Econ 21020 - Problem Set 2

Problem 1

Complete the proof of the biasedness of the sample variance that we began in class. That is, demonstrate that, for $X_1, ..., X_n$ iid $\sim X$

$$E[(\overline{X}_n - E[X])^2] = \frac{1}{n}Var(X)$$

Justify all steps. (Hint: We learned that for Y mean independent of X, $E[YX] = E[Y]E[X]$. We also learned that independence implies mean independence. Thus, the former property also holds for independent X and Y).

SOLUTION:

$$\begin{aligned}
E[(\overline{X}_n - E[X])^2] &= E[\overline{X}_n^2 - 2\overline{X}_n E[X] + E[X]^2] \\
&= E[\overline{X}_n^2] - E[\overline{X}_n]E[X] + E[X]^2 \\
&= E[(\frac{1}{n}\sum_{i=1}^{n} X_i)^2] - 2E[X]E[X] + E[X]^2 \qquad (\overline{X}_n \text{ unbiased}) \\
&= \frac{1}{n^2}E[\sum_{i=1}^{n} X_i^2 + \sum_{i\neq j} X_i X_j] - E[X]^2 \\
&= \frac{1}{n^2}(\sum_{i=1}^{n} E[X_i^2] + \sum_{i\neq j} E[X_i X_j]) - E[X]^2 \\
&= \frac{1}{n^2}(\sum_{i=1}^{n} E[X^2] + \sum_{i\neq j} E[X_i]E[X_j]) - E[X]^2 \\
&\qquad (X_1, ..., X_n \text{ iid}) \\
&= \frac{1}{n^2}(nE[X^2] + \sum_{i\neq j} E[X]E[X]) - E[X]^2 \quad (X_1, ..., X_n \sim X) \\
&= \frac{1}{n^2}(nE[X^2] + n(n-1)E[X]^2) - E[X]^2 \\
&= \frac{1}{n}E[X^2] + \frac{n-1}{n}E[X]^2 - E[X]^2 \\
&= \frac{1}{n}E[X^2] - \frac{1}{n}E[X]^2 \\
&= \frac{1}{n}Var(X)
\end{aligned}$$

The trickiest parts of this proof come in line 4, where we make use of an algebraic property regarding the square of a sum ($(\sum_{i=1}^{n} a_i)^2 = \sum_{i=1}^{n} a_i^2 + \sum_{i\neq j}^{n} a_i a_j$), in line 6, where we make use of the hint, and in line 8, where we have to be careful about how many elements there are in a sum $\sum_{i\neq j}$ (each of the n X_i is multiplied by each of the $n-1$ *other* X_j).

Problem 2

Suppose that Y_a and Y_b are Bernoulli random variables where $Y_a \sim \text{Bernoulli}(p_a)$ and $Y_b \sim \text{Bernoulli}(p_b)$. An iid sample of size n_a is drawn from $\sim Y_a$ and another iid sample of size n_b is drawn from $\sim Y_b$. Say that $\hat{p}_a$ is the proportion of the first sample that has a value of 1 and that $\hat{p}_b$ is the proportion of the second sample that has a value of 1. Assume that both samples are also independent of one another.

(a) Show that $\hat{p}_a - \hat{p}_b$ is an unbiased estimator for $p_a - p_b$.

SOLUTION: We're told that $\hat{p}_a$ and $\hat{p}_b$ are the proportions of their respective samples that take a value of 1. This would imply:

$$\hat{p}_a = \frac{1}{n_a}\sum_{i=1}^{n} Y_{a,i} \;\; \hat{p}_b = \frac{1}{n_b}\sum_{i=1}^{n} Y_{b,i}$$

Then, taking an expectation of the difference of these two would yield:

$$\begin{aligned}
E[\hat{p}_a - \hat{p}_b] &= E[\frac{1}{n_a}\sum_{i=1}^{n_a} Y_{a,i} - \frac{1}{n_b}\sum_{i=1}^{n_b} Y_{b,i}] \\
&= \frac{1}{n_a}\sum_{i=1}^{n_a} E[Y_{a,i}] - \frac{1}{n_b}\sum_{i=1}^{n_b} E[Y_{b,i}] \\
&= \frac{1}{n_a}\sum_{i=1}^{n_a} E[Y_a] - \frac{1}{n_b}\sum_{i=1}^{n_b} E[Y_b] \\
&\qquad (Y_{a,1}, ..., Y_{a,n_a} \sim Y_a;\, Y_{b,1}, ..., Y_{b,n_b} \sim Y_b) \\
&= \frac{1}{n_a}\sum_{i=1}^{n_a} p_a - \frac{1}{n_b}\sum_{i=1}^{n_b} p_b \\
&= \frac{1}{n_a} n_a p_a - \frac{1}{n_b} n_b p_b \\
&= p_a - p_b
\end{aligned}$$

where we used the fact that the expectation of a Bernoulli variable is the probability of taking on a value of 1 in the fourth line. Thus, we'll see that

$$Bias(\hat{p}_a - \hat{p}_b) = E[\hat{p}_a - \hat{p}_b] - p_a - p_b = 0$$

(b) Derive $Var(\hat{p}_a - \hat{p}_b)$

SOLUTION: We can solve this making use of the fact that the variance of a sum of independent random variables is equal to the sum of the variances:

$$\begin{aligned}
Var(\hat{p}_a - \hat{p}_b) &= Var(\frac{1}{n_a}\sum_{i=1}^{n_a} Y_{a,i} - \frac{1}{n_b}\sum_{i=1}^{n_b} Y_{b,i}) \\
&= \frac{1}{n_a^2}\sum_{i=1}^{n_a} Var(Y_{a,i}) - \frac{1}{n_b^2}\sum_{i=1}^{n_b} Var(Y_{b,i}) \quad \text{(Ind of samples)} \\
&= \frac{1}{n_a^2}\sum_{i=1}^{n_a} Var(Y_a) - \frac{1}{n_b^2}\sum_{i=1}^{n_b} Var(Y_b) \\
&\qquad (Y_{a,1}, ..., Y_{a,n_a} \sim Y_a;\, Y_{b,1}, ..., Y_{b,n_b} \sim Y_b) \\
&= \frac{1}{n_a^2} n_a p_a(1-p_a) - \frac{1}{n_b^2} n_b p_b(1-p_b) \\
&= \frac{1}{n_a} p_a(1-p_a) - \frac{1}{n_b} p_b(1-p_b)
\end{aligned}$$

where, in the fourth line, we make use of the variance of a Bernoulli random variable.

(c) Assume that both n_a and n_b are large. Show what a 95% confidence interval for $p_a - p_b$ would look like in terms of $\hat{p}_a$, $\hat{p}_b$, n_a, n_b, and numbers.

SOLUTION: (FOR 5 BONUS POINTS) It will be the case that our 95% confidence interval will look like:

$$
\begin{aligned}
C_n &= \hat{p}_a - \hat{p}_b \pm \Phi^{-1}(1 - \frac{0.05}{2})\sqrt{\frac{\hat{p}_a(1-\hat{p}_a)}{n_a} + \frac{\hat{p}_b(1-\hat{p}_b)}{n_b}} \\
&= \hat{p}_a - \hat{p}_b \pm 1.96\sqrt{\frac{\hat{p}_a(1-\hat{p}_a)}{n_a} + \frac{\hat{p}_b(1-\hat{p}_b)}{n_b}}
\end{aligned}
$$

We lack some of the tools necessary to rigorously derive this expression, but for now we can note the similarities between this expression and the confidence interval we saw in class. In both cases, we have our estimate plus or minus the inverse CDF of a standard normal evaluated at such a point as to achieve α significance times *something*. The point of difference is that in class, the *something* was the sample standard deviation divided by the square root of the sample size. Here, we see something related but different: the square root of the sum of the variances of each sub-estimator, divided by their respective sample sizes. We can also note that this is equal to the square root of the variance of the estimator that we got in part (b). The difference in the exact form, however, is attributable to the need to account for multiple sample sizes, as we will discuss later in the course.

Problem 3

Consider a sample $X_1, ..., X_n$ that is iid $\sim X$.

(a) Suppose that $X \sim N(\mu, \sigma^2)$. What is the sampling distribution for $\overline{X}_{10}$ (the sample mean when we have a sample of 10)?

SOLUTION: We know that a finite sum of independent, normal random variables will also be normally distributed. Thus, $\overline{X}_{10}$ will have a normal distribution. Moreover, we know that the sum of the normals will have a particular form to its distribution, as given by the equation we saw in class:

$$
\begin{aligned}
\frac{1}{10}\sum_{i=1}^{10} X_i &= \sum_{i=1}^{10} \frac{1}{10} X_i \\
&\sim N(\sum_{i=1}^{10} \frac{1}{10}\mu, \sum_{i=1}^{10} \frac{1}{100}\sigma^2) \\
&\sim N(\mu, \frac{1}{10}\sigma^2)
\end{aligned}
$$

(b) Suppose instead that $X \sim \text{Bernoulli}(p)$. Would the sampling distribution from part a) still apply for $\overline{X}_{10}$? (If not, no need to show what it would be instead).

SOLUTION: The result that the sum of a finite number of independent, *normal* random variables is normal, requires the normality of the elements of the sum. As these X_i are Bernoulli, this result will not apply, and we'll have a different sampling distribution that in part a).

(c) Find the limiting distribution for $\sqrt{n}(\overline{X}_n - E[X])$ - that is consider the distribution for this expression as n grows arbitrarily large - for both $X \sim N(\mu, \sigma^2)$ and $X \sim \text{Bernoulli}(p)$.

SOLUTION: We are no longer considering specific, finite sums of random variables, but sums over arbitrarily large numbers of random variables. Thus, we are no longer appealing to any results about sums of independent normals, but instead to the Central Limit Theorem. The CLT requires that the X_i are iid, which we have by assumption, and that the second moment exists, which will be the case for either of the distributions under consideration. Note the CLT applies for samples drawn from any distribution, not just the normal distribution. Thus, for both of the cases we are considering, we can directly apply the CLT to say that,

$$\sqrt{n}(\overline{X}_n - E[X]) \xrightarrow{d} N(0, Var(X))$$

where we need simply determine $Var(X)$ for each of our two cases. The variance of the normal X_i is given as σ^2. For a Bernoulli variable, X, we have previously shown that $Var(X) = p(1-p)$. Thus, the CLT will specifically tell us that:

$$
\begin{aligned}
X \sim N(\mu, \sigma^2) &\Rightarrow \sqrt{n}(\overline{X}_n - E[X]) \xrightarrow{d} N(0, \sigma^2) \\
X \sim \text{Bernoulli}(p) &\Rightarrow \sqrt{n}(\overline{X}_n - E[X]) \xrightarrow{d} N(0, p(1-p))
\end{aligned}
$$

Problem 4

Let $(X_1, Y_1), ..., (X_n, Y_n)$ be iid $\sim (X, Y)$. Assume that $E[X^2] < \infty$, $E[Y^2] < \infty$, and $E[(XY)^2] < \infty$. Show that the sample covariance

$$\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \overline{X}_n \overline{Y}_n$$

is a consistent estimator for $Cov(X, Y)$ using the WLLN and CMT. (Hint 1: This will look quite a bit like the proof of consistency for the sample variance). (Hint 2: Remember that the CMT can be applied with any finite number of sequences of random variables that converge to any finite number of scalars. Alternatively, you can consider applying the CMT twice in a row).

SOLUTION: We have that our sample is iid and that the second moments of X and Y exist, and that $E[(XY)^2] < \infty$. Thus, applying WLLN tells us that:

$$\overline{X}_n \xrightarrow{p} E[X]$$
$$\overline{Y}_n \xrightarrow{p} E[Y]$$
$$\frac{1}{n}\sum_{i=1}^{n} X_i Y_i \xrightarrow{p} E[XY]$$

where for the third statement, we are applying the WLLN to the "sample mean" of $X_i Y_i$. Thus, we have three sequences of random variables converging to three scalars. Moreover, the function:

$$g(a, b, c) = a - bc$$

is continuous everywhere. Thus, we can apply the Continuous Mapping Theorem to say:

$$\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \overline{X}_n \overline{Y}_n \xrightarrow{p} E[XY] - E[X]E[Y] = Cov(X, Y)$$

Problem 5

This problem will walk through the development of one-sided hypothesis tests, p-values, and confidence sets that are analogous to the two-sided versions we saw in class. For everything that follows, assume that $X_1, ..., X_n$ are iid $\sim X$ and that $0 < \sigma_X^2 < \infty$.

(a) We have a null hypothesis, $H_0 : E[X] = \mu_0$, alternative hypothesis, $H_1 : E[X] > \mu_0$, and a test statistic:

$$T_n = \frac{\sqrt{n}}{\hat{\sigma}_n}(\overline{X}_n - \mu_0)$$

We want to achieve the significance level α for our test. Determine the critical value c that will allow us to achieve the desired significance level, if our decision rule is to reject H_0 when $T_n > c$. (NB: We could use the same critical value you'll find here if we had the null $H_0 : E[X] \leq \mu_0$ instead).

SOLUTION: Assume for the purposes of determining the critical value that $E[X] = \mu_0$. Then, given the iid nature of the sample and the fact that the second moment of X exists (implied by the finiteness of σ_X^2), we can directly apply the Central Limit Theorem to say:

$$\sqrt{n}(\overline{X}_n - \mu_0) \xrightarrow{d} N(0, \sigma_X^2)$$

Further using the consistency of the sample variance and $\sigma_X^2 > 0$, Slutsky's Lemma implies:

$$T_n = \frac{\sqrt{n}}{\hat{\sigma}_n}(\overline{X}_n - \mu_0) \xrightarrow{d} N(0, 1)$$

Then, for large n, we'll say it is approximately true that:

$$P(T_n > c) = 1 - \Phi(c)$$

Wanting a significance level of α means that we want the probability on the left-hand side above to be equal to α, if $E[X] = \mu_0$ (which we assumed at the beginning of the section). Thus, we want

$$\begin{aligned} 1 - \Phi(c) &= \alpha \\ 1 - \alpha &= \Phi(c) \\ \Phi^{-1}(1 - \alpha) &= c \end{aligned}$$

If we have a specific value of α in mind, then we can look up $\Phi^{-1}(1 - \alpha)$ somewhere to get our desired critical value, c.

(b) We are using a (slightly) different test statistic than we saw in class. Explain why in words.

SOLUTION: This test statistic has parentheses where the test statistic for a two-sided test has an absolute value. Because of this, $T_n^{\text{One-sided}} = \frac{\sqrt{n}}{\hat{\sigma}_n}(\overline{X}_n - \mu_0)$ will only get "large" for $\overline{X}_n$ that are much *larger* than μ_0. By contrast, $T_n^{\text{Two-sided}} = \frac{\sqrt{n}}{\hat{\sigma}_n}|\overline{X}_n - \mu_0|$ will get large for $\overline{X}_n$ that are either much larger *or* much smaller than μ_0. Unlike in a two-sided test, we only want to reject when we have $\overline{X}_n$ much larger than μ_0 because our alternative hypothesis is that the mean of X is larger than μ_0. In a two-sided test, we want to reject if $\overline{X}_n$ is far away from μ_0 in either direction, because our alternative hypothesis is simply that the mean of X is different than μ_0.

(c) Our decision rule is that we will reject the null if $T_n > c$. Using the given test statistic and the critical value you derived in part a), express the one-sided p-value in terms of numbers and functions that we already know (we "know" μ_0 as we chose this in advance) and things we can estimate from the data.

SOLUTION: Our p-value is the smallest value of α at which we would reject the null hypothesis. Given our work in part (a), we know that we reject if:

$$\begin{aligned} T_n &> c \\ \frac{\sqrt{n}}{\hat{\sigma}_n}(\overline{X}_n - \mu_0) &> \Phi^{-1}(1-\alpha) \\ \Phi(\frac{\sqrt{n}}{\hat{\sigma}_n}(\overline{X}_n - \mu_0)) &> 1-\alpha \\ \alpha &> 1 - \Phi(\frac{\sqrt{n}}{\hat{\sigma}_n}(\overline{X}_n - \mu_0)) \end{aligned}$$

We will thus reject for any α greater than the object on the right-hand side of the inequality above. If α is any smaller than the object on the right-hand side, then is must be the case that we fail to reject. Thus, the exact value on the right-hand side is the smallest value at which we reject, which is exactly what we want the p-value to equal. As such,

$$\hat{p}_n = 1 - \Phi(\frac{\sqrt{n}}{\hat{\sigma}_n}(\overline{X}_n - \mu_0))$$

(d) Is this one-sided p-value weakly greater or weakly smaller than the two-sided p-value we saw in class? Explain the intuition for the answer.

SOLUTION: The answer here is that the one-sided p-value is neither weakly smaller or weakly greater - if it is larger or smaller will depend on the value of $\overline{X}_n$. Compare the p-value we saw in class to this one:

$$\begin{aligned} \hat{p}_n^{\text{One-sided}} &= 1 - \Phi(T_n^{\text{One-sided}}) \\ \hat{p}_n^{\text{Two-sided}} &= 2(1 - \Phi(T_n^{\text{Two-sided}})) \end{aligned}$$

Let's consider this in two cases. First, say $\overline{X}_n \geq \mu_0$. This implies that:

$$\begin{aligned} |\overline{X}_n - \mu_0| &= (\overline{X}_n - \mu_0) \\ \Rightarrow T_n^{\text{Two-sided}} &= T_n^{\text{One-sided}} \\ 1 - \Phi(T_n^{\text{Two-sided}}) &= 1 - \Phi(T_n^{\text{One-sided}}) \\ \hat{p}_n^{\text{Two-sided}} = 2(1 - \Phi(T_n^{\text{Two-sided}})) &> 1 - \Phi(T_n^{\text{One-sided}}) = \hat{p}_n^{\text{One-sided}} \end{aligned}$$

Thus, when $\overline{X}_n \geq \mu_0$, the two-sided p-value is always larger. In both types of test, we consider a large $\overline{X}_n$ evidence against the null, as the alternative hypothesis for both types of test includes $E[X] > \mu_0$. This is reflected in the equivalence of the test-statistics. However, our two-sided p-value is larger, as we are more "cautious" in using this type of evidence against the null in the two-sided case. In the two-sided case, Type 1 Error occurs if $E[X] = \mu_0$ but we happen to pull an unusually small *or* unusually large $\overline{X}_n$. Because we have to account for both "types" of Type-1 Error, this will increase the magnitude of the p-value when the test-statistic is identical. In the one-sided cases, Type 1 Error *only* occurs if $E[X] = \mu_0$ but we happen to pull an unusually large $\overline{X}_n$. Thus, for a fixed sensitivity to Type 1 Error across both types of test, we can be more "aggressive" in interpreting large $\overline{X}_n$ as evidence against the null in the one-sided case.

Now consider $\overline{X}_n < \mu_0$. This implies that:

$$
\begin{aligned}
|\overline{X}_n - \mu_0| &> (\overline{X}_n - \mu_0) \\
\Rightarrow T_n^{\text{Two-sided}} &> T_n^{\text{One-sided}} \\
1 - \Phi(T_n^{\text{Two-sided}}) &< 1 - \Phi(T_n^{\text{One-sided}}) \\
\hat{p}_n^{\text{Two-sided}} = 2(1 - \Phi(T_n^{\text{Two-sided}})) &\ ?\ 1 - \Phi(T_n^{\text{One-sided}}) = \hat{p}_n^{\text{One-sided}}
\end{aligned}
$$

As it turns out, the two-sided p-value will now be smaller if $\overline{X}_n$ is sufficiently far below μ_0, otherwise the two-sided p-value will still be larger. Say that $\overline{X}_n$ is sufficiently far below μ_0 and the two-sided p-value is smaller. What's going on? In the one-sided case the alternative hypothesis is that $E[X] > \mu_0$. Thus, getting a very small $\overline{X}_n$ does not provide evidence against the null. However, in a two-sided case, any $\overline{X}_n$ that is far away from μ_0, including very small $\overline{X}_n$ comports with the alternative hypothesis that $E[X] \neq \mu_0$. Thus, we'll have a smaller p-value in the two-sided case, as the very small $\overline{X}_n$ suggests we reject the two-sided null while it does not suggest we reject the one-sided null.

In conclusion, there will be some $\overline{X}_n' < \mu_0$. If $\overline{X}_n < \overline{X}_n'$, the two-sided p-value is smaller. If $\overline{X}_n > \overline{X}_n'$ the two-sided p-value is larger.

As this question is fairly challenging, characterizing one of the two "regions" accurately and with good intuition will be sufficient for full credit.

Problem 6

Complete the following in R or another language of your choice (providing that you have cleared you choice with the TA first).

(a) Generate 100 samples of $n = 5$ from a Uniform[-1,1] distribution. For each sample, compute the sample mean. What proportion of the sample means lie within the range [-0.1,0.1]?

(b) Repeat the above with sample sizes of $n = 10$ and $n = 100$.

(c) Interpret what you find through the lens of results from the class.

(d) For each of the three sample sizes, create histograms plotting 100 values of $\sqrt{n}\overline{X}_n$ (so three histograms - one each showing 100 values of $\sqrt{5}\overline{X}_5$, $\sqrt{10}\overline{X}_{10}$, and $\sqrt{100}\overline{X}_{100}$). Interpret what you find through the lens of results from the class.