# MATH 369/650 Fall 2021 Homework #2

## Professor Adam Kapelner

### Due by email 11:59PM October 31, 2021

(this document last updated Sunday 17th October, 2021 at 5:07pm)

**Instructions and Philosophy**

The path to success in this class is to do many problems. Unlike other courses, exclusively doing reading(s) will not help. Coming to lecture is akin to watching workout videos; thinking about and solving problems on your own is the actual "working out." Feel free to "work out" with others; **I want you to work on this in groups.**

Reading is still *required.* For this homework set, read about two-sample hypothesis testing, creation of estimators via the method of moments, likelihood, log likelihood, the CRLB, Fisher Information and the score function.

The problems below are color coded: green problems are considered *easy* and marked "[easy]"; yellow problems are considered *intermediate* and marked "[harder]", red problems are considered *difficult* and marked "[difficult]" and purple problems are extra credit. The *easy* problems are intended to be "giveaways" if you went to class. Do as much as you can of the others; I expect you to at least attempt the *difficult* problems. "[MA]" are for those registered for 621 and extra credit otherwise.

This homework is worth 100 points but the point distribution will not be determined until after the due date. See syllabus for the policy on late homework.

Up to 7 points are given as a bonus if the homework is typed using LaTeX. Links to instaling LaTeX and program for compiling LaTeX is found on the syllabus. You are encouraged to use `overleaf.com`. If you are handing in homework this way, read the comments in the code; there are two lines to comment out and you should replace my name with yours and write your section. The easiest way to use overleaf is to copy the raw text from hwxx.tex and preamble.tex into two new overleaf tex files with the same name. If you are asked to make drawings, you can take a picture of your handwritten drawing and insert them as figures or leave space using the "\vspace" command and draw them in after printing or attach them stapled.

The document is available with spaces for you to write your answers. If not using LaTeX, print this document and write in your answers. I do not accept homeworks which are *not* on this printout. Keep this first page printed for your records.

NAME: _____

In lecture, we did two-sided two-sample z and t tests. We will repeat these tests now but do them one sided. For extra practice, I will make them left-sided. To do this, I will switch the indexing of the two populations. The female population is now considered population #1 and the male population is now considered population #2. We assume the DGP for female height measurements is $\overset{iid}{\sim} \mathcal{N}(\theta_1,\, \sigma_1^2)$ independent of the DGP for male height measurements assumed to be $\overset{iid}{\sim} \mathcal{N}(\theta_2,\, \sigma_2^2)$.

The sample sizes, point estimates for the mean and point estimates for the variance computed from an in-class student survey are:

$$
\begin{aligned}
n_1 &= 6 \\
\bar{x}_1 &= 62.3 \\
s_1^2 &= 2.25^2 \\
n_2 &= 10 \\
\bar{x}_2 &= 70.5 \\
s_2^2 &= 2.07^2
\end{aligned}
$$

The theory we wish to prove is that females are shorter than males. We will do so via hypothesis testing at size $\alpha = 5\%$. We will test this under many different assumption scenarios about the variances in the DGP. The results should be approximately the same.

(a) [easy] Write the alternative and null hypotheses. Remember that $\theta_1$ and $\theta_2$ are now switched from class. We will use these hypotheses for the four scenarios in the rest of the problem.

_____

We will first assume that the variances are known to be $\sigma_1^2 = 3.5^2$ and $\sigma_2^2 = 4^2$.

(b) [easy] Write the exact or approximate distribution of the standardized estimator under the null hypothesis which we denote $(\hat{\theta}_1 - \hat{\theta}_2)/SE \mid H_0$. You should first write the SE as a mathematical expression. Then compute SE to three decimal places. The answer is in lecture 6.

(c) [easy] Illustrate the distribution from the previous problem. Label the x-axis and provide tick marks on the x-axis. Leave lots of room on the left side of the x-axis away from the bulk of the sampling distribution!

(d) [harder] Will this test be an *exact test* or instead an *approximate test*? Explain.

(e) [harder] Compute the retainment region and rejection region (remember $\Theta = \mathbb{R}$) and denote these two regions in your illustration in (c). To compute these regions, I'll provide you with the following fact: $\Phi(-1.645) = 5\%$ where $\Phi$ is the CDF of the standard normal rv. (These are the kind of facts that will be provided to you on exams).

(f) [easy] Why is this test named the *left-sided two-sample z test of unequal variances*?

(g) [easy] Run the test and write your conclusion using an English sentence.

(h) [easy] What type of error could you have made?

(i) [harder] Find the p-value of our estimate as a function of $\Phi$. Illustrate the p-value in the illustration in (c).

(j) [easy] Without computing the p-value explicitly, would it be above or below $\alpha = 5\%$? Is the estimate *statistically significant*?

(k) [E.C.] In the general case of $\theta_\Delta := \theta_1 - \theta_2$, $\sigma_1^2$, $\sigma_2^2$, $n_1$, $n_2$ and $\alpha$, find the power function $POW(\theta_\Delta, \sigma_1^2, \sigma_2^2, n_1, n_2, \alpha)$.

We will now assume that the variances are equal and known to be $\sigma_1^2 = \sigma_2^2 = \sigma^2 = 3.76^2$.

(l) [easy] Write the exact or approximate distribution of the standardized estimator under the null hypothesis which we denote $(\hat{\theta}_1 - \hat{\theta}_2)/SE \mid H_0$. You should first write the SE as a mathematical expression. Then compute SE to three decimal places. The answer is in lecture 6.

(m) [easy] Illustrate the distribution from the previous problem. Label the x-axis and provide tick marks on the x-axis. Leave lots of room on the left side of the x-axis away from the bulk of the sampling distribution!

(n) [harder] Will this test be an *exact test* or instead an *approximate test*? Explain.

(o) [harder] Compute the retainment region and rejection region (remember $\Theta = \mathbb{R}$) and denote these two regions in your illustration in (m). To compute these regions, I'll provide you with the following fact: $\Phi(-1.645) = 5\%$ where $\Phi$ is the CDF of the standard normal rv. (These are the kind of facts that will be provided to you on exams).

(p) [easy] Why is this test named the *left-sided two-sample z test of equal variances*?

(q) [easy] Run the test and write your conclusion using an English sentence.

(r) [easy] What type of error could you have made?

(s) [harder] Find the p-value of our estimate as a function of $\Phi$. Illustrate the p-value in the illustration in (m).

(t) [easy] Without computing the p-value explicitly, would it be above or below $\alpha = 5\%$? Is the estimate *statistically significant*?

We will now assume that the variances are equal i.e. $\sigma_1^2 = \sigma_2^2 = \sigma^2$ but its value is *unknown.*

(u) [easy] Write the exact or approximate distribution of the standardized estimator under the null hypothesis which we denote $(\hat{\theta}_1 - \hat{\theta}_2)/SE \mid H_0$. You should first write the SE as a mathematical expression. Then compute SE to three decimal places. The answer is in lecture 6.

(v) [easy] Illustrate the distribution from the previous problem. Label the x-axis and provide tick marks on the x-axis. Leave lots of room on the left side of the x-axis away from the bulk of the sampling distribution!

(w) [harder] Will this test be an *exact test* or instead an *approximate test*? Explain.

(x) [harder] Compute the retainment region and rejection region (remember $\Theta = \mathbb{R}$) and denote these two regions in your illustration in (v). To compute these regions, I'll provide you with the following fact: $\mathbb{P}(T_{14} \leq -1.76) = 5\%$ where $T_{14}$ denotes a standard Student's t rv with 14 degrees of freedom. (These are the kind of facts that will be provided to you on exams).

(y) [easy] Why is this test named the *left-sided two-sample t test of equal variances*?

(z) [easy] Run the test and write your conclusion using an English sentence.

(aa) [easy] What type of error could you have made?

(bb) [harder] Find the p-value of our estimate by writing a statement like $\mathbb{P}(T_{df} < t)$ or $\mathbb{P}(T_{df} > t)$. You need to solve for $df$, $t$. Illustrate the p-val in the illustration in (v).

(cc) [easy] Without computing the p-value explicitly, would it be above or below $\alpha = 5\%$? Is the estimate *statistically significant*?

We will now assume that the variances are unequal i.e. $\sigma_1^2 \neq \sigma_2^2$ and both values are *unknown*. This is known as the Behrens-Fisher problem.

(dd) [easy] Write the exact or approximate distribution of the standardized estimator under the null hypothesis which we denote $(\hat{\theta}_1 - \hat{\theta}_2)/SE \mid H_0$. You should first write the SE as a mathematical expression. Then compute SE to three decimal places. The answer is in lecture 7.

(ee) [easy] Illustrate the distribution from the previous problem. Label the x-axis and provide tick marks on the x-axis. Leave lots of room on the left side of the x-axis away from the bulk of the sampling distribution!

(ff) [harder] Will this test be an *exact test* or instead an *approximate test*? Explain. (Hint: the exact distribution was solved in 2018 and can be found in this paper).

(gg) [harder] Compute the retainment region and rejection region (remember $\Theta = \mathbb{R}$) and denote these two regions in your illustration in (ee). To compute these regions, I'll provide you with the following fact: $\mathbb{P}(T_{9.94} \leq -1.81) = 5\%$ where $T_{9.94}$ denotes a standard Student's t rv with 9.94 degrees of freedom. (These are the kind of facts that will be provided to you on exams).

(hh) [easy] Why is this test named the *left-sided two-sample t test of unequal variances*?

(ii) [easy] Run the test and write your conclusion using an English sentence.

(jj) [easy] What type of error could you have made?

(kk) [harder] Find the p-value of our estimate by writing a statement like $\mathbb{P}(T_{df} < t)$ or $\mathbb{P}(T_{df} > t)$. You need to solve for $df$, $t$. Illustrate the p-val in the illustration in (ee).

(ll) [easy] Without computing the p-value explicitly, would it be above or below $\alpha = 5\%$? Is the estimate *statistically significant*?

10

## Problem 2

These questions will be about the Method of Moments (MM) procedure for generating esti-
mators/estimates.

(a) [easy] Define the $k$th moment of a rv.

(b) [easy] For sample size $n$, define the of the $k$th sample moment *estimator* of a rv.

(c) [easy] For sample size $n$, define the of the $k$th sample moment *estimate* of a rv.

(d) [easy] Give an example of a DGP with two parameters.

(e) [easy] For a DGP with $K = 3$ parameters, write the system of equations that relates
each moment to parameters. There should be 3 equations with 3 unknowns. Then
write the system of equations that relates each parameter to moments. There should
be another 3 equations with 3 unknowns.

(f) [easy] For any iid DGP with finite mean, find an MM estimator for the mean.

(g) [easy] For any iid DGP with finite variance, find an MM estimator for the variance.

(h) [easy] Consider the $X_1, \ldots, X_n \overset{iid}{\sim} \text{Binomial}(\theta_1, \theta_2)$ DGP and derive the MM estimators $\hat{\theta}_1^{\text{MM}}$ and $\hat{\theta}_2^{\text{MM}}$ for $\theta_1$ and $\theta_2$ and express them in terms of $\bar{X}$ and $\hat{\sigma}^2$.

(i) [easy] Provide an example dataset (different from the one in class) where the MM estimates $\hat{\theta}_1^{\text{MM}}$ and $\hat{\theta}_2^{\text{MM}}$ in (h) are illegal and explain why they're illegal.

(j) [harder] Imagine you are a NYPD officer at precinct 100 in Queens. You want to estimates of the number of crimes in your precinct and people's propensity to phone in crimes. The number of daily phone reports for two weeks are: 13, 21, 25, 21, 15, 19, 15, 15, 17, 23, 16, 15, 19, 15. Estimate the true mean number of daily total crimes in the precinct and probability of the crime being phoned in.

(k) [difficult] What exactly what you need to know if you wanted to test if the true mean number of crimes daily exceeds 20? This is conceptual and should be answered with a sentence or two.

(l) [harder] Derive the MM estimator $\hat{\theta}^{\mathrm{MM}}$ for $\theta$ in the $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{U}\,(\theta,\,17)$ DGP.

(m) [easy] Provide an example dataset where the MM estimate $\hat{\theta}^{\mathrm{MM}}$ in the previous question is illegal and explain why it is illegal.

(n) [difficult] [MA] Derive the MM estimators $\hat{\theta}_1^{\mathrm{MM}}$ and $\hat{\theta}_2^{\mathrm{MM}}$ for the $\overset{iid}{\sim} \mathrm{U}\,(\theta_1,\,\theta_2)$ DGP. Then estimate $\theta_1$ and $\theta_2$ given the dataset 9.8, 3.1, 1.2, -0.1, 12.1, 15.9, -9, 3.4, 14.9, -3.6, 16.5, -9.6, 11.2, 11.6, -3.9, -9.3, -1. Remember that $\theta_1$ is defined to be $< \theta_2$!

(o) [harder] Consider the DGP $\overset{iid}{\sim}$ Beta $(\theta_1, \theta_2)$. Below are some facts about this distribution that I took from wikipedia:

$$X \;\sim\; \text{Beta}\,(\theta_1,\,\theta_2) := \underbrace{\frac{1}{B(\theta_1,\theta_2)}x^{\theta_1-1}(1-x)^{\theta_2-1}}_{f(x)}$$

$$\text{Supp}\,[X] \;=\; [0,1], \quad \theta_1, \theta_2 \in (0,\infty)$$

$$\mathbb{E}\,[X] \;=\; \int_0^\infty x f(x)dx = \frac{\theta_1}{\theta_1+\theta_2},$$

$$\mathbb{V}\text{ar}\,[X] \;=\; \int_0^\infty (x-\mathbb{E}\,[X])^2 f(x)dx = \frac{\theta_1\theta_2}{(\theta_1+\theta_2)^2(\theta_1+\theta_2+1)}$$

Specify the two equations that relate moments to parameters, i.e. $\mu_1 = \alpha_1(\theta_1,\theta_2)$ and $\mu_2 = \alpha_2(\theta_1,\theta_2)$. Do not simplify.

(p) [easy] You have two equations and two unknowns. It turns out after much algebra you can solve for the MM estimators in terms of $\bar{X}$ and $\hat{\sigma}^2$ as:

$$\hat{\theta}_1^{\text{MM}} = \bar{x}\left(\frac{\bar{X}(1-\bar{X})}{\hat{\sigma}^2} - 1\right), \quad \hat{\theta}_2^{\text{MM}} = (1-\bar{X})\left(\frac{\bar{X}(1-\bar{X})}{\hat{\sigma}^2} - 1\right)$$

Estimate $\theta_1$ and $\theta_2$ given the dataset 0.393, 0.29, 0.428, 0.117, 0.482, 0.524, 0.413, 0.226, 0.264, 0.567, 0.374.

(q) [harder] Consider the DGP $\overset{iid}{\sim} \text{Gamma}(\theta_1, \theta_2)$. Below are some facts about this distribution that I took from wikipedia:

$$X \sim \text{Gamma}(\theta_1, \theta_2) := \underbrace{\frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)} x^{\theta_1 - 1} e^{-\theta_2 x}}_{f(x)}$$

$$\begin{aligned}
\text{Supp}[X] &= [0, \infty), \quad \theta_1, \theta_2 \in (0, \infty) \\
\mathbb{E}[X] &= \int_0^\infty x f(x) dx = \frac{\theta_1}{\theta_2}, \\
\mathbb{V}\text{ar}[X] &= \int_0^\infty (x - \mathbb{E}[X])^2 f(x) dx = \frac{\theta_1}{\theta_2^2}
\end{aligned}$$

Specify the two equations that relate moments to parameters, i.e. $\mu_1 = \alpha_1(\theta_1, \theta_2)$ and $\mu_2 = \alpha_2(\theta_1, \theta_2)$ and then solve for the MM estimators $\hat{\theta}_1^{\text{MM}}$ and $\hat{\theta}_2^{\text{MM}}$ in terms of $\bar{X}$ and $\hat{\sigma}^2$. Hint: leave expressions in terms of $\hat{\sigma}^2$.

(r) [easy] Provide point estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ for the unknown parameters $\theta_1$ and $\theta_2$ given the dataset 10.8, 8.5, 13.2, 9.1, 13.5, 11.2, 7.1 for the $\overset{iid}{\sim} \text{Gamma}(\theta_1, \theta_2)$ DGP.

(s) [difficult] In Math 241 you learned about expectation and variance where expectation was a measure of central tendency of a distribution and variance is a measure of dispersion around that central tendency. The next most important metric for rv's is probably its *skewness* defined as

$$\gamma := \mathbb{S}\text{kew}[X] := \mathbb{E}\left[\left(\frac{X - \mathbb{E}[X]}{\mathbb{S}\text{D}[X]}\right)^3\right]$$

where SD refers to standard deviation. Skewness is technically the third standardized moment since $\frac{X - \mathbb{E}[X]}{\mathbb{SD}[X]}$ is the distribution standardized and then the third power is taken. Skewness is a metric of which tail of the distribution is longer and by how much as seen in this figure By Diva Jain.
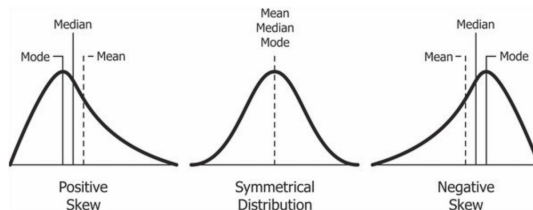


Since third powers are both positive and negative, skewness can be both positive and negative (and zero if the distribution is symmetric with right and left tails the same).

In class, when we were looking at the normal DGP, we derived nonparametric MM estimators $\bar{X}$ and $\hat{\sigma}^2$ for the expectation and variance (nonparametric meaning that the derivation for them was for all DGP's). Show that the nonparametric MM estimator for skewness is:

$$\hat{\gamma} = \sqrt{n} \frac{\sum_{i=1}^{n}(X_i - \bar{X})^3}{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^{3/2}}$$

Hint: assume a iid DGP with density / mass function $f(\theta_1, \theta_2, \theta_3)$ where $\theta_1$ is the expectation, $\theta_2$ is the variance and $\theta_3$ is the skewness.

## Problem 3

We will prove some of the main theorems of this class (and some other relevant facts) here.

(a) [easy] Prove the CRLB from scratch. Justify each step. List assumptions.

(b) [easy] State Thm 5.5.4 p233 C&B.

(c) [easy] Use Thm 5.5.4 to prove that if $X \xrightarrow{P} c$ then $\frac{X}{c} \xrightarrow{P} 1$ and $\frac{c}{X} \xrightarrow{P} 1$.

(d) [easy] State Slutsky's Theorem.

(e) [difficult] Assume an iid DGP with mean $\theta$ and variance $\sigma^2$. Recall the CLT,

$$\frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Let $S^2$ be the estimator for $\sigma^2$ we discussed in class. In a more advanced probability class, you can show that $S \xrightarrow{P} \sigma$. Use Thm 5.5.4 and Slutsky's Theorem to show that the usual t statistic is asymptotically normal i.e. prove that

$$\frac{\bar{X} - \theta}{\frac{S}{\sqrt{n}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

Hint: it is two steps. Don't overthink it!

(f) [easy] Prove that the MLE is asymptotically normal and asymptotically efficient from scratch. Justify each step.

(g) [difficult] Prove that the score function over $n$ is asymptotically normal using the CLT. The facts should be found in the previous problem which you got from the notes.

(h) [difficult] Prove that Fisher Information which is defined as $I(\theta) := \mathbb{E}\left[\ell'(\theta; X)^2\right]$, the expected score squared, is equal to $\mathbb{E}\left[-\ell''(\theta; X)\right]$. If you make any assumptions proving this, indicate it so. This is not easy. There will be lots of hints on slack given.