Math 369 / 650 Fall 2020
Midterm Examination Two

Professor Adam Kapelner

Wednesday, November 11, 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 75 minutes (variable time per question) and closed-book. You are allowed **one** page (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 [5min] (and 5min will have elapsed) These are questions on method of moments estimators.

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

(a) For any DGP, $\hat{\theta}^{\text{MM}}$ will be a function of the data.

(b) For any DGP, $\hat{\theta}^{\text{MM}}$ will be a function of the data and of θ.

(c) For any DGP, its fifth moment is $\mathbb{E}\left[X^5\right]$ if it exists.

(d) For any iid DGP with a finite expectation, $\hat{\theta} = \bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$ is an MM estimator for the DGP's expectation.

(e) For any iid DGP with a finite variance, $\hat{\theta} = \hat{\sigma}^2 := \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ is an MM estimator for the DGP's variance.

(f) Let $\hat{\theta}^{\text{MM}}$ be an estimator for θ. Its estimate $\hat{\hat{\theta}}^{\text{MM}}$ must be a legal value in the parameter space of θ.

(g) For an iid DGP with a finite variance where $\theta := \mathbb{E}\left[X\right]/\mathbb{SD}\left[X\right]$, then $\hat{\theta}^{\text{MM}} = \bar{X}/\sqrt{\hat{\sigma}^2}$ where $\bar{X}$ and $\hat{\sigma}^2$ are defined in (d) and (e).

Consider a DPG $X_1, \ldots, X_n \overset{iid}{\sim} \text{ShiftedParetoI}(1, \theta) := \theta(x+1)^{-\theta-1}$ which has support in positive numbers only. Using calculus you can show that $\mathbb{E}\left[X\right] = \frac{1}{\theta - 1}$.

(h) $\hat{\theta}^{\text{MM}} = \bar{X}$

(i) $\hat{\theta}^{\text{MM}} = 1/\bar{X}$

(j) $\hat{\theta}^{\text{MM}} = 1/\bar{X} + 1$

(k) $\hat{\theta}^{\text{MM}} = (\bar{X} + 1)/\bar{X}$

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

Problem 2 [13min] (and 18min will have elapsed) Consider the DPG from the previous problem, $\overset{iid}{\sim}$ ShiftedParetoI$(1, \theta) := \theta(x+1)^{-\theta-1}$ which has support in positive numbers only and consider a dataset $x_1, \ldots, x_n$ to be realizations from this DGP.

- [19 pt / 30 pts] Record the letter(s) of all the following that are **true**. At least one will be true.

(a) The likelihood for all the data is $= \theta(x_1, \ldots, x_n + 1)^{-\theta-1}$.

(b) The likelihood for all the data is $= \prod_{i=1}^n \theta(x_i + 1)^{-\theta-1}$.

(c) The log likelihood for all the data is $= \prod_{i=1}^n \theta(\ln(x_i) + 1)^{-\theta-1}$.

(d) The log likelihood for all the data is $= \ln\left(\prod_{i=1}^n \theta(x_i + 1)^{-\theta-1}\right)$.

(e) The log likelihood for all the data is $= n\ln(\theta) + \ln\left((-\theta - 1)\prod_{i=1}^n (x_i + 1)\right)$.

(f) The log likelihood for all the data is $= n\ln(\theta) + (-\theta - 1)\sum_{i=1}^n \ln(x_i + 1)$.

(g) The score function for all the data is the derivative of the log likelihood function.

(h) The deriviative of the log likelihood for all the data is $= \frac{n}{\theta} - \sum_{i=1}^n \ln(x_i + 1)$.

(i) Assuming (h) is true, $\hat{\hat{\theta}}^{\text{MLE}} = n \Big/ \sum_{i=1}^n \ln(x_i + 1)$.

(j) The Fisher information for this DGP is $I(\theta) = 1$.

(k) The Fisher information for this DGP is $I(\theta) = n/\theta^2$.

(l) The Fisher information for this DGP is $I(\theta) = -1/\theta^2$.

(m) The Fisher information for this DGP is $I(\theta) = 1/\theta^2$.

(n) The variance of any unbiased estimator for θ for this DGP must be at least $1/n$.

(o) The variance of any unbiased estimator for θ for this DGP must be at least θ/n.

(p) The variance of any unbiased estimator for θ for this DGP must be at least θ^2/n.

(q) To show that $\hat{\theta}^{\text{MLE}}$ is/isn't the UMVUE for θ involves simple algebra / calculus on information available to you on this page.

(r) To show that $\hat{\theta}^{\text{MLE}}$ is/isn't the UMVUE for θ involves a lot of algebra / calculus but the information you require is available to you on this page.

(s) To show that $\hat{\theta}^{\text{MLE}}$ is/isn't the UMVUE for θ is impossible given the information available to you on this page.

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

Problem 3 [6min] (and 24min will have elapsed) Consider the DPG from the previous problem, $\overset{iid}{\sim}$ ShiftedParetoI$(1, \theta) := \theta(x+1)^{-\theta-1}$ which has support in positive numbers only and consider a dataset $x_1, \ldots, x_n$ to be realizations from this DGP. Let σ^2 denote the variance of this DGP model. You can show that $\hat{\hat{\theta}}^{\text{MLE}} = n \Big/ \sum_{i=1}^{n} \ln(x_i + 1)$ and $I(\theta) = 1/\theta^2$.

- [13 pt / 43 pts] Record the letter(s) of all the following that are **true**. At least one will be true.

  (a) $\mathbb{V}\text{ar}\left[\hat{\theta}^{\text{MLE}}\right] = \theta^2/n$.

  (b) $\hat{\theta}^{\text{MLE}}$ is normally distributed.

  (c) $\hat{\theta}^{\text{MLE}}$ is distributed as a Student's t distribution.

  (d) $\hat{\theta}^{\text{MLE}}$ is an asymptotically normally estimator.

  (e) $\hat{\theta}^{\text{MLE}} \overset{.}{\sim} \mathcal{N}(0, 1)$.

  (f) $\hat{\theta}^{\text{MLE}} \overset{.}{\sim} \mathcal{N}(\theta, \sigma^2/n)$.

  (g) $\hat{\theta}^{\text{MLE}} \overset{.}{\sim} \mathcal{N}\left(\theta, \frac{1}{n\theta^2}\right)$.

  (h) You can use the fact in (g) to create a confidence interval for θ (i.e. a function of $x_1, \ldots, x_n$ and constants).

  (i) $\hat{\theta}^{\text{MLE}} \overset{.}{\sim} \mathcal{N}\left(\theta, \left(\frac{\theta}{\sqrt{n}}\right)^2\right)$.

  (j) You can use the fact in (i) to create a confidence interval for θ (i.e. a function of $x_1, \ldots, x_n$ and constants).

  (k) $\hat{\theta}^{\text{MLE}} \overset{.}{\sim} \mathcal{N}\left(\hat{\hat{\theta}}^{\text{MLE}}, \left(\frac{\hat{\hat{\theta}}^{\text{MLE}}}{\sqrt{n}}\right)^2\right)$.

  (l) You can use the fact in (k) to create a confidence interval for θ (i.e. a function of $x_1, \ldots, x_n$ and constants).

  (m) $\hat{\theta}^{\text{MLE}}$ can provide arbitrary precision to measure θ given a higher sample size n.

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

Problem 4 [6min] (and 30min will have elapsed) Consider $x_1, \ldots, x_n$ to be realizations from an iid DGP, $\hat{\theta}$ to be an unbiased estimator for θ and assume the conditions needed to prove the Cramer-Rao Lower Bound (CRLB). A quantity primed (e.g. ℓ') denotes differentiation with respect to θ.

- [7 pt / 50 pts] Record the letter(s) of all the following that are **true**. At least one will be true.

(a) $\ell'\left(\hat{\theta}; x_1, \ldots, x_n\right) = 0$

(b) $\mathbb{E}\left[\ell\left(\theta; X_1, \ldots, X_n\right)\right] = 0$

(c) $\mathbb{E}\left[\ell\left(\theta; X_1, \ldots, X_n\right)\right] = n\mathbb{E}\left[\ell\left(\theta; X\right)\right]$

(d) $\mathbb{E}\left[\ell\left(\theta; X_1, \ldots, X_n\right)^2\right] = 0$

(e) $\mathbb{E}\left[\ell'\left(\theta; X_1, \ldots, X_n\right)^2\right] = 0$

(f) $\dfrac{\partial}{\partial\theta}\left[\mathbb{E}\left[\ell\left(\theta; X_1, \ldots, X_n\right)\right]\right] = 0$

(g) $\mathbb{V}\text{ar}\left[\ell\left(\theta; X_1, \ldots, X_n\right)^2\right] = 0$

(h) $\mathbb{V}\text{ar}\left[\ell'\left(\theta; X_1, \ldots, X_n\right)^2\right] = nI(\theta)$

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

Problem 5 [5min] (and 35min will have elapsed) These are some questions about some theorems we discussed in class. Let $X_1, X_2, ...$ be rv's indexed also by n, the sample size, and $a_1, a_2, ...$ be positive constants.

- [6 pt / 56 pts] Record the letter(s) of all the following that are **true**. At least one will be true.

(a) If $X_1 \xrightarrow{p} a_1$, then $\ln(X_1) \xrightarrow{p} \ln(a_1)$.

(b) If $X_1 \xrightarrow{p} a_1$, then $\dfrac{1}{X_1} \xrightarrow{p} \dfrac{1}{a_1}$.

(c) If $X_1 \xrightarrow{p} a_1$ and $\dfrac{X_2 - \mu}{a_1} \xrightarrow{d} \mathcal{N}(0, 1)$ then $\dfrac{X_2 - \mu}{X_1} \xrightarrow{d} \mathcal{N}(0, 1)$

(d) If $\dfrac{X_1 - \mu}{a_1} \sim \mathcal{N}(0, 1)$ then $\dfrac{(X_1 - \mu)^2}{a_1^2} \sim \mathcal{N}(0, 1)$

(e) If $\dfrac{X_1 - \mu}{a_1} \sim \mathcal{N}(0, 1)$ then $\dfrac{(X_1 - \mu)^2}{a_1^2} \sim \chi^2_n$

(f) If you do a one-sample t test with $n = 20$ and get a standardized estimate of -2.30. If the square of this estimate is less than 95% quantile of the $F_{1,19}$ distribution, then you retain H_0.

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

Problem 6 [7min] (and 42min will have elapsed) According to Benford's Law in Base 10, in any measurement setting that spans orders of magnitude, the first digit of measurements is more likely to be a one than a two, a two than a three, etc. This phenomenon is ubiquitous and can be used to describe measurements for city populations, surface area of rivers, dollar amounts on tax returns, etc. The distribution is on the first row of the table below.

Here we investigate possible fraud in the 2009 Iranian election, a topic of many academic papers. We examine the vote counts in the 366 districts for one of the five main candidates and count the number of districts who's count had first digit = 1, first digit = 2, ..., first digit = 9. The tally is the second row of the table below.

	First Digit is $x =$									
	1	2	3	4	5	6	7	8	9	Total
Benford's Law in Base 10 $p_X(x) =$	.301	.176	.125	.097	.079	.067	.058	.051	.046	1.000
Observed Count	125	57	44	29	24	16	41	13	17	366
Row 3 Name	110.17	64.42	45.75	35.50	28.91	24.52	21.23	18.67	16.84	?
Row 4 Name	2.00	0.85	0.07	1.19	0.83	2.96	18.42	1.72	0.00	?

We wish to test if the voting counts *differ* from Benford's Law at $\alpha = 5\%$. Let θ_1 denote the true probability of a count having a first digit = 1, θ_2 denote the true probability of a count having a first digit = 2, ..., θ_9 denote the true probability of a count having a first digit = 9. Note: $F_{\chi^2_8}(15.51) = F_{\chi^2_9}(16.92) = F_{\chi^2_{24}}(36.42) = F_{\chi^2_{27}}(40.11) = 95\%$.

- [9 pt / 65 pts] Record the letter(s) of all the following that are **true**. At least one will be true.

(a) This test is a "goodness of fit" test.

(b) This test is an "independence" of ≥ 2 events test.

(c) The number of parameters in Benford's Law in Base 10 is 8.

(d) $H_0 : \theta_1 = \theta_2 = \ldots = \theta_9 = 1/9$.

(e) H_0 : The θ_k's are equal to the values in the table's first row.

(f) $H_a : \theta_1 = \theta_2 = \ldots = \theta_9 = 1/9$.

(g) H_a : at least one θ_k is not equal to the value in the table's first row.

(h) The name of row 3 could be "expected number of first digits in the district counts under Benford's Law in Base 10".

(i) The name of row 4 could be "expected number of first digits in the district counts under Benford's Law in Base 10".

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

Problem 7 [7min] (and 49min will have elapsed) **This header is the same as in the previous problem.** According to Benford's Law in Base 10, in any measurement setting that spans orders of magnitude, the first digit of measurements is more likely to be a 1 than a 2, a 2 than a 3, etc. This phenomenon is ubiquitous and can be used to describe measurements for city populations, surface area of rivers, dollar amounts on tax returns, etc. The distribution is on the first row of the table below.

Here we investigate possible fraud in the 2009 Iranian election, a topic of many academic papers. We examine the vote counts in the 366 districts for one of the five main candidates and count the number of districts who's count had first digit = 1, first digit = 2, ..., first digit = 9. The tally is the second row of the table below.

	First Digit is $x =$									
	1	2	3	4	5	6	7	8	9	Total
Benford's Law in Base 10 $p_X(x) =$	.301	.176	.125	.097	.079	.067	.058	.051	.046	1.000
Observed Count	125	57	44	29	24	16	41	13	17	366
Row 3 Name	110.17	64.42	45.75	35.50	28.91	24.52	21.23	18.67	16.84	?
Row 4 Name	2.00	0.85	0.07	1.19	0.83	2.96	18.42	1.72	0.00	?

We wish to test if the voting counts *differ* from Benford's Law at $\alpha = 5\%$. Let θ_1 denote the true probability of a count having a first digit = 1, θ_2 denote the true probability of a count having a first digit = 2, ..., θ_9 denote the true probability of a count having a first digit = 9. Note: $F_{\chi^2_8}(15.51) = F_{\chi^2_9}(16.92) = F_{\chi^2_{24}}(36.42) = F_{\chi^2_{27}}(40.11) = 95\%$.

- [7 pt / 72 pts] Record the letter(s) of all the following that are **true**. At least one will be true.

(a) The Pearson chisq statistic has an estimate = 366.00.

(b) The Pearson chisq statistic has an estimate = 28.04.

(c) The Pearson chisq statistic estimator here is asymptotically χ^2_{24}.

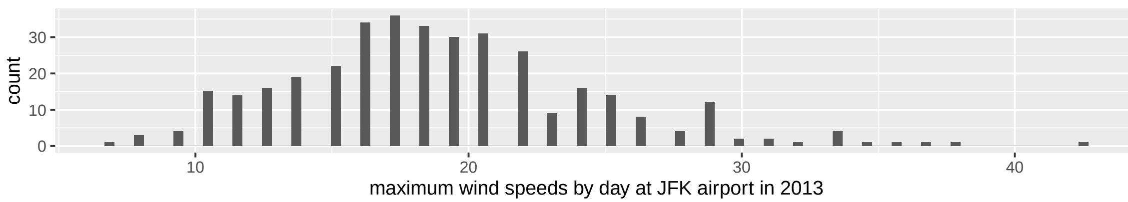
(d) The Pearson chisq statistic estimator here is asymptotically χ^2_8.

(e) If (d) were to be true, the retainment region for the estimate would be $[-15.51, +15.51]$.

(f) H_0 is rejected and we conclude that the counts for this specific candidate is *not* Benford-Law-distributed.

(g) Assuming (f), the digit that is most incongruent with Benford's law is $x = 9$.

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



Problem 8 [10min] (and 59min will have elapsed) Above is a histogram of wind speeds at JFK airport measured at midnight for every day in 2013. We fit eight different iid DGPs / models to this data. Below are the models, the maximum likelihood estimates for their parameter(s) and their log likelihood value estimates:

	Name of Model / DGP							
	Exponential	Normal	Weibull	Gamma	Gumbel	Gompertz	Frechet	Generalized Logistic
$\hat{\hat{\theta}}_1^{\text{MLE}}$	0.05	18.97	3.52	11.71	15.17	0.14	3.18	3.49
$\hat{\hat{\theta}}_2^{\text{MLE}}$		5.59	21.01	0.62	4.84	0.05	15.59	12.22
$\hat{\hat{\theta}}_3^{\text{MLE}}$								4.03
$\ell(\hat{\hat{\theta}}_1^{\text{MLE}}, \ldots, \hat{\hat{\theta}}_{K_m}^{\text{MLE}}; x_1, \ldots, x_n)$	-1435.12	-1142.72	-1148.11	-1129.28	-1143.31	-1200.26	-1174.37	-1129.65

- [9 pt / 81 pts] Record the letter(s) of all the following that are **true**. At least one will be true.

(a) All of these models have the same number of parameters K.

(b) The true model will be one of these eight models candidates.

(c) According to the log-likelihood estimates, the best fitting model candidate is the Gamma.

(d) According to the AICC metric, the best fitting model candidate is the Generalized Logistic.

(e) The asymptotic bias on the true log-likelihood for the normal model is the same as the bias for the Gompertz model.

(f) The AIC metric for the generalized logistic model candidate is 2263.30.

(g) The AIC metric for the exponential model candidate is 2872.24.

(h) Assuming (b), the probability the exponential model is the true model is negligibly small.

(i) AICC values will not significantly differ from the AIC values.

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

Problem 9 [16min] (and 75min will have elapsed) In the "Topiramate for the Treatment of Binge Eating Disorder Associated With Obesity: A Placebo-Controlled Study", the researchers were also interested in adverse effects of the drug Topiramate. One such bad effect is "Upper Respiratory Tract Infection" which we call *infection*. The experimental results are below.

- In the $n_T = 202$ Topiramate group, there were 75 cases of infection and
- in the $n_C = 202$ placebo group that did not take Topiramate, there were 40 cases of infection.

Let θ_T be the true proportion of infection in Topiramate-takers and θ_C be the true proportion of infection among non Topiramate-takers. All numbers below are rounded to the nearest 3 digits.

- [13 pt / 94 pts] Record the letter(s) of all the following that are **true**. At least one will be true.

  (a) When testing $H_a : \theta_T \neq \theta_C$ at $\alpha = 5\%$, a retainment region for H_0 is equal to or is approximately $[-0.086, 0.086]$.

  (b) When testing $H_a : \theta_T \neq \theta_C$ at $\alpha = 5\%$, a retainment region for H_0 is equal to or is approximately $[-0.088, 0.088]$.

  (c) $CI_{\theta_T - \theta_C, 95\%}$ is equal to or is approximately $[0.087, 0.260]$.

  (d) There is a 95% chance that the CI estimate in (c) contains the true mean difference $\theta_T - \theta_C$.

  (e) The estimate of the odds against infection in the Topiramate group is 1.693.

  (f) The estimate of the odds against infection in the Topiramate group is 0.591.

  (g) When testing $H_a : \dfrac{1 - \theta_T}{\theta_T} \neq \dfrac{1}{2}$ at $\alpha = 5\%$, a retainment region for H_0 is equal to or is approximately $[0.224, 0.776]$.

  (h) The estimate of the risk ratio of infection in the Topiramate group vs the control group is 0.173.

  (i) The estimate of the risk ratio of infection in the Topiramate group vs the control group is 1.875.

  (j) $CI_{\theta_T/\theta_C, 95\%}$ is equal to or is approximately $[1.255, 2.495]$.

  (k) $CI_{\theta_T/\theta_C, 95\%}$ is equal to or is approximately $[-0.088, 0.088]$.

  (l) It would be possible to create an exact CI for the risk ratio of infection in the Topiramate group vs the control group given the data here and the concepts learned in class.

  (m) The CI estimator in (j) is more approximate in coverage probability than the CI estimator in (c).

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