Problems for Topic 14

Problem 14.1 Suppose $X_0, X_1, \ldots$ is an iid sequence of Bernoulli trials with success probability $p$. Suppose $X_i$ is the indicator of your team’s success on rally $i$ in a volleyball game. (Note: This is a completely unrealistic model, since the serving team is always at a disadvantage when evenly matched teams play.) Your team scores a point each time it has a success that follows another success. Let $S_n = \sum_{i=1}^{n} X_i - 1$ denote the number of points your team scores by time $n$.

(a) Find the asymptotic distribution of $S_n$.

(b) Simulate a sequence $X_0, X_1, \ldots, X_{1000}$ as above and calculate $S_{1000}$ for $p = .4$. Repeat this process 100 times, then graph the empirical distribution of $S_{1000}$ obtained from simulation on the same axes as the theoretical asymptotic distribution from (a). Comment on your results.

Problem 14.2 Suppose $Z_0, Z_1, \ldots$ are iid with mean 0 and variance $\sigma^2$, and let

$$X_{i+1} = \xi + \beta (X_i - \xi) + (Z_i - 1)$$

for $i > 1$, where $\beta$ is a constant with $|\beta| < 1$. Assume that $\text{Cov}(X_i, Z_j) = 0$ for all $j \geq i$.

(a) Find $\text{Var} X_1$ in terms of $\sigma^2$ and $\beta$ if $X_1, X_2, \ldots$ is a stationary sequence.

(b) If the $X_i$ are normally distributed in addition to being a stationary sequence, find $\tau^2$ such that

$$\sqrt{n}(\bar{X}_n - \xi) \xrightarrow{L} N(0, \tau^2).$$

Problem 14.3 Let $X_0, X_1, \ldots$ be iid random variables from a continuous distribution $F(x)$. Define $Y_i = I\{X_i < X_{i-1} \text{ and } X_i < X_{i+1}\}$. Thus, $Y_i$ is the indicator that $X_i$ is a relative minimum. Let $S_n = \sum_{i=1}^{n} Y_i$.

(a) Find the asymptotic distribution of $S_n$.

(b) For a sample of size 5000 from the uniform $(0, 1)$ random number generator in Splus (or whatever language you use), compute an approximate p-value based on the observed value of $S_n$ and the answer to part (a). The null hypothesis is that the sequence is iid, of course.

Topic 15. Multivariate notions of convergence

Lehmann §5.1; Ferguson §1

We now consider vectors $x$ and random vectors $X$ in $R^k$, $k > 1$. Note that we adopt Lehmann’s convention of underlining vectors, a practice that is not always used in the literature. It should be clear from context and definitions when a variable is vector-valued, but nonetheless the adoption of special notation can be helpful when these ideas are being learned for the first time.

We must first define a norm on $R^k$. We are interested primarily in the norm of a vector going to zero, a
concept for which any norm will suffice, so we may as well take the Euclidean norm:

$$\|\mathbf{x}\|\stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{k} x_i^2}$$

Since subscripts now denote the components of a vector, we no longer refer to sequences of real vectors as, say, \(x_1, x_2, \ldots\); instead, we use superscripts in parentheses to denote sequences of vectors, as in 

\[x^{(1)}, x^{(2)}, \ldots \quad \text{and} \quad X^{(1)}, X^{(2)}, \ldots \]

A random vector \(X\) still has a cdf:

$$F(x) = P(X_1 \leq x_1, \ldots, X_k \leq x_k).$$

As in the univariate case, we must define a continuity point of a cdf. First, we generalize the definition of continuity given in Definition 5.4.

**Definition 15.1** A function \(f: \mathbb{R}^k \rightarrow \mathbb{R}^\ell\) is continuous at \(a \in \mathbb{R}^k\) if \(f(x) \rightarrow f(a)\) as \(x \rightarrow a\). In other words, for every \(\epsilon > 0\) there exists a \(\delta > 0\) such that

$$\|f(x) - f(a)\|_\ell < \epsilon \quad \text{whenever} \quad \|x - a\|_k < \delta. \quad (28)$$

Notice that in expression (28), the subscripts on the norms indicate the dimension of the space. We usually omit these subscripts, since it is clear from the context what the correct dimension is.

Many of the concepts relating to univariate sequences transfer to the multivariate situation with little modification:

**Definition 15.2** \(x^{(n)} \Rightarrow x\) means that \(\|x^{(n)} - x\| \rightarrow 0\).

**Definition 15.3** \(X^{(n)} \xrightarrow{P} X\) means that \(P(\|X^{(n)} - X\| > \epsilon) \rightarrow 0\).

**Definition 15.4** \(X^{(n)} \xrightarrow{\text{qm}} X\) means that \(E\left\{\left(\left(X^{(n)} - X\right)^t \left(X^{(n)} - X\right)\right)\right\} \rightarrow 0\).

**Definition 15.5** Assuming that \(F_n(x)\) and \(F(x)\) denote the cumulative distribution functions of \(X^{(n)}\) and \(X\), respectively, \(X^{(n)} \xrightarrow{L} X\) means that \(F_n(a) \rightarrow F(a)\) for all continuity points \(a\) of \(F\).

**Theorem 15.1** Generalization of Theorem 9.2: If \(X^{(n)} \xrightarrow{L} X \in \mathbb{R}^k\) and \(Y^{(n)} \xrightarrow{P} c \in \mathbb{R}^\ell\), then

$$f\left(X^{(n)}\right) \xrightarrow{L} f\left(X\right)$$

for any continuous function \(f: \mathbb{R}^{k+\ell} \rightarrow \mathbb{R}^m\). Expression (29) will sometimes be written

$$f\left(X^{(n)}, Y^{(n)}\right) \xrightarrow{L} f\left(X, c\right).$$

Since \(f(x) = x\) is a continuous function, we obtain the following corollary of Theorem 15.1, which says that multivariate convergence in distribution implies the univariate convergence of each of the marginals:

**Corollary 15.1** If \(X^{(n)} \xrightarrow{L} X\), then \(X_i^{(n)} \xrightarrow{L} X_i\) for all \(i, 1 \leq i \leq k\).

Furthermore, it is obvious from Theorem 15.1 that \(c^t X^{(n)} \xrightarrow{L} c^t X\) for any constant vector \(c\). However, the converse is also true:

36
Theorem 15.2 \(X^{(n)} \xrightarrow{P} X\) if and only if \(x^tX^{(n)} \xrightarrow{P} c^tx\) for all \(c \in \mathbb{R}^k\).

Theorem 15.1 may be strengthened if \(X^{(n)}\) is known to be independent of \(Y^{(n)}\), as follows:

**Theorem 15.3** If \(X^{(n)} \xrightarrow{P} X\) and \(Y^{(n)} \xrightarrow{P} Y\), where \(X^{(n)}\) is independent of \(Y^{(n)}\), then for a continuous function \(f\),

\[
f(X^{(n)}, Y^{(n)}) \xrightarrow{P} f(X, Y),
\]

where \(X\) and \(Y\) are taken to be independent.

The proof is easy, since the independence of random vectors implies that their joint cdf may be factored and that \((a, b)\) is a continuity point of the joint distribution if and only if \(a\) is a continuity point of \(X\) and \(b\) is a continuity point of \(Y\). Thus, for continuity points \(a\) of \(X\) and \(b\) of \(Y\),

\[
F_{(X^{(n)}, Y^{(n)})}(a, b) = F_{X^{(n)}}(a)F_{Y^{(n)}}(b) \rightarrow F_X(a)F_Y(b) = F_{(X, Y)}(a, b).
\]

Therefore, \((X^{(n)}, Y^{(n)}) \xrightarrow{P} (X, Y)\), and so \(f(X^{(n)}, Y^{(n)}) \xrightarrow{P} f(X, Y)\) follows by Theorem 15.1.

The delta method may be extended to the multivariate case as well. However, to do this will require the notion of a derivative of a function that maps \(\mathbb{R}^k\) into \(\mathbb{R}^\ell\).

**Definition 15.6** Write \(g : \mathbb{R}^k \rightarrow \mathbb{R}^\ell\) as \(g(x) = [g_1(x), \ldots, g_\ell(x)]^t\) and suppose that for \(i = 1, \ldots, \ell\), \(g_i(x)\) has partial derivatives with respect to \(x_1, \ldots, x_k\). Then the derivative of \(g(x)\) at \(a\), denoted \(\dot{g}(a)\), is the \(\ell \times k\) matrix

\[
\left(\begin{array}{ccc}
\frac{\partial g_1(a)}{\partial x_1} & \cdots & \frac{\partial g_1(a)}{\partial x_k} \\
\vdots & \ddots & \vdots \\
\frac{\partial g_\ell(a)}{\partial x_1} & \cdots & \frac{\partial g_\ell(a)}{\partial x_k}
\end{array}\right)
\]

With this concept of a derivative now in hand, we may now obtain the multivariate version of the delta method. This method hinges on a multivariate version of Taylor’s theorem:

**Theorem 15.4** If \(f : \mathbb{R}^k \rightarrow \mathbb{R}^\ell\) is differentiable at the point \(x\), then as \(\Delta \rightarrow 0\),

\[
f(x + \Delta) = f(x) + \dot{f}(x)\Delta + o(\|\Delta\|)1_\ell,
\]

where \(1_\ell\) denotes the \(\ell\)-vector of all ones.

It is straightforward using Theorem 15.4 to prove the following theorem:

**Theorem 15.5** Multivariate delta method: If \(g : \mathbb{R}^k \rightarrow \mathbb{R}^\ell\) has a derivative \(\dot{g}(a)\) at \(a \in \mathbb{R}^k\) and

\[
r^k \left(X^{(n)} - a\right) \xrightarrow{P} Y
\]

for some \(k\)-vector \(Y\) and some sequence \(X^{(1)}, X^{(2)}, \ldots\) of \(k\)-vectors, where \(b > 0\), then

\[
r^b \left\{g \left(X^{(n)}\right) - g(a)\right\} \xrightarrow{P} \dot{g}(a)Y.
\]

The proof of Theorem 15.5 involves a simple application of the multivariate version of Taylor’s theorem (see Theorem 5.2.2 on page 295 of Lehmann). To use the delta method, it will often be necessary to take advantage of the following well-known theorem:

**Theorem 15.6** If \(Y\) is any random \(\ell\)-vector with mean \(\mu\) and covariance matrix \(\Sigma\), then for any \(k \times \ell\) matrix \(M\) (of constants),

\[
E( MY) = M\mu \quad \text{and} \quad \text{Var} (MY) = M\Sigma M^t.
\]
Example 15.1 Two-sample z statistic: Here we prove the asymptotic normality of the statistic

\[ Z = \frac{(\bar{Y}_n - \bar{X}_m) - (\eta - \xi)}{\sqrt{\frac{\tau^2}{m} + \frac{\sigma^2}{n}}} , \]

where \(X_1, \ldots, X_m\) are iid with mean \(\xi\) and variance \(\sigma^2 < \infty\) and \(Y_1, \ldots, Y_n\) are iid with mean \(\eta\) and variance \(\tau^2 < \infty\). The two samples are independent.

The (univariate) CLT gives

\[ \sqrt{n}(\bar{Y}_n - \eta) \xrightarrow{L} N(0, \tau^2) \quad \text{and} \quad \sqrt{n}(\bar{X}_n - \xi) \xrightarrow{L} N(0, \sigma^2) \]

as \(n \to \infty\) and \(m \to \infty\). By Theorem 15.3,

\[ \left( \frac{\sqrt{n}(\bar{Y}_n - \eta)}{\sqrt{n}(\bar{X}_n - \xi)} \right) \xrightarrow{L} \left( \frac{\tau Z_1}{\sigma Z_2} \right) , \]

where \(Z_1\) and \(Z_2\) are independent standard normal random variables. We make only one additional assumption, which is that there exists \(\lambda \in (0, 1)\) such that \(m/(m+n) \to \lambda\). This gives

\[ \left( \frac{\sqrt{m+n}(\bar{Y}_n - \eta)}{\sqrt{m+n}(\bar{X}_n - \xi)} \right) \xrightarrow{L} \left( \frac{\tau Z_1/\sqrt{1-\lambda}}{\sigma Z_2/\sqrt{\lambda}} \right) , \]

which in turn implies

\[ \sqrt{m+n} \left\{ (\bar{Y}_n - \bar{X}_m) - (\eta - \xi) \right\} \xrightarrow{L} \frac{\tau Z_1}{\sqrt{1-\lambda}} - \frac{\sigma Z_2}{\sqrt{\lambda}} . \]

The random variable on the right side above has a \(N \left\{ 0, \tau^2/(1-\lambda) + \sigma^2/\lambda \right\} \) distribution. In other words,

\[ \sqrt{\frac{m+n}{\tau^2 + \sigma^2}} \left\{ (\bar{Y}_n - \bar{X}_m) - (\eta - \xi) \right\} \xrightarrow{L} N(0,1) . \]

Finally, one may easily verify that

\[ \sqrt{\frac{m+n}{\tau^2 + \sigma^2}} \sim \frac{1}{\sqrt{\frac{\tau^2}{m} + \frac{\sigma^2}{n}}} \]

so applying Slutsky’s theorem (the univariate version) gives

\[ Z = \frac{(\bar{Y}_n - \bar{X}_m) - (\eta - \xi)}{\sqrt{\frac{\tau^2}{m} + \frac{\sigma^2}{n}}} \xrightarrow{L} N(0,1) . \]

Notice in the previous example that multivariate techniques were employed to determine the univariate distribution of the random variable \(Z\). This is a common trick: We often obtain univariate results using multivariate techniques.

Problems for Topic 15

Problem 15.1 For a distribution function \(F(x)\), prove or disprove the following statement: A continuity point is a point \(a\) such that \(P(X = a) = 0\).
Problem 15.2  To illustrate a situation that can arise in the multivariate setting that cannot arise
in the univariate setting, construct an example of a sequence \((X_n, Y_n)\), a joint distribution
\((X, Y)\), and a connected subset \(S \subseteq \mathbb{R}^2\) such that

(i) \((X_n, Y_n) \overset{d}{\rightarrow} (X, Y)\);
(ii) every point of \(\mathbb{R}^2\) is a continuity point of the distribution function of \((X, Y)\);
(iii) \(P[(X_n, Y_n) \in S]\) does not converge to \(P[(X, Y) \in S]\).

Hint: Condition (ii) may be satisfied even if the distribution of \((X, Y)\) is concentrated on a
line.

Problem 15.3 (a) Prove that if \(f(x)\) is continuous at \(0\), then \(f_i(t)\) is continuous at \(0\) for each \(i\),
where

\[ f_i(t) \overset{\text{def}}{=} f(te_i^{(i)}) \]

and \(e_i^{(i)}\) is the \(i\)th standard basis vector (i.e., \(te_i^{(i)}\) is the vector with \(t\) in the \(i\)th component and
zeros elsewhere).

(b) Prove that the converse of (a) is not true by inventing a function \(f(x)\) that is not continuous
at \(0\) but such that \(f_i(t)\) is continuous at \(0\) for each \(i\).

---

### Topic 16. Joint distributions of extreme order
statistics

*Lehmann §5.1; Ferguson §15*

In Example 10.1, we derived the asymptotic distribution of the maximum from a random sample from a
uniform distribution. We did this using only the definition of convergence in distribution without relying on
any results other than the fact that

\[ \left(1 + \frac{c_n}{b_n}\right)^{b_n} \rightarrow e^c \quad (30) \]

if \(c_n \rightarrow c\) and \(b_n \rightarrow \infty\). In a similar way, we may derive the joint asymptotic distribution of several order
statistics, as seen in the following example.

**Example 16.1 Range of uniform sample:** Let \(X_1, \ldots, X_n\) be iid from Uniform(0, 1). Let \(R_n =
X_{(n)} - X_{(1)}\) denote the range of the sample. What is the asymptotic distribution of \(R_n\)?

We begin to answer this question by finding the joint asymptotic distribution of \((X_{(n)}, X_{(1)})\),
as follows. For certain sequences \(k_n\) and \(\ell_n\), as yet unspecified, consider

\[
P(k_nX_{(1)} > x \text{ and } \ell_n(1 - X_{(n)}) > y) = P(X_{(1)} > x/k_n \text{ and } X_{(n)} < 1 - y/\ell_n) =
P(x/k_n < X_{(1)} < \cdots < X_{(n)} < 1 - y/\ell_n),
\]