Math 340 / 640 Fall 2024
Midterm Examination Two

Professor Adam Kapelner

November 14, 2024

Full Name ____

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.

I acknowledge and agree to uphold this Code of Academic Integrity.

____ ____________________

signature date

Instructions

This exam is 110 minutes (variable time per question) and closed-book. You are allowed **one** page (front and back) of a "cheat sheet", blank scrap paper (provided by the proctor) and a graphing calculator (which is not your smartphone). Please read the questions carefully. Within each problem, I recommend considering the questions that are easy first and then circling back to evaluate the harder ones. No food is allowed, only drinks.

Problem 1 Let $X_1, X_2, \ldots \overset{iid}{\sim}$ some continuous random variable with $0 < \mu < \infty$ and $0 < \sigma^2 < \infty$. We also use the standard notation

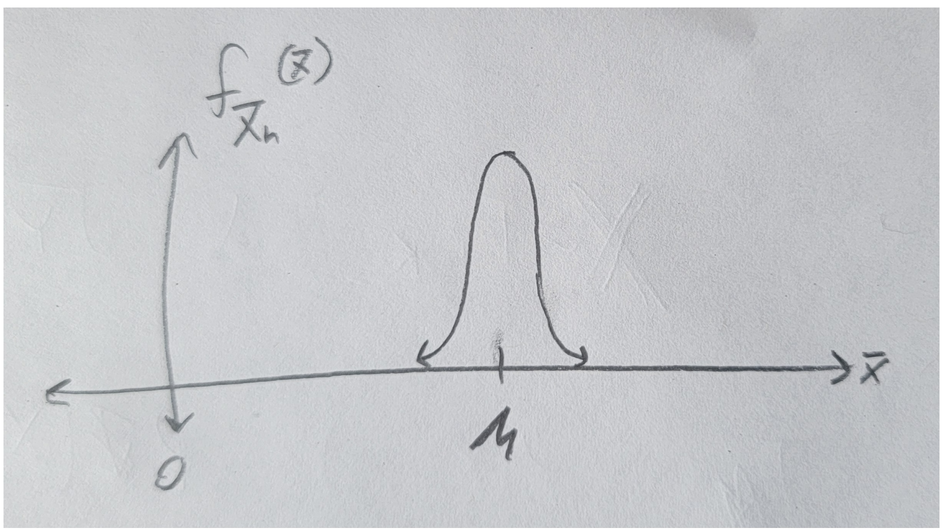
$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S_n^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

(a) [6 pt / 6 pts] Use Chebyshev's inequality to prove the WLLN, i.e., that $\bar{X}_n \xrightarrow{p} \mu$.

$$\forall \epsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}\left(|\bar{X}_n - \mu| \geq \epsilon\right) \leq \lim_{n\to\infty} \frac{\mathbb{V}\text{ar}\left[\bar{X}_n\right]}{\epsilon} = \lim_{n\to\infty} \frac{\sigma^2/n}{\epsilon} = \frac{\sigma^2}{\epsilon}\lim_{n\to\infty}\frac{1}{n} = 0 \quad \checkmark$$

(b) [6 pt / 12 pts] Assume n is large. Plot the approximate PDF of $\bar{X}_n$ denoted $f_{\bar{X}_n}(\bar{x})$. Label the horizontal and vertical axes. Indicate critical value(s) on the axes.

By the WLLN, $\bar{X}_n \xrightarrow{p} \mu$, which means that as n gets larger, all the probability "piles up" near μ and since we assumed $\mu > 0$, we should get something like this:



(c) [8 pt / 20 pts] Using the fact that $S_n^2 \xrightarrow{p} \sigma^2$ and the theorems we learned about in class, show the following fact about the ratio Student was investigating at the turn of the century:

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

Using the fact given above, $S \xrightarrow{p} \sigma$ by the CMT where $g(t) = \sqrt{t}$. Then, $\frac{\sigma}{S} \xrightarrow{p} 1$ by the CMT where $g(t) = t/\sigma$.

$$\frac{\bar{X}_n - \mu}{\frac{S}{\sqrt{n}}} = \overbrace{\underbrace{\frac{\sigma}{S}}_{\text{see above}} \underbrace{\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}}_{\text{by the CLT this term } \xrightarrow{d} \mathcal{N}(0,1)} \xrightarrow{d} \mathcal{N}(0,\, 1)}^{\text{by Slutsky's A theorem}}$$

For the remainder of the questions in this problem, assume that

$$X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu,\, \sigma^2\right)$$

(d) [3 pt / 23 pts] Let $Z_i := (X_i - \mu)/\sigma$. How are following distributed?

$$Z_1, \ldots, Z_n \overset{iid}{\sim} \mathcal{N}(0,\, 1)$$

(e) [4 pt / 27 pts] The following has a brand name distribution. Find this distribution and its parameter(s).

The ratio of two independent standard normals was proved in class to be a Cauchy(0,1). Below is just a linear transformation yielding

$$17 + 37\frac{Z_7}{Z_{11}} \sim \text{Cauchy}\,(17,\, 37)$$

(f) [4 pt / 31 pts] The following has a brand name distribution. Find this distribution and its parameter(s).

$$\frac{\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{\frac{n-1}{\sigma^2} S_n^2}{n-1}}} = \frac{\bar{X}_n - \mu}{\frac{S_n}{\sqrt{n}}} \sim T_{n-1}$$

Where the equality follows from algebraic simplification.

(g) [7 pt / 38 pts] The following has a brand name distribution. Find this distribution and its parameter(s).

$$\frac{\frac{n-1}{\sigma^2} S_n^2}{\left(\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2} =: R = \frac{U}{V^2}$$

We know from class that $U \sim \chi^2_{n-1} = \text{Gamma}\left(\frac{n-1}{2},\, \frac{1}{2}\right)$ and $V \sim \mathcal{N}(0,\, 1)$ which imples $V^2 \sim \chi^2_1 = \text{Gamma}\left(\frac{1}{2},\, \frac{1}{2}\right)$. By Cochran's theorem, U and V are independent. Thus, R is a ratio of two independent gamma rv's that share their β parameter value. We proved in class that such a ratio is beta prime with first parameter given by the first parameter of the numerator and second parameter given by the first parameter of the denominator,

$$R \sim \text{BetaPrime}\left(\frac{n-1}{2}, \frac{1}{2}\right)$$

(h) [7 pt / 45 pts] The following expression is distributed as a brand name variable. Find the brand name variable and its parameter values:

$$\begin{aligned} n\bar{X}_n^2 - 2n\mu\bar{X}_n + n\mu^2 + nS_n^2 - S_n^2 &= n\left(\bar{X}_n - \mu\right)^2 + (n-1)S_n^2 \\ &= \frac{\sigma^2}{\sigma^2}\left(n\left(\bar{X}_n - \mu\right)^2 + (n-1)S_n^2\right) \\ &= \sigma^2\left(\left(\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 + \frac{n-1}{\sigma^2}S_n^2\right) \\ &\sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2\sigma^2}\right) \end{aligned}$$

Let's examine the term inside the parentheses on the third line. By Cochran's theorem, the first term is χ^2_1 and the second term is χ^2_{n-1} and both terms are independent. Thus the entire parenthesis is distributed as $\chi^2_n = \text{Gamma}\left(\frac{n}{2},\, \frac{1}{2}\right)$. Hence we are scaling a gamma by σ^2 which yields another gamma distribution with its β parameter divided by the scaling value.

Problem 2 Some theoretical problems.

(a) [7 pt / 52 pts] Let $X \sim \text{Weibull}(k, \lambda)$. Find the most succinct expression as possible for $\mathbb{E}[X]$. I have started you off below. Hint: the solution contains a gamma function.

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x) dx = \int_{\mathbb{R}} x \left(k\lambda(\lambda x)^{k-1} e^{-(\lambda x)^k}\right) \mathbb{1}_{x>0}\, dx = k\lambda^k \int_0^\infty x^k e^{-\lambda^k x^k} dx$$

Let $u = x^k$ thus $x = u^{1/k}$, $du/dx = kx^{k-1}$, $dx = 1/(kx^{k-1})du$ and the bounds of the integration do not change. Using this substitution,

$$\mathbb{E}[X] = k\lambda^k \int_0^\infty x^k e^{-\lambda^k u} 1/(kx^{k-1}) du = \lambda^k \int_0^\infty x e^{-\lambda^k u} du = \lambda^k \int_0^\infty u^{1/k} e^{-\lambda^k u} du$$

Since $u^{1/k} = u^{1/k+1-1}$, we have a gamma-like integral with $\alpha = 1/k + 1$ and $c = \lambda^k$:

$$\mathbb{E}[X] = \lambda^k \int_0^\infty u^{\alpha-1} e^{-cu} du = \lambda^k \frac{\Gamma(\alpha)}{c^\alpha} = \lambda^k \frac{\Gamma(1/k+1)}{(\lambda^k)^{1/k+1}} = \lambda^k \frac{\Gamma(1/k+1)}{\lambda^{1+k}} = \frac{1}{\lambda}\Gamma(1/k+1)$$

(b) [7 pt / 59 pts] Let $X \sim \text{ParetoI}(k, \lambda)$. Find the distribution of $Y = X \mid X > c$ where $c > k$. If it's a brand-name rv, mark it as its brand name and find its parameters.

The trick here is to use the survival function of the ParetoI, $S_X(x) := 1 - F_X(x) = \left(\frac{k}{x}\right)^\lambda$

$$\begin{aligned} S_Y(y) &:= \mathbb{P}(Y > y) = \mathbb{P}(X > y \mid X > c) = \frac{\mathbb{P}(X > y, X > c)}{\mathbb{P}(X > c)} = \frac{\mathbb{P}(X > y)}{\mathbb{P}(X > c)} \\ &= \frac{S_X(y)}{S_X(c)} = \frac{\left(\frac{k}{y}\right)^\lambda}{\left(\frac{k}{c}\right)^\lambda} = \left(\frac{c}{y}\right)^\lambda \quad \Rightarrow \quad Y \sim \text{ParetoI}(c, \lambda) \end{aligned}$$

Survival functions characterize a distribution (just like CDF's, PDF's, PMF's, chf's).

(c) [3 pt / 62 pts] Let $\boldsymbol{X} \sim \text{Mult}_3\left(9, \frac{1}{3}\mathbf{1}_3\right)$. Find the distribution of $Y = \mathbf{1}_k^\top \boldsymbol{X}$ where k is the appropriate dimension to make the vector operation legal.

$$Y = \mathbf{1}_3^\top \boldsymbol{X} = X_1 + X_2 + X_3 \sim \text{Deg}(9)$$

(d) [5 pt / 67 pts] Let $\boldsymbol{X} \sim \text{Mult}_3\left(9, \frac{1}{3}\mathbf{1}_3\right)$. Find $\mathbb{V}\text{ar}\left[\boldsymbol{X}\right]$. Simplify your answer so that it is a function of $\boldsymbol{I}_3$, the identity matrix and $\boldsymbol{J}_3$, the matrix of all ones.

Note: $\boldsymbol{p} = \frac{1}{3}\mathbf{1}_3$ which means $p_1 = p_2 = p_3 = \frac{1}{3}$ thus we just plug in these values to the formula from class and simplify:

$$\mathbb{V}\text{ar}\left[\boldsymbol{X}\right] = n \begin{bmatrix} p_1(1-p_1) & -p_1p_2 & -p_1p_3 \\ -p_2p_1 & p_2(1-p_2) & -p_2p_3 \\ -p_3p_1 & -p_3p_2 & p_3(1-p_3) \end{bmatrix} = \frac{9}{9}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} = 3\boldsymbol{I}_3 - \boldsymbol{J}_3$$

(e) [7 pt / 74 pts] If $X \sim \text{Laplace}(0,1)$ and $Y = e^X$. Find $f_Y(y)$ in simplest form.

$$\begin{aligned} X &\sim \text{Laplace}(0,1) := \frac{1}{2}e^{-|x|}\mathbb{1}_{x\in\mathbb{R}}, \quad X = \ln(Y) = g^{-1}(Y) \quad \Rightarrow \quad \frac{d}{dy}\left[g^{-1}(y)\right] = \frac{1}{y} \\ f_Y(y) &= f_X(g^{-1}(y))\left|\frac{d}{dy}\left[g^{-1}(y)\right]\right| = \frac{1}{2}e^{-|(\ln(y))|}\mathbb{1}_{\ln(y)\in\mathbb{R}}\frac{1}{|y|} = \frac{1}{2y}e^{-|\ln(y)|}\mathbb{1}_{y\in(0,\infty)} \\ &= \frac{1}{2}\begin{cases} y^{-2} & \text{if} \quad y \geq 1 \\ 1 & \text{if} \quad y \in (0,1) \\ 0 & \text{if} \quad y \leq 0 \end{cases} \end{aligned}$$

(The piecewise function notation is the simplest form but it is not needed for full credit).

(f) [4 pt / 78 pts] If $X \sim F_{2\alpha,2\beta}$, find $k_X(x)$ in simplest form.

$$f_X(x) \propto x^{\frac{2\alpha}{2}-1}\left(1+\frac{2\alpha}{2\beta}x\right)^{-\frac{2\alpha+2\beta}{2}} \mathbb{1}_{x>0} = x^{\alpha-1}\left(1+\frac{\alpha}{\beta}x\right)^{-(\alpha+\beta)} \mathbb{1}_{x>0} = k_X(x)$$

(g) [7 pt / 85 pts] If $X \sim F_{2\alpha,2\beta}$, find the distribution of $Y = \frac{\alpha}{\beta}X$. If it is a brand-name distribution, indicate which one and the values of its parameter(s). Hint: use kernels.

Because the parameter space of the F distribution is two positive parameters, we know $\alpha > 0$ and $\beta > 0$. We make use of $k_X(x)$ from the previous problem

$$\begin{aligned} f_Y(y) &= \frac{1}{\frac{\alpha}{\beta}} f_X\left(\frac{1}{\frac{\alpha}{\beta}}y\right) \propto f_X\left(\frac{\beta}{\alpha}y\right) \propto k_X\left(\frac{\beta}{\alpha}y\right) = \left(\frac{\beta}{\alpha}y\right)^{\alpha-1}\left(1+\frac{\alpha}{\beta}\left(\frac{\beta}{\alpha}y\right)\right)^{-(\alpha+\beta)} \mathbb{1}_{\left(\frac{\beta}{\alpha}y\right)>0} \\ &\propto y^{\alpha-1}(1+y)^{-(\alpha+\beta)}\mathbb{1}_{y>0} = \frac{y^{\alpha-1}}{(1+y)^{\alpha+\beta}}\mathbb{1}_{y>0} \propto \text{BetaPrime}(\alpha,\beta) \end{aligned}$$

(h) [5 pt / 90 pts] If $X \sim \text{Bernoulli}(p) := p^x(1-p)^{1-x}\mathbb{1}_{x\in\{0,1\}}$, find the PMF of $Y = e^X$ in simplest form. Your answer should contain a $\mathbb{1}_{y\in\mathbb{S}_Y}$ term where $\mathbb{S}_Y$ is an explicit set.

$$\begin{aligned} X &= \ln(Y) = g^{-1}(Y) \\ p_Y(y) &= p_X(g^{-1}(y)) \\ &= p^{\ln(y)}(1-p)^{1-\ln(y)}\mathbb{1}_{\ln(y)\in\{0,1\}} \\ &= p^{\ln(y)}(1-p)^{1-\ln(y)}\mathbb{1}_{y\in\{1,e\}} \end{aligned}$$

(i) [3 pt / 93 pts] If $X \sim \text{ExtNegBin}(k, p)$ where $k > 0$ and $p \in (0, 1)$, what is $\mathbb{S}_X$? $\mathbb{N}_0$

(j) [7 pt / 100 pts] Each of the following distributions will be either "waiting time distributions", "error distributions" or neither (but not both). Underline those that are waiting time distributions. Draw a rectangle box around all the that are error distributions. Do not mark those that are neither waiting nor nor error distributions.

$\underline{\text{Erlang}(4,\ 0.3)}$ $\underline{\text{Gamma}(\pi,\ \text{e})}$ $\underline{\chi^2_{17}}$ $\mathcal{N}(1,\ 2^2)$ $\boxed{T_7}$ $\underline{F_{3,6}}$ $\underline{\text{Weibull}(1.6,\ 0.9)}$

$\text{Mult}_3(18,\ \frac{1}{3}\mathbf{1}_3)$ $\underline{\text{BetaPrime}(6.1{,}9.7)}$ $\boxed{\text{Logistic}(0{,}1)}$ $\boxed{\text{Laplace}(0,\ \pi)}$

$\underline{\text{LogNormal}(1{,}2)}$ $\text{ParetoI}(7{,}17)$ $\text{Gumbel}(0{,}1)$