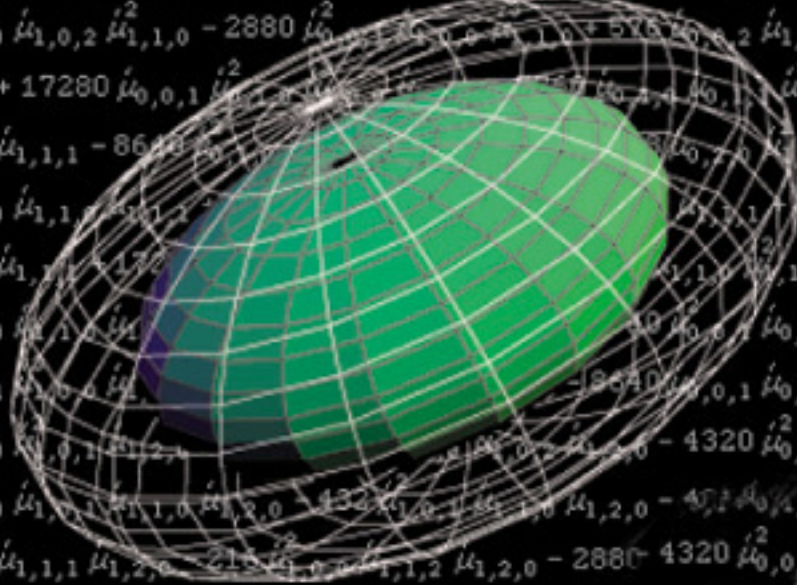
SPRINGER TEXTS IN STATISTICS

MATHEMATICAL STATISTICS *with* *Mathematica*®



COLIN ROSE
MURRAY D. SMITH

Mathematical Statistics with *Mathematica*

Chapter 8 – Asymptotic Theory

Please reference this 2002 edition as:

Rose, C. and Smith, M.D. (2002)
Mathematical Statistics with Mathematica, Springer-Verlag, New York.

Latest edition

For the latest up-to-date edition, please visit: www.mathStatica.com

Chapter 8

Asymptotic Theory

8.1 Introduction

Asymptotic theory is often used to justify the selection of particular estimators. Indeed, it is commonplace in modern statistical practice to base inference upon a suitable asymptotic theory. Desirable asymptotic properties—*consistency* and *limiting Normality*—can sometimes be ascribed to an estimator, even when there is relatively little known, or assumed known, about the population in question. In this chapter, we focus on both of these properties in the context of the sample mean,

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

and the sample sum,

$$S_n = \sum_{i=1}^{n} X_i$$

where symbol n denotes the sample size. We have especially attached n as a subscript to emphasise that $\overline{X}_n$ and S_n are random variables that depend on n. In subsequent chapters, we shall examine the asymptotic properties of estimators with more complicated structures than $\overline{X}_n$ and S_n. Our discussion of asymptotic theory centres on asking: What happens to an estimator (such as the sample mean) as n becomes large (in fact, as $n \to \infty$)? Thus, our presentation of asymptotic theory can be viewed as a theory relevant to increasing sample sizes. Of course, we require that the random variables used to form $\overline{X}_n$ and S_n must exist at each and every value of n. Accordingly, for an asymptotic theory to make sense, infinite-length sequences of random variables must be allowed to exist. For example, for $\overline{X}_n$ and S_n, the sequence of underlying random variables would be

$$(X_1, X_2, \ldots, X_i, X_{i+1}, \ldots) = \{X_n\}_{n=1}^{\infty}.$$

Throughout this chapter, apart from one or two exceptions, we shall work with examples dealing with the simplest of cases; namely, when all variables in the sequence are independent and identically distributed. Our treatment is therefore pitched at an elementary level.

The asymptotic properties of consistency and asymptotic normality are due to, respectively, the concepts of *convergence in probability* (§8.5) and *convergence in distribution* (§8.2). Moreover, these properties can often be established in a variety of situations through application of two fundamental theorems of asymptotic theory: *Khinchine's Weak Law of Large Numbers* and *Lindeberg–Lévy's Central Limit Theorem*.

The *Mathematica* tools needed in a chapter on asymptotic theory depend, not surprisingly, in large part on the built-in `Limit` function; however, we will also use the add-on package ``Calculus`Limit` ``. The add-on *removes* and *replaces* the built-in `Limit` function with an alternate algorithm for computing limits. As its development ceased a few years ago, we would ideally prefer to ignore this package altogether and use only the built-in `Limit` function, for the latter *is* subject to ongoing research and development.[1] Unfortunately, the world is not ideal! The built-in `Limit` function in Version 4 of *Mathematica* is unable to perform some limits that are commonplace in statistics, whereas if ``Calculus`Limit` `` is implemented, a number of these limits can be computed correctly. The solution that we adopt is to load and unload the add-on as needed. To illustrate our approach, consider the following limit (see *Example 2*) which cannot be solved by built-in `Limit` (try it and see!):

$$\lim_{n\to\infty} \text{Binomial}[n, x] \left(\frac{\theta}{n}\right)^x \left(1 - \frac{\theta}{n}\right)^{n-x}.$$

With ``Calculus`Limit` `` loaded, a solution to the limit is reported. Enter the following:

```
<< Calculus`Limit`
Limit[Binomial[n, x] (θ/n)^x (1 - θ/n)^(n-x), n → ∞];
Unprotect[Limit]; Clear[Limit];
```

The limit is computed correctly — we suppress the output here — what is important to see is the procedure for loading and unloading the ``Calculus`Limit` `` add-on.

Asymptotic theory is so widespread in its application that there is already an extensive field of literature in probability and statistics that contributes to its development. Accordingly, we shall cite only a select collection of works that we have found to be of particular use in preparing this chapter: Amemiya (1985), Bhattacharya and Rao (1976), Billingsley (1995), Chow and Teicher (1978), Hogg and Craig (1995), McCabe and Tremayne (1993) and Mittelhammer (1996).

8.2 Convergence in Distribution

The cumulative distribution function (cdf) has three attractive properties associated with it, namely (i) all random variables possess a cdf, (ii) the cdf has a range that is bounded within the closed unit interval [0,1], and (iii) the cdf is monotonic increasing. So when studying the behaviour of a sequence of random variables, we may, possibly just as easily, consider the behaviour of the infinite sequence of associated cdf's. This leads to the concept of convergence in distribution, a definition of which follows.

Let the random variable X_n have cdf F_n at each value of $n = 1, 2, \ldots$. Also, let the random variable X have cdf F, where X and F do not depend upon n. If it can be shown that

$$\lim_{n\to\infty} F_n(x) = F(x) \tag{8.1}$$

for all points x at which $F(x)$ is continuous, then X_n is said to *converge in distribution* to X.[2] A common notation to denote convergence in distribution is

$$X_n \xrightarrow{d} X. \tag{8.2}$$

F is termed the *limit distribution* of X_n.

⊕ ***Example 1:*** The Limit Distribution of a Sample Mean

In this example, the limiting distribution of the sample mean is derived, assuming that the population from which random samples are drawn is $N(0, 1)$. For a random sample of size n, the sample mean $\overline{X}_n \sim N(0, \frac{1}{n})$ (established in *Example 24* of Chapter 4). Therefore, the pdf and support of $\overline{X}_n$ are:

$$\mathbf{f} = \frac{e^{-\frac{\bar{x}^2}{2/n}}}{\sqrt{2\pi/n}}; \qquad \mathbf{domain[f]} = \{\bar{x}, -\infty, \infty\} \,\&\&\, \{n > 0\};$$

while the cdf (evaluated at a point x) is:

```
Fn = Prob[x, f]
```

$$\frac{1}{2}\left(1 + \mathrm{Erf}\left[\frac{\sqrt{n}\,x}{\sqrt{2}}\right]\right)$$

The limiting behaviour of the cdf depends on the sign of x. Here, we evaluate $\lim_{n\to\infty} F_n(x)$ when x is negative (say $x = -1$), zero, and positive (say $x = 1$):

```
<< Calculus`Limit`

Limit[Fn /. x → {-1, 0, 1}, n → ∞]

Unprotect[Limit]; Clear[Limit];
```

$$\left\{0, \frac{1}{2}, 1\right\}$$

The left-hand side of (8.1) is, in this case, a step function with a discontinuity at the origin, as the left panel of Fig. 1 shows.

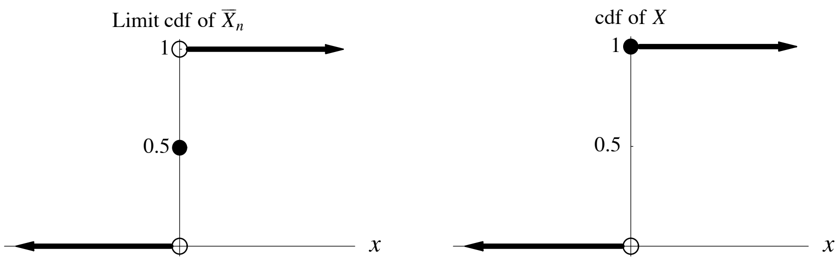


Fig. 1: Limit cdf of $\overline{X}_n$, and cdf of X

Now consider a random variable X whose cdf evaluated at a point x is given by

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

Comparing the graph of the cdf of X (given in the right panel of Fig. 1) to the graph of the limit of the cdf of $\overline{X}_n$, we see that both are identical at all points apart from when $x = 0$. However, because both graphs are discontinuous at $x = 0$, it follows that definition (8.1) holds, and so

$$\overline{X}_n \xrightarrow{d} X.$$

F is the limiting distribution function of $\overline{X}_n$. Now, focusing upon the random variable X and its cdf F, notice that F assigns all probability to a single point at the origin. Since X takes only one value, 0, with probability one, then X is a degenerate random variable, and F is termed a *degenerate distribution*. This is one instance where the limiting distribution provides information about the probability space of the underlying random variable. ■

⊕ ***Example 2:*** The Poisson as the Limit Distribution of a Binomial

It is sometimes possible to show convergence in distribution by deriving the limiting behaviour of functions other than the cdf, such as the pdf/pmf, the mgf, or the cf. This means that convergence in distribution becomes an issue of convergence of an infinite-length sequence of pdf/pmf, mgf, or cf.

In this example, convergence in distribution is illustrated by deriving the limit of a sequence of pmf. Recall that the Binomial(n, p) distribution has mean $n\,p$. Suppose that $X_n \sim$ Binomial(n, θ/n) (then $0 < \theta < n$); furthermore, assume that θ remains finite as n increases. To interpret the assumption on θ, note that $E[X_n] = n\,\theta/n = \theta$; thus, for every sample size n, the mean remains fixed and finite at the value of θ. Let f denote the pmf of X_n. Then:

```
f  = Binomial[n, x] (θ/n)^x (1 - θ/n)^(n-x);
domain[f] =
  {x, 0, n} && {0 < θ < n, n > 0, n ∈ Integers} && {Discrete};
```

```
<< Calculus`Limit`
Limit[f, n → ∞]
Unprotect[Limit]; Clear[Limit];
```

$$\frac{e^{-\theta}\,\theta^{x}}{\Gamma[1+x]}$$

Because $\Gamma[1+x] = x!$ for integer $x \geq 0$, this expression is equivalent to the pmf of a variable which is Poisson distributed with parameter θ. Therefore, under our assumptions,

$$X_n \xrightarrow{d} X \sim \text{Poisson}(\theta).$$

The limiting distribution of the Binomial random variable X_n is thus Poisson(θ).

⊕ ***Example 3:*** The Normal as the Limit Distribution of a Binomial

In the previous example, both the limit distribution and the random variables in the sequence were defined over a discrete sample space. However, this equivalence need not always occur: the limit distribution of a discrete variable may be continuous, or a continuous random variable may have a discrete limit distribution, as seen in *Example 1* (albeit that it was a degenerate limit distribution).

In this example, convergence in distribution is illustrated by deriving the limit of a sequence of moment generating functions (mgf). Suppose that $X_n \sim$ Binomial(n, θ), where $0 < \theta < 1$. Unlike the previous example where the probability of a 'success' diminished with n, in this example the probability stays fixed at θ for all n. Let f once again denote the pmf of X_n:

```
      f  = Binomial[n, x] θ^x (1 - θ)^(n-x);
domain[f] =
  {x, 0, n} && {0 < θ < 1, n > 0, n ∈ Integers} && {Discrete};
```

Then, the mgf of X_n is derived as:

```
mgf_X = Expect[e^(t x), f]
```

$$(1 + (-1 + e^{t})\,\theta)^{n}$$

Now consider the standardised random variable Y_n defined as

$$Y_n = \frac{X_n - E[X_n]}{\sqrt{\text{Var}(X_n)}} = \frac{X_n - n\,\theta}{\sqrt{n\,\theta(1-\theta)}}.$$

Y_n necessarily has a mean of 0 and a variance of 1. The mgf of Y_n can be obtained using the MGF Theorem (§2.4 D), setting a and b in that theorem equal to:

```
a = -n θ / Sqrt[n θ (1 - θ)];   b = 1 / Sqrt[n θ (1 - θ)];
```

to find:

```
mgf_Y = e^(a t) (mgf_X /. t → b t)
```

$$e^{-\frac{n t \theta}{\sqrt{n (1-\theta) \theta}}} \left(1 + \left(-1 + e^{\frac{t}{\sqrt{n (1-\theta) \theta}}}\right) \theta\right)^n$$

Executing built-in `Limit`, we find the limit mgf of the infinite sequence of mgf's equal to:

```
Limit[mgf_Y, n → ∞]
```

$$e^{\frac{t^2}{2}}$$

As this last expression is equivalent to the mgf of a $N(0, 1)$ variable, it follows that

$$Y_n \xrightarrow{d} Z \sim N(0, 1).$$

Thus, the limiting distribution of a standardised Binomial random variable is the standard Normal distribution. ■

8.3 Asymptotic Distribution

Suppose, for example, that we have established the following limiting distribution for a random variable X_n:

$$X_n \xrightarrow{d} Z \sim N(0, 1). \tag{8.3}$$

Let n_* denote a fixed and finite sample size; for example, n_* might correspond to the sample size of the data set with which we are working. In the absence of any knowledge about the exact distribution of X_n, it makes sense to use the limiting distribution of X_n as an approximation to the distribution of X_{n_*}, for if n_* is *sufficiently large,* the discrepancy between the exact distribution and the posited approximation must be small due to (8.3). This approximation is referred to as the *asymptotic distribution*. A commonly used notation for the asymptotic distribution is

$$X_{n_*} \overset{a}{\sim} N(0, 1) \tag{8.4}$$

which reads literally as 'the asymptotic distribution of X_{n_*} is $N(0, 1)$', or 'the approximate distribution of X_{n_*} is $N(0, 1)$'.

Of course, the variable that is of interest to us need not necessarily be X_{n_*}. However, if we know the relationship between X_{n_*} and the variable of interest, Y_{n_*} say, it is often possible to derive the asymptotic distribution for the latter. For example, if

$Y_{n_*} = \mu + \sigma X_{n_*}$, where $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$, then the asymptotic distribution of Y_{n_*} may be obtained directly from (8.4) using the properties of the Normal distribution:

$$Y_{n_*} \overset{a}{\sim} N(\mu, \sigma^2).$$

As a second example, suppose that W_{n_*} is related to X_{n_*} by the transformation $W_{n_*} = X_{n_*}^2$. Once again, the asymptotic distribution of W_{n_*} may be deduced by using the properties of the Normal distribution:

$$W_{n_*} \overset{a}{\sim} \text{Chi-squared}(1).$$

Typically, the distinction between arbitrary n and a specific value n_* is made implicit by dropping the * subscript. We too shall adopt this convention from now on.

⊕ ***Example 4:*** The Asymptotic Distribution of a Method of Moments Estimator

Let $X \sim$ Chi-squared(θ), where $\theta \in \mathbb{R}_+$ is unknown. Let $(X_1, X_2, \ldots, X_n)$ denote a random sample of size n drawn on X. The *method of moments* (§5.6) estimator of θ is the sample mean $\overline{X}_n$. Further, let Z_n be related to $\overline{X}_n$ by the following location shift and scale change,

$$Z_n = \frac{\overline{X}_n - \theta}{\sqrt{2\theta/n}} \tag{8.5}$$

Since it can be shown that $Z_n \xrightarrow{d} Z \sim N(0, 1)$, it follows that the asymptotic distribution of the estimator is

$$\overline{X}_n \overset{a}{\sim} N\left(\theta, \frac{2\theta}{n}\right).$$ ■

○ ***van Beek Bound***

One way to assess the accuracy of the asymptotic distribution is to calculate an upper bound on the approximation error of its cdf. Such a bound has been derived by van Beek,[3] and generally applies when the limiting distribution is the standard Normal. The relevant result is typically expressed in the form of an inequality.

Let $(W_1, \ldots, W_n)$ be a set of n independent variables, each with *zero mean* and finite third absolute moment. Define

$$\begin{aligned} \mu_2 &= \frac{1}{n} \sum_{i=1}^{n} E[W_i^2] \\ \mu_3^+ &= \frac{1}{n} \sum_{i=1}^{n} E\left[\,|W_i|^3\right] \qquad (8.6) \\ B &= \frac{1}{\sqrt{n}}\, 0.7975\, \mu_3^+\, \mu_2^{-3/2} \end{aligned}$$

and let

$$W_* = \frac{\overline{W}}{\sqrt{\mu_2 / n}}$$

where $\overline{W}$ denotes the sample mean $\frac{1}{n} \sum_{i=1}^{n} W_i$. Then van Beek's inequality holds for all w_* in the support of the variable W_*, namely,

$$\left| F_n(w_*) - \Phi(w_*) \right| \leq B \tag{8.7}$$

where $F_n(w_*)$ is the cdf of W_* evaluated at w_*, and $\Phi(w_*)$ is the cdf of a $N(0, 1)$ variable evaluated at the same point.[4] Some features of this result that are worth noting are: (i) the variables $(W_1, \ldots, W_n)$ need not be identically distributed, nor does their distribution need to be specified; (ii) van Beek's bound B decreases as the sample size increases, eventually reaching zero in the limit; and (iii) if $(W_1, \ldots, W_n)$ are identical in distribution to a random variable W, then $\mu_2 = E[W^2] = \text{Var}(W)$ and $\mu_3^+ = E\left[\left| W \right|^3\right]$. These simplifications will be useful in the next example.

⊕ ***Example 5:*** van Beek's Bound for the Method of Moments Estimator

We shall derive van Beek's bound B on the error induced by using the $N(0, 1)$ distribution to approximate the distribution of Z_n, where Z_n is the scaled method of moments estimator given in (8.5) in *Example 4*. Recall that $Z_n = (\overline{X}_n - \theta) \big/ \sqrt{2\,\theta / n}$, where $\overline{X}_n$ is the sample mean of n independent and identically distributed Chi-squared(θ) random variables, each with pdf $f(x)$:

$$\mathbf{f} = \frac{\mathbf{x}^{\theta/2-1}\ \mathbb{e}^{-\mathbf{x}/2}}{\Gamma[\theta/2]\ 2^{\theta/2}}; \quad \mathbf{domain[f] = \{x, 0, \infty\}\ \&\&\ \{\theta > 0\};}$$

Note that van Beek's bound assumes a zero mean, whereas X has mean θ. To resolve this difference, we shall work *about the mean* and take $W = X - \theta$. We now derive $\mu_2 = E[W^2]$:

$$\mathbf{w = x - \theta;} \quad \mu_2 = \mathbf{Expect[w^2, f]}$$

$$2\,\theta$$

To derive $\mu_3^+ = E\left[\left| W \right|^3\right] = E\left[\left| X - \theta \right|^3\right]$, note that *Mathematica* has difficulty integrating expressions with absolute values. Fortunately, **mathStatica** allows us to replace $\left| W \right|$ with an `If[]` statement. The calculation takes about 30 seconds on our reference machine:[5]

$$\mu_3^+ = \mathbf{Expect[\,If[x < \theta, -w, w]^3, f]}$$

$$\theta \left(-8 + \frac{1}{\Gamma[4 + \frac{\theta}{2}]} \left((2\,\mathbb{e})^{-\theta/2}\, \theta^{\frac{4+\theta}{2}}\ (6 + \theta) \left(2 + \theta + \mathbb{e}^{\theta/2}\, \theta\ \text{ExpIntegralE}\left[-2 - \frac{\theta}{2}, \frac{\theta}{2}\right] \right) \right) \right)$$

Since $\mu_2 = 2\,\theta$, we have

$$Z_n = \frac{\overline{X}_n - \theta}{\sqrt{2\,\theta / n}} = \frac{\overline{W}}{\sqrt{\mu_2 / n}} = W_*$$

allowing us to apply van Beek's bound (8.7):

```
B = 0.7975/√n  μ3⁺/μ2^(3/2);
```

$$B = \frac{0.7975}{\sqrt{n}} \frac{\mu_3^+}{\mu_2^{3/2}};$$

which depends on θ and n. To illustrate, we select a sample size of $n = 20$ and set $\theta = 1$, to find:

```
B /. {n → 20, θ → 1} // N
0.547985
```

At our chosen point, van Beek's bound is particularly large, and so will not be of any real use in judging the effectiveness of the asymptotic distribution in this case. Fortunately, with **mathStatica**, it is reasonably straightforward to evaluate the exact value of the approximation error by computing the left-hand side of (8.7). Recalling that $S_n = \sum_{i=1}^n X_i$, we have

$$\begin{aligned} F_n(w_*) &= P(Z_n \le w_*) \\ &= P\Big(\frac{n^{-1}\,S_n - \theta}{\sqrt{2\,\theta / n}} \le w_*\Big) \\ &= P\big(S_n \le w_*\,\sqrt{2\,\theta\,n} + n\,\theta\big). \end{aligned}$$

Example 23 of Chapter 4 shows that the random variable $S_n \sim$ Chi-squared$(n\,\theta)$. Its pdf $g(s_n)$ is thus:

$$g = \frac{s_n^{\frac{n\theta}{2}-1}\, e^{-\frac{s_n}{2}}}{2^{\frac{n\theta}{2}}\,\Gamma[\frac{n\theta}{2}]}; \quad \text{domain}[g] = \{s_n, 0, \infty\} \,\&\&\, \{\theta > 0, n > 0\};$$

Then, $F_n(w_*)$ is:

$$F_n = \text{Prob}\big[w_*\,\sqrt{2\,\theta\,n} + n\,\theta,\; g\big]$$

$$1 - \frac{\text{Gamma}\left[\frac{n\theta}{2}, \frac{n\theta}{2} + \frac{\sqrt{n}\,\theta\,w_*}{\sqrt{2}}\right]}{\Gamma\left[\frac{n\theta}{2}\right]}$$

After evaluating $\Phi(w_*)$, we can plot the *actual* error caused by approximating F_n with a Normal distribution, as shown in Fig. 2.

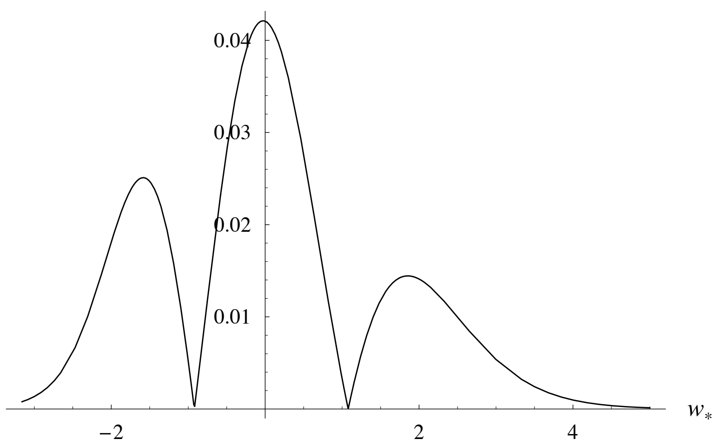


Fig. 2: Actual approximation error ($n = 20$, $\theta = 1$) in absolute value

It is easy to see from this diagram that at our selected values of n and θ, the discrepancy (in absolute value) between the exact cdf and the cdf of the asymptotic distribution is no larger than approximately 0.042. This is considerably lower than the reported van Beek bound of approximately 0.548. The error the asymptotic distribution induces is nevertheless fairly substantial in this case. Of course, as sample size increases, the size of the error must decline. ■

8.4 Central Limit Theorem

§8.2 discussed the convergence in distribution of a sequence of random variables whose distribution was known. In practice, such information is often not available, thus jeopardising the derivation of the limiting distribution. In such cases, if the variables in the sequence are used to form sums and averages, such as S_n and $\overline{X}_n$, the limiting distribution can often be derived by applying the famous *Central Limit Theorem*. Since many estimators *are* functions of sums of random variables, the Central Limit Theorem is of considerable importance in statistics. See Le Cam (1986) for an interesting discussion of the history of the Central Limit Theorem.

We consider random variables constructed in the following manner,

$$\frac{S_n - a_n}{b_n} \tag{8.8}$$

where $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ represent sequences of real numbers. The random variables appearing in the sum S_n, namely, $\{X_i\}_{i=1}^{n}$, are the first n elements of the infinite-length sequence $\{X_n\}_{n=1}^{\infty}$. If we set

$$a_n = \sum_{i=1}^{n} E[X_i] \quad \text{and} \quad b_n^2 = \sum_{i=1}^{n} \text{Var}(X_i) \tag{8.9}$$

then (8.8) would be a standardised random variable — it has mean 0 and variance 1. Notice that this construction necessarily requires that the mean and variance of every random variable in the sequence $\{X_n\}_{n=1}^{\infty}$ exists. The Central Limit Theorem states the conditions for $\{X_n\}$, $\{a_n\}$ and $\{b_n\}$ in order that

$$\frac{S_n - a_n}{b_n} \xrightarrow{d} Z \tag{8.10}$$

for some random variable Z. We shall only consider cases for which $Z \sim N(0, 1)$.

We present the *Lindeberg–Lévy* version of the Central Limit Theorem, which applies when the variables $\{X_n\}_{n=1}^{\infty}$ are mutually independent and identically distributed (iid). The Lindeberg–Lévy version is particularly relevant for determining asymptotic properties of estimators such as $\overline{X}_n$, where $\overline{X}_n$ is constructed from size n random samples collected on some variable which we may label X. Assuming that $E[X] = \mu$ and $\text{Var}(X) = \sigma^2$, under the iid assumption, each variable in $\{X_n\}_{n=1}^{\infty}$ may be viewed as a copy of X. Hence $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$. The constants in (8.9) therefore become

$$a_n = n\,\mu \quad \text{and} \quad b_n^2 = n\,\sigma^2$$

and the theorem states the conditions that μ and σ^2 must satisfy in order that the limiting distribution of $(S_n - n\,\mu)\big/\sqrt{n\,\sigma^2}$ is $Z \sim N(0, 1)$.

Theorem (Lindeberg–Lévy): Let the random variables in the sequence $\{X_n\}_{n=1}^{\infty}$ be independent and identically distributed, each with finite mean μ and finite variance σ^2. Then the random variable

$$\frac{S_n - n\,\mu}{\sigma\sqrt{n}} \tag{8.11}$$

converges in distribution to a random variable $Z \sim N(0, 1)$.

Proof: See, for example, Mittelhammer (1996, p. 270).

The strength of the Central Limit Theorem is that the distribution of X need not be known. Of course, if $X \sim N(\mu, \sigma^2)$, then the theorem holds trivially, since the sampling distribution of the sample sum is also Normal. On the other hand, for any non-Normal random variable X that possesses a finite mean and variance, the theorem permits us to construct an approximation to the sampling distribution of the sample sum which will become increasingly accurate with sample size. Thus, for the sample sum,

$$S_n \overset{a}{\sim} N(n\,\mu, n\,\sigma^2) \tag{8.12}$$

and, for the sample mean,

$$\overline{X}_n \overset{a}{\sim} N\Big(\mu, \frac{\sigma^2}{n}\Big). \tag{8.13}$$

⊕ ***Example 6:*** The Sample Mean and the Uniform Distribution

Let $X \sim$ Uniform(0, 1), the Uniform distribution on the interval (0, 1). Enter its pdf $f(x)$ as:

```
f = 1;       domain[f] = {x, 0, 1};
```

The mean μ and the variance σ^2 of X are, respectively:

```
Expect[x, f]
```

$$\frac{1}{2}$$

```
Var[x, f]
```

$$\frac{1}{12}$$

Let $\overline{X}_3$ denote the sample mean of a random sample of size $n = 3$ collected on X. Now suppose, for some reason, that we wish to obtain the probability:

$$p = P\left(\tfrac{1}{6} < \overline{X}_3 < \tfrac{5}{6}\right).$$

As the conditions of the Central Limit Theorem are satisfied, it follows from (8.13) that the asymptotic distribution of $\overline{X}_3$ is:

$$\overline{X}_3 \overset{a}{\sim} N\left(\tfrac{1}{2}, \tfrac{1}{36}\right).$$

We may therefore use this asymptotic distribution to find an approximate solution for p. Let $g(\overline{x})$ denote the pdf of the asymptotic distribution:

$$g = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\bar{x}-\mu)^2}{2\sigma^2}} \;/.\; \left\{\mu \to \frac{1}{2},\ \sigma \to \frac{1}{6}\right\};$$

```
domain[g] = {x̄, -∞, ∞};
```

Then p is approximated by:

$$\text{Prob}\left[\frac{5}{6}, g\right] - \text{Prob}\left[\frac{1}{6}, g\right] \text{ // N}$$

```
0.9545
```

Just as we were concerned about the accuracy of the asymptotic distribution in *Example 5*, it is quite reasonable to be concerned about the accuracy of the asymptotic approximation for the probability that we seek; after all, a sample size of $n = 3$ is far from large! Generally speaking, the answer to 'How large does n need to be?' is context dependent. Thus, our answer when $X \sim$ Uniform(0, 1) may be quite inadequate under different distributional assumptions for X.

○ ***Small Sample Accuracy***

In this subsection, we wish to compare the exact solution for p, with our asymptotic approximation 0.9545. For the exact solution, we require the sampling distribution of $\overline{X}_3$. More generally, if $X \sim$ Uniform(0, 1), the sampling distribution of $\overline{X}_n$ is known as Bates's distribution; for a derivation, see Bates (1955) or Stuart and Ord (1994, Example 11.9). The Bates(n) distribution has an n-part piecewise structure:

$$\texttt{Bates[x\_, n\_] := Table}\Big[\Big\{\frac{k}{n} \le x < \frac{k+1}{n},\ \texttt{Expand}\Big[\frac{n^n \sum_{i=0}^{k} (-1)^i\, \texttt{Binomial[n, i]}\, (x - \frac{i}{n})^{n-1}}{(n-1)\,!}\Big]\Big\},\ \{k, 0, n-1\}\Big]$$

For instance, when $n = 3$, the pdf of $Y = \overline{X}_3$ has the 3-part form:

```
Bates[y, 3]
```

$$\begin{pmatrix} 0 \le y < \frac{1}{3} & \frac{27 y^2}{2} \\ \frac{1}{3} \le y < \frac{2}{3} & -\frac{9}{2} + 27 y - 27 y^2 \\ \frac{2}{3} \le y < 1 & \frac{27}{2} - 27 y + \frac{27 y^2}{2} \end{pmatrix}$$

This means if $0 \le y < \frac{1}{3}$, the pdf of Y is given by $h(y) = \frac{27 y^2}{2}$, and so on. In the past, we have used `If` statements to represent 2-part piecewise functions. However, for functions with at least three parts, a `Which` statement is required. Given $Y = \overline{X}_n \sim$ Bates(n) with pdf $h(y)$, we may create the `Which` structure as follows:

```
h[y_] = Which @@ Flatten[Bates[y, 3]]
domain[h[y]] = {y, 0, 1};
```

$$\texttt{Which}\Big[0 \le y < \frac{1}{3},\ \frac{27 y^2}{2},\ \frac{1}{3} \le y < \frac{2}{3},\ -\frac{9}{2} + 27 y - 27 y^2,\ \frac{2}{3} \le y < 1,\ \frac{27}{2} - 27 y + \frac{27 y^2}{2}\Big]$$

Then, the natural way to find p with **mathStatica** would be to evaluate `Prob[`$\frac{5}{6}$`, h[y]] - Prob[`$\frac{1}{6}$`, h[y]]`. Unfortunately, at present, neither *Mathematica* nor **mathStatica** can perform integration on `Which` statements. However, implementation of this important feature is already being planned for version 2 of **mathStatica**. Nevertheless, we can still compute the exact value of p manually, as follows:

$$\int_{\frac{1}{6}}^{\frac{1}{3}} \frac{27 y^2}{2}\, dy + \int_{\frac{1}{3}}^{\frac{2}{3}} \left(-\frac{9}{2} + 27 y - 27 y^2\right) dy + \int_{\frac{2}{3}}^{\frac{5}{6}} \left(\frac{27}{2} - 27 y + \frac{27 y^2}{2}\right) dy$$

$$\frac{23}{24}$$

where $23/24 \simeq 0.958333$. By contrast, the approximation based on the asymptotic distribution was 0.9545. Thus, asymptotic theory is doing fairly well here — especially when we remind ourselves that the sample size is only three! Figure 3 illustrates the pdf of $\overline{X}_3$, which certainly has that nice 'bell-shaped' look associated with the Normal distribution.

```
PlotDensity[h[y]];
```

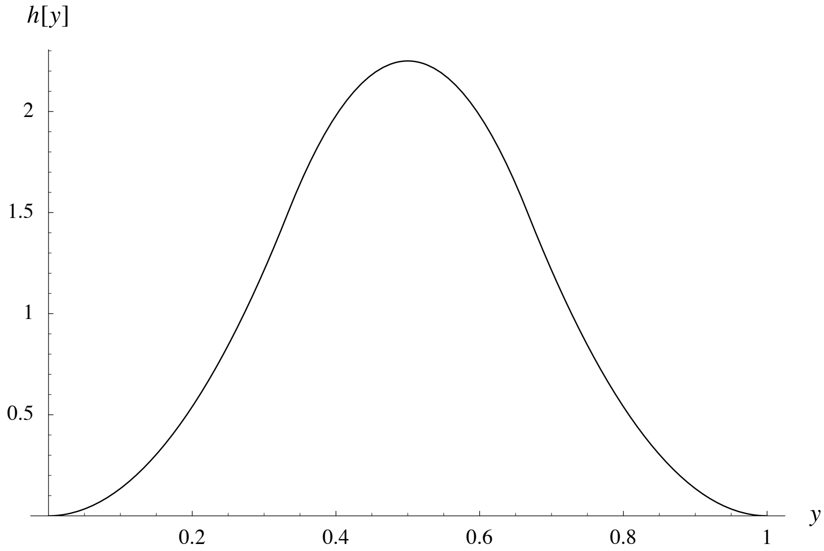


Fig. 3: Density of $\overline{X}_3$ — the Bates(3) distribution

Next, we examine the approximation provided by the cdf of the asymptotic $N(0, 1)$ distribution. In *Example 5*, a similar exercise was undertaken using the van Beek bound, as well as plotting the absolute difference of the exact to the asymptotic distribution. This time, however, we shall take a different route. We now conduct a *Monte Carlo* exercise to compare an artificially generated distribution with the asymptotic distribution. To do so, we generate a pseudo-random sample of size $n = 3$ from the Uniform(0, 1) distribution using *Mathematica*'s internal pseudo-random number generator: `Random[]`. The sample mean $\overline{X}_3$ is then computed. This exercise is repeated $T = 2000$ times. Here then are T realisations of the random variable $\overline{X}_3$:

```
realisations =
  Table[Plus @@ Table[Random[], {3}] / 3, {2000}];
```

We now standardise these realisations using the true mean ($\frac{1}{2}$) and the true standard deviation ($\frac{1}{6}$):

```
Sdata = (realisations - 1/2) / (1/6);
```

We may use a *quantile–quantile plot* to examine the closeness of the realised standardised sample means to the $N(0, 1)$ distribution. If the plot lies close to the 45° line, it suggests that the distribution of the standardised realisations is close to the $N(0, 1)$. The **mathStatica** function `QQPlot` constructs this quantile–quantile plot.

```
QQPlot[Sdata];
```

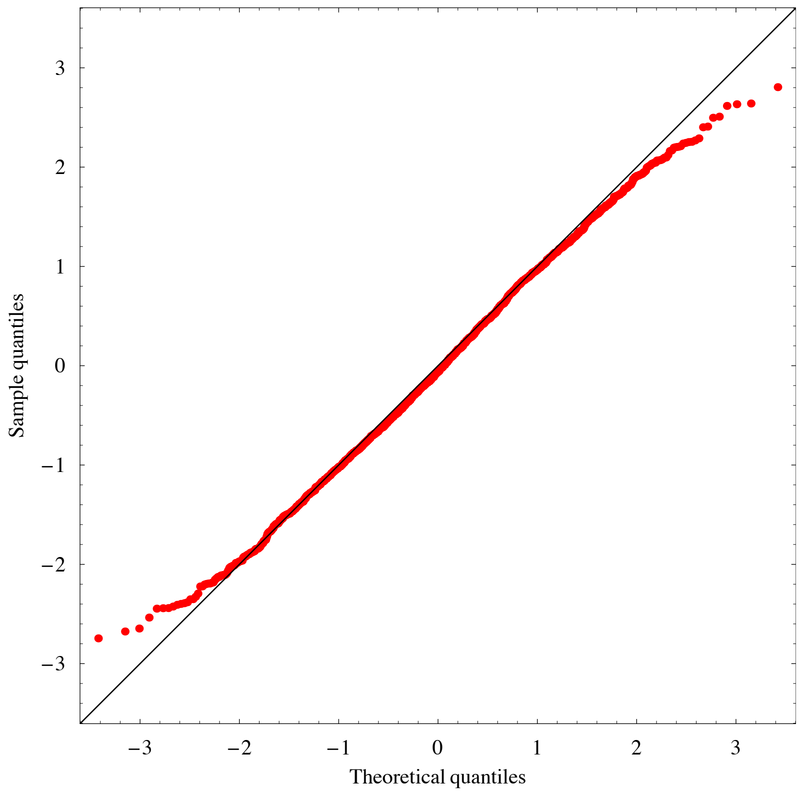


Fig. 4: Quantiles of $\overline{X}_3$ against the quantiles of $N(0, 1)$

The plotted points appear slightly S-shaped, with the elongated centre portion (from values of about -2 to $+2$ along the horizontal axis) closely hugging the 45° line. However, in the tails of the distribution (values below -2, and above $+2$), the accuracy of the Normal approximation to the true cdf weakens. The main reason for this is that the standardised statistic $6\,(\overline{X}_3 - \frac{1}{2})$ is bounded between -3 and $+3$ (notice that the plot stays within this interval of the vertical axis), whereas the Normal is unbounded. Evidently, the asymptotic distribution provides an accurate approximation except in the tails.

These ideas have practical value: they can be used to construct a pseudo-random number generator for standard Normal random variables. The Normal pseudo-random number generators considered previously were based on the inverse cdf method (see §2.6 B and §2.6 C) and the rejection method (see §2.6 D). By appealing to the Central Limit Theorem, a third possibility arises. We have seen that the cdf of $6\,(\overline{X}_3 - \frac{1}{2})$ performs fairly well in mimicking the cdf of the $N(0, 1)$ distribution, apart from in the tails. This suggests, due to the Central Limit Theorem, that an increase in sample size might improve tail behaviour; in this respect, using a sample size of $n = 12$ is a common choice. When $n = 12$, the statistic with a limiting $N(0, 1)$ distribution is

$$12\,(\overline{X}_{12} - \tfrac{1}{2}) = S_{12} - 6$$

which is now bounded between −6 and +6. The generator works by taking 12 pseudo-random drawings from the Uniform(0, 1) distribution, and then subtracts 6 from their sum — easy!

```
N01RNG := Plus @@ Table[Random[], {12}] - 6
```

The function `N01RNG` returns a single number each time it is executed. For example:

```
N01RNG
```

```
-0.185085
```

The suitability of this generator can be investigated by using `QQPlot`.[6] ■

8.5 Convergence in Probability

8.5 A Introduction

For a sequence of random variables $\{X_n\}_{n=1}^{\infty}$, *convergence in probability* is concerned with establishing whether or not the outcomes of those variables become increasingly close to the outcomes of another random variable X with high probability. A formal definition follows:

Let the sequence of random variables $\{X_n\}_{n=1}^{\infty}$ and the random variable X be defined on the same probability space. If for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P\big(\,|X_n - X| \geq \varepsilon\big) = 0 \tag{8.14}$$

then X_n is said to converge in probability to X, written $X_n \xrightarrow{p} X$.

The implication of the definition is that, if indeed $\{X_n\}_{n=1}^{\infty}$ is converging in probability to X, then for a fixed and finite value of n, say n_*, the outcomes of X can be used to approximate the outcomes of X_{n_*}. As we are now referring to outcomes of random variables, it is necessary to insist that all random variables in $\{X_n\}_{n=1}^{\infty}$ be measured in the same sample space as X.[7] This was not the case when we considered convergence in distribution, for this property concerned only the cdf function, and variables measured in different sample spaces are not generally restricted from having equivalent cdf's. Accordingly, convergence in probability is a stronger concept than convergence in distribution.

The following rule establishes the relationship between convergence in probability and convergence in distribution. If $X_n \xrightarrow{p} X$, then it follows that the limiting cdf of X_n must be identical to that of X, and hence,

$$X_n \xrightarrow{p} X \quad \text{implies} \quad X_n \xrightarrow{d} X. \tag{8.15}$$

On the other hand, by the argument of the preceding paragraph, the converse is not generally true. The situation when the converse is true occurs only when X is a degenerate random variable, for then convergence in distribution specifies exactly what that value must be. For a fixed constant c,

$$X_n \xrightarrow{p} X = c \quad \text{implies and is implied by} \quad X_n \xrightarrow{d} X = c. \tag{8.16}$$

The following two examples show the use of **mathStatica** in establishing convergence in probability.

⊕ ***Example 7:*** Convergence in Probability to a Normal Random Variable

Suppose that the random variable $X_n = (1 + \frac{1}{n})X$, where $n = 1, 2, \ldots$. Clearly, X_n and X must lie within the same sample space for all n, as they are related by a simple scaling transformation. Moreover, it is easy to see that $|X_n - X| = \frac{1}{n}|X|$. Therefore,

$$P(|X_n - X| \geq \varepsilon) = P(|X| \geq n\,\varepsilon). \tag{8.17}$$

For any random variable X, and any scalar $\alpha > 0$, we may express the event $\{|X| \geq \alpha\}$ as the union of two disjoint events, $\{X \geq \alpha\} \bigcup \{X \leq -\alpha\}$. Therefore, the occurrence probability can be written as

$$P(|X| \geq \alpha) = P(X \geq \alpha) + P(X \leq -\alpha). \tag{8.18}$$

Now if we suppose that $X \sim N(0, 1)$, and take $\alpha = n\,\varepsilon$, the right-hand side of (8.17) becomes

$$(1 - \Phi(n\,\varepsilon)) + \Phi(-n\,\varepsilon) = 2\,(1 - \Phi(n\,\varepsilon))$$

where Φ denotes the cdf of X, and the symmetry of the pdf of X about zero has been exploited. This can be entered into *Mathematica* as:

$$\mathtt{f} = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}; \quad \mathtt{domain[f] = \{x, -\infty, \infty\};}$$

```
sol = 2 (1 - Prob[n ε, f]) // Simplify
```

$$1 - \mathrm{Erf}\left[\frac{n\,\varepsilon}{\sqrt{2}}\right]$$

In light of definition (8.14), we now show that X_n converges in probability to X because the following limit is equal to zero:

```
<< Calculus`Limit`
Limit[sol, n → ∞]
Unprotect[Limit]; Clear[Limit];
```

```
0
```

As the limit of (8.17) is zero, $X_n \xrightarrow{p} X$. Of course, this outcome should be immediately obvious by inspection of the relationship between X_n and X; the transforming scalar $(1 + \frac{1}{n}) \to 1$ as $n \to \infty$. ■

Showing convergence in probability often entails complicated calculations, for as definition (8.14) shows, the joint distribution of the random variables X_n and X must typically be known for all n. This, fortunately, was not necessary in the previous example because the relation $X_n = (1 + \frac{1}{n})X$ was known. In any case, from now on, our concern lies predominantly with convergence in probability to a *constant*. Although this type of convergence is easier to deal with, this does not mean that it is less important. In fact, when it comes to determining properties of estimators, it is of vital importance to establish whether or not the estimator converges in probability to the (constant) parameter for which it is proposed. Under this scenario, we take X to be constant in (8.14). Then X can be thought of as representing a parameter θ, while X_n may be viewed as the estimator proposed to estimate it. Under these conditions, if (8.14) holds, X_n is said to be *consistent* for θ, or X_n is a *consistent estimator* of θ.

⊕ ***Example 8:*** Convergence in Probability to a Constant

For a random sample of size n from a $N(\theta, \sigma^2)$ population, the sample mean $\overline{X}_n$ is proposed as an estimator of θ. We shall show, using definition (8.14), that $\overline{X}_n$ is a consistent estimator of θ; that is, we shall show, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left| \overline{X}_n - \theta \right| \geq \varepsilon \right) = 0.$$

Input into *Mathematica* the pdf of $\overline{X}_n$, which we know to be exactly $N\left(\theta, \frac{\sigma^2}{n}\right)$:

```
f = 1/(σ √(2 π)) ⅇ^(-(x̄ - μ)^2/(2 σ^2)) /. {μ → θ, σ → σ/√n};
domain[f] =
  {x̄, -∞, ∞} && {θ ∈ Reals, σ > 0, n > 0, n ∈ Integers};
```

Now, by (8.18),

$$\begin{aligned} P\left(\left| \overline{X}_n - \theta \right| \geq \varepsilon \right) &= P(\overline{X}_n - \theta \geq \varepsilon) + P(\overline{X}_n - \theta \leq -\varepsilon) \\ &= P(\overline{X}_n \geq \varepsilon + \theta) + P(\overline{X}_n \leq -\varepsilon + \theta) \end{aligned}$$

which is equal to:

```
sol = 1 - Prob[ε + θ, f]  +  Prob[-ε + θ, f] // FullSimplify
```

$$\text{Erfc}\left[\frac{\sqrt{n}\ \varepsilon}{\sqrt{2}\ \sigma}\right]$$

Taking the limit, we find:

```
<< Calculus`Limit`
lsol = Limit[sol, n → ∞]
Unprotect[Limit]; Clear[Limit];
```

$$\frac{e^{\frac{-\infty\ \text{Sign}[\varepsilon]^2}{\text{Sign}[\sigma]^2}}\ \sigma}{\varepsilon}$$

The output is not zero as we had hoped for, but if we apply `Simplify` along with the conditions on ε and σ:

```
Simplify[lsol, {ε > 0, σ > 0}]
```

0

Thus, $\overline{X}_n \xrightarrow{p} \theta$; that is, $\overline{X}_n$ is a consistent estimator of θ. ■

8.5 B Markov and Chebyshev Inequalities

In the previous example, the sample mean was shown to be a consistent estimator of the population mean (under Normality) by applying the definition of convergence in probability (8.14). Essentially, this requires deriving the cdf of the estimator, followed by taking a limit. This procedure may become less feasible in more complicated settings. Fortunately, it is often possible to establish consistency (or otherwise) of an estimator by only knowing its first two moments. This is done using probability inequalities. Consider, initially, *Markov's Inequality*

$$P\left(\,|\,X\,|\geq\alpha\right)\ \leq\ \alpha^{-k}\ E\left[\,|\,X\,|^k\,\right] \tag{8.19}$$

valid for $\alpha > 0$ and provided the k^{th} moment of X exists. Notice that the inequality holds for X having any distribution. For a proof of Markov's Inequality, see Billingsley (1995). A special case of Markov's Inequality is obtained by replacing $|\,X\,|$ with $|\,X-\mu\,|$, where $\mu = E[X]$, and setting $k = 2$. Doing so yields

$$P\left(\,|\,X-\mu\,|\geq\alpha\right)\ \leq\ \alpha^{-2}\ E[(X-\mu)^2]\ =\ \alpha^{-2}\ \text{Var}(X) \tag{8.20}$$

which is usually termed *Chebyshev's Inequality*.

⊕ ***Example 9:*** Applying the Inequalities

Let X denote the number of customers using a particular gas pump on any given day. What can be said about $P(150 < X < 250)$ when it is known that:

(i) $E[X] = 200$ and $E[(X - 200)^2] = 400$, and

(ii) $E[X] = 200$ and $E[(X - 200)^4] = 10^6$?

Solution (i): We have $\mu = 200$ and $\mathrm{Var}(X) = 400$. Note that

$$P(150 < X < 250) \;=\; P\big(\,|\,X - 200\,| \,<\, 50\big) \;=\; 1 - P\big(\,|\,X - 200\,| \,\geq\, 50\big).$$

By Chebyshev's Inequality (8.20), with $\alpha = 50$,

$$P\big(\,|\,X - 200\,| \,\geq\, 50\big) \;\leq\; \tfrac{400}{2500} \;=\; 0.16.$$

Thus, $P(150 < X < 250) \geq 0.84$. The probability that the gas pump will be used by between 150 and 250 customers each day is at least 84%.

Solution (ii): Applying Markov's Inequality (8.19) with X replaced by $X - 200$, with α set to 50 and k set to 4, finds

$$P\big(\,|\,X - 200\,| \,\geq\, 50\big) \;\leq\; \tfrac{10^6}{50^4} \;=\; 0.16.$$

In this case, the results from (i) and (ii) are equivalent. ■

8.5 C Weak Law of Large Numbers

There exist general conditions under which estimators such as $\overline{X}_n$ converge in probability, as sample size n increases. Inequalities such as Chebyshev's play a vital role in this respect, as we now show.

In Chebyshev's Inequality (8.20), replace X, μ and α with the symbols $\overline{X}_n$, θ and ε, respectively. That is,

$$P\big(\,|\,\overline{X}_n - \theta\,| \geq \varepsilon\big) \;\leq\; \varepsilon^{-2}\; E\big[(\overline{X}_n - \theta)^2\big] \tag{8.21}$$

where we interpret θ to be a parameter, and given constant $\varepsilon > 0$. Let MSE denote the expectation on the right-hand side of (8.21). Under the assumption that $(X_1, \ldots, X_n)$ is a random sample of size n drawn on a random variable X, it can be shown that:[8]

$$\text{MSE} \;=\; E\big[(\overline{X}_n - \theta)^2\big] \;=\; \tfrac{1}{n}\, E[(X - \theta)^2] \;+\; \tfrac{n-1}{n}\,(E[X] - \theta)^2. \tag{8.22}$$

In the following example, MSE is used to show that the sample mean $\overline{X}_n$ is a consistent estimator of the population mean.

⊕ ***Example 10:*** Consistent Estimation

Let $X \sim$ Uniform(0, 1) with pdf:

```
f = 1;   domain[f] = {x, 0, 1};
```

Let parameter $\theta \in (0, 1)$. We may evaluate MSE (8.22) as follows:

```
MSE = 1/n Expect[(x - θ)^2, f] +
   (n - 1)/n (Expect[x, f] - θ)^2 // Simplify
```

$$\frac{1}{4} + \frac{1}{12\,n} - \theta + \theta^2$$

Accordingly, the right-hand side of (8.21) is given simply by $\varepsilon^{-2}\left(\frac{1}{4} + \frac{1}{12n} - \theta + \theta^2\right)$, when $X \sim$ Uniform(0, 1).

Taking limits of both sides of (8.21) yields

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - \theta\right| \geq \varepsilon\right) \;\leq\; \varepsilon^{-2}\left(\tfrac{1}{4} - \theta + \theta^2\right).$$

Figure 5 plots the limit of MSE across the parameter space of θ:

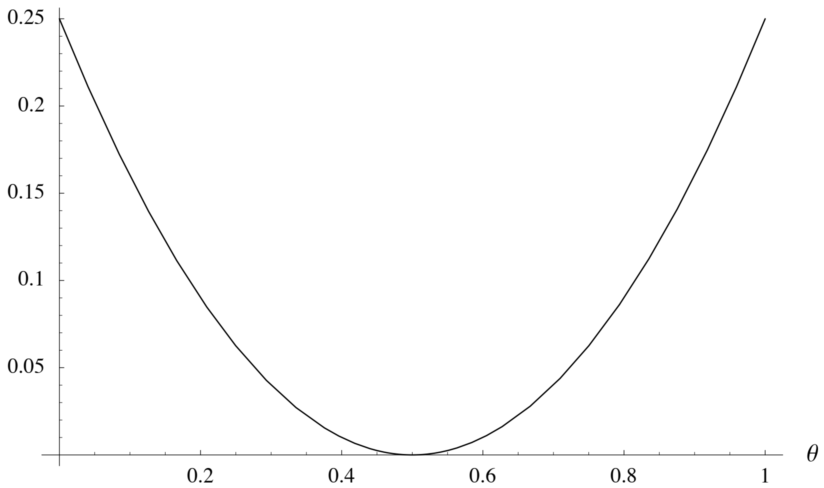


Fig. 5: Limit of MSE against θ

Since the plot is precisely 0 when $\theta = \theta_0 = \frac{1}{2}$, for every $\varepsilon > 0$, it follows from the definition of convergence in probability (8.14) that $\overline{X}_n \xrightarrow{p} \frac{1}{2}$, and ensures, due to uniqueness, that $\overline{X}_n$ cannot converge in probability to any other point in the parameter space. $\overline{X}_n$ is a consistent estimator of $\theta_0 = \frac{1}{2}$. What, if anything, is special about $\frac{1}{2}$ here? Put simply, $E[X] = \frac{1}{2}$. Thus, the sample mean $\overline{X}_n$ is a consistent estimator of the population mean. ■

Example 10 is suggestive of a more general result encapsulated in a set of theorems known as *Laws of Large Numbers*. These laws are especially relevant when trying to establish consistency of parameter estimators. We shall present just one — *Khinchine's Weak Law of Large Numbers*:

Theorem (Khinchine): Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of mutually independent and identically distributed random variables with finite mean μ. The sample mean:

$$\overline{X}_n \xrightarrow{p} \mu. \tag{8.23}$$

Proof: See, for example, Mittelhammer (1996, pp. 259–260).

In Khinchine's theorem, existence of a finite variance σ^2 for the random variables in the sequence is not required. If σ^2 is known to exist, a simple proof of (8.23) is to use Chebyshev's Inequality, because $E\left[(\overline{X}_n - \mu)^2\right] = \text{Var}(\overline{X}_n) = \frac{\sigma^2}{n}$.

8.6 Exercises

1. Let $X_n \sim \text{Bernoulli}(\frac{1}{2} + \frac{1}{2n})$, for $n \in \{1, 2, \ldots\}$. Show that $X_n \xrightarrow{d} X \sim \text{Bernoulli}(\frac{1}{2})$ using (i) the pmf of X_n, (ii) the mgf of X_n, and (iii) the cdf of X_n.

2. Let $X \sim \text{Poisson}(\lambda)$. Derive the cf of $X_\lambda = (X - \lambda)\big/\sqrt{\lambda}$. Then, use it to show that $X_\lambda \xrightarrow{d} Z \sim N(0, 1)$ as $\lambda \to \infty$.

3. Let $X \sim \text{Uniform}(0, \theta)$, where $\theta > 0$. Define $X_{(j)}$ as the j^{th} order statistic from a random sample of size n drawn on X, for $j \in \{1, \ldots, n\}$; see §9.4 for details on order statistics. Consider the transformation of $X_{(j)}$ to Y_j such that $Y_j = n(\theta - X_{(j)})$. By making use of **mathStatica**'s `OrderStat`, `OrderStatDomain`, `Transform`, `TransformExtremum` and `Prob` functions, derive the limit distribution of Y_j when (i) $j = n$, (ii) $j = n - 1$, and (iii) $j = n - 2$. From this pattern, can you deduce the limit distribution of Y_{n-k}, where constant k is a non-negative integer k?

4. Let $X \sim \text{Cauchy}$, and let $(X_1, \ldots, X_n)$ denote a random sample of size n drawn on X. Derive the cf of X. Then, use it to show that the sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ cannot converge in probability to a constant.

5. Let $X \sim \text{Uniform}(0, \pi)$, and let $(X_1, X_2, \ldots, X_n)$ denote a random sample of size n drawn on X. Determine a_n and b_n such that $\frac{S_n - a_n}{b_n} \xrightarrow{d} Z \sim N(0, 1)$, where $S_n = \sum_{i=1}^n \cos(X_i)$. Then evaluate van Beek's bound.

6. **Simulation I:** At the conclusion of *Example 6*, the function

```
N01RNG := Plus @@ Table[Random[], {12}] - 6
```

was proposed as an approximate pseudo-random number generator for a random variable $X \sim N(0, 1)$. Using `QQPlot`, investigate the performance of `N01RNG`.

7. **Simulation II:** Let $X \sim N(0, 1)$, and let $Y = X^2 \sim$ Chi-squared(1). From the relation between X and Y, it follows that $\texttt{N01RNG}^2$ is an approximate pseudo-random number generator for Y. That is, if

$$\texttt{N01RNG} \xrightarrow{d} X, \text{ then } \texttt{N01RNG}^2 \xrightarrow{d} Y.$$

(i) Noting that the sum of m independent Chi-squared(1) random variables is distributed Chi-squared(m), propose an approximate pseudo-random number generator for $Z \sim$ Chi-squared(m) based on `N01RNG`.

(ii) Provided that X and Z are independent, $T = X \big/ \sqrt{Z/m} \sim$ Student's $t(m)$. Hence, propose an approximate pseudo-random number generator for T based on `N01RNG`, and investigate its performance when $m = 1$ and 10.

8. **Simulation III:** Let $(W_1, W_2, \ldots, W_m)$ be mutually independent random variables such that $W_i \sim N(\mu_i, 1)$. Define $V = \sum_{i=1}^m W_i^2 \sim$ Noncentral Chi-squared (m, λ), where $\lambda = \sum_{i=1}^m \mu_i^2$.

(i) Use the relationship between V and $\{W_i\}$ to propose an approximate pseudo-random number generator for V based on `N01RNG`, as a *Mathematica* function of m and λ.

(ii) Use `N01RNG` and `DiscreteRNG` to construct an approximate pseudo-random number generator for V based on the parameter-mix

$$\text{Noncentral Chi-squared}\,(m, \lambda) \;=\; \text{Chi-squared}\,(m + 2\,K) \bigwedge_K \text{Poisson}\left(\tfrac{\lambda}{2}\right)$$

as a *Mathematica* function of m and λ.

9. For a random variable X with mean $\mu \neq 0$ and variance σ^2, reformulate the Chebyshev Inequality (8.20) in terms of the *relative mean deviation* $\frac{|X-\mu|}{|\mu|} = \left|\frac{X-\mu}{\mu}\right|$. That is, using pen and paper, show that

$$P\left(\left|\frac{X-\mu}{\mu}\right| \geq \beta\right) \leq (r\,\beta)^{-2}$$

where $\beta > 0$, and r denotes the signal-to-noise ratio $|\mu|/\sigma$. Then evaluate r^2 for the Binomial(n, p), Uniform(a, b), Exponential(λ) and Fisher $F(a, b)$ distributions.

10. Let X denote a random variable with mean μ and variance σ^2. In Chebyshev's Inequality, show (you need only use pencil and paper) that if $\alpha \geq 10\,\sigma$, then $P(|X-\mu| \geq \alpha) \leq 0.01$. Next, suppose there is more known about X; namely, $X \sim N(\mu, \sigma^2)$. By evaluating $P(|X-\mu| \geq \alpha)$, show that the assumption of Normality has the effect of allowing the inequality to hold over a larger range of values for α.

11. Let $X \sim$ Binomial(n, p), and let $a \leq b$ be non-negative integers. The Normal approximation to the Binomial is given by

$$P(a \leq X \leq b) \;\simeq\; \Phi(d) - \Phi(c)$$

where Φ denotes the cdf of a standard Normal distribution, and

$$c = \frac{a - n\,p - \frac{1}{2}}{\sqrt{n\,p\,(1-p)}} \quad \text{and} \quad d = \frac{b - n\,p + \frac{1}{2}}{\sqrt{n\,p\,(1-p)}}.$$

Investigate the accuracy of the approximation by plotting the error of the approximation when $a = 20$, $b = 80$ and $p = 0.1$, against values of n from 100 to 500 in increments of 10.

12. Let $X \sim \text{Binomial}(n, p)$ and $Y \sim \text{Poisson}(n\,p)$, and let $a \le b$ be non-negative integers. The Poisson approximation to the Binomial is given by

$$P(a \le X \le b) \;\simeq\; P(a \le Y \le b).$$

Investigate the accuracy of the approximation.