CHAPTER SIX

some fundamental limit theorems

Limit theorems characterize the mass behavior of experimental outcomes resulting from a large number of performances of an experiment. These theorems provide the connection between probability theory and the measurement of the parameters of probabilistic phenomena in the real world.

Early in this chapter, we discuss stochastic convergence, one important type of convergence for a sequence of random variables. This concept and an easily derived inequality allow us to establish one form of the law of large numbers. This law provides clarification of our earlier speculations (Sec. 3-6) regarding the relation, for large values of n, between the sum of n independent experimental values of random variable x and the quantity $nE(x)$.

We then discuss the Gaussian PDF. Subject to certain restrictions, we learn that the Gaussian PDF is often an excellent approximation to the actual PDF for the sum of many random variables, regardless of the forms of the PDF's for the individual random variables included in the sum.

This altogether remarkable result is known as the *central limit theorem*. A proof is presented for the case where the sum is composed of independent identically distributed random variables. Finally, we investigate several practical approximation procedures based on limit theorems.

6-1 The Chebyshev Inequality

The *Chebyshev inequality* states an upper bound on the probability that an experimental value of any random variable x will differ by at least any given positive quantity t from $E(x)$. In particular, the inequality will provide an upper bound on the quantity

$$\text{Prob}[|x - E(x)| \geq t]$$

in terms of t and σ_x. As long as the value of the standard deviation σ_x is known, other details of the PDF $f_x(x_0)$ are not relevant.

The derivation is simple. With $t > 0$, we have

$$\sigma_x^2 = \int_{x_0=-\infty}^{\infty} [x_0 - E(x)]^2 f_x(x_0)\,dx_0 \geq \int_{x_0=-\infty}^{E(x)-t} [x_0 - E(x)]^2 f_x(x_0)\,dx_0 + \int_{x_0=E(x)+t}^{\infty} [x_0 - E(x)]^2 f_x(x_0)\,dx_0$$

To obtain the above inequality, we note that the integrand in the leftmost integration is always positive. By removing an interval of length $2t$ from the range of that integral, we cannot increase the value of the integral. Inside the two integrals on the right-hand side of the above relation, it is always true that $|x - E(x)| \geq t$. We now replace $[x - E(x)]^2$ by t^2, which can never increase the value of the right-hand side, resulting in

$$\sigma_x^2 \geq \int_{x_0=-\infty}^{E(x)-t} t^2 f_x(x_0)\,dx_0 + \int_{x_0=E(x)+t}^{\infty} t^2 f_x(x_0)\,dx_0$$

After we divide both sides by t^2 and recognize the physical interpretation of the remaining quantity on the right-hand side, we have

$$\text{Prob}[|x - E(x)| \geq t] \leq \left(\frac{\sigma_x}{t}\right)^2$$

which is the Chebyshev inequality. It states, for instance, that the probability that an experimental value of any random variable x will be further than $K\sigma_x$ from $E(x)$ is always less than or equal to $1/K^2$

Since it is a rather weak bound, the Chebyshev inequality finds most of its applications in general theoretical work. For a random variable described by any particular PDF, better (though usually more complex) bounds may be established. We shall use the Chebyshev bound in Sec. 6-3 to investigate one form of the law of large numbers.

6-2 Stochastic Convergence

A *deterministic sequence* $\{x_n\} = x_1, x_2, \ldots$ is said to converge to the limit C if for every $\epsilon > 0$ we can find a finite n_0 such that

$$|x_n - C| < \epsilon \qquad \text{for all } n > n_0$$

If deterministic sequence $\{x_n\}$ does converge to the limit C, we write

$$\lim_{n\to\infty} x_n = C$$

Only for pathological cases would we expect to be able to make equally strong nonprobabilistic convergence statements for sequences of random variables. Several different types of convergence are defined for sequences of random variables. In this section we introduce and discuss one such definition, namely, that of *stochastic convergence*. We shall use this definition and the Chebyshev inequality to establish a form of the law of large numbers in the following section. (We defer any discussion of other forms of convergence for sequences of random variables until Sec. 6-9.)

A sequence of random variables, $\{y_n\} = y_1, y_2, y_3, \ldots$, is said to be *stochastically convergent* (or to *converge in probability*) to C if, for every $\epsilon > 0$, the condition $\lim_{n\to\infty} \text{Prob}(|y_n - C| > \epsilon) = 0$ is satisfied.

When a sequence of random variables, $\{y_n\}$, is known to be stochastically convergent to C, we must be careful to conclude *only* that the probability of the event $|y_n - C| > \epsilon$ vanishes as $n \to \infty$. We cannot conclude, for any value of n, that this event is impossible.

We may use the definition of a limit to restate the definition of stochastic convergence. Sequence $\{y_n\}$ is stochastically convergent to C if, for any $\epsilon > 0$ and any $\delta > 0$, it is possible to state a finite value of n_0 such that

$$\text{Prob}(|y_n - C| > \epsilon) < \delta \qquad \text{for all } n > n_0$$

Further discussion will accompany an application of the concept of stochastic convergence in the following section and a comparison with other forms of probabilistic convergence in Sec. 6-9.

6-3 The Weak Law of Large Numbers

A sequence of random variables, $\{y_n\}$, with finite expected values, is said to obey a *law of large numbers* if, in some sense, the sequence defined by

$$M_n = \frac{1}{n}\sum_{i=1}^{n} y_i$$

converges to its expected value. The type of convergence which applies determines whether the law is said to be *weak* or *strong*.

Let $y_1, y_2, \ldots$ form a sequence of independent identically distributed random variables with finite expected values $E(y)$ and finite variances σ_y^2. In this section we prove that the sequence

$$M_n = \frac{y_1 + y_2 + \cdots + y_n}{n}$$

is stochastically convergent to its expected value, and therefore the sequence $\{y_n\}$ obeys a (weak) law of large numbers.

For the conditions given above, random variable M_n is the average of n independent experimental values of random variable y. Quantity M_n is known as the *sample mean*.

From the definition of M_n and the property of expectations of sums, we have

$$E(M_n) = \frac{nE(y)}{n} = E(y)$$

and, because multiplying a random variable y by c defines a new random variable with a variance equal to $c^2\sigma_y^2$, we have

$$\sigma_{M_n}^2 = \frac{n\sigma_y^2}{n^2} = \frac{\sigma_y^2}{n} \qquad \sigma_{M_n} = \frac{\sigma_y}{\sqrt{n}}$$

To establish the weak law of large numbers for the case of interest, we simply apply the Chebyshev inequality to M_n to obtain

$$\text{Prob}[|M_n - E(M_n)| \geq \epsilon] \leq \left(\frac{\sigma_{M_n}}{\epsilon}\right)^2$$

and, substituting for M_n, $E(M_n)$, and σ_{M_n}, we find

$$\text{Prob}\left[\left|\frac{1}{n}\sum_{i=1}^{n} y_i - E(y)\right| \geq \epsilon\right] \leq \frac{\sigma_y^2}{n\epsilon^2}$$

(See Prob. 6.02 for a practical application of this relation.) Upon

taking the limit as $n \to \infty$ for this equation there finally results

$$\lim_{n\to\infty} \text{Prob}\left[\left|\frac{1}{n}\sum_{i=1}^{n} y_i - E(y)\right| \geq \epsilon\right] = 0$$

which is known as the *weak law of large numbers* (in this case for random variable y). The law states that, as $n \to \infty$, the probability that the average of n independent experimental values of random variable y differs from $E(y)$ by more than any nonzero ϵ goes to zero. We have shown that as long as the variance of a random variable is finite, the random variable obeys the weak law of large numbers.

Neither the independence of the y_i's nor the finite variance σ_y^2 conditions are necessary for the $\{y_n\}$ sequence to obey a law of large numbers. Proof of these statements is beyond the scope of this book.

Let's apply our result to a situation where the y_i's are independent Bernoulli random variables with parameter P. Suppose that there are n trials and k is the number of successes. Using our result above, we have

$$\lim_{n\to\infty} \text{Prob}\left(\left|\frac{k}{n} - P\right| > \epsilon\right) = 0$$

which is known as the *Bernoulli law of large numbers*. This relation is one of the bases of the *relative-frequency* interpretation of probabilities. Often people read into it far more than it says.

For instance, let the trials be coin flips and the successes heads. If we flip the coin any number of times, it is still possible that all outcomes will be heads. If we know that P is a valid parameter of a coin-flipping process and we set out to estimate P by the experimental value of k/n, there is no value of n for which we could be *certain* that our experimental value was within an arbitrary $\pm\epsilon$ of the true value of parameter P.

The Bernoulli law of large numbers does *not* imply that k converges to the limiting value nP as $n \to \infty$. We know that the standard deviation of k, in fact, becomes infinite as $n \to \infty$ (see Sec. 3-6).

6-4 The Gaussian PDF

We consider a very important PDF which describes a vast number of probabilistic phenomena.

The *Gaussian* (or *normal*) *PDF* is defined to be

$$f_x(x_0) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x_0-m)^2/2\sigma^2} \qquad -\infty \leq x_0 \leq \infty$$

where we have written a PDF for random variable x with parameters m and σ. The s transform of this PDF is obtained from

$$f_x^T(s) = E(e^{-sx}) = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{x_0=-\infty}^{\infty} e^{-sx_0}e^{-(x_0-m)^2/2\sigma^2}\,dx_0 = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{x_0=-\infty}^{\infty} e^{-A(x_0)}\,dx_0$$

where the expression $A(x_0)$ is given by

$$A(x_0) = sx_0 + \frac{x_0^2}{2\sigma^2} + \frac{m^2}{2\sigma^2} - \frac{2mx_0}{2\sigma^2} = \frac{1}{2\sigma^2}\{[x_0 + (s\sigma^2 - m)]^2 - s^2\sigma^4 + 2ms\sigma^2\}$$

In the above equation, we have carried out the algebraic operation known as "completing the square." We may substitute this result into the expression for $f_x^T(s)$,

$$f_x^T(s) = e^{(s^2\sigma^2/2)-sm}\int_{x_0=-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-(x_0+s\sigma^2-m)^2/2\sigma^2}\,dx_0$$

Note that the integral is equal to the total area under a Gaussian PDF for which m has been replaced by the expression $m - s\sigma^2$. This change displaces the PDF with regard to the x_0 axis but does not change the *total* area under the curve. Since the total area under any PDF must be unity, we have found the transform of a Gaussian PDF with parameters m and σ to be

$$f_x^T(s) = e^{(s^2\sigma^2/2)-sm}$$

Using the familiar relations

$$E(x) = -\left[\frac{d}{ds}f_x^T(s)\right]_{s=0}$$

$$\sigma_x^2 = \left\{\frac{d^2}{ds^2}f_x^T(s) - \left[\frac{d}{ds}f_x^T(s)\right]^2\right\}_{s=0}$$

we evaluate the expected value and the variance of the Gaussian PDF to be

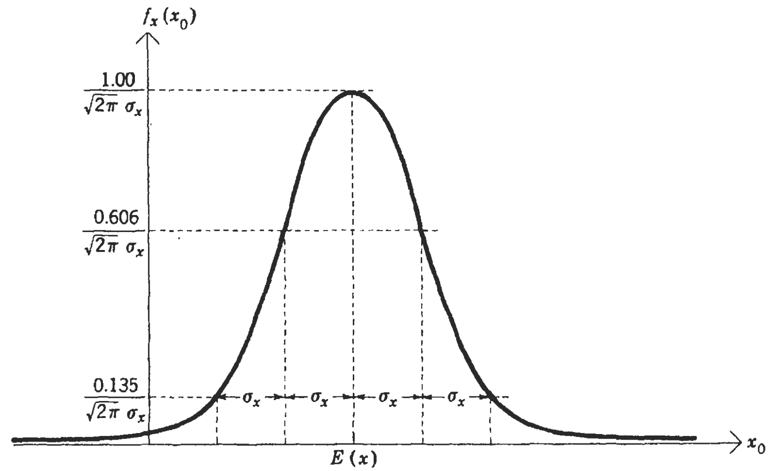
$$E(x) = m \qquad \sigma_x^2 = \sigma^2$$

We have learned that a normal PDF is specified by two parameters, the expected value $E(x)$ and the standard deviation σ_x. The normal PDF and its s transform are

$$f_x(x_0) = \frac{1}{\sqrt{2\pi}\,\sigma_x}e^{-[x_0-E(x)]^2/2\sigma_x^2} \qquad -\infty \le x_0 \le \infty$$

$$f_x^T(s) = e^{(s^2\sigma_x^2/2)-sE(x)}$$

A sketch of the normal PDF, in terms of $E(x)$ and σ_x, follows:



Suppose that we are interested in the PDF for random variable w, defined to be the sum of two *independent* Gaussian random variables x and y. We use the s transforms

$$f_x^T(s) = e^{(s^2\sigma_x^2/2)-sE(x)} \qquad \text{and} \qquad f_y^T(s) = e^{(s^2\sigma_y^2/2)-sE(y)}$$

and the relation for the transform of a PDF for the sum of independent random variables (Sec. 3-5) to obtain

$$f_w^T(s) = f_x^T(s)f_y^T(s) = e^{(s^2/2)(\sigma_x^2+\sigma_y^2)-s[E(x)+E(y)]}$$

We recognize this to be the transform of another Gaussian PDF. Thus we have found that

The PDF for a sum of independent Gaussian random variables is itself a Gaussian PDF.

A similar property was obtained for Poisson random variables in Sec. 4-7.

When one becomes involved in appreciable numerical work with PDF's of the same form but with different parameters, it is often desirable to suppress the parameters $E(x)$ and σ_x^2 of any particular situation and work in terms of a *unit* (or *normalized*, or *standardized*) random variable. A unit random variable has an expected value of zero, and its standard deviation is equal to unity. The unit random variable y for random variable x is obtained by subtracting $E(x)$ from x

and dividing this difference by σ_x

$$y = \frac{x - E(x)}{\sigma_x}$$

and we may demonstrate the properties,

$$E(y) = E\left[\frac{x - E(x)}{\sigma_x}\right] = \frac{E(x)}{\sigma_x} - \frac{E(x)}{\sigma_x} = 0$$

$$\sigma_y^2 = E\{[y - E(y)]^2\} = E\left\{\left[\frac{x - E(x)}{\sigma_x} - 0\right]^2\right\} = \frac{E\{[x - E(x)]^2\}}{\sigma_x^2} = 1$$

It happens that the cumulative distribution function, $p_{x\le}(x_0)$, for a Gaussian random variable cannot be found in closed form. We shall be working with tables of the CDF for a Gaussian random variable, and it is for this reason that we are interested in the discussion in the previous paragraph. We shall be concerned with the unit normal PDF, which is given by

$$f_y(y_0) = \frac{1}{\sqrt{2\pi}} e^{-y_0^2/2} \qquad -\infty \le y_0 \le \infty, \quad E(y) = 0, \quad \sigma_y = 1$$

We define the function $\Phi(y_0)$ to be the CDF for the unit normal PDF,

$$\Phi(y_0) \equiv p_{y\le}(y_0) = \int_{\alpha_0=-\infty}^{y_0} f_y(\alpha_0)\, d\alpha_0 = \frac{1}{\sqrt{2\pi}} \int_{\alpha_0=-\infty}^{y_0} e^{-\alpha_0^2/2}\, d\alpha_0$$

and extensive tables of $\Phi(y_0)$ exist.

To make use of tables of $\Phi(y_0)$ for a Gaussian (but unstandardized) random variable x, we need only recall the relation

$$y = \frac{x - E(x)}{\sigma_x}$$

and thus, for the CDF of Gaussian random variable x, we have

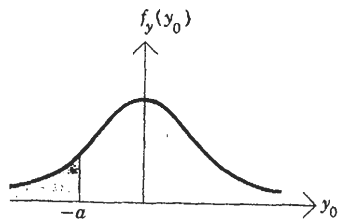
$$p_{x\le}(x_0) = p_{y\le}\left[\frac{x_0 - E(x)}{\sigma_x}\right] = \Phi\left[\frac{x_0 - E(x)}{\sigma_x}\right]$$

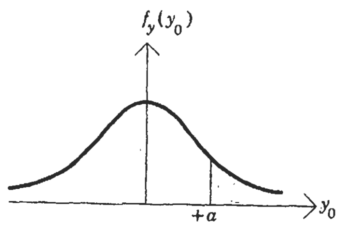
The argument on the right side of the above equation is equal to the number of standard deviations σ_x by which x_0 exceeds $E(x)$. If values of x_0 are measured in units of standard deviations from $E(x)$, tables of the CDF for the unit normal PDF may be used directly to obtain values of the CDF $p_{x\le}(x_0)$.

Since a Gaussian PDF is symmetrical about its expected value, the CDF may be fully described by tabulating it only for values above (or below) its expected value. The following is a brief four-place table of $\Phi(y_0)$, the CDF for a unit normal random variable:

y_0	$\Phi(y_0)$	y_0	$\Phi(y_0)$	y_0	$\Phi(y_0)$	y_0	$\Phi(y_0)$
0.00	0.5000	1.00	0.8413	2.00	0.9772	3.00	0.9987
0.10	0.5398	1.10	0.8643	2.10	0.9821	3.10	0.9990
0.20	0.5793	1.20	0.8849	2.20	0.9861	3.20	0.9993
0.30	0.6179	1.30	0.9032	2.30	0.9893	3.30	0.9995
0.40	0.6554	1.40	0.9192	2.40	0.9918	3.40	0.9997
0.50	0.6915	1.50	0.9332	2.50	0.9938	3.60	0.9998
0.60	0.7257	1.60	0.9452	2.60	0.9953		
0.70	0.7580	1.70	0.9554	2.70	0.9965		
0.80	0.7881	1.80	0.9641	2.80	0.9974		
0.90	0.8159	1.90	0.9713	2.90	0.9981		

To obtain $\Phi(y_0)$ for $y_0 < 0$, we note that the area under $f_y(y_0)$ is equal to unity and that the shaded areas in the following sketches are equal





and we may use the relation

$$\Phi(-y_0) = 1 - \Phi(y_0)$$

We present a brief example. Suppose that we wish to determine the probability that an experimental value of a Gaussian random variable x falls within $\pm k\sigma_x$ of its expected value, for $k = 1$, 2, and 3. Thus, we wish to evaluate

$$\text{Prob}[|x - E(x)| \le k\sigma_x] \qquad \text{for } k = 1, 2, \text{ and } 3$$

Since k is already in units of standard deviations, we may use the $\Phi(y_0)$ table directly according to the relation

$$\text{Prob}[|x - E(x)| \le k\sigma_x] = \Phi(k) - \Phi(-k) = \Phi(k) - [1 - \Phi(k)]$$

$$= 2\Phi(k) - 1 = \begin{cases} 0.682 & k = 1 \\ 0.954 & k = 2 \\ 0.997 & k = 3 \end{cases}$$

Our result states, for instance, that the probability that an experimental value of any Gaussian random variable x falls within an interval of total length $4\sigma_x$ which is centered on $E(x)$ is equal to 0.954.

6-5 Central Limit Theorems

Let random variable r be defined to be the sum of n independent identically distributed random variables, each of which has a finite expected value and a finite variance. It is an altogether remarkable fact that, *as $n \to \infty$, the CDF $p_{r\le}(r_0)$ approaches the CDF of a Gaussian random variable, regardless of the form of the PDF for the individual random variables in the sum.* This is one simple case of a *central limit theorem.*

Every central limit theorem states some particular set of conditions for which the CDF for a sum of n random variables will approach the CDF of a Gaussian random variable as $n \to \infty$.

As long as it is not always true that a particular few of the member random variables dominate the sum, the *identically distributed* (same PDF) condition is not essential for a central limit theorem to apply to the sum of a large number of independent random variables. The *independence* condition may also be relaxed, subject to certain other restrictions which are outside the scope of this presentation.

Since many phenomena may be considered to be the result of a large number of factors, central limit theorems are of great practical significance, especially when the effects of these factors are either purely additive ($r = x_1 + x_2 + \cdots$) or purely multiplicative ($\log r = \log x_1 + \log x_2 + \cdots$).

We first undertake a brief digression to indicate why we are stating central limit theorems in terms of the CDF $p_{r\le}(r_0)$ rather than the PDF $f_r(r_0)$.

There are several ways in which the members of a sequence of deterministic functions $\{g_n(x_0)\} = g_1(x_0), g_2(x_0), \ldots$ can "approach" the corresponding members of a sequence of functions $\{h_n(x_0)\}$ in the limit as $n \to \infty$. However, the simplest and most easily visualized manner is *point-by-point* convergence; namely, if for any particular x_0 and for any $\epsilon > 0$ we can always find an n_0 for which

$$|g_n(x_0) - h_n(x_0)| < \epsilon \qquad \text{for all } n > n_0$$

This is the type of convergence we wish to use in our statement of the central limit theorem.

Consider a case where random variable r is defined to be the sum of n independent experimental values of a Bernoulli random variable with parameter P. For any value of n we know that r will be a binomial random variable with PMF

$$p_r(r_0) = \binom{n}{r_0} P^{r_0}(1-P)^{n-r_0} \qquad r_0 = 0, 1, 2, \ldots, n$$

This PMF, written as a PDF, will always include a set of $n + 1$ impulses and be equal to zero between the impulses. Thus, $f_r(r_0)$ can never approach a Gaussian PDF on a point-by-point basis. However, it is possible for the CDF of r to approach the CDF for a Gaussian random variable on a point-by-point basis as $n \to \infty$, and the central limit theorem given in the first paragraph of this section states that this is indeed the case.

We now present a proof of the form of the central limit theorem stated in the opening paragraph of this section.

1 Let $x_1, x_2, \ldots, x_n$ be independent identically distributed random variables, each with finite expected value $E(x)$ and finite variance σ_x^2. We define random variable r to be $r = x_1 + x_2 + \cdots + x_n$, and we wish to show that the CDF $p_{r\le}(r_0)$ approaches the CDF of a Gaussian random variable as $n \to \infty$.

2 From the independence of the x_i's and the definition of r, we have

$$f_r^T(s) = [f_x^T(s)]^n$$

3 Note that for any random variable y, defined by $y = ar + b$, we may obtain $f_y^T(s)$ in terms of $f_r^T(s)$ from the definition of the s transform as follows:

$$f_y^T(s) = E(e^{-sy}) = E(e^{-ars}e^{-bs}) = e^{-sb}\int_{-\infty}^{\infty} e^{-asr_0}f_r(r_0)\,dr_0$$

We may recognize the integral in the above equation to obtain

$$f_y^T(s) = e^{-sb}f_r^T(as)$$

We shall apply this relation to the case where y is the standardized random variable for r,

$$y = \frac{r - E(r)}{\sigma_r} = \frac{r - nE(x)}{\sqrt{n}\,\sigma_x} \qquad a = \frac{1}{\sqrt{n}\,\sigma_x} \qquad b = -\frac{\sqrt{n}\,E(x)}{\sigma_x}$$

to obtain $f_y^T(s)$ from the expression for $f_r^T(s)$ of step 2.

$$f_y^T(s) = \left[e^{sE(x)/\sigma_x\sqrt{n}}\, f_x^T\left(\frac{s}{\sqrt{n}\,\sigma_x}\right)\right]^n$$

So far, we have found the s transform for y, the standardized sum of n independent identically distributed random variables,

$$y = \frac{x_1 + x_2 + \cdots + x_n - nE(x)}{\sqrt{n}\,\sigma_x}$$

4 The above expression for $f_y^T(s)$ may be written with $e^{sE(x)/\sigma_x\sqrt{n}}$ and $f_x^T\left(\frac{s}{\sqrt{n}\,\sigma_x}\right)$ each approximated suitably near to $s = 0$. These approximations are found to be

$$e^{sE(x)/\sigma_x\sqrt{n}} \approx 1 + \frac{E(x)}{\sigma_x}\left(\frac{s}{\sqrt{n}}\right) + \frac{[E(x)]^2}{2\sigma_x^2}\left(\frac{s}{\sqrt{n}}\right)^2$$

$$f_x^T\left(\frac{s}{\sqrt{n}\,\sigma_x}\right) \approx 1 - \frac{E(x)}{\sigma_x}\left(\frac{s}{\sqrt{n}}\right) + \frac{E(x^2)}{2\sigma_x^2}\left(\frac{s}{\sqrt{n}}\right)^2$$

When we multiply and collect terms, for suitably small s (or, equivalently, for suitably large n) we have

$$f_y^T(s) \approx \left[1 + \frac{1}{2}\frac{s^2}{n}\right]^n$$

5 We use the relation

$$\lim_{n\to\infty}\left(1 + \frac{a}{n}\right)^n = e^a$$

to take the limit as $n \to \infty$ of the approximation for $f_y^T(s)$ obtained in step 4. This results in

$$\lim_{n\to\infty} f_y^T(s) = e^{s^2/2}$$

and we have shown that the s transform of the PDF for random variable y approaches the transform of a unit normal PDF. This does not tell us *how* (or if) the PDF $f_y(y_0)$ approaches a Gaussian PDF on a point-by-point basis. But a relation known as the *continuity theorem* of transform theory may be invoked to assure us that

$$\lim_{n\to\infty} f_{y\le}(y_0) = \Phi(y_0)$$

[This theorem assures us that, if $\lim_{n\to\infty} f_{y_n}{}^T(s) = f_w{}^T(s)$ *and* if $f_w{}^T(s)$ is a continuous function, then the CDF for random variable y_n converges (on a point-by-point basis) to the CDF of random variable w. This convergence need not be defined at discontinuities of the limiting CDF.]

6 Since y is the standardized form of r, we simply substitute into the above result for $f_{y\le}(y_0)$ and conclude the following.

If $r = x_1 + x_2 + \cdots + x_n$ and $x_1, x_2, \ldots, x_n$ are independent identically distributed random variables each with finite expected value $E(x)$ and finite standard deviation σ_x, we have

$$\lim_{n\to\infty} p_{r\le}(r_0) = \Phi\left[\frac{r_0 - E(r)}{\sigma_r}\right] \qquad E(r) = nE(x) \qquad \sigma_r = \sqrt{n}\,\sigma_x$$

This completes our proof of one central limit theorem.

6-6 Approximations Based on the Central Limit Theorem

We continue with the notation $r = x_1 + x_2 + \cdots + x_n$, where the random variables $x_1, x_2, \ldots, x_n$ are mutually independent and identically distributed, each with finite expected value $E(x)$ and finite variance σ_x^2. If every member of the sum happens to be a Gaussian random variable, we know (from Sec. 6-4) that the PDF $f_r(r_0)$ will also be Gaussian for any value of n. *Whatever* the PDF for the individual members of the sum, one central limit theorem states that, as $n \to \infty$, we have

$$p_{r\le}(r_0) \to \Phi\left[\frac{r_0 - E(r)}{\sigma_r}\right]$$

As $n \to \infty$, the CDF for r approaches the CDF for that Gaussian random variable which has the same mean and variance as r.

If we wish to use the approximation

$$p_{r\le}(r_0) \approx \Phi\left[\frac{r_0 - E(r)}{\sigma_r}\right]$$

for "large" but finite values of n, and the individual x_i's are not Gaussian random variables, there are no simple general results regarding the precision of the approximation.

If the individual terms in the sum are described by any of the more common PDF's (with finite mean and variance), $f_r(r_0)$ rapidly ($n \approx 5$ or 10) approaches a Gaussian curve in the vicinity of $E(r)$. Depending on the value of n and the degree of symmetry expected in $f_r(r_0)$, we generally expect $f_r(r_0)$ to be poorly approximated by a Gaussian curve in ranges of r_0 more than some number of standard deviations distant from $E(r)$. For instance, even if the x_i's can take on only positive experimental values, the use of an approximation based on the central limit theorem will always result in some nonzero probability that the experimental value of $x_1 + x_2 + \cdots + x_n$ will be negative.

The discussion of the previous paragraph, however crude it may be, should serve to emphasize the *central* property of approximations based on the central limit theorem.

As one example of the use of an approximation based on the central limit theorem, let random variable r be defined to be the sum of 48 independent experimental values of random variable x, where the PDF for x is given by

$$f_x(x_0) = \begin{cases} 1 & \text{if } 0 < x_0 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

We wish to determine the probability that an experimental value of r falls in the range $22.0 < r \leq 25.0$. By direct calculation we easily obtain

$$E(x) = 0.5 \qquad \sigma_x^2 = 1/12$$
$$E(r) = 48E(x) = 24.0 \qquad \sigma_r^2 = 48\sigma_x^2 = 4.0$$

In using the central limit theorem to approximate

$$\text{Prob}(22.0 < r \leq 25.0)$$

we are approximating the true PDF $f_r(r_0)$ in the range $22 < r \leq 25$ by the Gaussian PDF

$$f_r(r_0) \approx \frac{1}{\sqrt{2\pi}\,\sigma_r} e^{-[r_0-E(r)]^2/2\sigma_r^2} = \frac{1}{2\sqrt{2\pi}} e^{-(r_0-24)^2/8}$$

If we wish to evaluate $\text{Prob}(22.0 < r \leq 25.0)$ directly from the table for the CDF of the unit normal PDF, the range of interest for random variable r should be measured in units of σ_r from $E(r)$. We have

$$\begin{aligned}\text{Prob}(22.0 < r \leq 25.0) &= \text{Prob}\{-1.0\sigma_r < [r - E(r)] \leq 0.5\sigma_r\}\\ &= \Phi(0.5) - \Phi(-1.0)\\ &= \Phi(0.5) - [1 - \Phi(1.0)]\\ &= 0.6915 - 1.0000 + 0.8413\\ &= 0.5328\end{aligned}$$

It happens that this is a very precise approximation. In fact, by simple convolution (or by our method for obtaining derived distributions in sample space) one can show, for the given $f_x(x_0)$, that even for $n = 3$ or $n = 4$, $f_r(r_0)$ becomes very close to a Gaussian PDF over most of the possible range of random variable r (see Prob. 6.10). However, a similar result for such very small n may not exist for several other common PDF's (see Prob. 6.11).

6-7 Using the Central Limit Theorem for the Binomial PMF

We wish to use an approximation based on the central limit theorem to approximate the PMF for a discrete random variable. Assume we are interested in events defined in terms of k, the number of successes in n trials of a Bernoulli process with parameter P. From earlier work (Sec. 4-1) we know that

$$p_k(k_0) = \binom{n}{k_0} P^{k_0}(1 - P)^{n-k_0} \qquad k_0 = 0, 1, 2, \ldots, n$$

and if a and b are integers with $b > a$, there follows

$$\text{Prob}(a \leq k \leq b) = \sum_{k_0=a}^{b} \binom{n}{k_0} P^{k_0}(1 - P)^{n-k_0}$$

Should this quantity be of interest, it would generally require a very unpleasant calculation. So we might, for large n, turn to the central limit theorem, noting that

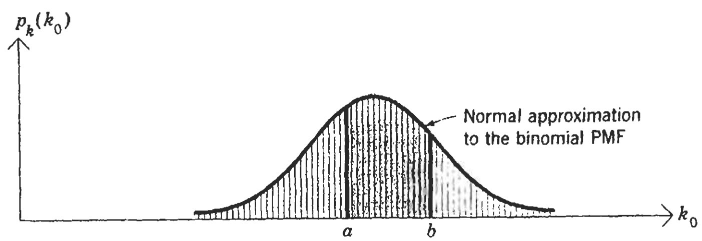
$$k = x_1 + x_2 + \cdots + x_n$$

where each x_i is an independent Bernoulli random variable.

If we applied the central limit theorem, subject to no additional considerations, we would have

$$\text{Prob}(a \leq k \leq b) \approx \Phi\left[\frac{b - E(k)}{\sigma_k}\right] - \Phi\left[\frac{a - E(k)}{\sigma_k}\right] \qquad E(k) = nP \qquad \sigma_k = \sqrt{nP(1 - P)}$$

We have approximated the probability that a binomial random variable k falls in the range $a \leq k \leq b$ by the area under a normal curve over this range. In many cases this procedure will yield excellent results. By looking at a picture of this situation, we shall suggest one simple improvement of the approximation.



The bars of $p_k(k_0)$ are shown to be about the same height as the approximating normal curve. This must be the case if n is large enough for the CDF's of $p_k(k_0)$ and the approximating normal curve to increase by about the same amount for each unit distance along the k_0 axis (as a result of the central limit theorem). The shaded area in this figure represents the approximation to $\text{Prob}(a \leq k \leq b)$ which results from direct substitution, where we use the CDF for a normal curve whose expected value and variance are the same as those of $p_k(k_0)$.

By considering the above sketch, we might expect that a more reasonable procedure could be suggested to take account of the discrete nature of k. In particular, it appears more accurate to associate the area under the normal curve between $k_0 - 0.5$ and $k_0 + 0.5$ with the probability of the event that random variable k takes on experimental value k_0. This not only seems better on a term-by-term basis than direct use of the central-limit-theorem approximation, but we can also show one extreme case of what may happen when this suggested improvement is not used. Notice (from the above sketch) that, if we have $b = a + 1$ [with a and b in the vicinity of $E(x)$], direct use of the CDF for the normal approximating curve will produce an approximation which is about 50% of the correct probability,

$$\text{Prob}(a \leq k \leq a+1) = p_k(a) + p_k(a+1)$$

When using the central limit theorem to approximate the binomial PMF, the adoption of our suggested improvement leads us to write,

$$\text{Prob}(a \leq k \leq b) \approx \Phi\left[\frac{b + \frac{1}{2} - nP}{\sqrt{nP(1-P)}}\right] - \Phi\left[\frac{a - \frac{1}{2} - nP}{\sqrt{nP(1-P)}}\right]$$

This result, a special case of the central limit theorem, is known as the *DeMoivre-Laplace limit theorem*. It can be shown to yield an improvement over the case in which the $\pm\frac{1}{2}$ terms are not used. These corrections may be significant when a and b are close $[(b-a)/\sigma_k < 1]$ or when a or b is near the peak of the approximating Gaussian PDF.

For example, suppose that we flipped a fair coin 100 times, and let k equal the number of heads. If we wished to approximate $\text{Prob}(48 \leq k \leq 51)$, the $\pm\frac{1}{2}$ corrections at the end of the range of k would clearly make a significant contribution to the accuracy of the approximation. On the other hand, for a quantity such as $\text{Prob}(23 \leq k \leq 65)$, the effect of the $\pm\frac{1}{2}$ is negligible.

One must always question the validity of approximations; yet it is surprising how well the DeMoivre-Laplace limit theorem applies for even a narrow range of k [near $E(k)$] when n is not very large. We shall do one such problem three ways. After obtaining these solutions, we shall comment on some limitations of this approximation technique.

Consider a set of 16 Bernoulli trials with $P = 0.5$. We wish to determine the probability that the number of successes, k, takes on an experimental value equal to 6, 7, or 8. First we do the exact calculation,

$$\text{Prob}(6 \leq k \leq 8) = \sum_{k_0=6}^{8} \binom{16}{k_0} P^{k_0}(1-P)^{16-k_0} = 0.49313$$

If we carelessly make direct use of the normal approximation, we have

$$\text{Prob}(6 \leq k \leq 8) \approx \Phi\left(\frac{8-8}{2}\right) - \Phi\left(\frac{6-8}{2}\right) = 0.34135$$

which, for reasons we have discussed, is poor indeed. Finally, if we use the $\pm\frac{1}{2}$ correction of the DeMoivre-Laplace theorem, we find

$$\text{Prob}(6 \leq k \leq 8) \approx \Phi\left(\frac{8 + \frac{1}{2} - 8}{2}\right) - \Phi\left(\frac{6 - \frac{1}{2} - 8}{2}\right) = 0.49306$$

which is within 0.02% of the correct value.

If P is too close either to zero or to unity for a given n, the resulting binomial PMF will be very asymmetric, with its peak very close to $k_0 = 0$ or to $k_0 = n$ and any Gaussian approximation will be poor. A reasonable (but arbitrary) rule of thumb for determining whether the DeMoivre-Laplace approximation to the binomial PMF may be employed [in the vicinity of $E(k)$] is to require that

$$nP > 3\sigma_k \qquad n(1-P) > 3\sigma_k \qquad \text{with } \sigma_k = \sqrt{nP(1-P)}$$

The better the margin by which these constraints are satisfied, the larger the range about $E(k)$ for which the normal approximation will yield satisfactory results.

6-8 The Poisson Approximation to the Binomial PMF

We have noted that the DeMoivre-Laplace limit theorem will not provide a useful approximation to the binomial PMF if either P or $1 - P$ is very small. When either of these quantities is too small for a given value of n, any Gaussian approximation to the binomial will be unsatisfactory. The Gaussian curve will remain symmetrical about $E(k)$, even though that value may be only a fraction of a standard deviation from the lowest or highest possible experimental value of the binomial random variable.

If n is large and P is small such that the DeMoivre-Laplace theorem may not be applied, we may take the limit of

$$p_k(k_0) = \binom{n}{k_0} P^{k_0}(1-P)^{n-k_0}$$

by letting $n \to \infty$ and $P \to 0$ while always requiring $nP = \mu$. Our result will provide a very good term-by-term approximation for the significant members [k_0 nonnegative and within a few σ_k of $E(k)$] of the binomial PMF for large n and small P.

First, we use the relation $nP = \mu$ to write

$$p_k(k_0) = \frac{n(n-1)\cdots(n-k_0+1)}{k_0!}\left(\frac{\mu}{n}\right)^{k_0}\left(1-\frac{\mu}{n}\right)^{n-k_0}$$

and, rearranging the terms, we have

$$p_k(k_0) = \frac{n(n-1)\cdots(n-k_0+1)}{n^{k_0}}\frac{\mu^{k_0}}{k_0!}\left(1-\frac{\mu}{n}\right)^{n-k_0}$$

Finally, we take the limit as $n \to \infty$ to obtain

$$\lim_{n\to\infty} p_k(k_0) = \frac{\mu^{k_0}e^{-\mu}}{k_0!} = \frac{(nP)^{k_0}e^{-(nP)}}{k_0!}$$

The above result is known, for obvious reasons, as the *Poisson approximation to the binomial PMF.*

For a binomial random variable k, we may note that, as $P \to 0$, the $E(k)/\sigma_k$ ratio is very nearly equal to $\sqrt{nP}$. For example, if $n = 100$ and $P = 0.01$, the expected value $E(k) = nP = 1.0$ is only one standard deviation from the minimum possible experimental value of k. Under these circumstances, the normal approximation (DeMoivre-Laplace) is poor, but the Poisson approximation is quite accurate for the small values of k_0 at which we find most of the probability mass of $p_k(k_0)$.

As an example, for the case $n = 100$ and $P = 0.01$, we find

	$k_0 = 0$	$k_0 = 1$	$k_0 = 3$	$k_0 = 10$
Exact value of $p_k(k_0)$	0.3660	0.3697	0.0610	$7 \cdot 10^{-8}$
Poisson approximation	0.3679	0.3679	0.0613	$10 \cdot 10^{-8}$
DeMoivre-Laplace approximation	0.2420	0.3850	0.0040	$<2 \cdot 10^{-18}$

6-9 A Note on Other Types of Convergence

In Sec. 6-2, we defined any sequence $\{y_n\}$ of random variables to be stochastically convergent (or to converge in probability) to C if, for every $\epsilon > 0$, the condition

$$\lim_{n\to\infty} \text{Prob}(|y_n - C| > \epsilon) = 0$$

is satisfied.

Let A_n denote the event $|y_n - C| < \epsilon$. By using the definition of a limit, an equivalent statement of the condition for stochastic convergence is that, for any $\epsilon > 0$ and any $\delta > 0$, we can find an n_0 such that

$$\text{Prob}(A_n) > 1 - \delta \qquad \text{for all } n > n_0$$

However, it does *not* follow from stochastic convergence that for any $\epsilon > 0$ and any $\delta > 0$ we can necessarily find an n_0 such that

$$\text{Prob}(A_{n_0+1}A_{n_0+2}A_{n_0+3} \cdots) > 1 - \delta$$

For a stochastically convergent sequence $\{y_n\}$, we would conclude that, *for all $n > n_0$, with n_0 suitably large:*

1 It is very probable that *any particular* y_n is within $\pm\epsilon$ of C.
2 It is not necessarily very probable that *every* y_n is within $\pm\epsilon$ of C.

A stronger form of convergence than stochastic convergence is known as *convergence with probability* 1 (or *convergence almost everywhere*). The sequence $\{y_n\}$ of random variables is defined to converge with probability 1 to C if the relation

$$\text{Prob}(\lim_{n\to\infty} y_n = C) = 1$$

is satisfied. Convergence with probability 1 implies stochastic convergence, but the converse is not true. Furthermore it can be shown (by using measure theory) that convergence with probability 1 does require that, for any $\epsilon > 0$ and any $\delta > 0$, we can find an n_0 such that

$$\text{Prob}(A_{n_0+1}A_{n_0+2}A_{n_0+3} \cdots) > 1 - \delta$$

For a sequence $\{y_n\}$ convergent to C with probability 1, we would conclude that, for one thing, the sequence is also stochastically convergent to C and also that, *for all $n > n_0$, with n_0 suitably large:*

1 It is very probable that *any particular* y_n is within $\pm\epsilon$ of C.
2 It is also very probable that *every* y_n is within $\pm\epsilon$ of C.

A third form of convergence, *mean-square convergence* (or *convergence in the mean*) of a sequence $\{y_n\}$ of random variables is defined by the relation

$$\lim_{n\to\infty} E[(y_n - C)^2] = 0$$

It is simple to show that mean-square convergence implies (but is not implied by) stochastic convergence (see Prob. 6.20). Mean-square convergence does not imply and is not implied by convergence with probability 1.

Determination of the necessary and sufficient conditions for sequences of random variables to display various forms of convergence, obey various laws of large numbers, and obey central limit theorems is well beyond the scope of our discussion. We do remark, however, that, because of the limited tools at our disposal, the law of large numbers obtained in Sec. 6-3 is unnecessarily weak.

PROBLEMS

6.01 Let x be a random variable with PDF $f_x(x_0) = \lambda e^{-\lambda x_0}$ for $x_0 > 0$. Use the Chebyshev inequality to find an upper bound on the quantity

$$\text{Prob}[|x - E(x)| \geq d]$$

as a function of d. Determine also the true value of this probability as a function of d.

6.02 In Sec. 6-3, we obtained the relation

$$\text{Prob}\left[\left|\frac{1}{n}\sum_{i=1}^{n} y_i - E(y)\right| \geq \epsilon\right] \leq \frac{\sigma_y^2}{n\epsilon^2}$$

where $y_1, y_2, \ldots$ were independent identically distributed random variables. If y is known to have a Bernoulli PMF with parameter P, use this relation to find how large n must be if we require that $(1/n)\sum_{i=1}^{n} y_i$ be within ± 0.01 of P with a probability of at least 0.95. (Remember that this is a loose bound and the resulting value of n may be unnecessarily large. See Prob. 6.09.)

6.03 Let $x_1, x_2, \ldots$ be independent experimental values of a random variable with PDF $f_x(x_0) = \mu_{-1}(x_0 - 0) - \mu_{-1}(x_0 - 1)$. Consider the sequence defined by

$$y_n = \max(x_1, x_2, \ldots, x_n)$$

Determine whether or not the sequence $\{y_n\}$ is stochastically convergent.

6.04 Consider a Gaussian random variable x, with expected value $E(x) = m$ and variance $\sigma_x^2 = (m/2)^2$.

a Determine the probability that an experimental value of x is negative.

b Determine the probability that the sum of four independent experimental values of x is negative.

c Two independent experimental values of x (x_1, x_2) are obtained. Determine the PDF and the mean and variance for:

i $ax_1 + bx_2$ **ii** $ax_1 - bx_2$ **iii** $|x_1 - x_2|$ $a, b > 0$

d Determine the probability that the product of four independent experimental values of x is negative.

6.05 A useful bound for the area under the *tails* of a unit normal PDF is obtained from the following inequality for $a \geq 0$:

$$p_{x\leq}(-a) = \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-x_0^2/2}\, dx_0 \leq \int_a^\infty \frac{1}{\sqrt{2\pi}} \frac{x_0}{a} e^{-x_0^2/2}\, dx_0$$

Use this result to obtain lower bounds on the probability that an experimental value of a Gaussian random variable x is within

a $\pm\sigma_x$ **b** $\pm 2\sigma_x$ **c** $\pm 4\sigma_x$

of its expected value. Compare these bounds with the true values of these probabilities and with the corresponding bounds obtained from the Chebyshev inequality.

6.06 A noise signal x may be considered to be a Gaussian random variable with an expected value of zero and a variance of σ_x^2. Assume that any experimental value of x will cause an error in a digital communication system if it is larger than $+A$.

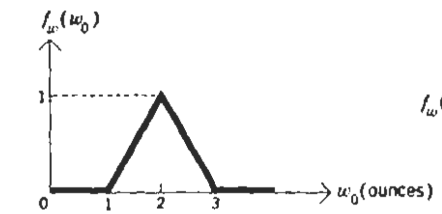
a Determine the probability that any particular experimental value of x will cause an error if:

i $\sigma_x^2 = 10^{-3}A^2$ **ii** $\sigma_x^2 = 10^{-1}A^2$

iii $\sigma_x^2 = A^2$ **iv** $\sigma_x^2 = 4A^2$

b For a given value of A, what is the largest σ_x^2 may be to obtain an error probability for any experimental value of x less than 10^{-3}? 10^{-6}?

6.07 The weight of a Pernotti Parabolic Pretzel, w, is a continuous random variable described by the probability density function



$$f_w(w_0) = \begin{cases} 0 & w_0 \leq 1 \\ w_0 - 1 & 1 \leq w_0 \leq 2 \\ 3 - w_0 & 2 \leq w_0 \leq 3 \\ 0 & 3 \leq w_0 \end{cases}$$

a What is the probability that 102 pretzels weigh more than 200 ounces?

b If we select 4 pretzels independently, what is the probability that exactly 2 of the 4 will each have the property of weighing more than 2 ounces?

c What is the smallest integer (the pretzels are not only inedible, they are also unbreakable) N for which the total weight of N pretzels will exceed 200 ounces with probability 0.990?

6.08 The energy of any individual particle in a certain system is an independent random variable with probability density function

$$f_E(E_0) = \begin{cases} 2e^{-2E_0} & E_0 \geq 0 \\ 0 & E_0 < 0 \end{cases}$$

The total system energy is the sum of the energies of the individual particles.

Numerical answers are required for parts (a), (b), (c), and (e).

a If there are 1,600 particles in the system, determine the probability that there are between 780 and 840 energy units in the system.
b What is the largest number of particles the system may contain if the probability that its total energy is less than 440 units must be at least 0.9725?
c Each particle will escape from the system if its energy exceeds $(\ln 3)/2$ units. If the system originally contained 4,800 particles, what is the probability that at least 1,700 particles will escape?
d If there are 10 particles in the system, determine an *exact* expression for the PDF for the total energy in the system.
e Compare the second and fourth moments of the answer to (d) with those resulting from a central-limit-theorem approximation.

6.09 Redo Prob. 6.02, using an approximation based on the central limit theorem rather than the Chebyshev inequality.

6.10 Determine and plot the precise PDF for r, the sum of four independent experimental values of random variable x, where the PDF for x is $f_x(x_0) = \mu_{-1}(x_0 - 0) - \mu_{-1}(x_0 - 1)$. Compare the numerical values of $f_r(r_0)$ and its Gaussian approximation at $r_0 = E(r) \pm K\sigma_r$, for $K = 0$, 1, and 2.

6.11 Let r be the sum of four independent experimental values of an exponential random variable. Compare the numerical values of $f_r(r_0)$ and its central limit theorem approximation at $r_0 = E(r) \pm K\sigma_r$, for $K = 0$, 1, and 2.

6.12 A certain town has a Saturday night movie audience of 600 who must choose between two comparable movie theaters. Assume that the movie-going public is composed of 300 couples, each of which independently flips a fair coin to decide which theater to patronize.
a Using a central limit theorem approximation, determine how many seats each theater must have so that the probability of exactly one theater running out of seats is less than 0.1.
b Repeat, assuming that each of the 600 customers makes an independent decision (instead of acting in pairs).

6.13 For 3,600 independent tosses of a fair coin:
a Determine a number n such that the probability is 0.5 that the number of heads resulting will be between 1,780 and n.
b Determine the probability that the number of heads is within $\pm 1\%$ of its expected value.

6.14 Reyab aspirin has exactly 10^7 users, all of them fitfully loyal. The number of tablets consumed by any particular customer on any one day is a random variable k, described by the probability mass function

$$p_k(k_0) = \frac{4 - k_0}{10} \qquad k_0 = 0, 1, 2, 3$$

Each customer's Reyab consumption is independent of that of all other customers. Reyab is sold only in 100-tablet bottles. On a day when a customer consumes exactly k_0 tablets, he purchases a new 100-tablet bottle of Reyab with probability $k_0/100$.
a Determine the mean and variance of k, the random variable describing the Reyab consumption of any particular customer in one day.
b Determine the mean and variance of t, the random variable describing the *total* number of Reyab tablets consumed in one day.
c Determine the probability that the *total* number of tablets consumed on any day will differ by more than $\pm 5{,}000$ tablets from the average daily total consumption.
d Determine the probability that a particular customer buys a new bottle of Reyab on a given day.
e What is the probability that a randomly selected tablet of Reyab (it was marked at the factory) gets consumed on a day when its owner consumes exactly two Reyab tablets?
f The Clip Pharmacy supplies exactly 30 Reyab customers with their entire requirements. What is the probability that this store sells exactly four bottles of aspirin on a particular day?

6.15 A population is sampled randomly (with replacement) to estimate S, the fraction of smokers in that population. Determine the sample size n such that the probability that the estimate is within ± 0.02 of the true value is at least 0.95. In other words, determine the smallest value of n such that

$$\text{Prob}\left(\left|\frac{\text{number of smokers counted}}{n} - S\right| < 0.02\right) \geq 0.95$$

6.16 Consider the following model for the weight gain of a prehistoric neopalenantioctipus. His (or her) weight gain in pounds on any particular day was an independent discrete random variable k, with the PMF

$$p_k(k_0) = \begin{cases} 0.74 & k_0 = 0.50 \\ 0.25 & k_0 = 4.00 \\ 0.01 & k_0 = 200.0 \end{cases}$$

Using this crude model, determine an approximation to the PDF for the weight of one such animal when it expired at the ripe age of 100,000 days.

6.17 Consider the number of 3s which result from 600 tosses of a fair six-sided die.

a Determine the probability that there are exactly 100 3s, using a form of Stirling's approximation for $n!$ which is very accurate for these values,

$$n! \approx \sqrt{2\pi}\, e^{-n} n^{n+0.5}$$

b Use the Poisson approximation to the binomial PMF to obtain the probability that there are exactly 100 3s.

c Repeat part (b), using the central limit theo.em intelligently.

d Use the Chebyshev inequality to find a lower bound on the probability that the number of 3s is between 97 and 103 inclusive, between 90 and 110 inclusive, and between 60 and 140 inclusive.

e Repeat part (d), using the central limit theorem and employing the DeMoivre-Laplace result when it appears relevant. Compare your answers with those obtained above, and comment.

6.18 A coin is tossed n times. Each toss is an independent Bernoulli trial with probability of heads P. Random variable x is defined to be the number of heads observed. For each of the following expressions, either find the value of K which makes the statement true for *all* $n \geq 1$, or state that no such value of K exists.

Example: $E(x) = An^k$ Answer: $k = 1$

a $E\left(\frac{x}{n}\right) = Bn^k$ **b** $E\{[x - E(x)]^2\} = Cn^k$ **c** $E(x^2) = Dn^k$

In the following part, consider only the case for large n:

d $\text{Prob}\left(\left|P - \frac{x}{n}\right| \leq Fn^k\right) = 0.15$

6.19 Each performance of a particular experiment is said to generate one "experimental value" of random variable x described by the probability density function

$$f_x(x_0) = \begin{cases} 1 & \text{if } 0 < x_0 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

The experiment is performed K times, and the resulting successive (and independent) experimental values are labeled $x_1, x_2, \ldots, x_K$.

a Determine the probability that x_2 and x_4 are the two largest of the K experimental values.

b Given that $x_1 + x_2 > 1.00$, determine the conditional probability that the smaller of these first two experimental values is smaller than 0.75.

c Determine the numerical value of

$$\lim_{K\to\infty} \text{Prob}\left(\left|\frac{K}{2} - \sum_{i=1}^{K} x_i\right| < \frac{K^{1/3}}{2\sqrt{3}}\right)$$

A formal proof is not required, but your reasoning should be fully explained.

d If $r = \min(x_1, x_2, \ldots, x_K)$ and $s = \max(x_1, x_2, \ldots, x_K)$, determine the joint probability density function $f_{r,s}(r_0, s_0)$ for all values of r_0 and s_0.

6.20 Use the Chebyshev inequality to prove that stochastic convergence is assured for any sequence of random variables which converges in the mean-square sense.

6.21 A simple form of the Cauchy PDF is given by

$$f_x(x_0) = [\pi(1 + x_0^2)]^{-1} \qquad -\infty \leq x_0 \leq \infty$$

a Determine $f_x^T(s)$. (You may require a table of integrals.)

b Let y be the sum of K independent samples of x. Determine $f_y^T(s)$.

c Would you expect $\lim_{K\to\infty} f_y(y_0)$ to become Gaussian? Explain.

6.22 The number of rabbits in generation i is n_i. Variable n_{i+1}, the number of rabbits in generation $i + 1$, depends on random effects of light, heat, water, food, and predator population, as well as on the number n_i. The relation is

$$p_{n_{i+1}}(jn_i) = \begin{cases} 0.2 & j = 1 \\ 0.3 & j = 2 \\ 0.5 & j = 3 \end{cases}$$

This states, for instance, that with probability 0.5 there will be three times as many rabbits in generation $i + 1$ as there were in generation i.

The rabbit population in generation 1 was 2. Find an approximation to the PMF for the number of rabbits in generation 12. (*Hint:* To use the central limit theorem, you must find an expression involving the *sum* of random variables.)