Math 369 / 650 Fall 2020
Final Examination

Professor Adam Kapelner

Wednesday, December 16, 2020

Code of Academic Integrity

Since the college is an academic community, its fundamental purpose is the pursuit of knowledge. Essential to the success of this educational mission is a commitment to the principles of academic integrity. Every member of the college community is responsible for upholding the highest standards of honesty at all times. Students, as members of the community, are also responsible for adhering to the principles and spirit of the following Code of Academic Integrity.

Activities that have the effect or intention of interfering with education, pursuit of knowledge, or fair evaluation of a student's performance are prohibited. Examples of such activities include but are not limited to the following definitions:

Cheating Using or attempting to use unauthorized assistance, material, or study aids in examinations or other academic work or preventing, or attempting to prevent, another from using authorized assistance, material, or study aids. Example: using an unauthorized cheat sheet in a quiz or exam, altering a graded exam and resubmitting it for a better grade, etc.

By taking this exam, you acknowledge and agree to uphold this Code of Academic Integrity.

Instructions

This exam is 110 minutes (variable time per question) and closed-book. You are allowed **three** pages (front and back) of a "cheat sheet", blank scrap paper and a graphing calculator. Please read the questions carefully. No food is allowed, only drinks.

Problem 1 [6min] (and 6min will have elapsed) Consider running $m > 1$ hypothesis tests each with level α. Following the notation from the lectures, below is a frequency table that tabulates every possible event in the course of these tests:

	Decision: Retain H_0	Decision: Reject H_0	total
H_0 true	u	v	m_0
H_a true	t	s	$m - m_0$
total	f	r	m

- [18 pt / 18 pts] Record the letter(s) of all the following that are always **true**. At least one will be true.

  (a) All quantities above are random realizations except for m and m_0

  (b) r are the number of Type I errors

  (c) v are the number of false discoveries

  (d) If you were to run these m tests using procedures that increased the power of each test, then the new value of t is less than or equal to the original value of t

  (e) If u is high, this means the measured effects have low practical significance

  (f) If you decrease α and you run all m tests again, the new value of r is less than or equal to the original value of r

  (g) If you decrease α and you run all m tests again, the new value of m_0 could be different than the original value of m_0

  (h) If $m = m_0$, all rejections of H_0 are errors

  (i) If $m = m_0$, then r is a realization from a Binomial(m, α) DGP

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.

Problem 2 [6min] (and 12min will have elapsed) Consider running $m > 1$ hypothesis tests. Following the notation from the lectures, below is a table that tabulates the random variables that model the possible events (denoted by uppercase letters) and the fixed constants (denoted by lowercase letters) in the course of these tests:

	Decision: Retain H_0	Decision: Reject H_0	total
H_0 true	U	V	m_0
H_a true	T	S	$m - m_0$
total	F	R	m

Assume each of the m tests are **independent from each other** and that the levels for each test are $\alpha_1, \alpha_2, \ldots, \alpha_m$.

- [10 pt / 28 pts] Record the letter(s) of all the following that are always **true**. At least one will be true.

(a) $R \sim \text{Binomial}\,(m,\, \alpha)$ if $\alpha = \alpha_1 = \ldots = \alpha_m$

(b) $V \sim \text{Binomial}\,(m_0,\, \alpha)$ if $\alpha = \alpha_1 = \ldots = \alpha_m$

(c) The familywise error rate (FWER) is defined as the probability that the number of false discoveries is one or more over all m tests

(d) If you begin with $\alpha_1 = \ldots = \alpha_m = 5\%$, then FWER is larger than 5%.

(e) If you set $\alpha_1 = \ldots = \alpha_m = 5\%/m$ then FWER is controlled at 5%.

Let α_B denote the Bonferroni-corrected individual test Type I error rate, α_{DS} denote the Dunn-Sidak-corrected individual test Type I error rate and α_S denote the Simes-corrected individual test Type I error rate.

(f) $\alpha_B > \alpha_{DS}$

(g) $\alpha_B > \alpha_S$

(h) You can compute the numeric value of α_B before running the m tests

(i) You can compute the numeric value of α_{DS} before running the m tests

(j) You can compute the numeric value of α_S before running the m tests

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.

Problem 3 [6min] (and 18min will have elapsed) Consider running $m > 1$ hypothesis tests. Following the notation from the lectures, below is a table that tabulates the random variables that model the possible events (denoted by uppercase letters) and the fixed constants (denoted by lowercase letters) in the course of these tests:

	Decision: Retain H_0	Decision: Reject H_0	total
H_0 true	U	V	m_0
H_a true	T	S	$m - m_0$
total	F	R	m

Also, now assume each of the m tests are independent from each other and that the levels for each test are $\alpha_1, \alpha_2, \ldots, \alpha_m$. Assume each test was run and Fisher's p-values were computed for each. The p-values are then sorted from smallest to largest and denoted $p_{(1)}, p_{(2)}, \ldots, p_{(m)}$.

- [7 pt / 35 pts] Record the letter(s) of all the following that are always **true**. At least one will be true.

(a) The values of $p_{(1)}, p_{(2)}, \ldots, p_{(m)}$ are realizations from an iid standard uniform DGP

(b) If $\alpha_1, \alpha_2, \ldots, \alpha_m$ are set via the Simes procedure at FWER = 5%, then the False Discovery Rate (FDR) is controlled at 5% for these m tests

(c) Given numeric values for $p_{(1)}, p_{(2)}, \ldots, p_{(m)}$, you can construct a procedure that will control the FDR in these m tests

(d) The FDR is a function of the estimators employed in the tests and thus beyond the experimenter's control

(e) Assuming FDR = 5% and $r = 800$, then $v = 40$

(f) Assuming FDR = 5% and $r = 800$, then $v \leq 40$

(g) If m is large, procedures that control the FDR at a certain threshold allow more discoveries to be made than procedures that control FWER at that same threshold

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.

Problem 4 [7min] (and 25min will have elapsed) Imagine you are trying to model the prices of diamonds using measurements on the diamonds denoted $x_1, x_2, \ldots$. So you fit a DGP model that looks like price = mean price + ε where ε is an iid $\mathcal{N}(0, \theta_e)$ model and for the mean price you fit two models of the same functional form:

(MOD 1) $\mathcal{N}(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4,\ \theta_e)$

(MOD 2) $\mathcal{N}(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 + \theta_5 x_5 + \theta_6 x_6 + \theta_7 x_7,\ \theta_e)$

- [11 pt / 46 pts] Record the letter(s) of all the following that are always **true**. At least one will be true.

  (a) Model 2 has 7 parameters

  (b) Model 2 has 8 parameters

  (c) Model 2 has 9 parameters

  (d) Model 1 is "nested" in model 2

  (e) Model 1 is the "reduced" model and model 2 is the "full" model

  (f) The maximum likelihood estimate of θ_e will be the same in both models.

  (g) The maximum likelihood estimate of θ_0 will be the same in both models.

  (h) If we were to test if the extra parameters in model 2 were "useful", the null hypothesis would be $H_0 : \theta_0 = \theta_1 = \theta_2 = \theta_3 = \theta_4 = 0$.

  (i) If we were to test if the extra parameters in model 2 were "useful", the null hypothesis would be $H_0 : \theta_5 = \theta_6 = \theta_7 = 0$.

  (j) Regardless of your answers to (g) and (h), the likelihood ratio test could be used to make an exact / approximate decision about H_0

  (k) Regardless of your answers to (g) and (h), the likelihood ratio test could be used to compute an approximate p-value

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.

Problem 5 [12min] (and 37min will have elapsed) Imagine you are trying to model the prices of diamonds using measurements on the diamonds denoted $x_1, x_2, \ldots$. So you fit a DGP model that looks like price = mean price + ε where ε is an iid $\mathcal{N}(0, \theta_e)$ model and for the mean price you fit two models of the same functional form:

(MOD 1) $\mathcal{N}(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4,\ \theta_e)$

(MOD 2) $\mathcal{N}(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 + \theta_5 x_5 + \theta_6 x_6 + \theta_7 x_7,\ \theta_e)$

In the first model, the log likelihood is -471866 and in the second model, the log likelihood is -471675.

- [15 pt / 61 pts] Record the letter(s) of all the following that are always **true**. At least one will be true.

(a) The likelihood ratio test statistic is a draw from a χ^2_1 distribution

(b) The likelihood ratio test statistic is a draw from a χ^2_3 distribution

(c) The likelihood ratio test statistic is a draw from a χ^2_6 distribution

(d) The likelihood ratio test statistic is 191

(e) The likelihood ratio test statistic is -191

(f) The likelihood ratio test statistic is 382

(g) You can conclude at any reasonable α that the extra complexity in model 2 is useful in fitting diamond prices

(h) You can conclude at any reasonable α that the extra complexity in model 2 is *not* useful in fitting diamond prices

(i) The AIC for model 1 is 943746

(j) The AIC for model 1 is 943734

(k) The AICC for model 1 is 943746

(l) The AICC for model 1 is 943736

(m) The AICC for model 1 cannot be computed given the information provided herein

(n) The AIC model selection procedure will select model 2

(o) The information herein can inform you of the practical / clinical significance of these models

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.

Problem 6 [9min] (and 46min will have elapsed) Consider the height data from class. We sampled $n_1 = 10$ men and measured heights in inches: 67, 68, 69, 70, 70, 71, 72, 72, 73 and 73 and $n_2 = 6$ females and measured heights in inches: 59, 60, 63, 64, 64 and 64.

- [12 pt / 73 pts] Record the letter(s) of all the following that are always **true**. At least one will be true.

(a) Given the data above, if we were to test against H_0 : mean of male heights = mean of female heights, then an appropriate test is the exact 2-sample z-test of unequal variances

(b) Given the data above, if we were to test against H_0 : mean of male heights = mean of female heights, then an appropriate test is the exact 2-sample t-test of equal variances

(c) Given the data above, if we were to test against H_0 : mean of male heights = mean of female heights, then an appropriate test is the approximate Welch-Satterthwaite test (the 2-sample t-test of uneqal variances)

(d) If we were to test against H_0 : mean of male heights = mean of female heights, then an appropriate test is the approximate Wald test (the 2-sample z-test)

(e) ~~If we were to test against H_0 : mean of male heights = mean of female heights, then an appropriate test is the approximate bootstrap test~~ Ignore (e)

(f) If we were to test against H_0 : mean of male heights = mean of female heights, then an appropriate test is the approximate permutation test

(g) If we were to test against H_0 : mean of male heights = mean of female heights, then an appropriate test is the approximate 1-sample Kolmogorov-Smirnov test

(h) If we were to test against H_0 : mean of male heights = mean of female heights, then an appropriate test is the approximate 2-sample Kolmogorov-Smirnov test

(i) We have enough data to create empirical CDF functions for both DGPs

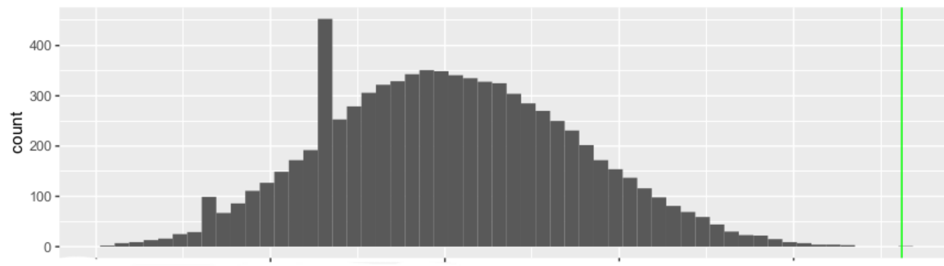
(j) Kolmogorov-Smirnov tests require data from iid normal DGPs

(k) All the tests above assumed sampling with replacement from the two populations

(l) All the tests above employed the population sampling assumption

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.

Problem 7 [8min] (and 54min will have elapsed) Consider the height data from class. We sampled $n_1 = 10$ men and measured heights in inches: 67, 68, 69, 70, 70, 71, 72, 72, 73 and 73 and $n_2 = 6$ females and measured heights in inches: 59, 60, 63, 64, 64 and 64. We would like to test if both samples come from the same DGP and we use a permutation test to do so. We employ the ratio of the sample averages statistic. Here is a frequency histogram of the values of this statistic calculated over $B = 1,000$ resamplings with the green line as the real data's statistic.



- [9 pt / 82 pts] Record the letter(s) of all the following that are always **true**. At least one will be true.

  (a) The ratio of the sample averages is appropriate to test against the H_0 specified

  (b) The center of the null distribution should be close to 1

  (c) The center of the null distribution should be close to 0

  (d) The B employed is too small to allow for valid inference

  (e) The B employed is too large to allow for valid inference

  (f) For all reasonable α, you can conclude the male height DGP is different from the female height DGP

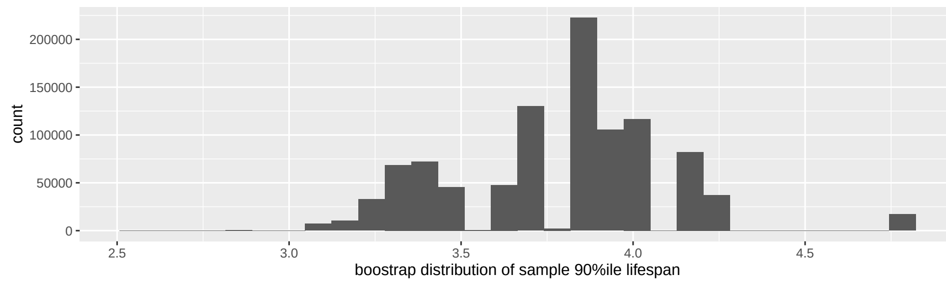
  (g) Given the plot above (and if the x-axis grid was provided), you can approximate the retainment region for this test for $\alpha = 5\%$ by eye

  (h) The p value is approximately 3%

  (i) The 2-sample Kolmogorov-Smirnov test would likely have give us a different decision about H_0

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.

Problem 8 [8min] (and 62min will have elapsed) On midterm I you were given an example of a study that measured the lifespan of rats. Here is the raw data: 1.09, 2.48, 3.08, 2.57, 1.04, 0.87, 4.18, 2.23, 3.22, 1.33, 2.49, 1.69, 3.18, 1.39, 2.52, 4.8, 2.44, 1.47, 2.64, 3.96, 3.08, 2.71, 2.8, 3.4, 3.86, 2.28, 3.65, 3.28, 1.54, 1.94. We are now interested in not the mean lifespan but the 90%ile lifespan. To investiage the 90%ile lifespan, we employ the nonparametric bootstrap. Here is a frequency histogram of the values of the statistic calculated over $B = 1,000,000$ resamplings.



- [8 pt / 90 pts] Record the letter(s) of all the following that are always **true**. At least one will be true.

  (a) The parameter of interest is the population 90%ile lifespan

  (b) The statistic employed is the population 90%ile lifespan

  (c) The boostrap is appropriate to study the 90%ile lifespan

  (d) The boostrap resamplings constitute a "null distribution"

  (e) The size of B is appropriate for this scenario

  (f) An approximate 95% CI for the population 90%ile lifespan is between 2.5yr and 5yr

  (g) If you were testing against H_0 : the population 90%ile lifespan is at most 3.5yr, this test would fail to reject H_0

  (h) If you were testing against H_0 : the population 90%ile lifespan is at most 3.0yr, this test would fail to reject H_0

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.

Problem 9 [11min] (and 73min will have elapsed) On midterm I you were given an example of a study that measured the lifespan of rats in two scenarios: (scenario 1) low dose of magnesium and (scenario 2) high dose of magnesium. In (1) the low dose, $n_1 = 30$, $\bar{x}_1 = 2.57$ and $s_1 = 1.00$. In (2) the high dose, $n_2 = 6$, $\bar{x}_2 = 2.87$ and $s_2 = 1.11$. Let $\theta_2 - \theta_1$ be the difference between the mean high dose lifespan and the mean low dose lifespan. Rats were randomized into the two groups using the completely randomized design. Assume an additive treatment effect.

- [14 pt / 104 pts] Record the letter(s) of all the following that are always **true**. At least one will be true.

(a) The study can be called an "experiment"

(b) This study has two arms

(c) Each rat has two potential outcomes

(d) The investigators are likely interested in a causal estimate for $\theta_2 - \theta_1$

(e) A causal point estimate for $\theta_2 - \theta_1$ is $\bar{x}_2 - \bar{x}_1$

(f) A 95% CI for the treatment effect is $\approx \left[(\bar{x}_2 - \bar{x}_1) \pm 1.96 \times \text{SE}\left[\bar{X}_2 - \bar{X}_1\right]\right]$

(g) Let d be a correctly-calculated point estimate for $\theta_2 - \theta_1$; you can conclude that the high dose of magnesium causes an approximately d difference in lifespan when compared to a low dose of magnesium

(h) Causal estimation is not possible here due to the low sample size n_2

(i) Causal estimation is not possible due to the high standard deviations

(j) If the rats were not randomized, there could be confounders that could bias the causal point estimate

(k) The mean point estimate over all randomizations will not have any bias due to confounders

(l) In this particular assignment, there is no bias due to confounders

(m) The number of assignments in this design is $\binom{30+6}{30}$

(n) The difference in average noise between the two samples is near zero

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.

Problem 10 [14min] (and 87min will have elapsed) Consider $X_1, \ldots, X_n \overset{iid}{\sim} \text{ShiftedParetoI}(1, \theta) := \theta(x+1)^{-\theta-1}$ which has support in positive real numbers only and a parameter space of $\theta > 0$. (You may remember this as the DPG from midterm II). Using calculus you can show that $\mathbb{E}[X] = \frac{1}{\theta-1}$. Let $u := \sum_{i=1}^n \ln(x_i+1)$ and $U := \sum_{i=1}^n \ln(X_i+1)$.

- [17 pt / 121 pts] Record the letter(s) of all the following that are always **true**. At least one will be true.

  (a) This DGP as stated has one parameter

  (b) $u > 0$

  (c) $u > n$

  (d) The likelihood is $u\theta^n$

  (e) The likelihood is $\theta^n \prod_{i=1}^n (x_i+1)^{-\theta-1}$

  (f) The log-likelihood is nu

  (g) The log-likelihood is $n\ln(\theta) - \theta u - u$

  (h) $\mathbb{E}[X] \approx \hat{\hat{\theta}}^{\text{MLE}}$

  (i) $\mathbb{E}[X] \approx \bar{x}$

  (j) $\hat{\hat{\theta}}^{\text{MLE}} = n/u$

  (k) $\hat{\theta}^{\text{MLE}} = n/U$

  (l) The Fisher information is a function of u

  (m) The Fisher information cannot be computed for this DGP since its parameter space is $\neq \mathbb{R}$

  (n) $I\left(\hat{\hat{\theta}}^{\text{MLE}}\right) = \left(\hat{\hat{\theta}}^{\text{MLE}}\right)^{-2}$

  (o) $\hat{\theta}^{\text{MLE}}$ is theoretically gauranteed to be unbiased based on properties of MLEs

  (p) $\ell(\hat{\theta}^{\text{MLE}}, x_1, \ldots, x_n)$ is theoretically gauranteed to be unbiased based on properties of MLEs

  (q) $\ell(\hat{\hat{\theta}}^{\text{MLE}}, x_1, \ldots, x_n)$ is the maximum value of the ℓ function

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.

Problem 11 [9min] (and 96min will have elapsed) Consider $X_1, \ldots, X_n \overset{iid}{\sim} \text{ShiftedParetoI}(1, \theta) := \theta(x+1)^{-\theta-1}$ which has support in positive real numbers only and a parameter space of $\theta > 0$. (You may remember this as the DPG from midterm II). Using calculus you can show that $\mathbb{E}[X] = \frac{1}{\theta-1}$. Let $u := \sum_{i=1}^{n} \ln(x_i + 1)$ and $U := \sum_{i=1}^{n} \ln(X_i + 1)$. The following can be shown for this DGP:

$$\begin{aligned}
\ell(\theta, x_1, \ldots, x_n) &= n \ln(\theta) - \theta u - u, \\
\hat{\hat{\theta}}^{\text{MLE}} &= n/u, \\
I(\theta) &= 1/\theta^2
\end{aligned}$$

Consider the scenario where you are testing against $H_0 : \theta = 1$ at $\alpha = 5\%$.

- [9 pt / 130 pts] Record the letter(s) of all the following that are always **true**. At least one will be true.

  (a) The log-likelihood under H_0 is $-2u$

  (b) The score function is $= n/\theta - u$

  (c) The Wald test rejects when $|n/u - 1| > 1.96$

  (d) The Score test rejects when $|n - u| > 1.96$

  (e) The Likelihood Ratio test rejects when $\sqrt{(n + u - n\ln(n/u))/u} > 1.96$

  (f) The Likelihood Ratio test rejects when $\sqrt{(n + u - n\ln(n/u))/(2u)} > 1.96$

  (g) The Likelihood Ratio test statistic is a draw from an approximate χ^2_2 distribution

  (h) The Likelihood Ratio test statistic is a draw from an approximate χ^2_n distribution

  (i) The Likelihood Ratio test statistic is a draw from an approximate χ^2_{n-1} distribution

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.

Problem 12 [14min] (and 110min will have elapsed) Consider $X_1, \ldots, X_n \stackrel{iid}{\sim} \text{ShiftedParetoI}(1, \theta) := \theta(x+1)^{-\theta-1}$ which has support in positive real numbers only and a parameter space of $\theta > 0$. (You may remember this as the DPG from midterm II). Using calculus you can show that $\mathbb{E}[X] = \frac{1}{\theta - 1}$. Let $u := \sum_{i=1}^n \ln(x_i + 1)$ and $U := \sum_{i=1}^n \ln(X_i + 1)$. The following can be shown for this DGP:

$$
\begin{aligned}
\ell(\theta, x_1, \ldots, x_n) &= n \ln(\theta) - \theta u - u, \\
\hat{\hat{\theta}}^{\text{MLE}} &= n/u, \\
I(\theta) &= 1/\theta^2
\end{aligned}
$$

Consider the scenario where you are testing against $H_0 : \theta = 1$ at $\alpha = 5\%$. You take a sample from the population of size $n = 58$ and you calculate $u = 43.56$. All the following answers are accurate to three decimal places.

- [13 pt / 143 pts] Record the letter(s) of all the following that are always **true**. At least one will be true.

  (a) The Wald test statistic is 2.525
  (b) The Score test statistic is 2.525
  (c) The Likelihood Ratio test statistic is 2.525
  (d) The Wald test statistic is 2.166
  (e) The Score test statistic is 2.166
  (f) The Likelihood Ratio test statistic is 2.166
  (g) The Wald test statistic is 1.896
  (h) The Score test statistic is 1.896
  (i) The Likelihood Ratio test statistic is 1.896
  (j) The Wald, Score and Likelihood Ratio test statistics are equal in this DGP
  (k) The Wald, Score and Likelihood Ratio test statistics are equal in all DGPs
  (l) $CI_{\theta,95\%} \approx [0.989, 1.674]$ to the nearest three decimals
  (m) $CI_{\theta,95\%} \approx [1.074, 1.589]$ to the nearest three decimals

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.