Suppose that \( X_1, X_2, \ldots \) is a stationary sequence. Recall that stationary is a stronger condition than identically distributed. Let \( \theta = E X_i \) (Lehmann switches from \( \xi \) to \( \theta \) here and we’ll do the same) and \( \sigma^2 = \text{Var} X_i \).

We have already seen, in equation (12), the expression for the variance of \( X_n \) for a stationary distribution. Letting \( \gamma_k = \text{Cov} (X_1, X_{1+k}) \), we conclude that

\[
\text{Var} \left\{ \sqrt{n} (X_n - \theta) \right\} = \sigma^2 + \frac{2}{n} \sum_{k=1}^{n-1} (n-k) \gamma_k. \tag{27}
\]

Suppose that

\[
\frac{2}{n} \sum_{k=1}^{n-1} (n-k) \gamma_k \to \gamma \tag{28}
\]

as \( n \to \infty \). Then based on equation (27), it seems reasonable to ask whether

\[
\sqrt{n} (X_n - \theta) \overset{\mathcal{L}}{\to} N(0, \sigma^2 + \gamma).
\]

The answer, in many cases, is yes. This topic explores some of these cases.

We first consider \( m \)-dependent sequences. Recall that \( X_1, X_2, \ldots \) is \( m \)-dependent for some \( m \geq 0 \) if the vector \( (X_1, \ldots, X_i) \) is independent of \( (X_{i+j}, X_{i+j+1}, \ldots) \) whenever \( j > m \). Therefore, for an \( m \)-dependent sequence we have \( \gamma_k = 0 \) for all \( k > m \), so equation (28) becomes (for \( n > m \))

\[
\frac{2}{n} \sum_{k=1}^{m} (n-k) \gamma_k \to 2 \sum_{k=1}^{m} \gamma_k.
\]

This leads to Theorem 15.1, which we will not prove rigorously but whose proof will be sketched.

**Theorem 15.1** If for some \( m \geq 0 \), \( X_1, X_2, \ldots \) is a stationary \( m \)-dependent sequence with \( E X_i = \theta \) and \( \text{Var} X_i = \sigma^2 < \infty \), then

\[
\sqrt{n} (X_n - \theta) \overset{\mathcal{L}}{\to} N(0, \tau^2),
\]

where

\[
\tau^2 = \sigma^2 + 2 \sum_{k=1}^{m} \text{Cov} (X_1, X_{1+k}).
\]

Although there are some technical details that make the proof a bit tedious—see Ferguson, Section 11 for details—we may sketch the idea of the proof as follows.

Let \( N = N(n) \) be an integer that goes to \( \infty \) as \( n \to \infty \), but at a slower rate than \( n \) so that \( n/N \to \infty \). Then \( \sum_{i=1}^{n} X_i \) may be broken into parts as follows:

\[
\sum_{i=1}^{n} X_i = (X_1 + \cdots + X_N) + (X_{N+1} + \cdots + X_{N+m}) + (X_{N+m+1} + \cdots + X_{2N+m}) + (X_{2N+m+1} + \cdots + X_{2N+2m}) + \cdots
\]

\[
= A_1 + B_1 + A_2 + B_2 + \cdots,
\]
where the $A_i$ each consist of $N$ terms and the $B_i$ each consist of $m$ terms. Thus, we obtain for $n = r(N + m)$

$$\sqrt{n}(\bar{X}_n - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{r} (A_i - \theta) + \frac{1}{\sqrt{n}} \sum_{i=1}^{r} (B_i - \theta).$$

The key is that the $A_i$ are iid, so a CLT-like result applies. On the other hand, the $B_i$ have an asymptotically negligible contribution to the sum. Lehmann, on page 109, points out the technical difficulty in completing the proof.

**Example 15.1** Runs of successes: Suppose $X_1, X_2, \ldots$ are iid Bernoulli($p$) variables. Let $T_n$ denote the number of runs of successes in $X_1, \ldots, X_n$, where a run of successes is defined as a sequence of consecutive $X_i$, all of which equal 1, that is both preceded and followed by zeros (unless the run begins with $X_1$ or ends with $X_n$). What is the asymptotic distribution of $T_n$?

We note that

$$T_n = \sum_{i=1}^{n} I\{\text{run starts at } i\text{th position}\}$$

$$= X_1 + \sum_{i=2}^{n} X_i(1 - X_{i-1}),$$

since a run starts at the $i$th position for $i > 1$ if and only if $X_i = 1$ and $X_{i-1} = 0$.

Letting $Y_i = X_{i+1}(1 - X_i)$, we see immediately that $Y_1, Y_2, \ldots$ is a stationary 1-dependent sequence with $E Y_i = p(1 - p)$, so that by Theorem 15.1, $\sqrt{n} \{ \bar{Y}_n - p(1 - p) \} \xrightarrow{D} N(0, \tau^2)$, where

$$\tau^2 = \text{Var} Y_1 + 2 \text{Cov} (Y_1, Y_2)$$

$$= E Y_1^2 - (E Y_1)^2 + 2 E Y_1 Y_2 - 2(E Y_1)^2$$

$$= E Y_1 - 3(E Y_1)^2 = p(1 - p) - 3p^2(1 - p)^2.$$

Since

$$\frac{T_n - np(1 - p)}{\sqrt{n}} = \sqrt{n} \{ \bar{Y}_n - p(1 - p) \} + \frac{X_1 - Y_n}{\sqrt{n}},$$

we conclude that

$$\frac{T_n - np(1 - p)}{\sqrt{n}} \xrightarrow{D} N(0, \tau^2).$$

We conclude this topic by considering a case in which asymptotic normality of $\bar{X}_n$ holds even though $\gamma_k \neq 0$ for all $k$.

**Example 15.2** First-order stationary autoregressive process: Suppose $U_1, U_2, \ldots$ are iid $N(0, \sigma_0^2)$. We’ll define, for $i \geq 1$,

$$X_{i+1} = \theta + \beta(X_i - \theta) + U_{i+1}$$

for some $\beta$ with $|\beta| < 1$. Let $X_1 \sim N(\theta, \sigma^2)$. Equation (29) gives an autoregressive sequence $X_1, X_2, \ldots$. Suppose we would also like to ensure that this sequence is stationary. Since this means $\text{Var} X_i = \sigma^2$ for all $i$, and equation (29) gives $\text{Var} X_{i+1} = \beta^2 \sigma^2 + \sigma_0^2$, we conclude that stationarity implies

$$\sigma^2 = \frac{\sigma_0^2}{1 - \beta^2}.$$
We may rewrite equation (29) to read
\[ X_{k+1} - \theta = \beta^k (X_1 - \theta) + \beta^{k-1} U_2 + \cdots + \beta U_k + U_{k+1}. \]
In this form, it is easy to see that \( \text{Cov} (X_1, X_{1+k}) = \beta^k \sigma^2 \). Therefore,
\[
\text{Var} \{ \sqrt{n} (X_n - \theta) \} = \sigma^2 + \frac{2}{n} \sum_{k=1}^{n-1} (n-k) \beta^k \sigma^2 = \sigma^2 \left\{ 1 + 2 \sum_{k=1}^{n-1} \beta_k - \frac{2}{n} \sum_{k=1}^{n-1} k \beta^k \right\}.
\]

Since \( \sum_{k=1}^{n-1} k \beta^k = \frac{\beta (1 + (n-1) \beta^n - n \beta^{n-1})}{(1-\beta)^2} \),
we see that \( \frac{2}{n} \sum_{k=1}^{n-1} k \beta^k \to 0 \). Thus,
\[
\text{Var} \{ \sqrt{n} (X_n - \theta) \} \to \sigma^2 \left( 1 + \frac{2 \beta}{1-\beta} \right) = \frac{1 + \beta}{1-\beta} \sigma^2.
\]

Since \( X_n \) is normal by assumption, this implies that
\[
\sqrt{n} (X_n - \theta) \xrightarrow{D} N \left( 0, \frac{1 + \beta}{1-\beta} \sigma^2 \right)
\]

by Lemma 15.1.

**Lemma 15.1** If \( X_n \sim N(0, \sigma_n^2) \) and \( \sigma_n^2 \to \sigma^2 \) where \( 0 < \sigma^2 < \infty \), then \( X_n \xrightarrow{D} N(0, \sigma^2) \).

The proof of Lemma 15.1 is immediate, since
\[
P(X_n \leq x) = \Phi(x/\sigma_n) \to \Phi(x/\sigma).
\]

**Problems**

**Problem 15.1** 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 an disadvantage when evenly matched teams play.) Your team scores a point each time it has a success that follows another success. Let \( S_n = \sum_{i=1}^{n} X_{i-1} X_i \) 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 15.2** Suppose \( Z_0, Z_1, \ldots \) are iid with mean 0 and variance \( \sigma^2 \), and let
\[ X_{i+1} = \xi + \beta (X_i - \xi) + (Z_{i-1} - Z_i) \]
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
\[
\sqrt{n} (\bar{X}_n - \xi) \xrightarrow{D} N(0, \tau^2).
\]
Problem 15.3  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.