Problem 1. [2 points] If $c$ is a constant such that $X_n \xrightarrow{L} c$, prove that $X_n \xrightarrow{P} c$.

By definition, $P(X_n \leq c + \epsilon) \to 1$ (which means $P(X_n > c + \epsilon) \to 0$) and $P(X_n \leq c - \epsilon) \to 0$ for all $\epsilon > 0$ since $c + \epsilon$ and $c - \epsilon$ are continuity points of the distribution constant at $c$. Thus,

\[
P(|X_n - c| \geq \epsilon) = P(X_n \leq c - \epsilon) + P(X_n \geq c + \epsilon) \to 0
\]

and $X_n \xrightarrow{P} c$ by definition.

Problem 2. [3 points] Suppose $X_1, X_2, \ldots$ are iid Bernoulli($p$) random variables. Let $S_n = \sum_{i=2}^{n} X_i X_{i-1}$. Find (with justification) the asymptotic distribution of $S_n/n$.

For $i = 2, 3, \ldots$, let $Y_{i-1} = X_i X_{i-1}$. Then $Y_1, Y_2, \ldots$ are a 1-dependent sequence with $E Y_i = p^2$ and $\text{Var} Y_i = p^2(1 - p^2)$ and $\text{Cov}(Y_i, Y_{i+1}) = p^3 - p^4$. The CLT for m-dependent sequences says that $\sqrt{n} (Y_n - E Y_i) \xrightarrow{L} N[0, \text{Var} Y_i + 2 \text{Cov}(Y_i, Y_{i+1})]$ so in this case

\[
\sqrt{n} (Y_n - p^2) \xrightarrow{L} N(0, p^2 + 2p^3 - 3p^4).
\]

Since $S_n/n \sim Y_n$,

\[
\sqrt{n} \left( \frac{S_n}{n} - p^2 \right) \xrightarrow{L} N(0, p^2 - 2p^3 + 3p^4).
\]

Problem 3. Suppose $X_1, \ldots, X_n$ are iid from Beta($\theta, 1$). That is, $f_\theta(x) = \theta x^{\theta-1} I\{0 < x < 1\}$.

(a) [3 points] Find the MLE of $\theta$ and its asymptotic distribution. You may assume without justifying it that the beta density is sufficiently regular, though you should justify any other assumptions you make.

We obtain

\[
l_n(\theta) = n \log \theta + (\theta - 1) \sum_{i=1}^{n} \log x_i,
\]

\[
l'_n(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i,
\]

\[
l''_n(\theta) = -\frac{n}{\theta^2}.
\]

This last equation gives $I(\theta) = 1/\theta^2$. There is clearly a unique solution to the likelihood equation, namely $\hat{\theta} = -n/\sum_{i=1}^{n} \log x_i$. Since there must exist a consistent root of the likelihood equation and that root must be efficient, we conclude that $\hat{\theta}$ is efficient. Therefore,

\[
\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{L} N(0, \theta^2).
\]
Problem 3 continued. Recall that $X_1, \ldots, X_n$ are iid Beta($\theta, 1$).

(b) [2 points] If we define $Y_i = -\log X_i$ for $i = 1, \ldots, n$, the Rao scores test of $H_0 : \theta = 2$ against $H_1 : \theta \neq 2$ at level $\alpha = .05$ rejects $H_0$ when

$$ |Y_n - c_1| > \frac{c_2}{\sqrt{n}}. $$

Find $c_1$ and $c_2$. [Use $u_{.025} = 1.96$.]

From part (a), $I(\theta) = 1/\theta^2$ and $I_n'(\theta) = (n/\theta) - n\bar{Y}_n$. Rao’s test rejects when $|R_n| > u_{.025} = 1.96$, where

$$ R_n = \frac{1}{\sqrt{nI(\theta_0)}} I_n'(\theta_0) = \theta \sqrt{n} \left[ \frac{1}{\theta_0} - \frac{Y_n}{\theta} \right]. $$

Thus we reject when

$$ |\frac{Y_n}{\theta_0} - \frac{1}{\theta_0}| > \frac{1.96}{\theta_0 \sqrt{n}}. $$

Since $\theta_0 = 2$ in this case, we have $c_1 = 1/2$ and $c_2 = .98$.

Problem 4. Either construct an example of each of the following or prove that no such example exists:

(a) [1 point] A sequence $X_1, X_2, \ldots$ of random variables, a random variable $X$, and a constant $x_0$ such that $X_n \xrightarrow{L} X$ and $F_n(x_0) \nrightarrow F(x_0)$.

Let $X_n$ be the constant $1/n$ and $X$ the constant 0. Then $X_n \xrightarrow{L} X$, but $F_n(0) = P(X_n \leq 0) = 0$ for all $n$, whereas $F(0) = P(X \leq 0) = 1$. Therefore, $F_n(0) = 0 \nrightarrow F(0) = 1$.

(b) [1 point] A sequence $X_1, X_2, \ldots$ of random variables such that $X_n \xrightarrow{P} 0$ and $E(X_n) \to 1$.

Let $P(X_n = n) = 1 - P(X_n = 0) = 1/n$. Then $E(X_n) = 1$ for all $n$, but

$$ P(|X_n - 0| < \epsilon) \geq P(X_n = 0) = 1 - (1/n) \to 1. $$

Thus $X_n \xrightarrow{P} 0$. 

Problem 5. [3 points] Let $X_1, \ldots, X_n$ be iid from a continuous symmetric distribution centered at 0. Suppose $(Y_1, \ldots, Y_n)$ is a permutation of $(X_1, \ldots, X_n)$ satisfying $|Y_1| < |Y_2| < \cdots < |Y_n|$; that is, the $Y_i$ are the $X_i$ arranged in order of increasing absolute value.

Let $W_n = \sum_{i=1}^n iI\{Y_i > 0\}$ be the usual signed-rank statistic. Derive the asymptotic distribution of $W_n$, justifying your steps.

[We have seen at least two ways to do this. You may of course use any valid method you choose.]

Here’s one solution: Let $Z_i = iI\{Y_i > 0\}$. Then the $Z_i$ are independent with $P(Z_i = 0) = P(Z_i = i) = 1/2$. Thus, $E Z_i = i/2$. The Lyapunov condition says that

$$\sum_{i=1}^n \frac{Z_i - i/2}{\sqrt{\text{Var} Z_i}} \xrightarrow{L} N(0, 1)$$

as long as

$$\sum_{i=1}^n E |Z_i - i/2|^{2+\delta} = o \left( \sum_{i=1}^n \text{Var} Z_i \right)^{1+\delta/2}$$

(1)

The left hand side of (1) equals $\sum_{i=1}^n (i/2)^{2+\delta} \geq n^{3+\delta}$. The right hand side equals $\left[ \sum_{i=1}^n (i/2)^2 \right]^{1+\delta/2} \leq (n^3)^{1+\delta/2} = n^{3+\delta/2}$. Therefore, the Lyapunov condition is satisfied, which implies that

$$\sum_{i=1}^n \frac{Z_i - i/2}{\sqrt{\text{Var} Z_i}} \xrightarrow{L} N(0, 1).$$

For another solution, see Example 6.1.1 on p. 370, where it is shown that

$$\sqrt{n} \left( \frac{W_n}{\binom{n}{2}} - \frac{1}{2} \right) \xrightarrow{L} N(0, 1/3).$$

A moment’s reflection should convince you that these two solutions are equivalent.

Problem 6. [2 points] Suppose $X_1, \ldots, X_n$ are iid random variables with cdf $F(x)$. Let $\hat{X}_n$ denote the sample median. Suppose we wish to estimate $h(F) = \text{Var}(\hat{X}_n) < \infty$. We use a bootstrap scheme in which we draw $B$ random samples of size $n$ from $\hat{F}_n$, the empirical cdf, and let $M_i$ be the sample median of the ith sample, $i = 1, \ldots, B$.

If $\overline{M}_B$ denotes $(1/B) \sum_{i=1}^B M_i$, explain (with justification) what happens to

$$\frac{1}{B} \sum_{i=1}^B (M_i - \overline{M}_B)^2$$

as $B \to \infty$.

Writing

$$\frac{1}{B} \sum_{i=1}^B (M_i - \overline{M}_B)^2 = \frac{1}{B} \sum_{i=1}^B M_i^2 - (\overline{M}_B)^2,$$

we obtain by the WLLN

$$\frac{1}{B} \sum_{i=1}^B (M_i - \overline{M}_B)^2 \xrightarrow{P} \text{E} M_i^2 - (\text{E} M_i)^2 = \text{Var} M_i.$$

Since $M_i$ is the sample median of a sample of size $n$ from $\hat{F}_n$, we get

$$\frac{1}{B} \sum_{i=1}^B (M_i - \overline{M}_B)^2 \xrightarrow{P} h(\hat{F}_n).$$
Problem 7. [3 points] Suppose that $X_1, \ldots, X_n$ are iid with
\[ P(X_i = 0) = \theta \quad \text{and} \quad P\left(X_i = -\sqrt{1 - \theta}\right) = P\left(X_i = \sqrt{1 - \theta}\right) = \frac{1 - \theta}{2}. \]

Define $Y_i = I\{X_i = 0\}$.

Let $Z^{(i)} = (X_i + \theta, Y_i)$. Find a function $g(x, y) = [g_1(x, y), g_2(x, y)]$ that is a variance-stabilizing transformation in the sense that the asymptotic distribution of $g\left[\overline{Z}\right] - g\left[E(Z^{(i)})\right]$ has a covariance matrix that doesn’t depend on $\theta$, where $\overline{Z}$ is the sample mean of the $Z^{(i)}$.

You may find it helpful to know that
\[ \frac{d}{dt} 2 \sin^{-1}(\sqrt{t}) = \frac{1}{\sqrt{t(1-t)}}. \]

Since $E Z_1 = \theta$, $E Z_2 = \theta$, $\text{Var} Z_1 = \theta$, $\text{Var} Z_2 = \theta(1 - \theta)$, and $\text{Cov}(Z_1, Z_2) = E(X_i Y_i + \theta Y_i) - \theta^2 = 0$, the CLT gives
\[ \sqrt{n} \left( \overline{Z} - \begin{pmatrix} \theta \\ \theta \end{pmatrix} \right) \overset{D}{\sim} N_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} (1 - \theta)^2 & 0 \\ 0 & \theta(1 - \theta) \end{pmatrix} \right). \]

Thus if
\[ \hat{g}(\theta, \theta) = \begin{pmatrix} \frac{1}{1 - \theta} & 0 \\ 0 & \frac{1}{\sqrt{\theta(1 - \theta)}} \end{pmatrix}, \]

then we have $\hat{g}(\theta, \theta) \Sigma \hat{g}(\theta, \theta)^t = I$ and we’re done. Thus, we may simply take $g_1(x, y) = -\log(1 - x)$ and $g_2(x, y) = 2 \sin^{-1}(\sqrt{y})$ and we’re done.