}$. $T_1 = 1$. For $1 \le k \le 364$, $T_{k+1} - T_k$ is geometric with success probability $1 - k/365$, since up to that point we have collected k birthdays. Using our formula for the mean and variance of the geometric:

$$\begin{aligned} ET_{365} &= 1 + \sum_{k=1}^{364} \frac{365}{k} = 365 \sum_{k=1}^{365} \frac{1}{k} = 2364.64 \\ \text{var}(T_{365}) &= \sum_{k=1}^{364} \frac{1 - k/365}{(k/365)^2} \\ &= (365)^2 \sum_{k=1}^{364} \frac{1}{k^2} - 365 \sum_{k=1}^{364} \frac{1}{k} = 216,418 \end{aligned}$$

The variance looks frighteningly large until you take square root to conclude that the standard deviation is 465.2.

General formulas. Suppose now that there are n objects to be collected. Let T_k be the time at which we see our kth different object, so $N = T_n$. $T_1 = 1$. For $1 \le k \le n-1$, $T_{k+1} - T_k$ is geometric with success probability $1 - k/n$, since up to that point we have collected k objects. Using our formula for the mean and of the geometric:

$$ET_n = 1 + \sum_{k=1}^{n-1} \frac{n}{k} = 365 \sum_{k=1}^{n-1} \frac{1}{k} \approx n \ln n$$

since $\sum_{i=1}^{n-1} 1/k \ge \int_1^n dx/x = \ln n$. Turning to the variance

$$\begin{aligned} \text{var}(T_n) &= \sum_{k=1}^{n-1} \frac{1 - k/n}{(k/n)^2} \\ &= n^2 \sum_{k=1}^{n-1} \frac{1}{k^2} - n \sum_{k=1}^{n-1} \frac{1}{k} \approx n^2 \pi^2/6 \end{aligned}$$

since $\sum_{k=1}^{\infty} 1/k^2 = \pi^2/6$. Note that both the mean and standard deviation are of order n. This holds because $T_n - T_{n-1}$ has mean n and variance $\approx n^2$.

Back to balls in boxes, Example 6.6. When $n = 100$, $n \ln n = 460.51$ and the standard deviation is roughly $n\pi/\sqrt{6} = 128.25$. This seems consistent with the previous computation that putting 500 balls into 100 boxes leaves an average of 2/3 empty boxes.

6.3 Laws of Large Numbers

To motivate the main result of this section, we turn to simulation. Figure 6.1 shows the fraction of heads versus time in a simulations of flipping a fair coin 5000 times. As dictated by the frequncy interpretation of probability, the fraction of heads seen at time n approaches $1/2$ as $n \to \infty$.

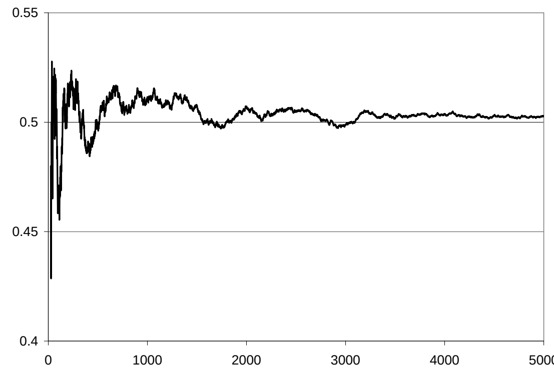


Figure 6.1: Fraction of heads after k coin flips for $25 \le k \le 5000$

To state the general result, we need some definitions. If $X_1, X_2, \ldots$ are independent and have the same distribution then we say the X_i are **independent and identically distributed**, or i.i.d. for short. Such sequences arise if we repeat some experiment such as flipping a coin or rolling a die, or if we stop people at random and measure their height or ask them how they will vote in an upcoming election.

Our first goal in this section will be to prove the **law of large numbers**, which says that if $X_1, X_2, \ldots$ are i.i.d. with $EX_i = \mu$ then when n is large, the average of the first n observations,

$$\bar{X}_n = (X_1 + \cdots + X_n)/n$$

will be close to EX with high probability.

$\bar{X}_n$ is called the **sample mean** because if we assigned probability $1/n$ to each of the first n observations then $\bar{X}_n$ would be the mean of that distribution. If we suppose that the X_i are i.i.d. with $EX_i = \mu$ then using the facts that $E(cY) = cEY$ and the expected value of the sum is the sum of the expected values, we have

$$E\bar{X}_n = \frac{1}{n}E(X_1 + \cdots + X_n) \tag{6.10}$$

$$= \frac{1}{n}\{EX_1 + \cdots + EX_n\} = \mu \tag{6.11}$$

If we suppose that $\text{var}(X_i) = \sigma^2$ then using the facts that $\text{var}(cY) = c^2 \text{var}(Y)$ and that for independent $X_1, \ldots, X_n$ the variance of the sum is the sum of the variances, we have

$$\text{var}(\bar{X}_n) = \frac{1}{n^2}\text{var}(X_1 + \cdots + X_n) \tag{6.12}$$

$$= \frac{1}{n^2}\{\text{var}(X_1) + \cdots + \text{var}(X_n)\} = \frac{\sigma^2}{n} \tag{6.13}$$

Taking square roots we see that the standard deviation of $\bar{X}_n$ is $\sigma/\sqrt{n}$. We have earlier called this the size of a typical deviation from the mean. The key to proving the law of large numbers is to show that if k is large then the probability of an observation more than k standard deviations from the mean is small. To motivate the inequality we will use to prove this, we consider a

Puzzle. Suppose $EX = 0$ and $EX^2 = 1$. How large can $P(|X| \geq 3)$ be?

Solution. On the set $\{|X| \geq 3\}$, $X^2 \geq 9$. Since $X^2 \geq 0$, EX^2 must be larger than what we get from considering only values with $|X| \geq 3$. That is,

$$1 = EX^2 \geq 9P(|X| \geq 3)$$

or $P(|X| \geq 3) \leq 1/9$. To see that this can be achieved we let $P(X = 3) = 1/18$, $P(X = -3) = 1/18$, $P(X = 0) = 8/9$ and note that

$$EX = 3 \cdot \frac{1}{18} + (-3) \cdot \frac{1}{18} = 0$$
$$EX^2 = 9 \cdot \frac{1}{18} + 9 \cdot \frac{1}{18} = 1$$

Generalizing leads to

Chebyshev's inequality. *If* $y > 0$ *then*

$$P(|Y - EY| \geq y) \leq \text{var}(Y)/y^2 \tag{6.14}$$

Proof. Again since $|Y - EY|^2 \geq 0$, $E|Y - EY|^2$ must be larger than what we get from considering only values with $|Y - EY| \geq y$ so

$$\text{var}(Y) = E|Y - EY|^2 \geq y^2 P(|Y - EY| \geq y)$$

and rearranging gives (6.14). □

If we let $\sigma^2 = \text{var}(Y)$ and take $y = k\sigma$ with $k \geq 1$ then (6.14) implies that

$$P(|Y - EY| \geq k\sigma) \leq 1/k^2 \tag{6.15}$$

This reinforces our notion that σ is the size of the typical deviation from the mean by showing that a deviation of k standard deviations has probability smaller than $1/k^2$.

Proof of the law of large numbers. Let $Y = \bar{X}_n$. (6.11) implies $EY = \mu$ and (6.13) implies $\text{var}(Y) = \sigma^2/n$ so using Chebyshev's inequality, we see that if $\epsilon > 0$ then

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\text{var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2 n} \to 0 \tag{6.16}$$

as $n \to \infty$. □

(6.16) could be called the **fundamental theorem of statistics** because it says that the sample mean is close to the mean μ of the underlying population when the sample is large. The last conclusion does not rule out the possibility that the sequence of sample means $\bar{X}_1$, $\bar{X}_2$, ... stays close to EX most of the time but occasionally wanders off because of a streak of bad luck. Our next result says that this does not happen.

Strong law of large numbers. *Suppose $X_1, X_2, \ldots$ are i.i.d. with $E|X_i| < \infty$. Then with probability one the sequence of numbers $\bar{X}_n$ converges to EX_i as $n \to \infty$.*

The first thing we have to explain is the phrase "with probability one." To do this we first consider flipping a coin and letting X_i be 1 if the ith toss results in Heads, and 0 otherwise. The strong law of large numbers says that with probability one

$$(X_1 + \cdots + X_n)/n \to 1/2 \qquad \text{as } n \to \infty$$

Tt is easy to write down sequences of tosses for which this is false:

$$H,\ H,\ T,\ H,\ H,\ T\,H,\ H,\ T\,H,\ H,\ T, \ldots$$

However, the strong law of large numbers implies that the collection of "bad sequences" (i.e., those for which the asymptotic frequency of Heads is not 1/2) has probability zero.

Example 6.15. Growth of a risky investment. In our simple model, each year the stock market either increases by 40% or decreases by 20% with probability 1/2 each. How fast will our money grow?

Reasoning naively, our expected value is $(1/2)40 + (1/2)(-20) = 10$ percent per year. However, this is not the right way to look at things. Let X_i be independent and equal to 1.4 and 0.8 with probability 1/2 each. If we start with M_0 dollars then after n years we have

$$M_n = M_0 X_1 X_2 \cdots X_n$$

To turn this into something that we can apply the law of large numbers to, we take logarithms:

$$\ln(M_n/M_0) = \sum_{i=1}^{n} \ln X_i \approx nE(\ln X_i)$$

Computing the expected value:

$$E(\ln X_i) = \frac{1}{2}\ln(1.4) + \frac{1}{2}\ln(0.8) = 0.0566643$$

so we have

$$M_n \approx M_0 \exp(n(\ln(1.4) + \ln(0.8))/2) = M_0(\sqrt{(1.4)(0.8)})^n$$

From this we see that the growth is not given by the arithmetic mean $(1.4 + 0.8)/2 = 1.1$ but by the geometric mean $\sqrt{(1.4)(0.8)} = 1.0583$. To connect with the value at 0 in Figure 6.2, we note that the exponenetial growth rate is $\ln(1.0583) = 0.0566$.

Example 6.16. Optimal investment. Suppose now that we can either invest in the (i) the stock market with either increases by 40% or decreases by 20% with probability 1/2 each, or (ii) buy a bond which always pays 4% interest. How should we allocate our money between these two alternatives to maximize our rate of return?

The answer is easier to understand if we consider a general problem. Let $\alpha = 1.4$ and $\beta = 0.8$ be the two outcomes for the stock and let $\gamma = 1.04$ be the growth for the bond. If we put a fraction p in the bond and $1 - p$ in the stock market then the expected return is

$$R(p) = \frac{1}{2}\ln[p\gamma + (1-p)\alpha] + \frac{1}{2}\ln[p\gamma + (1-p)\beta]$$

To optimize we take the derivative:

$$\begin{aligned} R'(p) &= \frac{1}{2}\frac{\gamma - \alpha}{\alpha + p(\gamma - \alpha)} + \frac{1}{2}\frac{\gamma - \beta}{\beta + p(\gamma - \beta)} \\ &= \frac{\beta(\gamma-\alpha) + \alpha(\gamma-\beta) + 2(\gamma-\beta)(\gamma-\alpha)p}{2(\alpha + p(\gamma-\alpha))(\beta + p(\gamma-\beta))} \end{aligned}$$

For the maximum to occur at a value $0 < p < 1$, we need $R'(0) > 0$ and $R'(1) < 0$. The denominator is positive so we only need examine the numerator

$$\begin{aligned} q(0) &= \gamma(\beta + \alpha) - 2\alpha\beta \\ q(1) &= \gamma(\gamma - \alpha) + \gamma(\gamma - \beta) \end{aligned}$$

For $q(1) < 0$, i.e., to have some of the stock in our portfolio, we need $\gamma < (\alpha + \beta)/2$. For $q(0) > 0$, i.e., to have some of the bond in our portfolio, we need

$$\gamma > \frac{\alpha\beta}{(\alpha + \beta)/2}$$

When both of these conditions hold there will be an optimal that can be found by setting $q(p^*) = 0$.

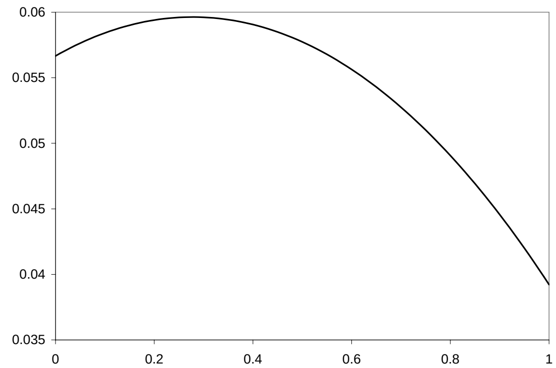


Figure 6.2: Return on investment as a function of p when $\alpha = 1.4$, $\beta = 0.8$, and $\gamma = 1.04$.

In our concrete example

$$\frac{\alpha+\beta}{2} = 1.1 > \gamma = 1.04 > \frac{\alpha\beta}{(\alpha+\beta)/2} = \frac{(1.4)(0.8)}{1.1} = 1.018$$

so the optimal fraction in the bond is

$$p^* = \frac{\gamma(\beta+\alpha)/2 - \alpha\beta}{(\gamma-\beta)(\alpha-\gamma)} = \frac{(1.04)(1.1)-(1.4)(0.8)}{(0.24)(0.36)} = 0.2777$$

6.4 Normal Distribution

To prepare for results in the next section we need to introduce the **standard normal distribution.**

$$f(x) = (2\pi)^{-1/2} e^{-x^2/2}$$

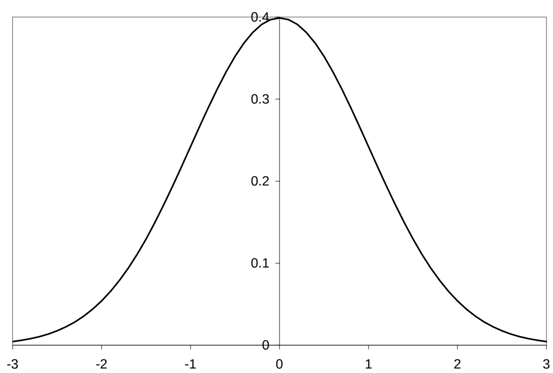


Figure 6.3: Density function for the standard normal distribution.

Since there is no closed form expression for the antiderivative of f, it takes some ingenuity to check:

Theorem 6.5. *$f(x) = (2\pi)^{-1/2} e^{-x^2/2}$ is a probability density.*

Proof. Let $I = \int e^{-x^2/2}\,dx$. To show that $\int f(x)\,dx = 1$, we want to show that $I = \sqrt{2\pi}$.

$$I^2 = \int e^{-x^2/2}\,dx \int e^{-y^2/2}\,dy = \iint e^{-(x^2+y^2)/2}\,dx\,dy$$

Changing to polar coordinates, the last integral becomes

$$\int_0^\infty \int_0^{2\pi} e^{-r^2/2}\, r\,d\theta\,dr = 2\pi \int_0^\infty e^{-r^2/2}\, r\,dr = 2\pi \left(-e^{-r^2/2}\right)\Big|_0^\infty = 2\pi$$

So $I^2 = 2\pi$ or $I = \sqrt{2\pi}$. □

There is, unfortunately, no closed form expression for the distribution function,

$$\Phi(x) = \int_{-\infty}^x (2\pi)^{-1/2} e^{-y^2/2}\,dy$$

so we have to use a table like the one given in the Appendix. In the next few calculations, we will only use a little of the table:

x	0	1	2	3
$\Phi(x)$	0.5000	0.8413	0.9972	0.9986

$P(a < X \le b) = F(b) - F(a)$. For example, when $b = 2$ and $a = 1$, we have

$$P(1 < X \le 2) = \Phi(2) - \Phi(1) = 0.9772 - 0.8413 = 0.1359$$

The table only gives the values of $\Phi(x)$ for $x \ge 0$. Values for $x < 0$ are computed by noting that the normal density function is symmetric about 0 ($f(x) = f(-x)$) so

$$P(X \le -x) = P(X \ge x) = 1 - P(X \le x)$$

since $P(X = x) = 0$. For an example of the use of symmetry, we note that

$$P(X \le -1) = 1 - P(X \le 1) = 1 - 0.8413 = 0.1587$$

so we have

$$P(-1 \le X \le 1) = \Phi(1) - \Phi(-1) = 0.8413 - 0.1587 = 0.6826$$

In the case of discrete random variables, it was important to keep track of the difference between $<$ and $\le$. Here it is not, since for the normal distribution $P(X = x) = 0$ for all x.

The standard normal distribution has mean 0 by symmetry.

$$\text{var}\,(X) = E(X^2) = \int (2\pi)^{-1/2} x^2 e^{-x^2/2}\, dx$$

To show that the variance is 1, we use integration by parts, (5.7), with $g(x) = (2\pi)^{-1/2}x$, $h'(x) = xe^{-x^2/2}$ and hence $g'(x) = (2\pi)^{-1/2}$, $h(x) = -e^{-x^2/2}$ to conclude that the integral above is

$$\begin{aligned} &= -(2\pi)^{-1/2} x e^{-x^2/2}\Big|_{-\infty}^{\infty} + \int (2\pi)^{-1/2} e^{-x^2/2}\, dx \\ &= 0 + 1 \end{aligned}$$

For the last step we observe $e^{-x^2/2}$ goes to 0 much faster than x goes to ∞, and the integral gives the total mass of the normal density.

To create a normal distribution with mean μ and variance σ^2, let $Y = \sigma X + \mu$ where $\sigma > 0$. The inverse of $r(x) = \sigma x + \mu$ is $s(y) = (y - \mu)/\sigma$, so (5.9) implies that Y has density function

$$\begin{aligned} f(s(y))s'(y) &= (2\pi)^{-1/2} e^{-\{(y-\mu)/\sigma\}^2/2} \frac{1}{\sigma} \\ &= (2\pi\sigma^2)^{-1/2} e^{-(y-\mu)^2/2\sigma^2} \end{aligned}$$

If $Y = \text{normal}(\mu, \sigma^2)$ then reversing the formula we used to define Y, $X = (Y - \mu)/\sigma$ has the standard normal distribution.

Example 6.17. Suppose that a man's height has a normal distribution with mean $\mu = 69$ inches and standard deviation $\sigma = 3$ inches. What is the probability a man is more than 6 feet tall (72 inches)?

The first step in the solution is to rephrase the question in terms of X

$$\begin{aligned} P(Y \geq 72) &= P(Y - 69 \geq 3) = P\left(\frac{Y - 69}{\sqrt{9}} \geq 1\right) \\ &= P(X \geq 1) = 1 - 0.8413 = 0.1587 \end{aligned}$$

from the little table above.

6.5 Central Limit Theorem

The limit theorem in this section gets its name not only from the fact that it is of central importance but also because it shows that if you add up a large number of random variables with a fixed distribution with finite variance then, if we subtract the mean and divide by the standard deviation the result has approximately a normal distribution.

Central limit theorem. *Suppose $X_1, X_2, \dots$ are i.i.d. and have $EX_i = \mu$ and $var(X_i) = \sigma^2$ with $0 < \sigma^2 < \infty$. Let $S_n = X_1 + \cdots + X_n$. As $n \to \infty$*

$$P\left(a \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \tag{6.17}$$

Example 6.18. Suppose we flip a coin 100 times. What is the probability we get at least 56 heads?

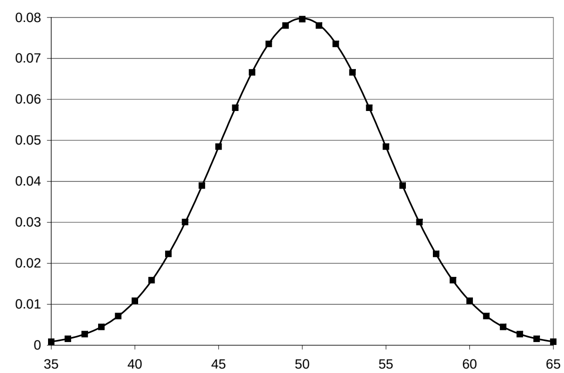


Figure 6.4: Distribution of the number of heads in 100 coin flips (squares) compared to normal approximation (line).

As Figure 6.4 shows, the distribution is approximately normal. To use (6.17) we note that $X_i = 1$ or 0 with probability 1/2 each so the X_i have mean 1/2 and standard deviation $\sqrt{0.5 \cdot 0.5} = 1/2$. The mean number of heads in 100 tosses is $100/2 = 50$ and the standard deviation $\sqrt{100}/2 = 5$, so (6.17) implies

$$\begin{aligned} P(S_{100} \ge 56) &= P\left(\frac{S_{100} - 50}{5} \ge \frac{6}{5}\right) \approx P(\chi \ge 1.2) \\ &= 1 - P(\chi \le 1.2) = 1 - 0.8849 = 0.1151 \end{aligned}$$

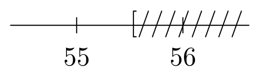
As we will now explain there is a small problem with this solution. If the question in the problem had been formulated as "What is the probability of at most 55 heads?" we would have computed

$$P(S_n \le 55) = P\left(\frac{S_n - 450}{5} \le \frac{5}{5}\right) \approx P(\chi \le 1.0) = 0.8413$$

which does not quite agree with our first answer since

$$0.8413 + 0.1151 = 0.9568 < 1$$

whereas $P(S_{100} \leq 55) + P(S_{100} \geq 56) = 1$. The solution to this problem is to regard $\{S_n \geq 56\}$ as $\{S_n \geq 55.5\}$, that is, the integers 55 and 56 split up the territory that lies between them.



When we do this, the answer to our original question becomes

$$P(S_n \geq 55.5) = P\left(\frac{S_n - 450}{55} \geq \frac{5.5}{15}\right)$$
$$\approx P(\chi \geq 1.1) = 1 - 0.8643 = 0.1357$$

which is a much better approximation of the exact probability 0.135627 than was our first answer, 0.1151.

The last correction, which is called the **histogram correction**, should be used whenever we apply (6.17) to integer-valued random variables. As we did in the last example, if k is an integer we regard $P(S_n \geq k)$ as $P(S_n \geq k - 0.5)$ and $P(S_n \leq k)$ as $P(S_n \leq k + 0.5)$. More generally, we replace each integer k in the set of interest by the interval $[k - 0.5, k + 0.5]$. The next example shows that the histogram correction is not only a device to get more accurate estimates, it also allows us to get answers in cases where a naive application of the central limit theorem would give a senseless answer.

Example 6.19. Suppose we flip 16 coins. Use (6.17) to estimate the probability that we get exactly 8 heads.

The mean number of heads is $n/2 = 16$ while the standard deviation is $\sqrt{n}/2 = 2$. To use the normal approximation we write $\{S_{16} = 8\}$ as $\{7.5 \leq S_{16} \leq 8.5\}$. By (6.17)

$$P\left(\frac{7.5 - 8}{2} \leq \frac{S_{16} - 8}{2} \leq \frac{8.5 - 8}{2}\right) \approx P(-0.25 \leq \chi \leq 0.25)$$

Since $P(\chi = -0.25) = 0$, the probability of interest is $P(\chi \leq 0.25) - P(\chi \leq -0.25)$. The table tells us that $P(\chi \leq 0.25) = 0.5987$. There are no negative numbers in the table but the normal distribution is symmetric so

$$P(\chi \leq -0.25) = P(\chi \geq 0.25) = 1 - P(\chi \leq 0.25) = 1 - 0.5987 = 0.4013$$

and we have

$$P(S_{16} = 8) \approx 0.5987 - 0.4013 = 0.1974$$

The exact answer is

$$2^{-16}\frac{16!}{8!\,8!} = 0.1964$$

Similar reasoning shows that

$$P(S_{16} = 8 + k) \approx P(\chi \leq (k + 1/2)/2) - P(\chi \leq (k - 1/2)/2)$$

As the next table shows these probabilities are fairly close to the exact answers obtained from the binomial distribution.

k	9	10	11	12
normal approx.	.1747	.1209	.0616	.0319
exact ans.	.1746	.1221	.0666	.0277

Example 6.20. Suppose we roll a die 24 times. What is the probability that the sum of the numbers $S_{24} \geq 100$?

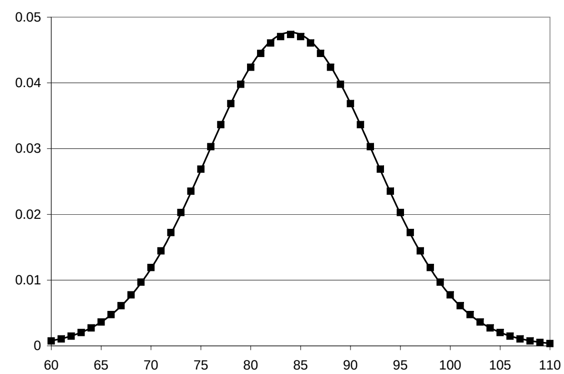


Figure 6.5: Distribution of the sum of 24 rolls of a die (squares) compared to normal approximation (line).

As the figure shows the distribution is approximately normal. To apply (6.17), the first step is to compute the mean and variance of S_{24}. $ES_{24} = 24 \cdot 7/2 = 84$. $\text{var}\,(S_{24}) = 24 \cdot 35/12 = 70$, so the standard deviation is $\sqrt{70} = 8.366$. Using (6.17) now we have

$$\begin{aligned} P(S_{24} \geq 99.5) &= P\left(\frac{S_{24} - 84}{\sqrt{70}} \geq \frac{15.5}{8.366}\right) \\ &\approx P(\chi \geq 1.85) = 1 - P(\chi \leq 1.85) = 0.0322 \end{aligned}$$

compared with the exact answer 0.031760.

z-score. The key to finding the solution of this and the previous two problems is to compute the number of standard deviations separating the observed value from the expected value. The z-score is defined by

$$z = \frac{\text{observed value} - \text{expected value}}{\text{standard deviation}}$$

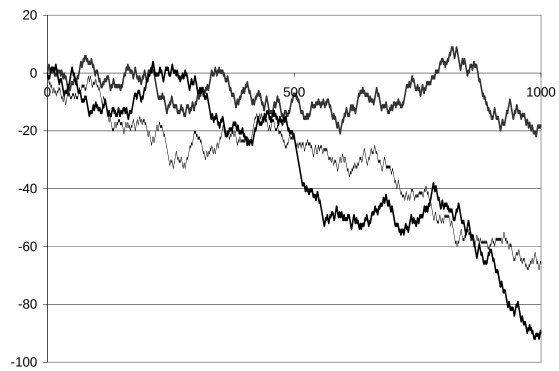
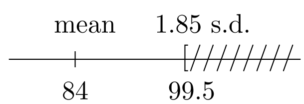


Figure 6.6: Three simulations of 1000 plays of roulette, betting \$1 on black.

In the preceding example this is

$$z = \frac{99.5 - 85}{8.366} = 1.85$$

so the normal approximation is $P(\chi \geq 1.85)$. Here a picture is worth a hundred words.

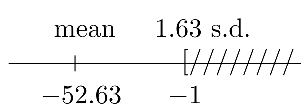


Example 6.21. Roulette. Consider a person playing roulette 1000 times and betting \$1 on black each time. What is the probability their net winnings are ≥ 0?

The outcome of the ith play $P(X_i = 1) = 18/38$ and $P(X_i = -1) = 20/38$ so $EX_i = -1/19 = -0.05263$, $EX_i^2 = 1$ and $\text{var}\,(X_i) = 1 - (1/19)^2 \approx 1$. The mean of 1000 plays is -52.63, while the standard deviation is $\sqrt{1000} = 31.62$. Writing ≥ 0 as ≥ -1 since only even numbers are possible values for S_{1000}, the z-score is

$$z = \frac{-1 - (-52.63)}{31.62} = 1.63$$

so the normal approximation is $P(\chi \geq 1.63) = 1 - P(\chi \leq 1.63) = 0.0516$



Example 6.22. Normal approximation to the Poisson. Each year in Mythica, an average of 64 letter carriers are bitten by dogs. In the past year, 88 incidents were reported. Is this number exceptionally high?

Assuming that dog bites are a rare event, we will use the Poisson distribution for the number of dog bites. As we observed in Example 6.3, a Poisson with mean 64 is the sum of 64 independent Poisson mean 1 random variables, so we can use the Normal to approximate the Poisson. The mean is 64 while the standard deviation is $\sqrt{64}$. Writing the observed event as ≥ 87.5 we see that this is $(87.5 - 64)/8 = 2.94$ standard deviations above the mean, so the normal approximation is $P(\chi \geq 2.94) = 1 - 0.9982 = 0.0018$ so this is an unusual event.

Example 6.23. A manufacturing plant produces boxes of biscuit mix that are one pound (454 grams). However due to the poor flow properties of the powder the standard deviaition of the weight of a box is 10 grams. A sample of 25 boxes had an average weight of 449.4 grams. Does this indicate a problem with the manufacturing process?

The average weight $\bar{X}_{25}$ has mean 454 grams and standard deviation $\sigma/\sqrt{n} = 10/\sqrt{25} = 2$. The observed weight is 4.6 grams below the mean, which is 2.3 standard deviations. From the normal table $P(\chi \geq 2.3) = 1 - 0.9893 = 0.0107$, so a deviation this large by chance is rather unlikely.

Note: Continuous random variables can take on any value so there is no "histogram correction."

Example 6.24. Suppose that the average weight of a person is 182 pounds with a standard deviation of 40 pounds. A large plane can hold 400 people. What is the probability the total weight of the people, S_{400} will be more than 75,000 pounds?

The expected value of S_{400} is $400 \cdot 182 = 72,800$. The standard deviation is $\sigma\sqrt{n} = 40 \cdot 20 = 800$.

$$\begin{aligned} P(S_{400} \geq 75,000) &= P\left(\frac{S_{420} - 72,800}{800} \geq 2.75\right) \\ &\approx P(\chi \geq 2.75) = 1 - P(\chi \leq 2.75) = 0.003 \end{aligned}$$

In designing airplanes one cannot afford to make a mistake 3 times out of a 1000 if the error will have disastrous consequences like a crash. Our table stops at 3.09. For larger values one can use the following approximation

$$\int_x^\infty e^{-y^2/2}\,dy \leq \frac{1}{x}e^{-x^2/2} \tag{6.18}$$

Proof. Multiplying by y/x which is ≥ 1 when $y \geq x$ we have

$$\begin{aligned} \int_x^\infty e^{-y^2/2}\,dy &\leq \frac{1}{x}\int_x^\infty ye^{-y^2/2}\,dy \\ &= \frac{1}{x}\left(-e^{-y^2/2}\right)\Big|_x^\infty = \frac{1}{x}e^{-x^2/2} \end{aligned}$$

which proves the desired estimate. □

From this we see that the probability that the total weight is more than 77,600 pounds, which is six standard deviations above the mean is at most

$$\frac{1}{\sqrt{2\pi}} \cdot \frac{1}{6} e^{-18} \approx 1 \times 10^{-9}$$

6.6 Applications to Statistics

The central limit theorem is the key to several topics in statistics. In this section, we will briefly discuss its use in hypothesis testing and confidence intervals. In the former we have a hypothesis about how the world works and ask if the data is consistent with that hypothesis.

Example 6.25. Is there a difference between baseball and flipping coins? In 2007, the Boston Red Sox won 96 games and lost 66. How likely is this result if the games were decided by flipping coins?

If they were flipping coins the mean number of wins would be $162/2 = 81$, the variance $162/4 = 40.5$ and standard deviation $\sqrt{40.5} = 6.36396$. The event $W \geq 96$ translates to $W \geq 95.5$ when we apply the histogram correction. This is 14.5 above the mean, or $14.5/6.36396 = 2.28$ standard deviations. The normal approximation for this probability is $P(\chi \geq 2.28) = 1 - 0.9887 = 0.0113$.

This probability is small but we must also remember that the Boston Red Sox are 1 of 30 teams in baseball ($1/30 = 0.033$) and we picked because they had the best record. Figure 6.7 compare the outcomes of the 2007 with the normal distribution that we would have if games were decided by flipping coins. Here, we have compared the number of teams with $\leq k$ wins with the normal cumulative distribution function, and the graph suggests that there is more variability in baseball win-loss records compared to flipping coins.

One final small point is that due to the histogram correction, we put the data point for k at $k+0.5$. To see why we do this, note that the normal approximation for ≤ 91 is not 1/2 but is $P(\chi \leq 0.5/6.36396) = P(\chi \leq 0.08) = 0.5319$.

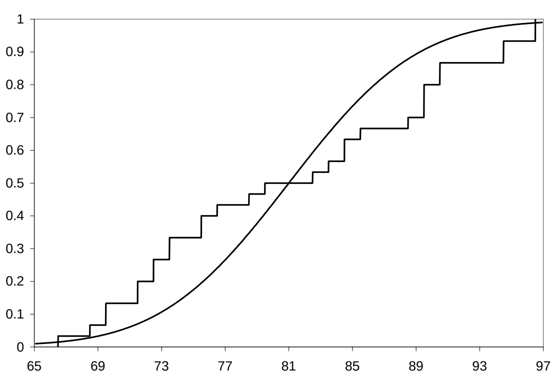


Figure 6.7: Wins for 30 major league baseball teams in 2007 compared with normal cumulative distribution function.

Example 6.26. Weldon's dice data. An English biologist named Weldon was interested in the "pip effect" in dice – the idea that the spots, or "pips," which on some dice are produced by cutting small holes in the surface, make the sides with more spots lighter and more likely to turn up. Weldon threw 12 dice 26,306 times for a total of 315,672 throws and observed that a 5 or 6 came up on 106,602 throws or with probability 0.33770. Is this significantly different from 1/3?

If the true value was 1/3 then the expected number of successes would be 105,224, the variance would be $87,686.666$ and the standard deviation 296.12. Ignoring the histogram correction the excess of 5's and 6's is 1378 or 4.6535 standard deviations, so we are very confident that the true probability is not 1/3.

To make a statement about what we think the real value of p is, we note that if the true probability is p, then the average in the sample $\hat{p} = 106,602/315,672$ then $\hat{p}$ has mean p and standard deviation $\sqrt{p(1-p)/n}$. Consulting the normal table we see that

$$P(-2 \le \chi \le 2) = 2(P(\chi \le 2) - 1/2) = 0.9544$$

Replacing p by the estimate $\hat{p}$ in the formula for the standard deviation we see that 95% of the time the true value of p will lie in

$$\left[\hat{p} - \frac{2\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \hat{p} + \frac{2\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right] \tag{6.19}$$

Plugging in our estimate $\hat{p}$ we have

$$0.33770 \pm 2\sqrt{\frac{0.3377 \cdot 0.6623}{315,672}} = 0.33770 \pm 0.00168 = [0.33602, 0.33938]$$

This difference is not enough to be noticeable by people who play dice games for amusement, but is, perhaps, large enough to be of concern for a casino that entertains tens of thousands of gamblers a year, and offers a wide variety of bets on dice games. For this reason most casinos use dice with no pips.

Example 6.27. Sample size selection. Suppose you want to forecast the outcome of an election and you are trying to figure out how many people to survey so that with probability 0.95 your guess does not differ from the true answer by more than 2%.

From (6.19) we see that if a fraction $\hat{p}$ of the people in a sample of size n are for candidate B then the 95% confidence interval will be

$$\hat{p} \pm \frac{2\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

To get rid of the $\hat{p}$'s in the width of the confidence interval we note that the function $g(x) = x(1-x) = x - x^2$ has derivative $g'(x) = 1 - 2x$, which is positive for $x < 1/2$ and negative for $x > 1/2$. So g is increasing for $x < 1/2$, decreasing for $x > 1/2$, and hence the maximum value occurs at $x = 1/2$. Noticing that $2\sqrt{x(1-x)} = 1$ when $x = 1/2$, we have

$$P\left(p \in \left[\hat{p} - \frac{1}{\sqrt{n}}, \hat{p} + \frac{1}{\sqrt{n}}\right]\right) \ge 0.95 \tag{6.20}$$

To see that this approximation is reasonable for elections, notice that if $0.4 \leq p \leq 0.6$ then $\sqrt{p(1-p)} \geq \sqrt{0.24} = 0.4899$ compared with our upper bound of $1/2$. Even when p is as small as 0.2, $\sqrt{p(1-p)} = \sqrt{0.16} = 0.4$.

To answer our original question now, we set $1/\sqrt{n} = 0.02$ and solve to get

$$n = (1/0.02)^2 = 50^2 = 2500$$

To get the error down to 1% we would need $n = 1/(0.01)^2 = 10,000$. Comparing the last two results and noticing that the radius of the confidence interval in (6.20) is $1/\sqrt{n}$, we see that to reduce the error by a factor of 2 requires a sample that is $2^2 = 4$ times as large.

Example 6.28. The Literary Digest poll. In order to forecast the outcome of the 1936 election, *Literary Digest* polled 2.4 million people and found that 57% of them were going to vote for Alf Landon and 43% were going to vote for F. D. Roosevelt. A 95% confidence interval for the true fraction of people voting for Landon based on this sample would be 0.57 ± 0.00064 but Roosevelt won, getting 62% of the vote to Landon's 38%.

To explain how this happened we have to look at the methods *Literary Digest* used. They sent 10 million questionnaires to people whose names came from telephone books and club membership lists. Since many of the 9 million unemployed did not belong to clubs or have telephones the sample was not representative of the population as a whole. A second bias came from the fact that only 24% of the people filled out the form. This problem was mentioned in our discussion of exit polls in Example 3.19. If, for example, 36% of Landon voters and 16.6% of Roosevelt voters responded then the fraction of people who responded would be $0.62(0.166) + 0.38(0.36) = 0.24$ and the fraction in the sample for Landon would be

$$\frac{0.38(0.36)}{0.62(0.166) + 0.38(0.36)} = \frac{0.1368}{0.24} = 0.57$$

in agreement with the data.

Finally, we would like to observe that *Literary Digest*, which soon after went bankrupt, could have saved a lot of money by taking a smaller sample. George Gallup, who was just then getting started in the polling business, predicted based on a sample of size 50,000 that Roosevelt would get 56% of the vote. His 95% confidence interval for the election result would be 0.56 ± 0.0045, compared with the election result of 62%. Again there could be some bias in his sample, or perhaps Landon voters, discouraged by the predicted outcome, were less likely to vote. The moral of our story is: It is much better to take a good sample than a large one.

6.7 Exercises

Expected value of sums

1. A man plays roulette and bets \$1 on black 19 times. He wins \$1 with probability 18/38 and loses \$1 with probability 20/38. What are his expected winnings?

2. Suppose we draw 13 cards out of a deck of 52. What is the expected value of the number of aces we get?

3. Suppose we pick 3 students at random from a class with 10 boys and 15 girls. Let X be the number of boys selected and Y be the number of girls selected. Find $E(X - Y)$.

4. Twelve ducks fly overhead. Each of 6 hunters picks one duck at random to aim at and kills it with probability 0.6. (a) What is the mean number of ducks that are killed? (b) What is the expected number of hunters who hit the duck they aim at?

5. 10 people get on an elevator on the first floor of a seven story building. Each gets off at one of the six higher floors chosen at random. What is the expected number of stops the elevator makes.

6. Suppose Noah started with n pairs of animals on the ark and m of them died. If we suppose that fate chose the m animals at random, what is the expected number of complete pairs that are left?

7. Suppose we draw 5 cards out of a deck of 52. What is the expected number of different suits in our hand? For example, if we draw $K\spadesuit$ $3\spadesuit$ $10\heartsuit$ $8\heartsuit$ $6\clubsuit$ there are three different suits in our hand.

8. Suppose we draw cards out of a deck without replacement. How many cards do we expect to draw out before we get an Ace? Hint: the locations of the four aces in the deck divide it into five groups $X_1, \ldots X_5$.

Variance and covariance

9. Roll two dice and let $Z = XY$ be the product of the two numbers obtained. What is the mean and variance of Z?

10. Suppose X and Y are independent with $EX = 1$, $EY = 2$, $\text{var}\,(X) = 3$ and $\text{var}\,(Y) = 1$. Find the mean and variance of $3X + 4Y - 5$.

11. In a class with 18 boys and 12 girls, boys have probability 1/3 of knowing the answer and girls have probability 1/2 of knowing the answer to a typical question the teacher asks. Assuming that whether or not the students know the answer are independent events, find the mean and variance of the number of students who know the answer.

12. Let N_k be the number of independent trials we need to get k successes when success has probability p. Find the mean and variance of N_k.

13. Suppose we roll a die repeatedly until we see each number at least once and let R be the number of rolls required. Find the mean and variance of R.

14. Suppose X takes on the values $-2, -1, 0, 1, 2$ with probability 1/5 each, and let $Y = X^2$. (a) Find $\text{cov}\,(X, Y)$. (b) Are X and Y independent?

Chebyshev's inequality

15. Suppose that it is known that the number of items produced at a factory per week is a random variable X with mean 50. (a) What can we say about the probability $X \geq 75$? (b) Suppose that the variance of X is 25. What can we say about $P(40 < X < 60)$?

16. Let $X = \text{binomial}(4, 1/2)$. Use Chebyshev's inequality to estimate $P(|X - 2| \geq 2)$ and compare with the exact probability.

17. Let $\bar{X}_{10000}$ be the fraction of heads in 10,000 tosses. Use Chebyshev's inequality to bound $P(|\bar{X}_n - 1/2| \geq 0.01)$ and the normal approximation to estimate this probability.

18. Let X have a Poisson distribution with mean 16. Estimate $P(X \geq 28)$ using (a) Chebyshev's inequality, (b) the normal approximation.

Central limit theorem

19. Suppose that each of 300 patients has a probability of 1/3 of being helped by a treatment. Find approximately the probability that more than 120 patients are helped by the treatment.

20. A person bets you that in 100 tosses of a fair coin the number of Heads will differ from 50 by 4 or more. What is the probability you will win this bet?

21. Suppose we toss a coin 100 times. Which is bigger, the probability of exactly 50 Heads or at least 60 Heads?

22. Suppose that 10% of a certain brand of jelly beans are red. Use the normal approximation to estimate the probability that in a bag of 400 jelly beans there are at least 45 red ones.

23. To estimate the percent of voters who oppose a certain ballot measure, a survey organization takes a random sample of 200 voters. If 45% of the voters 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.

24. A basketball player makes 80% of his free throws on the average. Use the normal approximation to compute the probability that in 25 attempts he will make at least 23.

25. In a 162 game season find the approximate probability that a team with a .5 chance of winning will win at least 87 games.

26. Suppose we roll a die 600 times. What is the approximate probability that the number of 1's obtained lies between 90 and 110?

27. British Airways and United offer identical service on two flights from New York to London that leave at the same time. Suppose that they are competing for the same pool of 400 customers who choose an airline at random. What is the probability United will have more customers than its 230 seats?

28. An insurance company has 10,000 automobile policyholders. The expected yearly claim per policyholder is \$240 with a standard deviation of \$800. Approximate the probability that the yearly claim exceeds \$2.7 million.

29. On each bet a gambler loses \$1 with probability 0.7, loses \$2 with probability 0.2, and wins \$10 with probability 0.1. Estimate the probability that the gambler will be losing after 100 bets.

30. Suppose we roll a die 10 times. What is the approximate probability that the sum of the numbers obtained lies between 30 and 40?

31. An airline knows that in the long run only 90% of passengers who book a seat show up for their flight. On a particular flight with 300 seats there are 324 reservations. (a) Assuming passengers make independent decisions what is the chance that the flight will be over booked? (b) Redo (a) assuming passengers travel in pairs and each pair flips a coin with probability 0.9 of heads to see if they will both show up or both stay home.

32. A student is taking a true/false test with 48 questions. (a) Suppose she has a probability $p = 3/4$ of getting each question right. What is the probability she will get at least 38 right? (b) Answer the last question if she knows the answers to half the questions and flips a coin to answer the other half. Notice that in each case the expected number of questions she gets right is 36.

33. The number of students who enroll in a psychology class is Poisson with mean 100. If the enrollment is > 120 then the class will be split into two sections. Estimate the probability that this will occur.

34. A gymnast has a difficult trick with a 10% chance of success. She tries the trick 25 times and wants to know the probability she will get exactly two successes. Compute the (a) exact answer, (b) Poisson approximation, (c) normal approximation.

35. Suppose that we roll two dice 180 times and we are interested in the probability that we get exactly 5 double sixes. Find (a) the normal approximation, (b) the exact answer, (c) the Poisson approximation.

36. A seed manufacturer sells seeds in packets of 50. Assume that each seed germinates with probability 0.99 independently of all the others. The manufacturer promises to replace, at no cost to the buyer, any packet with 3 or more seeds that do not germinate. (a) Use the Poisson to estimate the probability a packet must be replaced. (b) Use the normal to estimate the probability that the manufacturer has to replace more than 70 of the last 4000 packets sold.

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

38. A die is rolled repeatedly until the sum of the numbers obtained is larger than 200. What is the probability that you can do this in 66 rolls or fewer?

39. Suppose that the checkout time at a grocery store has a mean of 5 minutes and a standard deviation of 2 minutes. Estimate the probability that a checker will serve at least 49 customers during her 4-hour shift.

40. A fair coin is tossed 2500 times. Find a number m so that the chance that the number of heads is between 1250-m and 1250+m is approximately 2/3.

41. Members of the Beta Eta Zeta fraternity each drink a random number of beers with mean 6 and standard deviation 3. If there are 81 fraternity members, how much should they buy so that using the normal approximation they are 93.32% sure they will not run out?

42. An electronics company produces devices that work properly 95% of the time. The new devices are shipped in boxes of 400. The company wants to guarantee that k or more devices per box work. What is the largest k so that at least 95% of the boxes meet the warranty?