STA 711 Homework 4

Due: Friday, February 13, 11:59pm on Canvas.

Instructions: Submit your work as a single PDF. You may choose to either hand-write your work and submit a PDF scan, or type your work using LaTeX and submit the resulting PDF. For computational work, include all code and R output necessary to answer the questions, and submit as an html or pdf on Canvas.

See the course website for a homework template file and instructions on getting started with LaTeX and Overleaf.

Convergence of random variables

In this section, you will practice proving limits for sequences of random variables. As a reminder, here are some of the common techniques for proving convergence:

- For convergence in probability:
  - If you have a sequence of means, try to apply the WLLN
  - If you can easily calculate a mean or variance, try bounding probabilities with Markov's or Chebyshev's inequality
  - If calculating means or variances is hard, try calculating the probabilities directly for the convergence
- For convergence in distribution:
  - To a normal or χ^2: check if the central limit theorem applies
  - If CLT is not the right strategy, try calculating the cdfs directly
- Other key results:
  - Continuous mapping theorem is useful for functions of sequences: if a sequence $\{X_n\}$ converges and g is continuous, then $\{g(X_n)\}$ converges
  - Slutsky's theorem is useful for sums, products, and ratios of of multiple sequences

1. For each of the following sequences $\{Y_n\}$, show that $Y_n \xrightarrow{p} 1$. Then write a simulation in R demonstrating the convergence. (For your simulation, choose some $\varepsilon > 0$, and some n. Generate many samples of Y_n, and approximate $P(|Y_n - 1| \geq \varepsilon)$. Then repeat with different values of n, and plot $P(|Y_n - 1| \geq \varepsilon)$ as a function of n).

   (a) $Y_n = 1 + nX_n$, where $X_n \sim Bernoulli(1/n)$

   (b) $Y_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2$, where $X_i \overset{iid}{\sim} N(0,1)$

2. Suppose that $Y_1, Y_2, ... \overset{iid}{\sim} Beta(1, \beta)$. Find a value of ν such that $n^\nu(1 - Y_{(n)})$ converges in distribution. Then write a simulation in R demonstrating the convergence. (*Hint:* if you are struggling to find ν, starting with simulations may be helpful). For your simulations, start with some value of n, and generate many samples of $n^\nu(1 - Y_{(n)})$. Plot the empirical cdf (the `ecdf` function in R), and overlay the limiting cdf. Make the plot for different values of n, and show that the cdfs get closer as n increases.

3. Suppose that $Y_1, Y_2, ... \overset{iid}{\sim} Exponential(1)$. Find a sequence a_n such that $Y_{(n)} - a_n$ converges in distribution.

4. In this problem, we will prove part of the continuous mapping theorem. Let $\{Y_n\}$ be a sequence of real-valued random variables such that $Y_n \xrightarrow{p} Y$ for some random variable Y. Let g be a continuous function; recall that g is continuous if for all $\varepsilon > 0$, there exists some $\delta > 0$ such that $|g(x) - g(y)| < \varepsilon$ whenever $|x - y| < \delta$. Prove that $g(Y_n) \xrightarrow{p} g(Y)$.

5. Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where the X_i are known constants, and the ε_i are iid with $\mathbb{E}[\varepsilon_i] = 0$ and $Var(\varepsilon_i) = \sigma^2$. It can be shown that the least squares estimate of β_1 is

$$\widehat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)\varepsilon_i}{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}.$$

Show that if $\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \to \infty$ as $n \to \infty$, then $\widehat{\beta}_1 \xrightarrow{p} \beta_1$. (Note: no distribution for ε_i or Y_i has been assumed, so $\widehat{\beta}_1$ cannot be treated as a maximum likelihood estimator).

6. Let $(X_1, Y_1), ..., (X_n, Y_n)$ be an iid sample from the joint distribution of X and Y, such that $Var(X)$, $Var(Y)$, $Var(X^2)$, $Var(Y^2)$, $Var(XY)$, and $Var(X^2Y^2)$ are all finite, and all of these variances are greater than 0.

We are interested in estimating the *correlation* between X and Y:

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}.$$

The Pearson correlation estimates ρ by

$$\widehat{\rho} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}.$$

The goal of this question is to show that $\widehat{\rho} \xrightarrow{p} \rho$ as $n \to \infty$. (*Hint*: Both Slutsky's theorem and the continuous mapping theorem will be useful in this question).

(a) Show that $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \xrightarrow{p} Var(X)$ and $\frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \xrightarrow{p} Var(Y)$.

(b) Show that $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) \xrightarrow{p} Cov(X, Y)$.

(c) Show that $\widehat{\rho} \xrightarrow{p} \rho$.

7. Let $\{Y_n\}$ be a sequence of random variables. We say that the sequence converges in *quadratic mean* to a random variable Y, and write $Y_n \overset{q.m.}{\to} Y$, if $\mathbb{E}[(Y_n - Y)^2] \to 0$ as $n \to \infty$.

   (a) Show that if $Y_n \overset{q.m.}{\to} Y$, then $Y_n \overset{p}{\to} Y$. (That is, convergence in quadratic mean implies convergence in probability).

   (b) Show that for any constant c, $\mathbb{E}[(Y_n - c)^2] = (\mathbb{E}[Y_n] - c)^2 + Var(Y_n)$.

   (c) Using part (a) and part (b), show that if $\mathbb{E}[Y_n] \to c$ and $Var(Y_n) \to 0$ as $n \to \infty$, then $Y_n \overset{p}{\to} c$.