Discrete Multivariate Analysis,
Yvonne MM Bishop
Stephen E. Fienberg
Paul W. Holland
MIT Press, 1975.

14.2 The O, o Notation

The essence of asymptotic methods is approximation. We approximate functions, random variables, probability distributions, means, variances, and covariances. Careful use of approximations requires attention to their accuracy or *order*. A notation that is especially useful for keeping track of the order of an approximation is the "big O, little o" notation. In this section, we introduce and review this notation for the reader who may be unfamiliar with it. More advanced readers may wish to skim the material here.

There are really two big O, little o notations: one for nonstochastic variables, denoted O and o, and one for stochastic variables, denoted O_p and o_p. In this section we define and exemplify the O and o notation for nonstochastic variables. In the next section we generalize this to O_p and o_p for stochastic variables.

If $\{a_n\}$ and $\{b_n\}$ are two sequences of real numbers, then the following two formal definitions define the expressions $a_n = O(b_n)$ and $a_n = o(b_n)$, which are basic for understanding the O and o notation.

DEFINITION 14.2-1 $a_n = O(b_n)$ *(Read: a_n is big O of b_n) if the ratio $|a_n/b_n|$ is bounded for large n; in detail, if there exists a number K and an integer $n(K)$ such that if n exceeds $n(K)$ then $|a_n| < K|b_n|$.*

DEFINITION 14.2-2 $a_n = o(b_n)$ *(Read: a_n is little o of b_n) if the ratio $|a_n/b_n|$ converges to zero; in detail, if for any $\varepsilon > 0$, there exists an integer $n(\varepsilon)$ such that if n exceeds $n(\varepsilon)$ then $|a_n| < \varepsilon|b_n|$.*

The idea behind these two definitions is the comparison of the approximate size or *order of magnitude* of $\{a_n\}$ to that of $\{b_n\}$. In most applications, $\{a_n\}$ is the sequence of interest while $\{b_n\}$ is a comparison sequence. Some important examples of $\{b_n\}$ are: $b_n = n^{-1}$, $b_n = n^{-1/2}$, $b_n = n$, $b_n = n \log n$.

The interpretation of $a_n = O(b_n)$ is that the sequence of interest $\{a_n\}$ is of roughly the same size or order of magnitude as the comparison sequence $\{b_n\}$. The interpretation of $a_n = o(b_n)$ is that $\{a_n\}$ is of a smaller order of magnitude than is $\{b_n\}$. Of course, both of these comparisons refer to "large n" properties of the sequences and are unaffected by their initial behavior.

14.2.1 *Conventions in the use of the O, o notation*

A number of conventions are observed in the use of the O, o notation. We list here some of the more important ones.

(i) Definitions 14.2-1 and 14.2-2 above are unaffected if we allow a_n to be infinite or even undefined for a finite number of values of n. In a number of statistical applications this is a convenience, and so we will always assume this slight extension of the definitions.

(ii) Definitions 14.2-1 and 14.2-2 above may easily be extended to apply to a sequence $\{\mathbf{a}_n\}$ of vectors. If $\|\mathbf{a}_n\|$ denotes the length of the vector $\mathbf{a}_n$, i.e.

$$\|\mathbf{a}_n\| = \sqrt{\sum_i a_{ni}^2},$$

then $\mathbf{a}_n = O(b_n)$ means $\|\mathbf{a}_n\| = O(b_n)$, and $\mathbf{a}_n = o(b_n)$ means $\|\mathbf{a}_n\| = o(b_n)$.

(iii) The expressions $a_n = O(b_n)$ and $a_n = O(cb_n)$ are equivalent if c is a nonzero constant. The same is true for o. Hence multiplicative constants are ignored in the argument of O and o. For example, $o(2n^{-1})$ is written $o(n^{-1})$. The sign, positive or negative, of an O or o term is always ignored.

(iv) The expression $a_n = o(1)$ is used to signify that $a_n \to 0$, while $a_n = O(1)$ means that $|a_n| \leq K$ for some constant K if n is large enough, i.e., that $\{a_n\}$ is eventually bounded.

(v) We always have $a_n = O(a_n)$.

(vi) Products of O and o factors obey these easily proved rules:

(P1) $O(a_n)O(b_n) = O(a_nb_n)$,
(P2) $O(a_n)o(b_n) = o(a_nb_n)$,
(P3) $o(a_n)o(b_n) = o(a_nb_n)$.

(vii) It is often necessary to add together several O and o expressions to obtain a single order-of-magnitude term. The rule is that the order of magnitude of a sum is the largest order of magnitude of the summands. For example

$$o(1) + O(n^{-1/2}) + O(n^{-1}) = o(1).$$

This rule is not necessarily correct if the number of terms in the summation depends also on n. For example, if we have the n terms

$$\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n} = 1,$$

then the largest order of magnitude of the summands is n^{-1}, but this is not the order of magnitude of the summation.

The O and o notation also appears with a continuous variable in the argument rather than a sequence, especially in expansions of functions. For example, the first-order Taylor expansion of a function $f(\cdot)$ about the value a is stated as:

$$f(x) = f(c) + (x - c)f'(c) + o(|x - c|) \quad \text{as } x \to c. \tag{14.2-1}$$

In this example, the little o means that if x_n is any sequence such that $x_n \to c$ and if a_n and b_n are defined by

$$a_n = f(x_n) - f(c) - (x_n - c)f'(c), \tag{14.2-2}$$

$$b_n = x_n - c, \tag{14.2-3}$$

then

$$a_n = o(b_n). \tag{14.2-4}$$

In order for (14.2-1) to be true, (14.2-4) must be true for any choice of x_n such that $x_n \to c$. The following two definitions formalize the use of the o, O notation with a continuous argument in terms of definitions 14.2-1 and 14.2-2.

DEFINITION 14.2-3 *$a(x) = O(b(x))$ as $x \to L$ if for any sequence $\{x_n\}$ such that $x_n \to L$, we have $a(x_n) = O(b(x_n))$ in the sense of definition 14.2-1.*

DEFINITION 14.2-4 *$a(x) = o(b(x))$ as $x \to L$ if for any sequence $\{x_n\}$ such that $x_n \to L$, we have $a(x_n) = o(b(x_n))$ in the sense of definition 14.2-2.*

In the applications of definitions 14.2-3 and 14.2-4, we observe that when there is no ambiguity the condition "as $x \to L$" is not always explicitly stated. The value of L may be any real number, $+\infty$, or $-\infty$.

In the following section we go through an illustrative example that allows the reader to see the use of the O, o notation in a natural setting.

14.2.2 *An example of the use of the O, o notation*

Example 14.2-1 Approximating e

We begin by considering the sequence $\{e_n\}$ given by

$$e_n = (1 + n^{-1})^n. \tag{14.2-5}$$

Let $\log(x)$ denote the natural logarithm of x (to the base $e = 2.7182818\ldots$). We may find the limit of $\{e_n\}$ by first finding the limit of $\log(e_n) = n\log(1 + n^{-1})$ and then taking antilogs. We let

$$f(t) = \log(1 + t). \tag{14.2-6}$$

From elementary calculus, we recall that

$$f(0) = 0, \qquad f'(t) = (1 + t)^{-1}, \qquad f'(0) = 1. \tag{14.2-7}$$

Hence the Taylor expansion of f about $t = 0$ may be expressed as

$$f(t) = f(0) + tf'(0) + o(t)$$

$$= t + o(t) \quad \text{as } t \to 0. \tag{14.2-8}$$

Now we apply definition 14.2-3 to (14.2-8) with $x_n = n^{-1}$, and we obtain

$$\log(1 + n^{-1}) = n^{-1} + o(n^{-1}), \tag{14.2-9}$$

so that

$$\log(e_n) = n\log(1 + n^{-1}) = 1 + no(n^{-1}) = 1 + o(1) \tag{14.2-10}$$

and hence

$$\log(e_n) \to 1. \tag{14.2-11}$$

From (14.2-11) it follows that the limit we seek is

$$\lim_{n\to\infty}(1 + n^{-1})^n = \lim_{n\to\infty} e^{\log(e_n)} = e^1 = e. \tag{14.2-12}$$

The convergence of $\{e_n\}$ to e is often too slow for applications, but by considering higher-order Taylor series expansions of $f(t)$, we may find simple "corrections" to e_n that markedly improve the convergence to e. The remainder of this example is devoted to finding some of these corrections as further illustrations of the use of the O, o notation.

Define the sequence x_n by

$$x_n = \log(1 + n^{-1}). \tag{14.2-13}$$

Now we look at the sequence $\{(n + c)x_n\}$, where c is a constant to be determined. The second-order Taylor series expansion of $f(t) = \log(1 + t)$ is given by

$$f(t) = t - \tfrac{1}{2}t^2 + o(t^2) \quad \text{as } t \to 0. \tag{14.2-14}$$

Hence

$$x_n = f(n^{-1}) = n^{-1} - \tfrac{1}{2}n^{-2} + o(n^{-2}), \tag{14.2-15}$$

so that

$$(n + c)x_n = (n + c)(n^{-1} - \tfrac{1}{2}n^{-2} + o(n^{-2}))$$

$$= 1 + (c - \tfrac{1}{2})n^{-1} - \tfrac{1}{2}cn^{-2} + no(n^{-2}) + co(n^{-2})$$

$$= 1 + (c - \tfrac{1}{2})n^{-1} + o(n^{-1}). \tag{14.2-16}$$

By choosing $c = 1/2$, the order of the convergence of $(n + c)x_n$ to 1 improves from $o(1)$ to $o(n^{-1})$. Thus we define a new sequence e_n^* by

$$e_n^* = (1 + n^{-1})^{n+\frac{1}{2}}. \tag{14.2-17}$$

The distinction between e_n and e_n^* is that

$$e_n = e^{1+o(1)} \quad \text{while} \quad e_n^* = e^{1+o(n^{-1})}. \tag{14.2-18}$$

We need not stop here. We can find higher-order approximations by the same device. For example, consider the sequence $\{(n + c + dn^{-1})x_n\}$, for constants

c and d to be determined. As an exercise the reader is invited to use the third-order Taylor expansion of $f(t) = \log(1 + t)$ about $t = 0$ to show that if $c = 1/2$ and $d = -1/12$ then

$$(n + c + dn^{-1})x_n = 1 + o(n^{-2}). \tag{14.2-19}$$

This leads us to define e_n^{**} by

$$e_n^{**} = (1 + n^{-1})^{n + \frac{1}{2} - \frac{1}{12n}}, \tag{14.2-20}$$

which converges to e at a faster rate than either e_n or e_n^*.

Accuracy of the approximations to e

The careful reader will have one question that gnaws at him through the above manipulation of little o terms. He remembers that

$$10^6 n^{-2} = o(n^{-1})$$

and realizes that the multiplicative constant 10^6, while not affecting the order of magnitude of the convergence, can have a tremendous effect on the actual degree of the approximation for the values of n that concern him. In other words, after e_n^* and e_n^{**} have been invented it is still important to find out if the actual values of these quantities are really nearer e than e_n is, for small values of n. To answer this question, table 14.2-1 gives values of n, e_n, e_n^*, and e_n^{**} for $n = 1$ to 20. It is evident that for many purposes e_n is a totally inadequate approximation to e for these values of n. On the other hand, two- and three-decimal accuracy is quickly attained by e_n^* and e_n^{**}. ■■

Table 14.2-1

n	$\left(1 + \frac{1}{n}\right)^n$	$\left(1 + \frac{1}{n}\right)^{n+\frac{1}{2}}$	$\left(1 + \frac{1}{n}\right)^{n+\frac{1}{2}-\frac{1}{12n}}$
1	2	2.8284	2.6697
2	2.25	2.7557	2.7095
3	2.37	2.7371	2.7153
4	2.44	2.7296	2.7169
5	2.49	2.7258	2.7175
6	2.52	2.7237	2.7178
7	2.55	2.7223	2.7179986
8	2.57	2.7214	2.7180886
9	2.58	2.7207	2.7181441
10	2.59	2.7203	2.7181803
11	2.60	2.719997	2.7182048
12	2.61	2.719733	2.7182220
13	2.62	2.719526	2.7182345
14	2.627	2.719360	2.7182437
15	2.633	2.719225	2.7182507
16	2.638	2.719114	2.7182560
17	2.642	2.719022	2.7182602
18	2.646	2.718944	2.7182636
19	2.650	2.718878	2.7182663
20	2.653	2.718821	2.7182684
∞	2.718...	2.718281...	2.7182818...

14.2.3 *Exercises*

1. Show that $(2n^2 + 3n + 4)\sin n = O(n^2)$.
2. Show that $\log n = o(n^\alpha)$ for any power $\alpha > 0$.
3. Show that $\log x = o(x^{-1})$ as $x \to 0$.
4. Show that $\sin x = O(x)$ as $x \to 0$.
5. Show that if $c = 1/2$ and $d = -1/12$, then (14.2-19) is valid.
6. Prove the multiplicative rules for o and O given in (P1), (P2), and (P3).
7. Expand $(x + yn^{-1/2} + zn^{-1})^2$ to order n^{-1} by applying the rule for multiplying and adding o and O terms.
8. Show that we could have defined $O(b_n)$ and $o(b_n)$ by first defining $a_n = O(1)$ and $a_n = o(1)$ appropriately, and then defining $O(b_n) = b_n O(1)$ and $o(b_n) = b_n o(1)$.
9. Show that $(1 + (\lambda/n))^n \to e^\lambda$ for all λ.
10. Show that $(1 + \lambda/n + o(1/n))^n \to e^\lambda$ for all λ.
11. Does $f(t) = \sqrt{t}$ have a first-order Taylor expansion at $t = 0$?

Suppose that we are testing

$$H_0: \mathbf{p} = \boldsymbol{\pi}$$

with the test statistic X^2 of (14.3-49), but instead of H_0 being true, suppose that $\mathbf{p}$ is given by

$$\mathbf{p} = \boldsymbol{\pi} + n^{-1/2}\boldsymbol{\mu},$$

as in example 14.3-4. The reader can show that

$$\mathscr{L}[X^2] \to \mathscr{L}[\mathbf{U}\mathbf{D}_{\pi}^{-1}\mathbf{U}'], \tag{14.3-52}$$

where $\mathbf{U}$ has the multivariate normal distribution $\mathscr{N}(\boldsymbol{\mu}, \mathbf{D}_{\pi} - \boldsymbol{\pi}'\boldsymbol{\pi})$. If $Y = \mathbf{U}\mathbf{D}_{\pi}^{-1}\mathbf{U}'$, then it can be shown that Y has a noncentral chi square distribution with $T - 1$ degrees of freedom and noncentrality parameter

$$\psi^2 = \boldsymbol{\mu}\mathbf{D}_{\pi}^{-1}\boldsymbol{\mu}'. \tag{14.3-53}$$

If we write $\boldsymbol{\mu}$ as

$$\boldsymbol{\mu} = \sqrt{n}(\mathbf{p} - \boldsymbol{\pi}),$$

then this result is sometimes stated with the noncentrality parameter given as

$$\psi^2 = n(\mathbf{p} - \boldsymbol{\pi})\mathbf{D}_{\pi}^{-1}(\mathbf{p} - \boldsymbol{\pi})'. \quad ▮▮ \tag{14.3-54}$$

The effect of a small remainder term on convergence in distribution

The technique discussed in this section is actually a special case of theorem 14.3-6, but it is so important that we discuss it separately. The technique is based on the following theorem.

THEOREM 14.3-9 *If* $\mathscr{L}[X_n] \to \mathscr{L}[X]$ *and if* $Y_n \xrightarrow{p} 0$, *then* $\mathscr{L}[X_n + Y_n] \to \mathscr{L}[X]$.

For a proof see Rao [1965], p. 104.

The idea behind theorem 14.3-9 is as follows. Suppose we are interested in the limiting distribution of Z_n and we are able to express Z_n as $Z_n = X_n + Y_n$, where $\{X_n\}$ is a stochastic sequence with an asymptotic distribution we can find by some method and Y_n is a small remainder term that converges to 0 in probability, i.e., $Y_n \xrightarrow{p} 0$. Then the limiting distribution of Z_n is the same as that of X_n. We delay examples of this approach until Section 14.4, 14.6, 14.8, and 14.9, where it is applied extensively.

14.3.4 *Exercises*

1. Show that if $\mathbf{X}_n \xrightarrow{p} \mathbf{c}$, then $\mathscr{L}[\mathbf{X}_n] \to \mathscr{L}[\mathbf{C}]$, where $\mathbf{C}$ is a degenerate random variable concentrated at $\mathbf{c}$.
2. Show that if $\mathbf{X} = (X_1, \ldots, X_T)$ has the multinomial distribution $\mathscr{M}(n, \mathbf{p})$, where $\mathbf{p} = (p_1, \ldots, p_T)$, then

$$M_{\mathbf{X}}(\mathbf{t}) = \left(\sum_i p_i e^{t_i}\right)^n.$$

3. Use the result of problem 2 to show that $E(X_i) = np_i$, $\mathrm{Var}(X_i) = np_i(1 - p_i)$, and $\mathrm{Cov}(X_i, X_j) = -np_ip_j$, so that

$$E(\mathbf{X}) = n\mathbf{p},$$

$$\mathrm{Cov}(\mathbf{X}) = n(\mathbf{D}_p - \mathbf{p}'\mathbf{p}),$$

where $\mathbf{D}_p = \mathrm{diag}(\mathbf{p})$.

4. Fill in the details of example 14.3-6.
5. Use theorem 14.3-6 to prove theorem 14.3-9.
6. Let X have the negative binomial distribution given by

$$P\{X = x\} = \binom{r + x - 1}{r - 1} p^r(1 - p)^x \quad x = 0, 1, \ldots.$$

(i) Find the moment-generating function of X.

(ii) Find the mean and variance of X. (See Section 13.7.1.)

(iii) If $p = p_n = \lambda/n$ and r remains fixed, find the limiting distribution of $n^{-1}X$.

14.4 The O_p, o_p Notation for Stochastic Sequences

We now develop the O_p, o_p notation for stochastic sequences, which generalizes the O, o notation for nonstochastic sequences. In the first four sections we describe the notation and its elementary properties. In the fifth section, we deal with a more general version of this theory due to Chernoff and Pratt.

14.4.1 *Definition of* $X_n = o_p(b_n)$

We follow up the idea expressed in exercise 8 of Section 14.2.3, by defining $o_p(1)$ first and then setting $o_p(b_n) = b_n o_p(1)$. The notation $X_n = o_p(1)$ means exactly the same thing as $X_n \xrightarrow{p} 0$. More formally, we have:

DEFINITION 14.4-1 $X_n = o_p(1)$ *if for every* $\varepsilon > 0$,

$$\lim_{n\to\infty} P\{|X_n| \le \varepsilon\} = 1. \tag{14.4-1}$$

If $\mathbf{X}_n$ is a vector, we say that $\mathbf{X}_n = o_p(1)$ if $\|\mathbf{X}_n\| = o_p(1)$. Then we define the general o_p notation by:

DEFINITION 14.4-2 $X_n = o_p(b_n)$ *if* $X_n/b_n = o_p(1)$, *or equivalently*, $X_n = b_n o_p(1)$.

Again, if $\mathbf{X}_n$ is a vector, we say that $\mathbf{X}_n = o_p(b_n)$ if $\|\mathbf{X}_n\| = o_p(b_n)$.

14.4.2 *Definition of* $X_n = O_p(b_n)$

We continue the approach of Section 14.4.1 and define $X_n = O_p(1)$ first and then set $O_p(b_n) = b_n O_p(1)$. However, to motivate the definition of $O_p(1)$, it is useful to begin by amplifying definition 14.4-1 somewhat. Since this amplification is nothing more than a restatement of definition 14.4-1 with more technical detail, we call it definition 14.4-1a.

DEFINITION 14.4-1a $X_n = o_p(1)$ *if for every* $\varepsilon > 0$ *and every* $\eta > 0$ *there exists an integer* $n(\varepsilon, \eta)$ *such that if* $n \ge n(\varepsilon, \eta)$, *then*

$$P\{|X_n| < \varepsilon\} \ge 1 - \eta. \tag{14.4-2}$$

Informally, this definition means that with arbitrarily high probability (i.e., $\ge 1 - \eta$), $|X_n| = o(1)$. Taking our cue from this informal description, we want $X_n = O_p(1)$ to mean that with arbitrarily high probability, $|X_n| = O(1)$. In terms of a formal definition this becomes:

DEFINITION 14.4-3 *$X_n = O_p(1)$ if for every $\eta > 0$ there exist a constant $K(\eta)$ and an integer $n(\eta)$ such that if $n \geqq n(\eta)$, then*

$$P\{|X_n| \leqq K(\eta)\} \geqq 1 - \eta. \tag{14.4-3}$$

If $\mathbf{X}_n$ is a vector, we say $\mathbf{X}_n = O_p(1)$ if $\|\mathbf{X}_n\| = O_p(1)$. Finally, we define the general O_p notation by:

DEFINITION 14.4-4 *$X_n = O_p(b_n)$ if $X_n/b_n = O_p(1)$, or equivalently, $X_n = b_n O_p(1)$.* As before, if $\mathbf{X}_n$ is a vector, we say $\mathbf{X}_n = O_p(b_n)$ if $\|\mathbf{X}_n\| = O_p(b_n)$.

If we compare definitions 14.4-1a and 14.4-3 with their counterparts in the O, o notation (definitions 14.2-1 and 14.2-2, with $b_n = 1$) we see a number of parallels. In both O_p and o_p, events are required to hold with a probability that exceeds a preassigned limit $1 - \eta$, replacing the "certainty" of the O, o definitions. In both $o(1)$ and $o_p(1)$, the sequences are required to be less than any prescribed small positive ε if n is sufficiently large. In both $O(1)$ and $O_p(1)$, the sequences are required to be bounded by some constant K if n is sufficiently large.

It is sometimes useful to refer to $X_n = O_p(1)$ by saying that X_n is "bounded in probability," and to $X_n = o_p(1)$ by saying that X_n "converges to zero in probability." The information in the O_p, o_p notation is often referred to as the "stochastic order" of X_n.

14.4.3 *Methods for determining the stochastic order of a sequence*

In this section we discuss two easily applied methods for determining the stochastic order of a sequence of random variables $\{X_n\}$.

You are only as big as your standard deviation

The standard deviation of a random variable is often used as an index of the size or order of magnitude of the typical departure of the random variable from its expected value. This use of the standard deviation to measure typical deviations is usually motivated by the fact that for the normal distribution, the probability that an observation will lie within one standard deviation from the mean is approximately 0.66. However, even for nonnormal distributions, the standard deviation gives the order of magnitude of typical deviations. For example, Tchebychev's inequality asserts that if X is a random variable with mean μ and variance $\sigma^2 < \infty$ and h is any positive number, then

$$P\{|X - \mu| \leqq h\sigma\} \geqq 1 - h^{-2}. \tag{14.4-4}$$

We use Tchebychev's inequality to connect the O_p, o_p notion of stochastic order of magnitude with the standard deviation as an index of the order of magnitude of deviations from the expected value.

THEOREM 14.4-1 *If $\{X_n\}$ is a stochastic sequence with $\mu_n = E(X_n)$ and $\sigma_n^2 =$ $\mathrm{Var}(X_n) < \infty$, then*

$$X_n - \mu_n = O_p(\sigma_n).$$

Proof If in (14.4-4) we set $h = \eta^{-1/2}$ for any $0 < \eta < 1$ and apply (14.4-4) to X_n, μ_n, and σ_n, then we have

$$P\left\{\frac{|X_n - \mu_n|}{\sigma_n} < \eta^{-1/2}\right\} \geqq 1 - \eta. \tag{14.4-5}$$

Thus (14.4-5) holds for $n = 1, 2, \ldots$. Setting $K(\eta) = \eta^{-1/2}$, we apply definition 14.4-3 and conclude that

$$\frac{X_n - \mu_n}{\sigma_n} = O_p(1),$$

from which the desired result immediately follows. ■

Example 14.4-1 The order of a binomial variable

If X_n has the binomial distribution $\mathscr{B}(n, p)$, then $E(X_n) = np$, $\mathrm{Var}(X_n) = np(1 - p)$. Hence $\sigma_n = \sqrt{np(1 - p)} = O(\sqrt{n})$, from which we conclude that $X_n - np = O_p(\sqrt{n})$, or, as it is usually written,

$$X_n = np + O_p(\sqrt{n}). \tag{14.4-6}$$

Actually, this example is a special case of the more general result that if X_n is a sum of n independent and identically distributed random variables with mean μ and variance σ^2, then

$$X_n = n\mu + O_p(\sqrt{n}) \tag{14.4-7}$$

(see exercise 2 in Section 14.4.6). ■■

A sequence that converges in distribution is also bounded in probability

The subject of this short section is a theorem that is the most common tool used for showing that $X_n = O_p(1)$. This theorem also illustrates one of the many connections between convergence in distribution and the O_p, o_p notation.

THEOREM 14.4-2 *If $\mathscr{L}[X_n] \to \mathscr{L}[X]$, then $X_n = O_p(1)$.*

The proof of this result is left for exercise 3 in Section 14.4.6. The reader may wish to apply this result to example 14.4-1. The next example gives a further application of this theorem.

Example 14.4-2 The order of the minimum of exponential variables

Let $Y_1, \ldots, Y_n$ be independent random variables, all identically distributed with the exponential distribution

$$P\{Y_i \leqq t\} = F(t) = 1 - e^{-\lambda t}$$

for $t \geqq 0$. Now if $X_n = \min\{Y_1, \ldots, Y_n\}$, then the distribution function of X_n is

$$\begin{aligned} G_n(x) = P\{X_n \leqq x\} &= 1 - P\{X_n > x\} \\ &= 1 - P\{Y_1 > x, Y_2 > x, \ldots, Y_n > x\} \\ &= 1 - P\{Y_1 > x\}P\{Y_2 > x\} \ldots P\{Y_n > x\} \\ &= 1 - (1 - F(x))^n = 1 - e^{-n\lambda x}. \end{aligned}$$

We conclude that X_n has the exponential distribution with parameter $n\lambda$. From this it follows that nX_n has the same distribution as Y_1, from which we conclude that

$$\mathscr{L}[nX_n] \to \mathscr{L}[Y_1]. \tag{14.4-8}$$

Applying theorem 14.4-2 to (14.4-8), we obtain

$$nX_n = O_p(1), \tag{14.4-9}$$

or

$$X_n = O_p(n^{-1}). \quad ■■ \tag{14.4-10}$$

14.4.4 *Convergence in distribution and $o_p(1)$*

This section extends the discussion in Section 14.3.3 on the effect of a small remainder term on convergence in distribution. In the O_p, o_p notation, theorem 14.3-9 can be restated as

THEOREM 14.4-3 *If $\mathscr{L}[X_n] \to \mathscr{L}[X]$, then $\mathscr{L}[X_n + o_p(1)] \to \mathscr{L}[X]$.*

This theorem, in conjunction with the various tools we have for determining the stochastic order of a sequence, allows us to find the limiting distributions of fairly complicated random variables. We illustrate this point with two examples involving the binomial distribution.

Example 14.4-3 Asymptotic confidence intervals for p

We continue using the notation of example 14.4-1. From example 14.3-3 it follows that if $\hat{p} = n^{-1}X_n$, then

$$\mathscr{L}[\sqrt{n}(\hat{p} - p)] \to \mathscr{L}[W], \tag{14.4-11}$$

where W has the normal distribution $\mathscr{N}(0, p(1-p))$. Define Z_n by

$$Z_n = \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p(1-p)}}. \tag{14.4-12}$$

By theorem 14.3-6, $\mathscr{L}[Z_n] \to \mathscr{L}[Z]$, where Z has the unit normal distribution $\mathscr{N}(0, 1)$.

From (14.4-11) and theorem 14.4-2, we deduce that

$$\hat{p} = p + O_p(n^{-1/2}). \tag{14.4-13}$$

But if $Y_n = O_p(n^{-1/2})$, then $Y_n = o_p(1)$ (see exercise 6 in Section 14.4.6), so that

$$\hat{p} = p + o_p(1) \quad \text{as } n \to \infty.$$

Applying theorems 14.3-1 and 14.3-6, we deduce that if $p \neq 0$, and $p \neq 1$, then

$$\left(\frac{p(1-p)}{\hat{p}(1-\hat{p})}\right)^{1/2} = 1 + o_p(1). \tag{14.4-14}$$

Now define V_n by

$$\begin{aligned} V_n = \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{\hat{p}(1-\hat{p})}} &= Z_n\left(\frac{p(1-p)}{\hat{p}(1-\hat{p})}\right)^{1/2} = Z_n(1 + o_p(1)) \\ &= Z_n + Z_n o_p(1) = Z_n + O_p(1)o_p(1) \\ &= Z_n + o_p(1). \end{aligned}$$

We conclude from theorem 14.3-9 that

$$\mathscr{L}[V_n] \to \mathscr{L}[Z]. \tag{14.4-15}$$

The result (14.4-15) is often used to set asymptotic confidence intervals for p as follows. If n is large enough, (14.4-15) says that

$$P\{-a < V_n < a\} \approx P\{-a < Z < a\}.$$

If a is chosen so that

$$P\{-a < Z < a\} = 1 - \alpha,$$

then inverting the inequality $-a < V_n < a$, we obtain

$$P\left\{\hat{p} - \frac{a}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p})} < p < \hat{p} + \frac{a}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p})}\right\} \approx 1 - \alpha$$

and so obtain an asymptotic $1 - \alpha$ level confidence interval for p. ■■

Example 14.4-4 The square of a binomial proportion

We continue using the notation of the example 14.4-3. Our aim now is to find the asymptotic distribution of $\hat{p}^2$. We begin by observing that

$$\hat{p}^2 = (p + (\hat{p} - p))^2 = p^2 + 2p(\hat{p} - p) + (\hat{p} - p)^2,$$

or

$$\begin{aligned} \sqrt{n}(\hat{p}^2 - p^2) &= 2p\sqrt{n}(\hat{p} - p) + \sqrt{n}(\hat{p} - p)^2 \\ &= 2pW_n + n^{-1/2}W_n^2, \end{aligned}$$

where we set $W_n = \sqrt{n}(\hat{p} - p)$. We know that W_n converges in distribution, so from theorem 14.3-6 we know that W_n^2 also converges in distribution. Then from theorem 14.4-2 we know that $W_n^2 = O_p(1)$, and hence $n^{-1/2}W_n^2 = O_p(n^{-1/2}) = o_p(1)$. These facts give us

$$\sqrt{n}(\hat{p}^2 - p^2) = 2pW_n + o_p(1).$$

Since $\mathscr{L}[W_n] \to \mathscr{L}[W]$, where W has the normal distribution $\mathscr{N}(0, p(1-p))$, we now have

$$\mathscr{L}[\sqrt{n}(\hat{p}^2 - p^2)] \to \mathscr{L}[2pW].$$

But $2pW$ has the $\mathscr{N}(0, 4p^3(1-p))$ distribution, from which we conclude that $\hat{p}^2$ has an approximate normal distribution with mean p^2 and variance $n^{-1}4p^3(1-p)$. ■■

14.4.5 *Chernoff–Pratt theory of stochastic order*

The similarity of the O, o and O_p, o_p notation suggests that they be used in tandem. For instance, in example 14.4-1 we uncritically assumed that $O_p(O(\sqrt{n})) = O_p(\sqrt{n})$. Similarly, in using Taylor expansion arguments we often make use of the fact that

$$o(O_p(n^{-1/2})) = o_p(n^{-1/2}). \tag{14.4-16}$$

Also, if $f(x) = o(x)$ as $x \to 0$, and if $X_n = O_p(n^{-1/2})$, we would like to be sure that

$$f(X_n) = o_p(n^{-1/2}). \tag{14.4-17}$$

Furthermore, in example 14.4-3 we made uncritical use of the fact that $O_p(1)o_p(1) = o_p(1)$.

The approach to o_p and O_p discussed in this section is based on the work of Chernoff, as extended by Pratt [1959]. It requires a little attention to the formal structure that underlies a stochastic sequence; however, once this machinery has been set up, it provides a simple and mathematically rigorous way of turning nonstochastic results involving o and O into parallel stochastic results involving o_p and O_p.

Formal structure of a stochastic sequence

For each $n = 1, 2, \ldots$, let (Ω_n, P_n) be a probability space, where Ω_n is a sample space and P_n is a probability measure on the subsets of Ω_n. (The measurability of all subsets encountered is assumed but not explicitly stated.) Let ω_n be an abstract random variable taking values in Ω_n and distributed according to P_n. Next, suppose H_n is a (measurable) function mapping Ω_n into c-dimensional Euclidean space. Finally, suppose that the sequence of random vectors of interest is given by

$$\xi_n = H_n(\omega_n). \tag{14.4-18}$$

We let w_n denote the elements of Ω_n, i.e., possible values of ω_n.

Example 14.4-5 The multinomial distribution

The important examples in later sections of this chapter concern the multinomial distribution. We give the formal structure here that is used in the later sections. We let

$$\mathscr{S}_T = \left\{ \mathbf{p} = (p_1, \ldots, p_T) : p_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{T} p_i = 1 \right\}.$$

Then $\mathscr{S}_T$ is the set of all T-dimensional probability vectors. First, we set

$$\Omega_n = \mathscr{S}_T \qquad n = 1, 2, \ldots. \tag{14.4-19}$$

In $\mathscr{S}_T$, the multinomial distribution is the distribution of the vector of cell proportions rather than the vector of cell counts. For the $\mathscr{M}(n, \boldsymbol{\pi})$ distribution, the cell proportions are constrained to lie in the subset of $\mathscr{S}_T$ given by

$$T_n = \{\mathbf{p} \in \mathscr{S}_T : np_i \text{ is an integer for } i = 1, \ldots, T\}. \tag{14.4-20}$$

Hence if $\boldsymbol{\omega}_n$ has the $\mathscr{M}(n, \boldsymbol{\pi})$ distribution, then P_n is given by

$$P_n\{B\} = \sum_{\mathbf{p} \in B \cap T_n} \binom{n}{np_1, \ldots, np_T} \pi_1^{np_1} \cdots \pi_T^{np_T}, \tag{14.4-21}$$

where B is any subset of $\mathscr{S}_T$.

There are several functions of the $\boldsymbol{\omega}_n$-sequence that arise in later sections of this chapter.

1. If $H_n(\mathbf{w}_n) = n\mathbf{w}_n$ for $\mathbf{w}_n \in \mathscr{S}_T$, then $\xi_n = H_n(\boldsymbol{\omega}_n) = n\boldsymbol{\omega}_n$, so that $\{\xi_n\}$ is the ordinary sequence of multinomial distributions.
2. If

$$H_n(\mathbf{w}_n) = n(\mathbf{w}_n - \boldsymbol{\pi})\mathbf{D}_{\pi}^{-1}(\mathbf{w}_n - \boldsymbol{\pi})', \tag{14.4-22}$$

where $\mathbf{D}_{\pi}$ is the diagonal matrix based on $\boldsymbol{\pi}$, then $\{\xi_n\}$ is the sequence of chi square statistics used to test the hypothesis that $\boldsymbol{\pi}$ is the vector of true cell probabilities.

3. If $H_n(\mathbf{w}_n)$ is any value of a parameter $\boldsymbol{\theta}$ which maximizes $\pi_1(\boldsymbol{\theta})^{nw_{n1}} \cdots \pi_T(\boldsymbol{\theta})^{nw_{nT}}$, then $\xi_n = H_n(\boldsymbol{\omega}_n)$ is the maximum likelihood estimate of $\boldsymbol{\theta}$.

Chernoff–Pratt definitions of o_p and O_p

The following two theorems form the basis of the Chernoff–Pratt approach to the O_p, o_p notation. They can be viewed as providing alternative definitions of o_p and O_p to those given by definitions 14.4-2 and 14.4-4. These theorems make use of the notation developed in the previous section.

THEOREM 14.4-4 *We have $\xi_n = H_n(\omega_n) = o_p(b_n)$ if and only if for every $\eta > 0$, (measurable) subsets $S_n \subseteq \Omega_n$ can be found such that*

(i) $P_n\{\omega_n \in S_n\} \geq 1 - \eta$ *for all* n;
(ii) *if* $\{w_n\}$ *is a nonstochastic sequence such that* $w_n \in S_n$ *for every* n, *then* $x_n = H_n(w_n) = o(b_n)$.

THEOREM 14.4-5 *We have $\xi_n = H_n(\omega_n) = O_p(b_n)$ if and only if for every $\eta > 0$, (measurable) subsets $S_n \subseteq \Omega_n$ can be found such that*

(i) $P_n\{\omega_n \in S_n\} \geq 1 - \eta$ *for all* n;
(ii) *if* w_n *is a nonstochastic sequence such that* $w_n \in S_n$ *for every* n, *then* $x_n = H_n(w_n) = O(b_n)$.

See Pratt [1959] for a proof of these results. Exercises 7, 8, and 9 in Section 14.4.6 constitute proofs of theorems 14.4-4 and 14.4-5.

We note two implications of these new definitions of o_p and O_p. First, the two definitions of O_p and o_p are completely parallel. Second, the definitions separate out two distinct problems to be solved in assessing the stochastic order or ξ_n. One is stochastic, namely, finding events $S_n \subseteq \Omega_n$ with arbitrarily large P_n-probability (i.e., $\geq 1 - \eta$). The other is analytic, namely, establishing the order of magnitude of certain nonstochastic sequences $\{w_n\}$. We illustrate the power of these results in an initial application of the δ method which we discuss more extensively in Section 14.6.

Example 14.4-6 Delta method for the binomial distribution

We continue using the notation of example 14.4-4, in which we found the asymptotic distribution of the square $\hat{p}^2$ of a binomial proportion. We now consider finding the asymptotic distribution of an arbitrary function $g(\hat{p})$ of $\hat{p}$, where g has a first-order Taylor expansion about p, i.e.,

$$g(w) = g(p) + (w - p)g'(p) + o(w - p) \tag{14.4-23}$$

as $w \to p$. From (14.4-23) it follows that if w_n is any sequence of numbers such that

$$w_n = p + O(n^{-1/2}), \tag{14.4-24}$$

then

$$g(w_n) - g(p) - (w_n - p)g'(p) = o(n^{-1/2}). \tag{14.4-25}$$

Note that (14.4-24) and (14.4-25) are statements about nonstochastic sequences and that (14.4-25) can be verified by referring to the definitions and properties of the o, O notation.

Now set $\omega_n = \hat{p}$. From (14.4-13) we know that

$$\omega_n - p = O_p(n^{-1/2}). \tag{14.4-26}$$

Applying theorem 14.4-5 to (14.4-26) with

$$\xi_n^{(1)} = H_n^{(1)}(\omega_n) = \omega_n - p, \tag{14.4-27}$$

we deduce the existence of (measurable) subsets $S_n \subseteq \Omega_n$ such that

(i) $P\{\omega_n \in S_n\} \geq 1 - \eta$ for all n;

(ii) if w_n is a nonstochastic sequence such that $w_n \in S_n$ for every n, then $x^{(1)} = H_n^{(1)}(w_n) = O(n^{-1/2})$.

From (14.4-24) and (14.4-25), however, we see that these same subsets S_n also have the property that if $w_n \in S_n$ for all n, then (14.4-25) holds. If we set

$$\xi_n^{(2)} = H_n^{(2)}(\omega_n) = g(\omega_n) - g(p) - (\omega_n - p)g'(p),$$

then, applying the sufficiency of the conditions in theorem 14.4-4 to the S_n and $\xi_n^{(2)}$, we obtain

$$\xi_n^{(2)} = o_p(n^{-1/2}). \tag{14.4-28}$$

Expressing (14.4-28) in terms of $\hat{p}$ and g, we have the basic result:

$$g(\hat{p}) - g(p) = (\hat{p} - p)g'(p) + o_p(n^{-1/2}). \tag{14.4-29}$$

Multiplying both sides of (14.4-29) by $\sqrt{n}$ and setting $U_n = \sqrt{n}(\hat{p} - p)$, we have

$$\sqrt{n}(g(\hat{p}) - g(p)) = g'(p)U_n + o_p(1). \tag{14.4-30}$$

Now we can apply theorems 14.3-6 and 14.4-3 to (14.4-30), and we find that

$$\mathscr{L}[\sqrt{n}(g(\hat{p}) - g(p))] \to \mathscr{L}[V], \tag{14.4-31}$$

where V has the normal distribution $\mathcal{N}(0, (g'(p))^2 p(1-p))$. Setting $g(p) = p^2$ and differentiating, we see that the variance of the asymptotic normal distribution is $4p^3(1-p)$, which agrees with the result of example 14.4-4. ∎∎

The separation of the stochastic and analytic problems of o_p and O_p made in theorems 14.4-4 and 14.4-5 allows us to use more explicitly the O, o convention (described in (i) of Section 14.2.1) that for a finite number of values of n, the sequence may be infinite or undefined.

Pratt's theory of occurrence in probability

In this section we briefly describe a general theory in which certain aspects of a stochastic sequence "occur in probability." Occurrence in probability is a generalized notion that arises from the form of theorems 14.4-4 and 14.4-5. We use the notation of the earlier parts of this section.

We would like to describe all sequences $\{w_n\}$ such that $w_n \in \Omega_n$ for every n. Let Ω denote the set of such sequences. Then Ω can be expressed as

$$\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n \times \cdots = \mathop{\times}_{n=1}^{\infty} \Omega_n, \tag{14.4-32}$$

i.e., the infinite Cartesian product of the Ω_n. Similarly, if S_n is any subset of Ω_n, then

$$S = S_1 \times S_2 \times \cdots \times S_n \times \cdots = \mathop{\times}_{n=1}^{\infty} S_n \tag{14.4-33}$$

is a subset of Ω. There are subsets of Ω that cannot be described as in (14.4-33). When we describe the convergence properties of a sequence $\{w_n\}$, where $w_n \in \Omega_n$, we are defining a subset of Ω. Pratt [1959] gives the following general definition of occurrence in probability.

DEFINITION 14.4-5 *$S \subseteq \Omega$ occurs in probability if for every $\eta > 0$, (measurable) subsets $S_n(\eta) \subseteq \Omega_n$ can be found such that*

(i) *$P\{\omega_n \in S_n\} \geq 1 - \eta$ for every n;*

(ii) $\mathop{\times}_{n=1}^{\infty} S_n \subseteq S.$

This definition is clearly motivated by the form of theorems 14.4-4 and 14.4-5. We may easily show the relation between occurrence in probability and the O_p, o_p notation by constructing the proper subsets of Ω. We define

$$S^{(1)} = \{\{w_n\}:\ w_n \in \Omega_n \ \text{ and } \ x_n = H_n(w_n) = o(b_n)\},$$

$$S^{(2)} = \{\{w_n\}:\ w_n \in \Omega_n \ \text{ and } \ x_n H_n(w_n) = O(b_n)\}.$$

Theorem 14.4-4 is equivalent to saying that $\xi_n = o_p(b_n)$ if and only if $S^{(1)}$ occurs in probability. Theorem 14.4-5 is equivalent to saying that $\xi_n = O_p(b_n)$ if and only if $S^{(2)}$ occurs in probability.

The use of the occurrence in probability device is enhanced by the following two theorems.

THEOREM 14.4-6 *If S occurs in probability and T is another subset of Ω such that $S \subseteq T$, then T also occurs in probability.*

THEOREM 14.4-7 *Let $S^{(1)}, S^{(2)}, \ldots$ all be subsets of Ω. $S^{(1)}, S^{(2)}, \ldots$ all occur in probability if and only if their intersection $\bigcap_{j=1}^{\infty} S^{(j)}$ occurs in probability.*

Theorem 14.4-6 can be used to simplify the discussion in example 14.4-6. Theorem 14.4-7 is useful when the analytic parts of theorems 14.4-4 and 14.4-5 require the combination of a number of O, o terms. For example, a simple application of theorems 14.4-6 and 14.4-7 proves the following complicated theorem involving the simultaneous use of the O, o and O_p, o_p notation.

THEOREM 14.4-8 *Let $f_n^{(j)}(\cdot)$ $(j = 1, \ldots, J)$, $g_n^{(k)}(\cdot)$ $(k = 1, \ldots, K)$, and $h_n(\cdot)$ be functions such that if $\{x_n\}$ is a nonstochastic sequence and*

(i) $f_n^{(j)}(x_n) = O(r_n^{(j)})$ *for* $j = 1, \ldots, J$, (ii) $g_n^{(k)}(x_n) = o(s_n^{(k)})$ *for* $k = 1, \ldots, K$,

then $h_n(x_n) = O(t_n)$ (or $= o(t_n)$). Moreover, let $X_1, X_2, \ldots$ be a stochastic sequence such that

(i) $f_n^{(j)}(X_n) = O_p(r_n^{(j)})$ *for* $j = 1, \ldots, J$, (ii) $g_n^{(k)}(X_n) = o_p(s_n^{(k)})$ *for* $k = 1, \ldots, K$.

Then $h_n(X_n) = O_p(t_n)$ (or $= o_p(t_n)$).

While this result is sufficiently general to include many applications, it is generally easier in any specific case to apply theorems 14.4-6 and 14.4-7 directly.

14.4.6 *Exercises*

1. Show that o_p and O_p obey the same rules for products as do o, O.

   (P1) $O_p(a_n)O_p(b_n) = O_p(a_nb_n)$,

   (P2) $O_p(a_n)o_p(b_n) = o_p(a_nb_n)$,

   (P3) $o_p(a_n)o_p(b_n) = o_p(a_nb_n)$.

2. Show that if X_n is the sum of n independent and identically distributed random variables, each with mean μ and variance $\sigma^2 < \infty$, then $X_n = n\mu + O_p(\sqrt{n})$.
3. Prove theorem 14.4-2.
4. Prove theorem 14.4-6.
5. Prove theorem 14.4-7.
6. Show that if $Y_n = O_p(n^{-1/2})$, then $Y_n = o_p(1)$.
7. Show that $\xi_n = H_n(\omega_n) = O_p(b_n)$ if and only if for any $\eta > 0$ there is a sequence of (extended real) numbers $d_n = d_n(\eta)$ such that

   (i) $P_n\{\|\xi_n\| \leqq d_n\} \geqq 1 - \eta$ for all n;

   (ii) $d_n = O(b_n)$.

8. Show that $\xi_n = H_n(\omega_n) = O_p(b_n)$ if and only if for any $\eta > 0$ there is a sequence of Borel sets (in c-dimensional Euclidean space) $T_n = T_n(\eta)$ such that

   (i) $P_n\{\xi_n \in T_n\} \geqq 1 - \eta$ for all n;

   (ii) if $x_n \in T_n$ for every n, then $x_n = O(b_n)$.

9. Show that $\xi_n = H_n(\omega_n) = O_p(b_n)$ if and only if for any $\eta > 0$ there is a sequence of (measurable) sets $S_n = S_n(\eta) \subseteq \Omega_n$ such that

   (i) $P_n\{\omega_n \in S_n\} \geqq 1 - \eta$ for all n;

   (ii) if $w_n \in S_n$ for every n, then $x_n = H_n(w_n) = O(b_n)$.

10. Show that the results of problems 7, 8, and 9 remain true if O and O_p are replaced throughout by o and o_p.

14.5 Convergence of Moments

If $\mathscr{L}[X_n] \to \mathscr{L}[X]$, we may also want to know if

$$E(X_n^r) \to E(X^r) \tag{14.5-1}$$

for various choices of r, usually 1 and 2. It should be noted that convergence in distribution does not in general entail (14.5-1) for any value of $r \neq 0$. In most of the applications in this book, however, (14.5-1) does hold if $\mathscr{L}[X_n] \to \mathscr{L}[X]$. The remainder of this section is devoted to a short discussion of this convergence problem.

14.5.1 *Limits of moments and asymptotic moments*

If $\mathscr{L}[X_n] \to \mathscr{L}[X]$, then we refer to

$$E(X^r) \tag{14.5-2}$$

as the *asymptotic* rth moment of X_n, while

$$\lim_{n\to\infty} E(X_n^r) \tag{14.5-3}$$

is the *limit of the* rth *moment* of X_n, if it exists. Since (14.5-3) and (14.5-2) may be unequal, we have separate terminology for them. The next example gives a case when they are equal.

Example 14.5-1 Square of a binomial proportion

If X_n has the binomial distribution $\mathscr{B}(n, p)$ and $\hat{p} = n^{-1}X_n$, then from example 14.4-4 we have $\mathscr{L}[\sqrt{n}(\hat{p}^2 - p^2)] \to \mathscr{L}[V]$, where V has the normal distribution $\mathscr{N}(0, 4p^3(1-p))$. We calculate that the actual variance of $\sqrt{n}(\hat{p}^2 - p^2)$ is:

$$\begin{aligned}\operatorname{Var}(\sqrt{n}(\hat{p}^2 - p^2)) &= n\operatorname{Var}(\hat{p}^2)\\ &= 4p^3(1-p) + n^{-1}p^2(10p^2 - 16p + 6)\\ &\quad - n^{-2}p(6p^3 - 12p^2 + 7p - 1)\\ &= 4p^3(1-p) + O(n^{-1}).\end{aligned} \tag{14.5-4}$$

Hence we have

$$\lim_{n\to\infty} \operatorname{Var}(\sqrt{n}(\hat{p}^2 - p^2)) = \operatorname{Var}(V). \blacksquare\blacksquare$$

The general relationship between the limit of the variances and the asymptotic variance is given in the next theorem.

THEOREM 14.5-1 *If $\mathscr{L}[X_n] \to \mathscr{L}[X]$ and if we let $\operatorname{Var}(X_n)$ denote the variance of X_n when it exists and set it equal to $+\infty$ otherwise, then*

$$\liminf_{n\to\infty} \operatorname{Var}(X_n) \geqq \operatorname{Var}(X). \tag{14.5-5}$$

For a proof, see Zacks [1971], pp. 252–253.

From this theorem we see that the asymptotic variance of X_n can be no larger than the limit of the variances. Examples can be given to show that it can be strictly smaller (see exercise 1 of Section 14.5.3).

14.5.2 *The order of moments and o_p*

In example 14.5-1, we saw that

$$E(\hat{p}^2) = p^2 + o(1), \tag{14.5-6}$$

$$\operatorname{Var}(\hat{p}^2) = n^{-1}4p^3(1-p) + o(n^{-1}). \tag{14.5-7}$$

However, from the δ method discussed in Section 14.6 and example 14.4-4, we know that

$$\mathscr{L}[\sqrt{n}((\hat{p})^2 - p^2)] \to \mathscr{N}(0, 4p^3(1-p)).$$