Assignment 8

Exercise 4.7  Use the Cramér-Wold Theorem along with the univariate Central Limit Theorem (from Example 2.12) to prove Theorem 4.9.

Sketch of solution: This proof is actually given in the course notes, just before Exercise 4.8: Let \( X \sim N_k(0, \Sigma) \) and take any vector \( a \in \mathbb{R}^k \). We wish to show that 
\[
a^\top \left[ \sqrt{n} \left( X_n - \mu \right) \right] \xrightarrow{d} a^\top X.
\]
But this follows immediately from the univariate Central Limit Theorem, since \( a^\top (X_1 - \mu) \), \( a^\top (X_2 - \mu) \), \ldots are independent and identically distributed with mean 0 and variance \( a^\top \Sigma a \).

Exercise 4.11  In Example 4.18, we show that \( np_n(1 - p_n) \to \infty \) is a sufficient condition for (4.15) to hold. Prove that it is also a necessary condition. You may assume that \( p_n(1 - p_n) \) is always nonzero.

Hint: Use the Lindeberg-Feller Theorem.

Sketch of solution: In this case, \( s_n^2 = np_n(1 - p_n) \) and \( \sigma_{ni}^2 = p_n(1 - p_n) \) for all \( n \) and \( i \). Therefore, Condition (4.14) is satisfied. Thus, the Lindeberg-Feller Theorem tells us that (4.15) implies that the Lindeberg condition is satisfied, so it suffices to show that \( np_n(1 - p_n) \to \infty \) is necessary for the Lindeberg condition. In this example, the Lindeberg condition (4.16) reduces to
\[
\mathbb{E} \left( Y_{n1}^2 I \left\{ |Y_{n1}| \geq \epsilon \sqrt{np_n(1 - p_n)} \right\} \right) \xrightarrow{p_n(1 - p_n)} 0
\]
for all \( \epsilon > 0 \). If \( np_n(1 - p_n) \not\to \infty \), there exists \( M \) such that \( np_n(1 - p_n) < M \) for infinitely many \( n \). But this means there exists \( \epsilon > 0 \) such that \( \epsilon \sqrt{np_n(1 - p_n)} < 1/2 \) infinitely often. When \( \epsilon \sqrt{np_n(1 - p_n)} < 1/2 \),
\[
Y_{n1}^2 I\{|Y_{n1}| \geq \epsilon \sqrt{np_n(1 - p_n)} \} \geq Y_{n1}^2 I\{|Y_{n1}| \geq 1/2 \} \geq \frac{1}{2} Y_{n1} |I\{|Y_{n1}| \geq 1/2 \}|
\]
Since either \( p_n \geq 1/2 \) or \( 1 - p_n \geq 1/2 \), \( \mathbb{E} \left[ \frac{1}{2} |Y_{n1}| I\{|Y_{n1}| \geq 1/2 \} \right] \geq \frac{1}{2} p_n(1 - p_n) \). We conclude that whenever \( \epsilon \sqrt{np_n(1 - p_n)} < 1/2 \),
\[
\mathbb{E} \left( Y_{n1}^2 I \left\{ |Y_{n1}| \geq \epsilon \sqrt{np_n(1 - p_n)} \right\} \right) \geq \frac{1}{2} \frac{1}{2}.
\]
So if $\epsilon \sqrt{np_n(1-p_n)} < 1/2$ infinitely often, then the Lindeberg condition cannot be satisfied. This concludes the proof.

**Exercise 4.20**  Suppose $X_0, X_1, \ldots$ is an independent 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. 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 $S_n.$

**Sketch of solution:** If $Y_i = X_{i-1}X_i,$ then $Y_1, Y_2, \ldots$ is a stationary 1-dependent sequence. Furthermore,

\[
\begin{align*}
\mathbb{E} Y_i &= p^2, \\
\text{Var } Y_i &= p^2(1-p^2), \\
\text{Cov } (Y_1, Y_2) &= \mathbb{E}(Y_1Y_2) - p^4 = p^3 - p^4.
\end{align*}
\]

We conclude that $(1/\sqrt{n})(S_n - np) \xrightarrow{d} N[0, p^2(1-p^2) + 2p^3(1-p)],$ which may be rewritten as $N[0, p^2(1-p)(3p+1)].$

(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.

**Sketch of solution:** From part (a), we conclude that $S_n$ is approximately distributed as $N(np^2, np^2(1-p)(3p+1)).$ With $n = 1000$ and $p = 0.4,$ this becomes $N(160, 211.2).$ The plot below shows that the empirical distribution is quite close to the theoretical distribution.

```r
f <- function(x) sum(x[-1] * x[-length(x)]) # This will find S_n
out <- replicate(100, f(rbinom(1001,1,.4))) # 100 trials on samples of n=1001
plot(ecdf(out))
xx <- seq(110, 210, len=200)
lines(xx, pnorm(xx, mean=160, sd=sqrt(211.2)), lwd=3, lty=2)
```
Exercise 5.1 Let $\delta_n$ be defined as in Example 5.4. Find the asymptotic distribution of $\delta_n$ in the case $p = 1/2$. That is, find constant sequences $a_n$ and $b_n$ and a nontrivial random variable $X$ such that $a_n(\delta_n - b_n) \xrightarrow{d} X$.

**Hint:** Let $Y_n = X_n - (n/2)$. Apply the central limit theorem to $Y_n$, then transform both sides of the resulting limit statement so that a statement involving $\delta_n$ results.

**Sketch of solution:** In the case $p = 1/2$, we know that

$$2\sqrt{n} \left( \frac{X_n}{n} - \frac{1}{2} \right) \xrightarrow{d} N(0, 1)$$

by the central limit theorem. Squaring, we obtain

$$4n \left( \frac{X_n^2}{n^2} - \frac{X_n}{n} + \frac{1}{4} \right) \xrightarrow{d} \chi^2_1.$$

A bit of algebra then yields

$$-4n \left[ \frac{X_n}{n} \left( \frac{n - X_n}{n} \right) - \frac{1}{4} \right] = -4n \left( \delta_n - \frac{1}{4} \right) \xrightarrow{d} \chi^2_1.$$

Notice the negative sign in this case, which is sensible because $\delta_n$ is of the form $\hat{p}(1 - \hat{p})$ and therefore cannot possibly be larger than $1/4$. 

245
**Exercise 5.3** Suppose \(X_n \sim \text{binomial}(n, p)\), where \(0 < p < 1\).

(a) Find the asymptotic distribution of \(g(X_n/n) - g(p)\), where \(g(x) = \min\{x, 1-x\}\).

**Sketch of solution:** By the central limit theorem, we know that
\[
\sqrt{n}\left(\frac{X_n}{n} - p\right) \xrightarrow{d} N(0, p - p^2).
\] (5.16)

On the interval \((0, 1/2)\), we have \(g(p) = p\); on the interval \((1/2, 1)\), we have \(g(p) = 1-p\). In either case, \(g(p)\) is differentiable with \([g'(p)]^2 = 1\) and we obtain
\[
\sqrt{n}\left[g\left(\frac{X_n}{n}\right) - g(p)\right] \xrightarrow{d} N(0, p - p^2)
\]
immediately by the delta method. However, in the case \(p = 1/2\), the delta method does not apply because \(g(p)\) is not differentiable at \(p = 1/2\). Instead, we find that
\[
\sqrt{n}\left[g\left(\frac{X_n}{n}\right) - g\left(\frac{1}{2}\right)\right] = \sqrt{n}\left|\frac{X_n}{n} - \frac{1}{2}\right|.
\]
Therefore, since the absolute value is a continuous function, equation (5.16) implies that for \(p = 1/2\) we obtain
\[
\sqrt{n}\left[g\left(\frac{X_n}{n}\right) - g(p)\right] \xrightarrow{d} -\sqrt{p - p^2}|Z|,
\]
where \(Z\) is a standard normal random variable.

(b) Show that \(h(x) = \sin^{-1}(\sqrt{x})\) is a variance-stabilizing transformation for \(X_n/n\). This is called the **arcsine transformation** of a sample proportion.

**Hint:** \((d/du)\sin^{-1}(u) = 1/\sqrt{1 - u^2}\).

**Sketch of solution:** If \(h(x) = \sin^{-1}(\sqrt{x})\), then \(h'(x) = [x(1-x)]^{-1/2}/2\) by the chain rule. Therefore, starting from equation (5.16), the delta method implies that
\[
\sqrt{n}\left[h\left(\frac{X_n}{n}\right) - h(p)\right] \xrightarrow{d} N\left(0, \frac{1}{4}\right)
\]
because \([h'(p)]^2[p - p^2] = 1/4\).