STAT 510: Homework 03

David Dalpiaz

Due: Monday, February 14, 11:59 PM

Exercise 1 (Expectation Review)

Let X_1, X_2, and X_3 be independent Uniform$(0, 1)$ random variables. Define $Y = X_1 - 3X_2 + 2X_3$. Provide an upper bound for $P(|Y| \geq 2)$ using Chebyshev's inequality.

Exercise 2 (Creating a Confidence Interval)

(Based on **LW** 4.4) Let $X_1, X_2, \ldots X_n \sim$ Bernoulli(p). Let $\alpha > 0$ and define

$$\epsilon_n = \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}.$$

Define $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and

$$C_n = (\hat{p}_n - \epsilon_n, \hat{p}_n + \epsilon_n).$$

Show that

$$P(C_n \text{ contains } p) \geq 1 - \alpha.$$

Exercise 3 (Decreasing Rate Poissons)

(Based on **LW** 5.7) Let $\lambda_n = 1/n$ for $n = 1, 2, \ldots$ and let $X_n \sim$ Poisson(λ_n).

Also define $Y_n = nX_n$. Show that

$$Y_n \xrightarrow{p} 0.$$

Exercise 4 (More Classic Setup)

(**LW** 5.3) Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed and $\mu = \mathbb{E}[X_1]$. Give that the variance is finite, show that

$$\bar{X}_n \xrightarrow{qm} \mu.$$

Exercise 5 (The Sample Variance)

(**LW** 5.3) Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed and finite mean $\mu = \mathbb{E}[X_1]$ and finite variance $\sigma^2 = \mathbb{V}[X_1]$. Let $\bar{X}_n$ be the sample mean and let S_n^2 be the sample variance. Show that

$$S_n^2 \xrightarrow{p} \sigma^2.$$

Exercise 6 (Normal Approximations with the CLT)

(Based on **LW** 2.8) Suppose we have a computer program consisting of $n = 1000$ lines of code. (And somehow, someone wrote it without debugging along the way.) Let X_i be the number of errors on the i-th line of code. Suppose that the X_i are Poisson with mean 0.01 and that they are independent. Let Y be the sum of the X_i, that is, the total errors. Use the CLT to approximate the probability that there are 5 errors or less. Compare this to the exact probability.

Exercise 7 (CLT with Sample Variance)

Assuming the same conditions as the CLT, and knowing that the CLT exists, show that

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{D} N(0,1).$$

where S_n^2 is the sample variance.

Exercise 8 (Clever Titles are Hard)

(**LW** 2.14) Let $X_1, \ldots, X_n \sim \text{Uniform}(0,1)$. Let $Y_n = \bar{X}_n^2$. Find the limiting distribution of Y_n.

Exercise 9 (Coverage)

(Based on **LW** 4.4) Return to the results from Exercise 2. Set $\alpha = 0.2$ and $p = 0.4$. Use a simulation study to see how often this interval contains p. We call this quantity the interval's *coverage*. Do this for various sample sizes, n, between 1 and 10,000. Plot the coverage versus n. Note, for each n you will need to perform multiple simulations. Use enough values of n, and enough simulations for each, to create a reasonable looking plot.

Exercise 10 (Rate of Convergence)

So far, we have only been concerned with **if** a random variable converges, and to an extent, **how** a random variable converges, but we have not looked at the **rate** of convergence. To investigate this idea, consider random samples from two different distributions.

1. A Bernoulli like distribution with $P(X = -0.2) = P(X = 0.2) = 0.5$.
2. A t distribution with 2 degrees of freedom.

Note that both of these distributions have mean 0.

Generate a sample of size 10,000 from both and plot the sample mean against the sample size. Repeat this process three times and arrange the plots side-by-side. Comment on which distribute you believe converges faster.

Exercise 11 (Hodges' Estimator)

Let $X_1, \ldots, X_n \sim N(\theta, 1)$. Define

$$\hat{\theta}_n = \begin{cases} 0 & |\bar{X}_n| \leq n^{-1/4} \\ \bar{X}_n & |\bar{X}_n| > n^{-1/4} \end{cases}$$

Prove that

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} \begin{cases} 0 & \theta = 0 \\ N(0,1) & \theta \neq 0 \end{cases}$$