11. The nine players on a basketball team consist of 2 centers, 3 forwards, and 4 backcourt players. If the players are paired up at random into three groups of size 3 each, find the (a) expected value and the (b) variance of the number of triplets consisting of one of each type of player.

12. A deck of 52 cards is shuffled and a bridge hand of 13 cards is dealt out. Let X and Y denote, respectively, the number of aces and the number of spades in the dealt hand.
(a) Show that X and Y are uncorrelated.
(b) Are they independent?

13. Each coin in a bin has a value attached to it. Each time that a coin with value p is flipped it lands on heads with probability p. When a coin is randomly chosen from the bin, its value is uniformly distributed on (0, 1). Suppose that after the coin is chosen but before it is flipped, you must predict whether it will land heads or tails. You will win 1 if you are correct and will lose 1 otherwise.
(a) What is your expected gain if you are not told the value of the coin?
(b) Suppose now that you are allowed to inspect the coin before it is flipped, with the result of your inspection being that you learn the value of the coin. As a function of p, the value of the coin, what prediction should you make?
(c) Under the conditions of part (b), what is your expected gain?

14. In Self-Test Problem 1 we showed how to use the value of a uniform (0, 1) random variable (commonly called a *random number*) to obtain the value of a random variable whose mean is equal to the expected number of distinct names on a list. However, its use required that one chooses a random position and then determine the number of times that the name in that position appears on the list. Another approach, which can be more efficient when there is a large amount of name replication, is as follows. As before, start by choosing the random variable X as in Problem 3. Now identify the name in position X, and then go through the list starting at the beginning until that name appears. Let I equal 0 if you encounter that name before getting to position X, and let I equal 1 if your first encounter with the name is at position X. Show that $E[mI] = d$.

HINT: Compute $E[I]$ by using conditional expectation.

CHAPTER 8

Limit Theorems

8.1 INTRODUCTION

The most important theoretical results in probability theory are limit theorems. Of these, the most important are those that are classified either under the heading *laws of large numbers* or under the heading *central limit theorems.* Usually, theorems are considered to be laws of large numbers if they are concerned with stating conditions under which the average of a sequence of random variables converges (in some sense) to the expected average. On the other hand, central limit theorems are concerned with determining conditions under which the sum of a large number of random variables has a probability distribution that is approximately normal.

8.2 CHEBYSHEV'S INEQUALITY AND THE WEAK LAW OF LARGE NUMBERS

We start this section by proving a result known as Markov's inequality.

Proposition 2.1 Markov's inequality

If X is a random variable that takes only nonnegative values, then for any value $a > 0$,

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

Proof: For $a > 0$, let

$$I = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{otherwise} \end{cases}$$

and note that since $X \geq 0$,

$$I \leq \frac{X}{a}$$

Taking expectations of the above yields that

$$E[I] \leq \frac{E[X]}{a}$$

which, since $E[I] = P\{X \geq a\}$, proves the result.

As a corollary, we obtain Proposition 2.2.

Proposition 2.2 Chebyshev's inequality

If X is a random variable with finite mean μ and variance σ^2, then for any value $k > 0$,

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

Proof: Since $(X - \mu)^2$ is a nonnegative random variable, we can apply Markov's inequality (with $a = k^2$) to obtain

$$P\{(X - \mu)^2 \geq k^2\} \leq \frac{E[(X - \mu)^2]}{k^2} \tag{2.1}$$

But since $(X - \mu)^2 \geq k^2$ if and only if $|X - \mu| \geq k$, Equation (2.1) is equivalent to

$$P\{|X - \mu| \geq k\} \leq \frac{E[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

and the proof is complete.

The importance of Markov's and Chebyshev's inequalities is that they enable us to derive bounds on probabilities when only the mean, or both the mean and the variance, of the probability distribution are known. Of course, if the actual distribution were known, then the desired probabilities could be exactly computed and we would not need to resort to bounds.

Example 2a. Suppose that it is known that the number of items produced in a factory during a week is a random variable with mean 50.
(a) What can be said about the probability that this week's production will exceed 75?
(b) If the variance of a week's production is known to equal 25, then what can be said about the probability that this week's production will be between 40 and 60?

Solution Let X be the number of items that will be produced in a week:
(a) By Markov's inequality

$$P\{X > 75\} \leq \frac{E[X]}{75} = \frac{50}{75} = \frac{2}{3}$$

(b) By Chebyshev's inequality

$$P\{|X - 50| \geq 10\} \leq \frac{\sigma^2}{10^2} = \frac{1}{4}$$

Hence

$$P\{|X - 50| < 10\} \geq 1 - \tfrac{1}{4} = \tfrac{3}{4}$$

so the probability that this week's production will be between 40 and 60 is at least .75. ∎

As Chebyshev's inequality is valid for all distributions of the random variable X, we cannot expect the bound on the probability to be very close to the actual probability in most cases. For instance, consider Example 2b.

Example 2b. If X is uniformly distributed over the interval (0, 10), then, as $E[X] = 5$, $\text{Var}(X) = \frac{25}{3}$, it follows from Chebyshev's inequality that

$$P\{|X - 5| > 4\} \leq \frac{25}{3(16)} \approx .52$$

whereas the exact result is

$$P\{|X - 5| > 4\} = .20$$

Thus, although Chebyshev's inequality is correct, the upper bound that it provides is not particularly close to the actual probability.

Similarly, if X is a normal random variable with mean μ and variance σ^2, Chebyshev's inequality states that

$$P\{|X - \mu| > 2\sigma\} \leq \tfrac{1}{4}$$

whereas the actual probability is given by

$$P\{|X - \mu| > 2\sigma\} = P\left\{\left|\frac{X - \mu}{\sigma}\right| > 2\right\} = 2[1 - \Phi(2)] \approx .0456 \quad ∎$$

Chebyshev's inequality is often used as a theoretical tool in proving results. This is illustrated first by Proposition 2.3 and then, most importantly, by the weak law of large numbers.

Proposition 2.3

If $\text{Var}(X) = 0$, then

$$P\{X = E[X]\} = 1$$

In other words, the only random variables having variances equal to 0 are those that are constant with probability 1.

Proof: By Chebyshev's inequality we have, for any $n \geq 1$

$$P\left\{|X - \mu| > \frac{1}{n}\right\} = 0$$

Letting $n \to \infty$ and using the continuity property of probability yields

$$0 = \lim_{n\to\infty} P\left\{|X - \mu| > \frac{1}{n}\right\} = P\left\{\lim_{n\to\infty}\left\{|X - \mu| > \frac{1}{n}\right\}\right\} = P\{X \neq \mu\}$$

and the result is established.

Theorem 2.1 The weak law of large numbers

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean $E[X_i] = \mu$. Then, for any $\varepsilon > 0$,

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right\} \to 0 \quad \text{as} \quad n \to \infty$$

Proof: We shall prove the result only under the additional assumption that the random variables have a finite variance σ^2. Now, as

$$E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu \quad \text{and} \quad \operatorname{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

it follows from Chebyshev's inequality that

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2}$$

and the result is proved.

The weak law of large numbers was originally proved by James Bernoulli for the special case where the X_i are 0—1 (that is, Bernoulli) random variables. His statement and proof of this theorem were presented in his book *Ars Conjectandi,* which was published in 1713, 8 years after his death by his nephew Nicholas Bernoulli. It should be noted that as Chebyshev's inequality was not known in his time, Bernoulli had to resort to a quite ingenious proof to establish the result. The general form of the weak law of large numbers presented in Theorem 2.1 was proved by the Russian mathematician Khintchine.

8.3 THE CENTRAL LIMIT THEOREM

The central limit theorem is one of the most remarkable results in probability theory. Loosely put, it states that the sum of a large number of independent random variables has a distribution that is approximately normal. Hence it not only provides a simple method for computing approximate probabilities for sums of independent random variables, but it also helps explain the remarkable fact that the empirical frequencies of so many natural populations exhibit bell-shaped (that is, normal) curves.

In its simplest form the central limit theorem is as follows.

Theorem 3.1 The central limit theorem

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables each having mean μ and variance σ^2. Then the distribution of

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \to \infty$. That is, for $-\infty < a < \infty$,

$$P\left\{\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2/2}\,dx \quad \text{as} \quad n \to \infty$$

The key to the proof of the central limit theorem is the following lemma, which we state without proof.

Lemma 3.1

Let $Z_1, Z_2, \ldots$ be a sequence of random variables having distribution functions F_{Z_n} and moment generating functions M_{Z_n}, $n \geq 1$; and let Z be a random variable having distribution function F_Z and moment generating function M_Z. If $M_{Z_n}(t) \to M_Z(t)$ for all t, then $F_{Z_n}(t) \to F_Z(t)$ for all t at which $F_Z(t)$ is continuous.

If we let Z be a unit normal random variable, then, as $M_Z(t) = e^{t^2/2}$, it follows from Lemma 3.1 that if $M_{Z_n}(t) \to e^{t^2/2}$ as $n \to \infty$, then $F_{Z_n}(t) \to \Phi(t)$ as $n \to \infty$.

We are now ready to prove the central limit theorem.

Proof of the Central Limit Theorem: Let us assume at first that $\mu = 0$ and $\sigma^2 = 1$. We shall prove the theorem under the assumption that the moment generating function of the X_i, $M(t)$, exists and is finite. Now the moment generating function of $X_i/\sqrt{n}$ is given by

$$E\left[\exp\left\{\frac{tX_i}{\sqrt{n}}\right\}\right] = M\left(\frac{t}{\sqrt{n}}\right)$$

and thus the moment generating function of $\sum_{i=1}^{n} X_i/\sqrt{n}$ is given by $\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n$. Let

$$L(t) = \log M(t)$$

and note that

$$\begin{aligned} L(0) &= 0 \\ L'(0) &= \frac{M'(0)}{M(0)} \\ &= \mu \\ &= 0 \\ L''(0) &= \frac{M(0)M''(0) - [M'(0)]^2}{[M(0)]^2} \\ &= E[X^2] \\ &= 1 \end{aligned}$$

Now, to prove the theorem, we must show that $[M(t/\sqrt{n})]^n \to e^{t^2/2}$ as $n \to \infty$, or equivalently, that $nL(t/\sqrt{n}) \to t^2/2$ as $n \to \infty$. To show this, note that

$$\begin{aligned} \lim_{n\to\infty} \frac{L(t/\sqrt{n})}{n^{-1}} &= \lim_{n\to\infty} \frac{-L'(t/\sqrt{n})\, n^{-3/2}\, t}{-2n^{-2}} \qquad \text{by L'Hospital's rule} \\ &= \lim_{n\to\infty} \left[\frac{L'(t/\sqrt{n})t}{2n^{-1/2}}\right] \\ &= \lim_{n\to\infty} \left[\frac{-L''(t/\sqrt{n})\, n^{-3/2}\, t^2}{-2n^{-3/2}}\right] \qquad \text{again by L'Hospital's rule} \\ &= \lim_{n\to\infty} \left[L''\left(\frac{t}{\sqrt{n}}\right)\frac{t^2}{2}\right] \\ &= \frac{t^2}{2} \end{aligned}$$

Thus the central limit theorem is proved when $\mu = 0$ and $\sigma^2 = 1$. The result now follows in the general case by considering the standardized random variables $X_i^* = (X_i - \mu)/\sigma$ and applying the result above, since $E[X_i^*] = 0$, $\text{Var}(X_i^*) = 1$.

REMARK. Although Theorem 3.1 only states that for each a,

$$P\left\{\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le a\right\} \to \Phi(a)$$

it can, in fact, be shown that the convergence is uniform in a. [We say that $f_n(a) \to f(a)$ uniformly in a, if for each $\varepsilon > 0$, there exists an N such that $|f_n(a) - f(a)| < \varepsilon$ for all a whenever $n \ge N$.]

The first version of the central limit theorem was proved by DeMoivre around 1733 for the special case where the X_i are Bernoulli random variables with $p = \frac{1}{2}$. This was subsequently extended by Laplace to the case of arbitrary p. (Since a binomial random variable may be regarded as the sum of n independent and identically distributed Bernoulli random variables, this justifies the normal approximation to the binomial that was presented in Section 5.4.1.) Laplace also discovered the more general form of the central limit theorem given in Theorem 3.1. His proof, however, was not completely rigorous and, in fact, cannot easily be made rigorous. A truly rigorous proof of the central limit theorem was first presented by the Russian mathematician Liapounoff in the period 1901–1902.

This important theorem is illustrated by the central limit theorem module on the text diskette. This diskette plots the density function of the sum of n independent and identically distributed random variables that each take on one of the values 0, 1, 2, 3, 4. When using it, one enters the probability mass function and the desired value of n. Figure 8.1 shows the resulting plots for a specified probability mass function when (a) $n = 5$, (b) $n = 10$, (c) $n = 25$, and (d) $n = 100$.

Example 3a. An astronomer is interested in measuring, in light years, the distance from his observatory to a distant star. Although the astronomer has a measuring technique, he knows that, because of changing atmospheric conditions and normal error, each time a measurement is made it will not yield the exact distance but merely an estimate. As a result the astronomer plans to make a series of measurements and then use the average value of these measurements as his estimated value of the actual distance. If the astronomer believes that the values of the measurements are independent and identically distributed random variables having a common mean d (the actual distance) and a common variance of 4 (light years), how many measurements need he make to be reasonably sure that his estimated distance is accurate to within $\pm.5$ light year?

Solution Suppose that the astronomer decides to make n observations. If $X_1, X_2, \ldots, X_n$ are the n measurements, then, from the central limit theorem, it follows that

$$Z_n = \frac{\sum_{i=1}^{n} X_i - nd}{2\sqrt{n}}$$

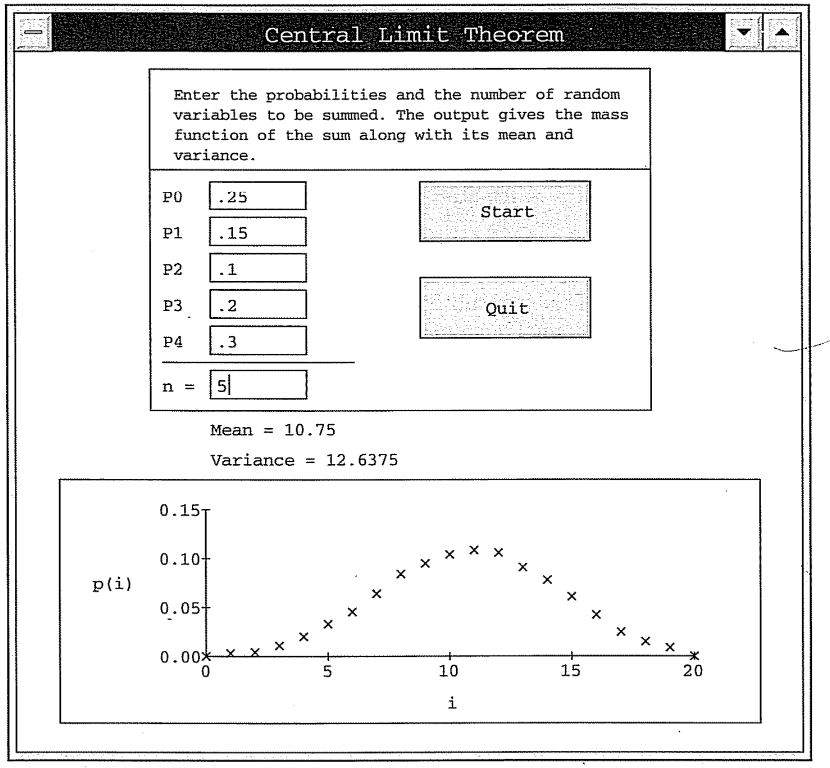

Figure 8.1(a)

has approximately a unit normal distribution. Hence

$$P\left\{-.5 \leq \frac{\sum_{i=1}^{n} X_i}{n} - d \leq .5\right\} = P\left\{-.5\frac{\sqrt{n}}{2} \leq Z_n \leq .5\frac{\sqrt{n}}{2}\right\}$$

$$\approx \Phi\left(\frac{\sqrt{n}}{4}\right) - \phi\left(-\frac{\sqrt{n}}{4}\right) = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1$$

Therefore, if the astronomer wanted, for instance, to be 95 percent certain that his estimated value is accurate to within .5 light year, he should make n^* measurements, where n^* is such that

$$2\Phi\left(\frac{\sqrt{n^*}}{4}\right) - 1 = .95 \qquad \text{or} \qquad \Phi\left(\frac{\sqrt{n^*}}{4}\right) = .975$$

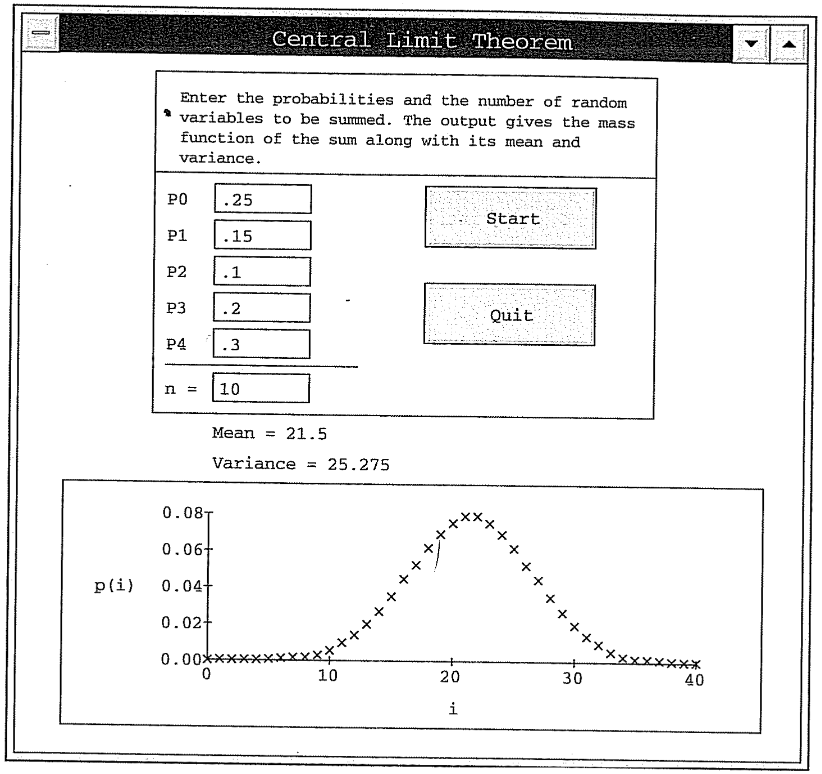

Figure 8.1(b)

and thus from Table 5.1 of Chapter 5,

$$\frac{\sqrt{n^*}}{4} = 1.96 \qquad \text{or} \qquad n^* = (7.84)^2 \approx 61.47$$

As n^* is not integral valued, he should make 62 observations.

It should, however, be noted that the preceding analysis has been done under the assumption that the normal approximation will be a good approximation when $n = 62$. Although this will usually be the case, in general the question of how large n need be before the approximation is "good" depends on the distribution of the X_i. If the astronomer was concerned about this point and wanted to take no chances, he could still solve his problem by using Chebyshev's inequality. Since

$$E\left[\sum_{i=1}^{n} \frac{X_i}{n}\right] = d \qquad \text{Var}\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{4}{n}$$

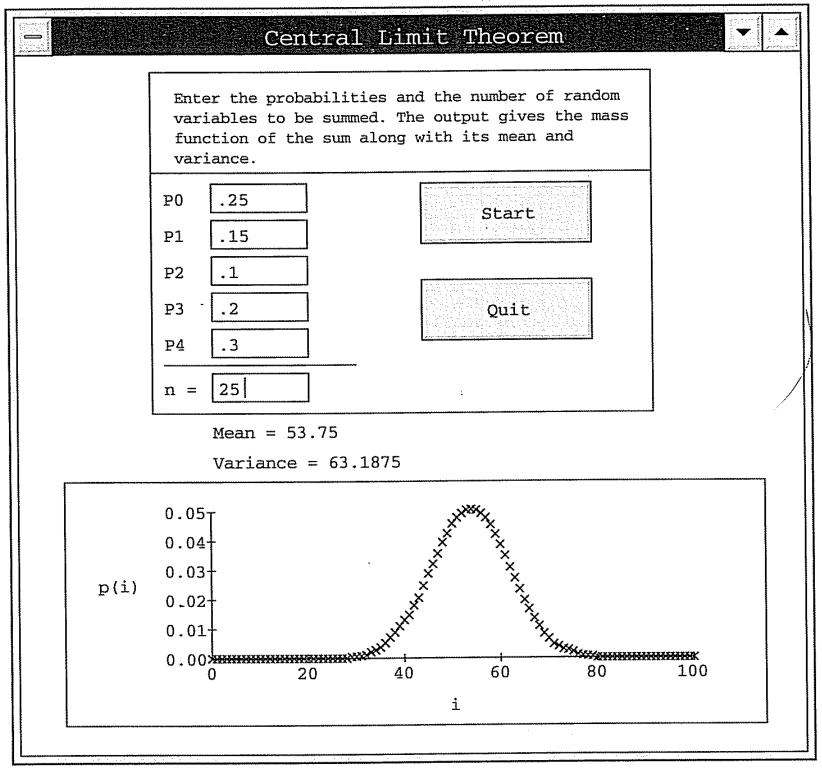

Figure 8.1(c)

Chebyshev's inequality yields that

$$P\left\{\left|\sum_{i=1}^{n} \frac{X_i}{n} - d\right| > .5\right\} \le \frac{4}{n(.5)^2} = \frac{16}{n}$$

Hence, if he makes $n = 16/.05 = 320$ observations, he can be 95 percent certain that his estimate will be accurate to within .5 light year. ∎

Example 3b. The number of students that enroll in a psychology course is a Poisson random variable with mean 100. The professor in charge of the course has decided that if the number enrolling is 120 or more he will teach the course in two separate sections, whereas if fewer than 120 students enroll he will teach all of the students together in a single section. What is the probability that the professor will have to teach two sections?

Solution The exact solution $e^{-100} \sum_{i=120}^{\infty} (100)^i/i!$ does not readily yield a numerical answer. However, by recalling that a Poisson random variable with mean 100 is the sum of 100 independent Poisson random variables

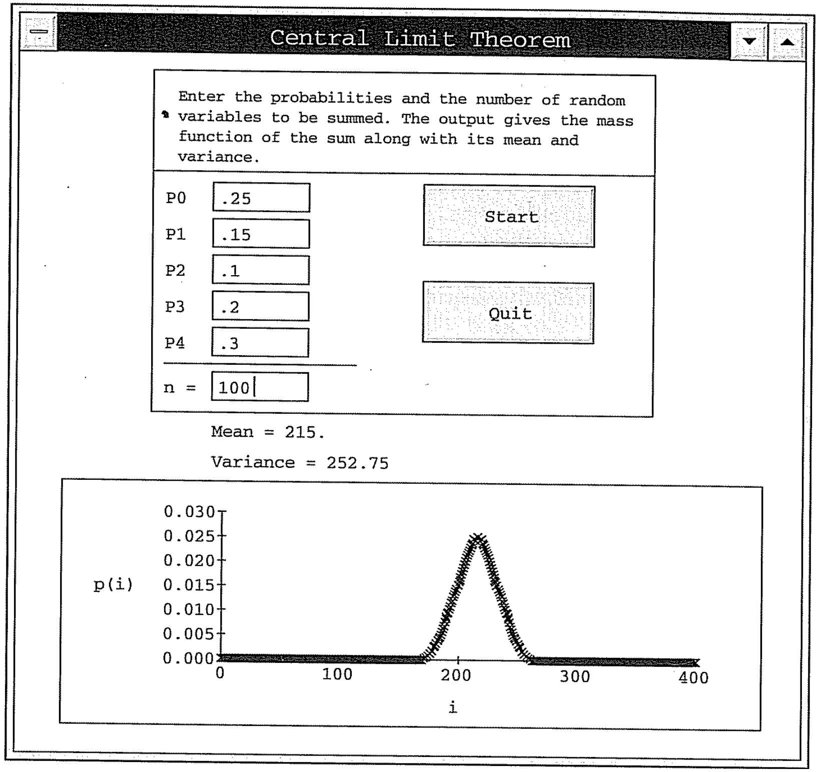

Figure 8.1(d)

each with mean 1, we can make use of the central limit theorem to obtain an approximate solution. If X denotes the number of students that enroll in the course, we have

$$\begin{aligned} P\{X \ge 120\} &= P\left\{\frac{X - 100}{\sqrt{100}} \ge \frac{120 - 100}{\sqrt{100}}\right\} \\ &\approx 1 - \Phi(2) \\ &\approx .0228 \end{aligned}$$

where we have used the fact that the variance of a Poisson random variable is equal to its mean. ∎

Example 3c. If 10 fair dice are rolled, find the approximate probability that the sum obtained is between 30 and 40.

Solution Let X_i denote the value of the ith die, $i = 1, 2, \ldots, 10$. Since $E(X_i) = \frac{7}{2}$, $\text{Var}(X_i) = E[X_i^2] - (E[X_i])^2 = \frac{35}{12}$, the central limit theorem yields

$$P\left\{30 \le \sum_{i=1}^{10} X_i \le 40\right\} = P\left\{\frac{30-35}{\sqrt{\frac{350}{12}}} \le \frac{\sum_{i=1}^{10} X_i - 35}{\sqrt{\frac{350}{12}}} \le \frac{40-35}{\sqrt{\frac{350}{12}}}\right\}$$

$$\approx 2\Phi\left(\sqrt{\tfrac{6}{7}}\right) - 1$$

$$\approx .65$$

Example 3d. Let X_i, $i = 1, \ldots, 10$ be independent random variables, each uniformly distributed over (0, 1). Calculate an approximation to $P\left\{\sum_{i=1}^{10} X_i > 6\right\}$.

Solution Since $E[X_i] = \frac{1}{2}$, $\operatorname{Var}(X_i) = \frac{1}{12}$, we have by the central limit theorem

$$P\left\{\sum_{1}^{10} X_i > 6\right\} = P\left\{\frac{\sum_{1}^{10} X_i - 5}{\sqrt{10\left(\frac{1}{12}\right)}} > \frac{6-5}{\sqrt{10\left(\frac{1}{12}\right)}}\right\}$$

$$\approx 1 - \Phi(\sqrt{1.2})$$

$$\approx .16$$

Hence only 16 percent of the time will $\sum_{i=1}^{10} X_i$ be greater than 6.

Central limit theorems also exist when the X_i are independent but not necessarily identically distributed random variables. One version, by no means the most general, is as follows.

Theorem 3.2 Central limit theorem for independent random variables

Let $X_1, X_2, \ldots$ be a sequence of independent random variables having respective means and variances $\mu_i = E[X_i]$, $\sigma_i^2 = \operatorname{Var}(X_i)$. If (a) the X_i are uniformly bounded; that is, if for some M, $P\{|X_i| < M\} = 1$ for all i, and (b) $\sum_{i=1}^{\infty} \sigma_i^2 = \infty$, then

$$P\left\{\frac{\sum_{i=1}^{n}(X_i - \mu_i)}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} \le a\right\} \to \Phi(a) \quad \text{as} \quad n \to \infty$$

HISTORICAL NOTE

Pierre Simon, Marquis de Laplace

The central limit theorem was originally stated and proved by the French mathematician Pierre Simon, the Marquis de Laplace, who came to this theorem from his observations that errors of measurement (which can usually be regarded as being the sum of a large number of tiny forces) tend to be normally distributed. Laplace, who was also a famous astronomer (and indeed was called "the Newton of France"), was one of the great early contributors to both probability and statistics. Laplace was also a popularizer of the uses of probability in everyday life. He strongly believed in its importance, as is indicated by the following quotations of his taken from his published book *Analytical Theory of Probability*. *"We see that the theory of probability is at bottom only common sense reduced to calculation; it makes us appreciate with exactitude what reasonable minds feel by a sort of instinct, often without being able to account for it. . . . It is remarkable that this science, which originated in the consideration of games of chance, should become the most important object of human knowledge. . . . The most important questions of life are, for the most part, really only problems of probability."*

The application of the central limit theorem to show that measurement errors are approximately normally distributed is regarded as an important contribution to science. Indeed, in the seventeenth and eighteenth centuries the central limit theorem was often called the "law of frequency of errors." The *law of frequency of errors* was considered a major advance by scientists. Listen to the words of Francis Galton (taken from his book *Natural Inheritance*, published in 1889): "I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the 'Law of Frequency of Error.' The Law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement amidst the wildest confusion. The huger the mob and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of unreason."

8.4 THE STRONG LAW OF LARGE NUMBERS

The *strong law of large numbers* is probably the best-known result in probability theory. It states that the average of a sequence of independent random variables having a common distribution will, with probability 1, converge to the mean of that distribution.

Theorem 4.1 The strong law of large numbers

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having a finite mean $\mu = E[X_i]$. Then, with probability 1,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu \quad \text{as} \quad n \to \infty^\dagger$$

As an application of the strong law of large numbers, suppose that a sequence of independent trials of some experiment is performed. Let E be a fixed event of the experiment and denote by $P(E)$ the probability that E occurs on any particular trial. Letting

$$X_i = \begin{cases} 1 & \text{if } E \text{ occurs on the } i\text{th trial} \\ 0 & \text{if } E \text{ does not occur on the } i\text{th trial} \end{cases}$$

we have by the strong law of large numbers that with probability 1,

$$\frac{X_1 + \cdots + X_n}{n} \to E[X] = P(E) \tag{4.1}$$

Since $X_1 + \cdots + X_n$ represents the number of times that the event E occurs in the first n trials, we may interpret Equation (4.1) as stating that, with probability 1, the limiting proportion of time that the event E occurs is just $P(E)$.

Although the theorem can be proven without this assumption, our proof of the strong law of large numbers will assume that the random variables X_i have a finite fourth moment. That is, we will suppose that $E[X_i^4] = K < \infty$.

Proof of the Strong Law of Large Numbers: To begin, assume that μ, the mean of the X_i, is equal to 0. Let $S_n = \sum_{i=1}^{n} X_i$ and consider

$$E[S_n^4] = E[(X_1 + \cdots + X_n)(X_1 + \cdots + X_n) \times (X_1 + \cdots + X_n)(X_1 + \cdots + X_n)]$$

Expanding the right side of the above will result in terms of the form

$$X_i^4, \quad X_i^3X_j, \quad X_i^2X_j^2, \quad X_i^2X_jX_k, \quad \text{and} \quad X_iX_jX_kX_l$$

† That is, the strong law of large numbers states that

$$P\left\{\lim_{n\to\infty} (X_1 + \cdots + X_n)/n = \mu\right\} = 1$$

where i, j, k, l are all different. As all the X_i have mean 0, it follows by independence that

$$\begin{aligned} E[X_i^3X_j] &= E[X_i^3]E[X_j] = 0 \\ E[X_i^2X_jX_k] &= E[X_i^2]E[X_j]E[X_k] = 0 \\ E[X_iX_jX_kX_l] &= 0 \end{aligned}$$

Now, for a given pair i and j there will be $\binom{4}{2} = 6$ terms in the expansion that will equal $X_i^2X_j^2$. Hence it follows upon expanding the preceding product and taking expectations term by term that

$$\begin{aligned} E[S_n^4] &= nE[X_i^4] + 6\binom{n}{2} E[X_i^2X_j^2] \\ &= nK + 3n(n-1)E[X_i^2]E[X_j^2] \end{aligned}$$

where we have once again made use of the independence assumption. Now, since

$$0 \le \text{Var}(X_i^2) = E[X_i^4] - (E[X_i^2])^2$$

we see that

$$(E[X_i^2])^2 \le E[X_i^4] = K$$

Therefore, from the preceding we have that

$$E[S_n^4] \le nK + 3n(n-1)K$$

which implies that

$$E\left[\frac{S_n^4}{n^4}\right] \le \frac{K}{n^3} + \frac{3K}{n^2}$$

Therefore, it follows that

$$E\left[\sum_{n=1}^{\infty} \frac{S_n^4}{n^4}\right] = \sum_{n=1}^{\infty} E\left[\frac{S_n^4}{n^4}\right] < \infty$$

But the preceding implies that with probability 1, $\sum_{n=1}^{\infty} S_n^4/n^4 < \infty$. (For if there is a positive probability that the sum is infinite, then its expected value is infinite.) But the convergence of a series implies that its nth term goes to 0; so we can conclude that with probability 1,

$$\lim_{n\to\infty} \frac{S_n^4}{n^4} = 0$$

But if $S_n^4/n^4 = (S_n/n)^4$ goes to 0, then so must S_n/n; so we have proven that with probability 1,

$$\frac{S_n}{n} \to 0 \quad \text{as} \quad n \to \infty$$

When μ, the mean of the X_i, is not equal to 0, we can apply the preceding argument to the random variables $X_i - \mu$ to obtain that with probability 1,

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{(X_i - \mu)}{n} = 0$$

or, equivalently,

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{X_i}{n} = \mu$$

which proves the result.

The strong law is illustrated by two modules on the text diskette that consider independent and identically distributed random variables which take on one of the values 0, 1, 2, 3, 4. The modules simulate the values of n such random variables; the proportions of time that each outcome occurs, as well as the resulting sample mean $\sum_{i=1}^{n} X_i/n$, are then indicated and plotted. When using these modules, which differ only in the type of graph presented, one enters the probabilities and the desired value of n. Figure 8.2 gives the results of a simulation using a specified probability mass function and (a) $n = 100$, (b) $n = 1000$, and (c) $n = 10{,}000$.

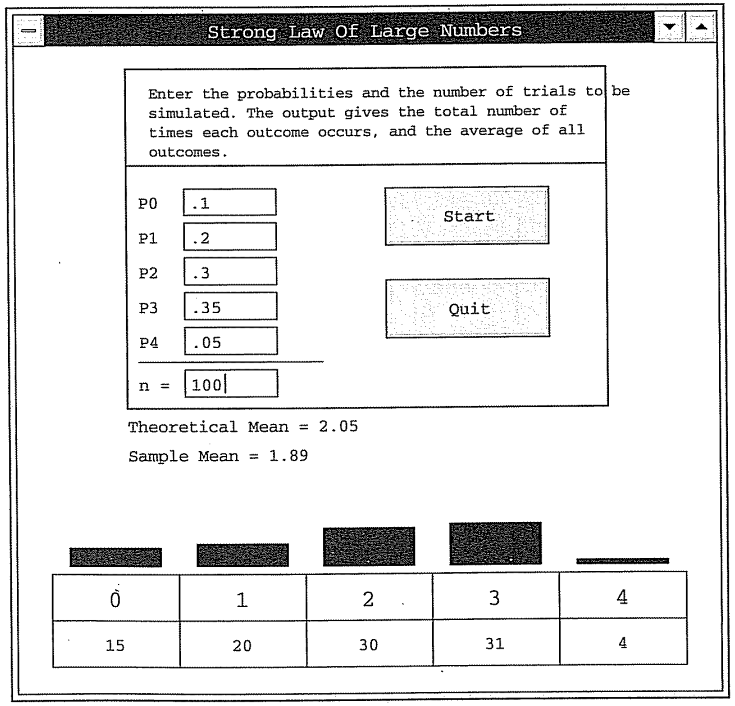

Figure 8.2(a)

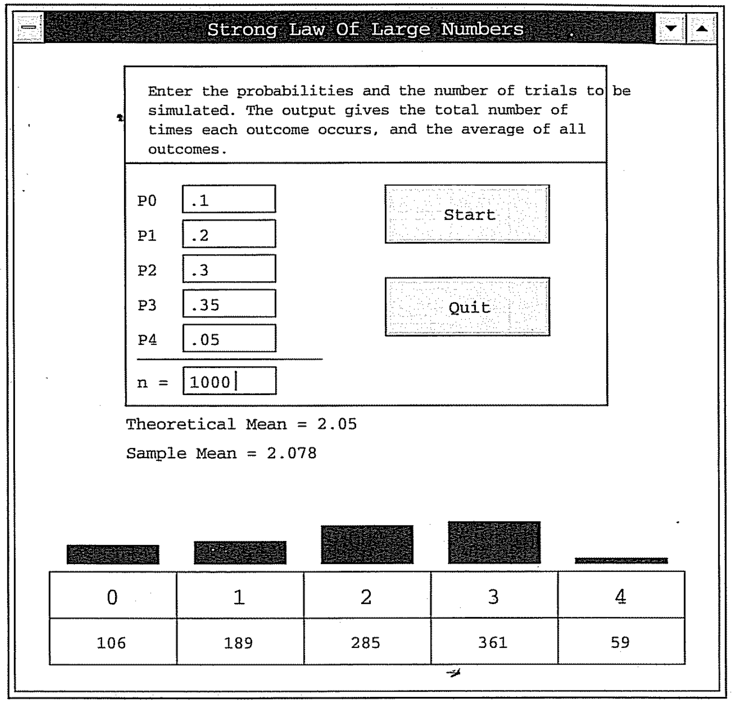

Figure 8.2(b)

Many students are initially confused about the difference between the weak and the strong law of large numbers. The weak law of large numbers states that for any specified large value n^*, $(X_1 + \cdots + X_{n^*})/n^*$ is likely to be near μ. However, it does not say that $(X_1 + \cdots + X_n)/n$ is bound to stay near μ for all values of n larger than n^*. Thus it leaves open the possibility that large values of $|(X_1 + \cdots + X_n)/n - \mu|$ can occur infinitely often (though at infrequent intervals). The strong law shows that this cannot occur. In particular, it implies that with probability 1, for any positive value ε,

$$\left| \sum_{1}^{n} \frac{X_i}{n} - \mu \right|$$

will be greater than ε only a finite number of times.

The strong law of large numbers was originally proved, in the special case of Bernoulli random variables, by the French mathematician Borel. The general form of the strong law presented in Theorem 4.1 was proved by the Russian mathematician A. N. Kolmogorov.

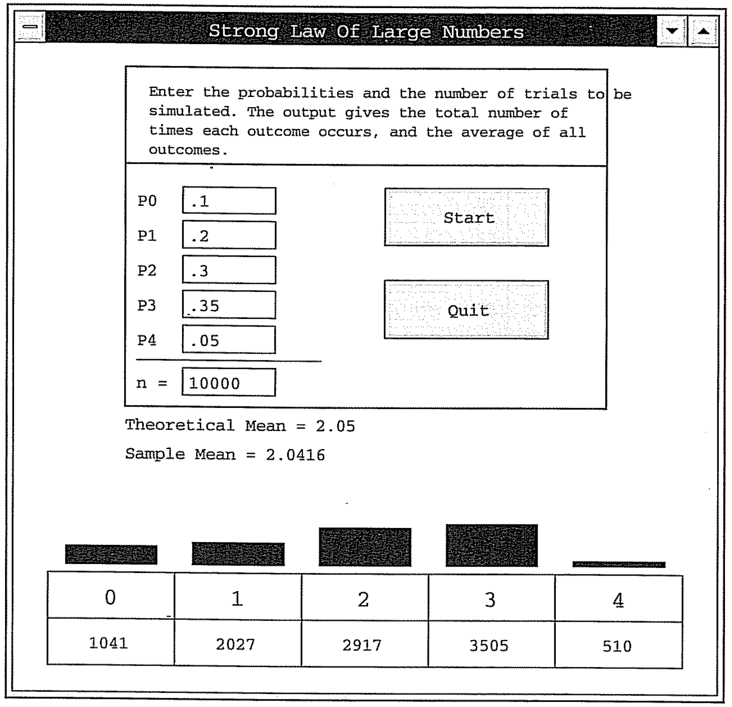

Figure 8.2(c)

8.5 OTHER INEQUALITIES

We are sometimes confronted with situations in which we are interested in obtaining an upper bound for a probability of the form $P\{X - \mu \geq a\}$, where a is some positive value and when only the mean $\mu = E[X]$ and variance $\sigma^2 = \text{Var}(X)$ of the distribution of X are known. Of course, since $X - \mu \geq a > 0$ implies that $|X - \mu| \geq a$, it follows from Chebyshev's inequality that

$$P\{X - \mu \geq a\} \leq P\{|X - \mu| \geq a\} \leq \frac{\sigma^2}{a^2} \qquad \text{when} \quad a > 0$$

However, as the following proposition shows, it turns out that we can do better.

Proposition 5.1 One-sided Chebyshev inequality

If X is a random variable with mean 0 and finite variance σ^2, then for any $a > 0$,

$$P\{X \geq a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

Proof: Let $b > 0$ and note that

$$X \geq a \quad \text{is equivalent to} \quad X + b \geq a + b$$

Hence

$$\begin{aligned} P\{X \geq a\} &= P\{X + b \geq a + b\} \\ &\leq P\{(X + b)^2 \geq (a + b)^2\} \end{aligned}$$

where the inequality above is obtained by noting that since $a + b > 0$, $X + b \geq a + b$ implies that $(X + b)^2 \geq (a + b)^2$. Upon applying Markov's inequality, the above yields that

$$P\{X \geq a\} \leq \frac{E[(X + b)^2]}{(a + b)^2} = \frac{\sigma^2 + b^2}{(a + b)^2}$$

Letting $b = \sigma^2/a$ [which is easily seen to be the value of b that minimizes $(\sigma^2 + b^2)/(a + b)^2$] gives the desired result.

Example 5a. If the number of items produced in a factory during a week is a random variable with mean 100 and variance 400, compute an upper bound on the probability that this week's production will be at least 120.

Solution It follows from the one-sided Chebyshev inequality that

$$P\{X \geq 120\} = P\{X - 100 \geq 20\} \leq \frac{400}{400 + (20)^2} = \frac{1}{2}$$

Hence the probability that this week's production will be 120 or more is at most $\frac{1}{2}$.

If we attempted to obtain a bound by applying Markov's inequality, then we would have obtained

$$P\{X \geq 120\} \leq \frac{E(X)}{120} = \frac{5}{6}$$

which is a far weaker bound than the preceding one. ∎

Suppose now that X has mean μ and variance σ^2. As both $X - \mu$ and $\mu - X$ have mean 0 and variance σ^2, we obtain from the one-sided Chebyshev inequality that for $a > 0$,

$$P\{X - \mu \geq a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

and

$$P\{\mu - X \geq a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

Thus we have the following corollary.

Corollary 5.1

If $E[X] = \mu$, $\text{Var}(X) = \sigma^2$, then for $a > 0$,

$$P\{X \geq \mu + a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

$$P\{X \leq \mu - a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

Example 5b. A set of 200 people, consisting of 100 men and 100 women, is randomly divided into 100 pairs of 2 each. Give an upper bound to the probability that at most 30 of these pairs will consist of a man and a woman.

Solution Number the men, arbitrarily, from 1 to 100 and let for $i = 1, 2, \ldots 100$,

$$X_i = \begin{cases} 1 & \text{if man } i \text{ is paired with a woman} \\ 0 & \text{otherwise} \end{cases}$$

Then X, the number of man–woman pairs, can be expressed as

$$X = \sum_{i=1}^{100} X_i$$

As man i is equally likely to be paired with any of the other 199 people, of which 100 are women, we have

$$E[X_i] = P\{X_i = 1\} = \frac{100}{199}$$

Similarly, for $i \neq j$,

$$\begin{aligned} E[X_iX_j] &= P\{X_i = 1, X_j = 1\} \\ &= P\{X_i = 1\}P\{X_j = 1 | X_i = 1\} = \frac{100}{199}\frac{99}{197} \end{aligned}$$

where $P\{X_j = 1 | X_i = 1\} = 99/197$ since, given that man i is paired with a woman, man j is equally likely to be paired with any of the remaining 197 people, of which 99 are women. Hence we obtain that

$$\begin{aligned} E[X] &= \sum_{i=1}^{100} E[X_i] \\ &= (100)\frac{100}{199} \\ &\approx 50.25 \end{aligned}$$

$$\begin{aligned} \text{Var}(X) &= \sum_{i=1}^{100} \text{Var}(X_i) + 2\sum_{i<j}\sum \text{Cov}(X_i, X_j) \\ &= 100\frac{100}{199}\frac{99}{199} + 2\binom{100}{2}\left[\frac{100}{199}\frac{99}{197} - \left(\frac{100}{199}\right)^2\right] \\ &\approx 25.126 \end{aligned}$$

The Chebyshev inequality yields that

$$P\{X \leq 30\} \leq P\{|X - 50.25| \geq 20.25\} \leq \frac{25.126}{(20.25)^2} \approx .061$$

and thus there are fewer than 6 chances in a hundred that fewer than 30 men will be paired with women. However, we can improve on this bound by using the one-sided Chebyshev inequality, which yields that

$$\begin{aligned} P\{X \leq 30\} &= P\{X \leq 50.25 - 20.25\} \\ &\leq \frac{25.126}{25.126 + (20.25)^2} \\ &\approx .058 \end{aligned}$$

∎

When the moment generating function of the random variable X is known, we can obtain even more effective bounds on $P\{X \geq a\}$. Let

$$M(t) = E[e^{tX}]$$

be the moment generating function of the random variable X. Then for $t > 0$,

$$\begin{aligned} P\{X \geq a\} &= P\{e^{tX} \geq e^{ta}\} \\ &\leq E[e^{tX}]e^{-ta} \qquad \text{by Markov's inequality} \end{aligned}$$

Similarly, for $t < 0$,

$$\begin{aligned} P\{X \leq a\} &= P\{e^{tX} \geq e^{ta}\} \\ &\leq E[e^{tX}]e^{-ta} \end{aligned}$$

Thus we have the following inequalities, known as *Chernoff bounds.*

Proposition 5.2 Chernoff bounds

$$P\{X \geq a\} \leq e^{-ta}M(t) \qquad \text{for all } t > 0$$
$$P\{X \leq a\} \leq e^{-ta}M(t) \qquad \text{for all } t < 0$$

Since the Chernoff bounds hold for all t in either the positive or negative quadrant, we obtain the best bound on $P\{X \geq a\}$ by using the t that minimizes $e^{-ta}M(t)$.

Example 5c. ***Chernoff bounds for the standard normal random variable.*** If Z is a standard normal random variable, then its moment generating function is $M(t) = e^{t^2/2}$, so the Chernoff bound on $P\{Z \geq a\}$ is given by

$$P\{Z \geq a\} \leq e^{-ta}e^{t^2/2} \qquad \text{for all } \quad t > 0$$

Now the value of t, $t > 0$, that minimizes $e^{t^2/2-ta}$ is the value that minimizes $t^2/2 - ta$, which is $t = a$. Thus for $a > 0$ we see that

$$P\{Z \geq a\} \leq e^{-a^2/2}$$

Similarly, we can show that for $a < 0$,

$$P\{Z \leq a\} \leq e^{-a^2/2}$$

■

Example 5d. ***Chernoff bounds for the Poisson random variable.*** If X is a Poisson random variable with parameter λ, then its moment generating function is $M(t) = e^{\lambda(e^t-1)}$. Hence the Chernoff bound on $P\{X \geq i\}$ is

$$P\{X \geq i\} \leq e^{\lambda(e^t-1)}e^{-it} \qquad t > 0$$

Minimizing the right side of the above is equivalent to minimizing $\lambda(e^t - 1) - it$, and calculus shows that the minimal value occurs when $e^t = i/\lambda$. Provided that $i/\lambda > 1$, this minimizing value of t will be positive. Therefore, assuming that $i > \lambda$ and letting $e^t = i/\lambda$ in the Chernoff bound yields that

$$P\{X \geq i\} \leq e^{\lambda(i/\lambda-1)}\left(\frac{\lambda}{i}\right)^i$$

or, equivalently,

$$P\{X \geq i\} \leq \frac{e^{-\lambda}(e\lambda)^i}{i^i}$$

■

Example 5e. Consider a gambler who on every play is equally likely, independent of the past, to either win or lose 1 unit. That is, if X_i is the gambler's winnings on the ith play, then the X_i are independent and

$$P\{X_i = 1\} = P\{X_i = -1\} = \tfrac{1}{2}$$

Let $S_n = \sum_{i=1}^{n} X_i$ denote the gambler's winnings after n plays. We will use the Chernoff bound on $P\{S_n \geq a\}$. To start, note that the moment generating function of X_i is

$$E[e^{tX}] = \frac{e^t + e^{-t}}{2}$$

Now, using the McLaurin expansions of e^t and e^{-t} we see that

$$e^t + e^{-t} = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots + \left(1 - t + \frac{t^2}{2!} - \frac{t^3}{3!} + \cdots\right)$$

$$= 2\left\{1 + \frac{t^2}{2!} + \frac{t^4}{4!} + \cdots\right\}$$

$$= 2\sum_{n=0}^{\infty} \frac{t^{2n}}{(2n)!}$$

$$\leq 2\sum_{n=0}^{\infty} \frac{(t^2/2)^n}{n!} \qquad \text{since } (2n)! \geq n!\,2^n$$

$$= 2e^{t^2/2}$$

Therefore,

$$E[e^{tX}] \leq e^{t^2/2}$$

Since the moment generating function of the sum of independent random variables is the product of their moment generating functions, we have that

$$E[e^{tS_n}] = (E[e^{tX}])^n$$
$$\leq e^{nt^2/2}$$

Using the result above along with the Chernoff bound gives that

$$P\{S_n \geq a\} \leq e^{-ta}e^{nt^2/2} \qquad t > 0$$

The value of t that minimizes the right side of the above is the value that minimizes $nt^2/2 - ta$, and this value is $t = a/n$. Supposing that $a > 0$ (so that this minimizing t is positive) and letting $t = a/n$ in the preceding inequality yields that

$$P\{S_n \geq a\} \leq e^{-a^2/2n} \qquad a > 0$$

For instance, this inequality yields that

$$P\{S_{10} \geq 6\} \leq e^{-36/20} \approx .1653$$

whereas the exact probability is

$$P\{S_{10} \geq 6\} = P\{\text{gambler wins at least 8 of the first 10 games}\}$$

$$= \frac{\binom{10}{8} + \binom{10}{9} + \binom{10}{10}}{2^{10}} = \frac{56}{1024} \approx .0547$$

■

The next inequality is one having to do with expectations rather than probabilities. Before stating it, we need the following definition.

Definition

A twice-differentiable real-valued function $f(x)$ is said to be *convex* if $f''(x) \geq 0$ for all x; similarly, it is said to be *concave* if $f''(x) \leq 0$.

Some examples of convex functions are $f(x) = x^2$, $f(x) = e^{ax}$, $f(x) = -x^{1/n}$ for $x \geq 0$. If $f(x)$ is convex, then $g(x) = -f(x)$ is concave, and vice versa.

Proposition 5.3 *Jensen's inequality*

If $f(x)$ is a convex function, then

$$E[f(X)] \geq f(E[X])$$

provided that the expectations exist and are finite.

Proof: Expanding $f(x)$ in a Taylor's series expansion about $\mu = E[X]$ yields

$$f(x) = f(\mu) + f'(\mu)(x - \mu) + \frac{f''(\xi)(x - \xi)^2}{2}$$

where ξ is some value between x and μ. Since $f''(\xi) \geq 0$, we obtain

$$f(x) \geq f(\mu) + f'(\mu)(x - \mu)$$

Hence

$$f(X) \geq f(\mu) + f'(\mu)(X - \mu)$$

Taking expectations yields

$$E[f(X)] \geq f(\mu) + f'(\mu)E[X - \mu] = f(\mu)$$

and the inequality is established.

Example 5f. An investor is faced with the following choices: She can either invest all of her money in a risky proposition that would lead to a random return X that has mean m; or she can put the money into a risk-free venture that will lead to a return of m with probability 1. Suppose that her decision will be made on the basis of maximizing the expected value of $u(R)$, where R is her return and u is her utility function. By Jensen's inequality it follows that if u is a concave function, then $E[u(X)] \leq u(m)$, so the risk-free alternative is preferable; whereas if u is convex, then $E[u(X)] \geq u(m)$, so the risk investment alternative would be preferred.

8.6 BOUNDING THE ERROR PROBABILITY WHEN APPROXIMATING A SUM OF INDEPENDENT BERNOULLI RANDOM VARIABLES BY A POISSON RANDOM VARIABLE

In this section we establish bounds on how closely a sum of independent Bernoulli random variables is approximated by a Poisson random variable with the same mean. Suppose that we want to approximate the sum of independent Bernoulli random variables with respective means $p_1, p_2, \ldots, p_n$. Starting with a sequence $Y_1, \ldots, Y_n$ of independent Poisson random variables, with Y_i having mean

p_i, we will construct a sequence of independent Bernoulli random variables $X_1, \ldots, X_n$ with parameters $p_1, \ldots, p_n$ such that

$$P\{X_i \neq Y_i\} \leq p_i^2 \qquad \text{for each } i$$

Letting $X = \sum_{i=1}^{n} X_i$ and $Y = \sum_{i=1}^{n} Y_i$, we will use the preceding fact to conclude that

$$P\{X \neq Y\} \leq \sum_{i=1}^{n} p_i^2$$

Finally, we will show that the inequality above implies that for any set of real numbers A,

$$|P\{X \in A\} - P\{Y \in A\}| \leq \sum_{i=1}^{n} p_i^2$$

Since X is the sum of independent Bernoulli random variables and Y is a Poisson random variable, the preceding inequality will yield the desired bound.

To show how the preceding is accomplished, let Y_i, $i = 1, \ldots, n$ be independent Poisson random variables with respective means p_i. Now let $U_1, \ldots, U_n$ be independent random variables that are also independent of the Y_i's and which are such that

$$U_i = \begin{cases} 0 & \text{with probability} \quad (1 - p_i)e^{p_i} \\ 1 & \text{with probability} \quad 1 - (1 - p_i)e^{p_i} \end{cases}$$

The preceding definition implicity makes use of the inequality

$$e^{-p} \geq 1 - p$$

in assuming that $(1 - p_i)e^{p_i} \leq 1$.

Now define the random variables X_i, $i = 1, \ldots, n$ by

$$X_i = \begin{cases} 0 & \text{if } Y_i = U_i = 0 \\ 1 & \text{otherwise} \end{cases}$$

Note that

$$P\{X_i = 0\} = P\{Y_i = 0\}P\{U_i = 0\} = e^{-p_i}(1 - p_i)e^{p_i} = 1 - p_i$$
$$P\{X_i = 1\} = 1 - P\{X_i = 0\} = p_i$$

Now if X_i is equal to 0, then so must Y_i equal 0 (by the definition of X_i). Therefore, we see that

$$\begin{aligned} P\{X_i \neq Y_i\} &= P\{X_i = 1, Y_i \neq 1\} \\ &= P\{Y_i = 0, X_i = 1\} + P\{Y_i > 1\} \\ &= P\{Y_i = 0, U_i = 1\} + P\{Y_i > 1\} \\ &= e^{-p_i}[1 - (1 - p_i)e^{p_i}] + 1 - e^{-p_i} - p_i e^{-p_i} \\ &= p_i - p_i e^{-p_i} \\ &\leq p_i^2 \qquad (\text{since } 1 - e^{-p} \leq p) \end{aligned}$$

Now let $X = \sum_{i=1}^{n} X_i$ and $Y = \sum_{i=1}^{n} Y_i$ and note that X is the sum of independent Bernoulli random variables and Y is Poisson with the expected value $E[Y] = E[X] = \sum_{i=1}^{n} p_i$. Note also that the inequality $X \neq Y$ implies that $X_i \neq Y_i$ for some i, so

$$\begin{aligned} P\{X \neq Y\} &\leq P\{X_i \neq Y_i \text{ for some } i\} \\ &\leq \sum_{i=1}^{n} P\{X_i \neq Y_i\} \qquad \text{(Boole's inequality)} \\ &\leq \sum_{i=1}^{n} p_i^2 \end{aligned}$$

For any event B, let I_B, the indicator variable for the event B, be defined by

$$I_B = \begin{cases} 1 & \text{if } B \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

Note that for any set of real numbers A,

$$I_{\{X \in A\}} - I_{\{Y \in A\}} \leq I_{\{X \neq Y\}}$$

The above follows since, as an indicator variable is either 0 or 1, the left-hand side equals 1 only when $I_{\{X \in A\}} = 1$ and $I_{\{Y \in A\}} = 0$. But this would imply that $X \in A$ and $Y \notin A$, which means that $X \neq Y$, so the right side would also equal 1. Upon taking expectations of the preceding inequality, we obtain that

$$P\{X \in A\} - P\{Y \in A\} \leq P\{X \neq Y\}$$

By reversing X and Y, we obtain in the same manner that

$$P\{Y \in A\} - P\{X \in A\} \leq P\{X \neq Y\}$$

and thus we can conclude that

$$|P\{X \in A\} - P\{Y \in A\}| \leq P\{X \neq Y\}$$

Therefore, we have proven that with $\lambda = \sum_{i=1}^{n} p_i$,

$$\left| P\left\{ \sum_{i=1}^{n} X_i \in A \right\} - \sum_{i \in A} \frac{e^{-\lambda}\lambda^i}{i!} \right| \leq \sum_{i=1}^{n} p_i^2$$

REMARK. When all the p_i are equal to p, X is a binomial random variable. Thus the above shows that for any set of nonnegative integers A,

$$\left| \sum_{i \in A} \binom{n}{i} p^i (1-p)^{n-i} - \sum_{i \in A} \frac{e^{-np}(np)^i}{i!} \right| \leq np^2$$

SUMMARY

Two useful probability bounds are provided by the *Markov* and *Chebyshev* inequalities. The Markov inequality is concerned with nonnegative random variables, and says that for X of that type

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

for every positive value a. The Chebyshev inequality, which is a simple consequence of the Markov inequality, states that if X has mean μ and variance σ^2, then for every positive k,

$$P\{|X - \mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

The two most important theoretical results in probability are the *central limit theorem* and the *strong law of large numbers.* Both are concerned with a sequence of independent and identically distributed random variables. The central limit theorem says that if the random variables have a finite mean μ and a finite variance σ^2, then the distribution of the sum of the first n of them is, for large n, approximately that of a normal random variable with mean $n\mu$ and variance $n\sigma^2$. That is, if X_i, $i \geq 1$, is the sequence, then the central limit theorem states that for every real number a,

$$\lim_{n \to \infty} P\left\{ \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}\, dx$$

The *strong law of large numbers* requires only that the random variables in the sequence have a finite mean μ. It states that with probability 1, the average of the first n of them will converge to μ as n goes to infinity. This implies that if A is any specified event of an experiment for which independent replications are performed, then the limiting proportion of experiments whose outcomes are in A will, with probability 1, equal $P(A)$. Therefore, if we accept the interpretation that "with probability 1" means "with certainty," we obtain the theoretical justification for the long-run relative frequency interpretation of probabilities.

PROBLEMS

1. Suppose that X is a random variable with mean and variance both equal to 20. What can be said about $P\{0 \leq X \leq 40\}$?

2. From past experience a professor knows that the test score of a student taking her final examination is a random variable with mean 75.

(a) Give an upper bound for the probability that a student's test score will exceed 85.

Suppose, in addition, the professor knows that the variance of a student's test score is equal to 25.

(b) What can be said about the probability that a student will score between 65 and 85?
(c) How many students would have to take the examination to ensure, with probability at least .9, that the class average would be within 5 of 75? Do not use the central limit theorem.

3. Use the central limit theorem to solve part (c) of Problem 2.

4. Let $X_1, \ldots, X_{20}$ be independent Poisson random variables with mean 1.
(a) Use the Markov inequality to obtain a bound on

$$P\left\{\sum_1^{20} X_i > 15\right\}$$

(b) Use the central limit theorem to approximate $P\left\{\sum_1^{20} X_i > 15\right\}$.

5. Fifty numbers are rounded off to the nearest integer and then summed. If the individual round-off errors are uniformly distributed over $(-.5, .5)$ what is the probability that the resultant sum differs from the exact sum by more than 3?

6. A die is continually rolled until the total sum of all rolls exceeds 300. What is the probability that at least 80 rolls are necessary?

7. One has 100 light bulbs whose lifetimes are independent exponentials with mean 5 hours. If the bulbs are used one at a time, with a failed bulb being replaced immediately by a new one, what is the probability that there is still a working bulb after 525 hours?

8. In Problem 7 suppose that it takes a random time, uniformly distributed over (0, .5), to replace a failed bulb. What is the probability that all bulbs have failed by time 550?

9. If X is a gamma random variable with parameters $(n, 1)$ how large need n be so that

$$P\left\{\left|\frac{X}{n} - 1\right| > .01\right\} < .01?$$

10. Civil engineers believe that W, the amount of weight (in units of 1000 pounds) that a certain span of a bridge can withstand without structural damage resulting, is normally distributed with mean 400 and standard deviation 40. Suppose that the weight (again, in units of 1000 pounds) of a car is a random variable with mean 3 and standard deviation .3. How many cars would have to be on the bridge span for the probability of structural damage to exceed .1?

11. Many people believe that the daily change of price of a company's stock on the stock market is a random variable with mean 0 and variance σ^2. That is, if Y_n represents the price of the stock on the nth day, then

$$Y_n = Y_{n-1} + X_n \qquad n \geq 1$$

where $X_1, X_2, \ldots$ are independent and identically distributed random variables with mean 0 and variance σ^2. Suppose that the stock's price today is 100. If $\sigma^2 = 1$, what can you say about the probability that the stock's price will exceed 105 after 10 days?

12. We have 100 components that we will put in use in a sequential fashion. That is, component 1 is initially put in use, and upon failure it is replaced by component 2, which is itself replaced upon failure by component 3, and so on. If the lifetime of component i is exponentially distributed with mean $10 + i/10$, $i = 1, \ldots, 100$, estimate the probability that the total life of all components will exceed 1200. Now repeat when the life distribution of component i is uniformly distributed over $(0, 20 + i/5)$, $i = 1, \ldots, 100$.

13. Student scores on exams given by a certain instructor have mean 74 and standard deviation 14. This instructor is about to give two exams, one to a class of size 25 and the other to a class of size 64.
(a) Approximate the probability that the average test score in the class of size 25 exceeds 80.
(b) Repeat part (a) for the class of size 64.
(c) Approximate the probability that the average test score in the larger class exceeds that of the other class by over 2.2 points.
(d) Approximate the probability that the average test score in the smaller class exceeds that of the other class by over 2.2 points.

14. A certain component is critical to the operation of an electrical system and must be replaced immediately upon failure. If the mean lifetime of this type of component is 100 hours and its standard deviation is 30 hours, how many of these components must be in stock so that the probability that the system is in continual operation for the next 2000 hours is at least .95?

15. 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 total yearly claim exceeds \$2.7 million.

16. Redo Example 5b under the assumption that the number of man–woman pairs is (approximately) normally distributed. Does this seem like a reasonable supposition?

17. Repeat part (a) of Problem 2 when it is known that the variance of a student's test score is equal to 25.

18. A lake contains 4 distinct types of fish. Suppose that each fish caught is equally likely to be any one of these types. Let Y denote the number of fish that need be caught to obtain at least one of each type.
(a) Give an interval (a, b) such that $P\{a \leq Y \leq b\} \geq .90$.
(b) Using the one-sided Chebyshev inequality, how many fish need we plan on catching so as to be at least 90 percent certain of obtaining at least one of each type?

19. If X is a nonnegative random variable with mean 25, what can be said about:
(a) $E[X^3]$;
(b) $E[\sqrt{X}]$;
(c) $E[\log X]$;
(d) $E[e^{-X}]$?

20. Let X be a nonnegative random variable. Prove that

$$E[X] \le (E[X^2])^{1/2} \le (E[X^3])^{1/3} \le \cdots$$

21. Would the results of Example 5f change if the investor were allowed to divide her money and invest the fraction α, $0 < \alpha < 1$ in the risky proposition and invest the remainder in the risk-free venture? Her return for such a split investment would be $R = \alpha X + (1 - \alpha)m$.

22. Let X be a Poisson random variable with mean 20.

(a) Use the Markov inequality to obtain an upper bound on

$$p = P\{X \ge 26\}$$

(b) Use the one-sided Chebyshev inequality to obtain an upper bound on p.
(c) Use the Chernoff bound to obtain an upper bound on p.
(d) Approximate p by making use of the central limit theorem.
(e) Determine p by running an appropriate program.

THEORETICAL EXERCISES

1. If X has variance σ^2, then σ, the positive square root of the variance, is called the *standard deviation.* If X has mean μ and standard deviation σ, show that

$$P\{|X - \mu| \ge k\sigma\} \le \frac{1}{k^2}$$

2. If X has mean μ and standard deviation σ, the ratio $r \equiv |\mu|/\sigma$ is called the *measurement signal-to-noise ratio* of X. The idea is that X can be expressed as $X = \mu + (X - \mu)$ with μ representing the signal and $X - \mu$ the noise. If we define $|(X - \mu)/\mu| \equiv D$ as the relative deviation of X from its signal (or mean) μ, show that for $\alpha > 0$,

$$P\{D \le \alpha\} \ge 1 - \frac{1}{r^2\alpha^2}$$

3. Compute the measurement signal-to-noise ratio—that is, $|\mu|/\sigma$ where $\mu = E[X]$, $\sigma^2 = \mathrm{Var}(X)$—of the following random variables:
(a) Poisson with mean λ;
(b) binomial with parameters n and p;
(c) geometric with mean $1/p$;
(d) uniform over (a, b);
(e) exponential with mean $1/\lambda$;
(f) normal with parameters μ, σ^2.

4. Let Z_n, $n \ge 1$ be a sequence of random variables and c a constant such that for each $\varepsilon > 0$, $P\{|Z_n - c| > \varepsilon\} \to 0$ as $n \to \infty$. Show that for any bounded continuous function g,

$$E[g(Z_n)] \to g(c) \quad \text{as} \quad n \to \infty$$

5. Let $f(x)$ be a continuous function defined for $0 \le x \le 1$. Consider the functions

$$B_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k(1 - x)^{n-k}$$

(called *Bernstein polynomials*) and prove that

$$\lim_{n\to\infty} B_n(x) = f(x)$$

HINT: Let $X_1, X_2, \ldots$ be independent Bernoulli random variables with mean x. Show and then use the fact (by making use of the result of Theoretical Exercise 4) that

$$B_n(x) = E\left[f\left(\frac{X_1 + \cdots + X_n}{n}\right)\right]$$

As it can be shown that the convergence of $B_n(x)$ to $f(x)$ is uniform in x, the above provides a probabilistic proof to the famous Weierstrass theorem of analysis that states that any continuous function on a closed interval can be approximated arbitrarily closely by a polynomial.

6. (a) Let X be a discrete random variable, whose possible values are $1, 2, \ldots$. If $P\{X = k\}$ is nonincreasing in $k = 1, 2, \ldots$, prove that

$$P\{X = k\} \le 2\frac{E[X]}{k^2}$$

(b) Let X be a nonnegative continuous random variable having a nonincreasing density function. Show that

$$f(x) \le \frac{2E[X]}{x^2} \quad \text{for all} \quad x > 0$$

7. Suppose that a fair die is rolled 100 times. Let X_i be the value obtained on the ith roll. Compute an approximation for

$$P\left\{\prod_{1}^{100} X_i \le a^{100}\right\} \qquad 1 < a < 6$$

8. Explain why a gamma random variable with parameters (t, λ) has an approximately normal distribution when t is large.

9. Suppose a fair coin is tossed 1000 times. If the first 100 tosses all result in heads, what proportion of heads would you expect on the final 900 tosses? Comment on the statement that "the strong law of large numbers swamps but does not compensate."

10. If X is a Poisson random variable with mean λ, show that for $i < \lambda$,

$$P\{X \le i\} \le \frac{e^{-\lambda}(e\lambda)^i}{i^i}$$

11. Let X be a binomial random variable with parameters n and p. Show that for $i > np$:

(a) $\underset{t>0}{\text{minimum}}\ e^{-ti}E[e^{tX}]$ occurs when t is such that $e^t = \dfrac{iq}{(n-i)p}$, where $q = 1 - p$.

(b) $P\{X \geq i\} \leq \dfrac{n^n}{i^i(n-i)^{n-i}} p^i(1-p)^{n-i}$.

12. The Chernoff bound on a standard normal random variable Z gives that $P\{Z > a\} \leq e^{-a^2/2}$, $a > 0$. Show, by considering the density of Z, that the right side of the inequality can be reduced by a factor 2. That is, show that

$$P\{Z > a\} \leq \tfrac{1}{2}e^{-a^2/2} \qquad a > 0$$

13. If $E[X] < 0$ and $\theta \neq 0$ is such that $E[e^{\theta X}] = 1$, show that $\theta > 0$.

SELF-TEST PROBLEMS AND EXERCISES

1. The number of automobiles sold weekly at a certain dealership is a random variable with expected value 16. Give an upper bound to the probability that
(a) next week's sales exceed 18;
(b) next week's sales exceed 25.

2. Suppose in Problem 1 that the variance of the number of automobiles sold weekly is 9.
(a) Give a lower bound to the probability that next week's sales are between 10 and 22 inclusively.
(b) Give an upper bound to the probability that next week's sales exceed 18.

3. If

$$E[X] = 75 \qquad E[Y] = 75 \qquad \text{Var}(X) = 10 \qquad \text{Var}(Y) = 12 \qquad \text{Cov}(X, Y) = -3$$

give an upper bound to
(a) $P\{|X - Y| > 15\}$;
(b) $P\{X > Y + 15\}$;
(c) $P\{Y > X + 15\}$.

4. Suppose that the number of units produced daily at factory A is a random variable with mean 20 and standard deviation 3 and the number produced at factory B is a random variable with mean 18 and standard deviation 6. Assuming independence, derive an upper bound for the probability that more units are produced today at factory B than at factory A.

5. The number of days that a certain type of component functions before failing is a random variable with probability density function

$$f(x) = 2x \qquad 0 < x < 1$$

Once the component fails it is immediately replaced by another one of the same type. If we let X_i denote the lifetime of the ith component to be put in use, then $S_n = \sum_{i=1}^{n} X_i$ represents the time of the nth failure. The long-term rate at which failures occur, call it r, is defined by

$$r = \lim_{n\to\infty} \frac{n}{S_n}$$

Assuming that the random variables X_i, $i \geq 1$ are independent, determine r.

6. In Self-Test Problem 5, how many components would one need to have on hand to be approximately 90 percent certain that the stock will last at least 35 days?

7. The servicing of a machine requires two separate steps, with the time needed for the first step being an exponential random variable with mean .2 hour and the time for the second step being an independent exponential random variable with mean .3 hour. If a repairperson has 20 machines to service, approximate the probability that all the work can be completed in 8 hours.

8. On each bet, a gambler loses 1 with probability .7, loses 2 with probability .2, or wins 10 with probability .1. Approximate the probability that the gambler will be losing after his first 100 bets.

9. Determine t so that the probability that the repairperson in Self-Test Problem 7 finishes the 20 jobs within time t is approximately equal to .95.