Econ 241A Probability, Statistics and Econometrics

Fall 2013

Final Exam

- You have 3 hrs to complete this exam
- The exam has two parts. Part I requires to solve all problems. Part II allows you to choose between two problems. Please solve just one problem in Part II. If you answer both, only the lowest grade out of the two will be taken into account.
- The last page of the exam has a list of pmf's and pdf's that you may (or may not) need to use throughout the exam.

Part I

1. (5) Let $X_1, ..., X_n$ be a random sample from a Pareto distribution with parameter $\alpha > 2$:

$$f_{X_i}(x) = \frac{\alpha}{x^{\alpha+1}} \text{ for } x \geq 1$$

with $\mathbb{E}(X_i) = \frac{\alpha}{\alpha-1}$ and $\text{Var}(X_i) = \frac{\alpha}{(\alpha-1)^2(\alpha-2)}$. What is the Cramér-Rao lower bound for the variance of an unbiased estimator of α?

Answer: Let $W_n = W(X_1, ..., X_n)$ be an unbiased estimator of α. The standard regularity conditions hold, and the sample is iid,

$$\text{Var}(W_n) \geq \frac{1}{-\mathbb{E}\left(\left(\frac{\partial^2}{\partial \lambda} \log f(X_1, ..., X_n|\alpha)\right)\right)}$$

Since

$$f(X_1, ..., X_n|\alpha) = \prod_{i=1}^{n} \frac{\alpha}{x_i^{\alpha+1}} = \frac{\alpha^n}{\left(\prod_{i=1}^{n} x_i\right)^{\alpha+1}}$$

$$\log f(X_1, ..., X_n|\alpha) = n \log \alpha - (\alpha + 1) \sum_{i=1}^{n} \log x_i$$

$$\frac{\partial}{\partial \alpha} \log f(X_1, ..., X_n|\alpha) = \frac{n}{\alpha} - \sum_{i=1}^{n} \log x_i$$

$$\frac{\partial^2}{\partial\alpha}\log f(X_1,...,X_n|\alpha) = -\frac{n}{\alpha^2}$$

and thus:

$$\text{Var}(W(X_1,...,X_n)) \geq \frac{1}{\frac{n}{\alpha^2}} = \frac{\alpha^2}{n}$$

2. (5) Assume $U \sim \text{uniform}(2,5]$ and X has a Pareto distribution (given in question 1) with parameter equal to random variable U (i.e. $\alpha = U$). What is the joint pdf of (X, U)?

Answer: use the definition of a conditional distribution to see:

$$\begin{aligned} f(x|u) =& \frac{u}{x^{u+1}} \text{ for } x \geq 1, \text{ and } u \in (2,5] \\ f(x,u) =& \frac{u}{x^{u+1}}\frac{1}{3} \text{ for } u \in (2,5], \text{ and } x \geq 1 \end{aligned}$$

3. (5) Assume $(X_1,Y_1),...,(X_n,Y_n)$ is a bivariate random sample. Show that $M_{11} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$ converges in probability to σ_{XY}. Hint: Note that you can write M_{11} as a function of $M_{11}^* = \frac{1}{n}\sum_{i=1}^n (X_i - \mu_X)(Y_i - \mu_Y)$ (no need to prove this statement):

$$M_{11} = M_{11}^* - (\bar{X} - \mu_X)(\bar{Y} - \mu_Y).$$

Answer: Note that M_{11}^* is an average of iid variables:

$$M_{11}^* = \frac{1}{n}\sum_{i=1}^n (X_i - \mu_X)(Y_i - \mu_Y),$$

where $(X_i - \mu_X)(Y_i - \mu_Y)$ is independent of $(X_j - \mu_X)(Y_j - \mu_Y)$ for $i \neq j$. Hence, we can apply directly the Law of Large Numbers:

$$M_{11}^* \to_p \mathbb{E}\left[(X_i - \mu_X)(Y_i - \mu_Y)\right] = \sigma_{XY}$$

So we have a consistency result for M_{11}^*. Note that

$$(\bar{X} - \mu_X) \to_p 0$$
$$(\bar{Y} - \mu_Y) \to_p 0$$

So we can use Slutsky's Theorem to prove that

$$(\bar{X} - \mu_X)(\bar{Y} - \mu_Y) \to_p 0,$$

and once more to prove that

$$M_{11} = M_{11}^* - (\bar{X} - \mu_X)(\bar{Y} - \mu_Y) \to_p \sigma_{XY}$$

4. (5) Show that $\sqrt{n}\,(M_{11} - \sigma_{XY})$ converges in distribution to a normal, where M_{11} is defined as in question 3. (You do <u>not</u> need to state the variance).

Answer: We cannot yet apply the CLT directly, because M_{11} is not quite an average of a sequence of iid random variables. That is, we need to address all components. Begin by decomposing:

$$\begin{aligned}
\sqrt{n}(M_{11} - \sigma_{XY}) &= \sqrt{n}(M_{11}^* - (\bar{X} - \mu_X)(\bar{Y} - \mu_Y) - \sigma_{XY}) \\
&= \sqrt{n}(M_{11}^* - \sigma_{XY}) - \sqrt{n}((\bar{X} - \mu_X)(\bar{Y} - \mu_Y))
\end{aligned}$$

The first term is a relatively straightforward application of the CLT: M_{11}^* is an average of iid random variables, so from the convergence results in the previous section, we have

$$\sqrt{n}(M_{11}^* - \sigma_{XY}) \to_d n(0, \xi)$$

Turning our attention to the second part, first note the $\bar{X}$ is an average of iid random variables, so that $\sqrt{n}(\bar{X} - \mu_X) \to_d n(0, \sigma_X^2)$. But because $(\bar{Y} - \mu_Y) \to_p 0$, we can use Slutsky's Theorem to assert that $\sqrt{n}(\bar{X} - \mu_X)(\bar{Y} - \mu_Y) \to_p 0$. So finally:

$$\sqrt{n}(M_{11} - \sigma_{XY}) \to_d n(0, \xi)$$

5. (5) Let $X_1, X_2, ..., X_n$ be a sequence of independent random variables. Each random variable is drawn from a an exponential distribution with parameter λ_i: $X_i \sim \exp(\lambda_i)$, where $\lambda_i > 0\ \forall i$. Define the statistic $X_{(1)} = \min\{X_1, X_2, ..., X_n\}$. Derive the CDF of $X_{(1)}$. What distribution does this CDF represent?

Answer: We will utilize the independence of the random variables. Begin:

$$\begin{aligned} F_{X_{(1)}}(x) &= P(X_{(1)} \leq x) \\ &= P(\min\{X_1, X_2, \ldots, X_n\} \leq x) \\ &= 1 - P(X_1 > x) \cdot P(X_2 > x) \cdot \cdots \cdot P(X_n > n) \end{aligned}$$

where the last line is possible by independence. Noting that $P(X_i > x) = 1 - P(X_i \leq x) = 1 - (1 - e^{\lambda_i x})$, we continue:

$$\begin{aligned} F_{X_{(1)}}(x) &= 1 - P(X_1 > x) \cdot P(X_2 > x) \cdot \cdots \cdot P(X_n > n) \\ &= 1 - (1 - P(X_1 \leq x)) \cdot (1 - P(X_2 \leq x)) \cdot \cdots \cdot (1 - P(X_n \leq n)) \\ &= 1 - (e^{\lambda_1 x}) \cdot (e^{\lambda_2 x}) \cdot \cdots \cdot (e^{\lambda_n x}) \\ &= 1 - e^{x \cdot \sum_{i=1}^{n} \lambda_i} \end{aligned}$$

That is to say, the $X_{(1)} \sim exp(\sum_{i=1}^{n} \lambda_i)$.

6. (Extra Credit) Prove that $\text{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}(X))Y\right]$

Answer: There are several ways of doing this. The first way (somewhat lazy): Starting with the definition of covariance:

$$
\begin{aligned}
Cov(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y]] \\
&= \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[XY - Y\mathbb{E}[X]] \\
&= \mathbb{E}[(X-\mathbb{E}[X])Y]
\end{aligned}
$$

The second way (a little more full bodied):

$$
\begin{aligned}
\mathbb{E}[(X-\mathbb{E}[X])Y] &= \mathbb{E}[\mathbb{E}[(X-\mathbb{E}[X])Y|X]] \\
&= \mathbb{E}[\mathbb{E}[XY|X]] - \mathbb{E}[\mathbb{E}[Y\mathbb{E}[X]|X]] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\cdot\mathbb{E}[\mathbb{E}[Y|X]] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\
&= Cov(X,Y)
\end{aligned}
$$

Part II

7. (15) Consider a random sample, $X_1, X_2, ..., X_n$, where X_i is distributed normal with mean $\mu > 0$, and variance of one.

$$
f_{X_i}(x) = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{(x_i-\mu)^2}{2}\right]
$$

a) What is the pdf of $W = n\left(\bar{X}-\mu\right)^2$, where $\bar{X}$ is the sample average? Hint: What is the distribution of $\sqrt{n}(\bar{X}-\mu)$?

Answer: The hint should remind one that $\sqrt{n}(\bar{X}-\mu) \sim n(0,1)$, and then recall that the square of a standard normal random variable is distributed chi-squared with df 1: thus, $W \sim \chi^2_1$ and:

$$
f_W(w) = \frac{1}{\sqrt{2\pi}} \left(we^w\right)^{-1/2}
$$

Fun side note, $\sqrt{\pi} = \Gamma(1/2)$.

b) Write the formula for the pdf of $Y = \ln\left(\bar{X}\right)$, which should be characterized by parameters μ and n.

Answer: First just right out the pdf of a random variable distributed $n(\mu, 1/n)$ (note that this does NOT need to be in the asymptote because the X's are distributed normal)

$$f_{\bar{X}}(\bar{x}) = \frac{\sqrt{n}}{\sqrt{2\pi}} \left[\frac{n(\bar{x}-\mu)^2}{2}\right]$$

Then use the transformation rule. The result is a log-normal distribution:

$$f_Y(y) = \frac{\sqrt{n}}{\sqrt{2\pi}} \left[\frac{n\left(e^y-\mu\right)^2}{2}\right] e^y$$

c) Assume you observe a random sample, $Y_1, Y_2, ..., Y_m$, i.e. a random sample from the distribution derived in part (b). What is the method of moments estimator for parameters μ and n? Hint: note that $\bar{X} = \exp(Y)$, and $\bar{X}$ has a simpler distribution than Y.

Answer: First note that $\exp(Y_i) \sim \text{normal}(\mu, \frac{1}{n})$, hence, the method of moments estimator for n can be obtained from solving the following two equations for $\hat{n}$ and $\hat{\mu}$.

$$\frac{1}{m}\sum_{i=1}^{m} \exp(Y_i) = \hat{\mu}$$

$$\frac{1}{m-1}\sum_{i=1}^{m} \left(\exp(Y_i) - \hat{\mu}\right)^2 = \frac{1}{\hat{n}}$$

or

$$\frac{1}{m}\sum_{i=1}^{m} \left(\exp(Y_i) - \hat{\mu}\right)^2 = \frac{1}{\hat{n}}$$

Thus,

$$\hat{n} = \frac{1}{\frac{1}{m-1}\sum_{i=1}^{m} \left(\exp(Y_i) - \hat{\mu}\right)^2}$$

d) Is the method of moments estimator for n you proposed in part (c) unbiased? If no, what is the direction of the bias.

Answer: First note that

$$\mathbb{E}\left[\frac{1}{m-1}\sum_{i=1}^{m}(\exp(Y_i)-\hat{\mu})^2\right]=\frac{1}{n}$$

We want to inspect $\mathbb{E}[\hat{n}]$. Because $1/X$ is strictly convex, by Jensen's inequality:

$$\mathbb{E}[\hat{n}]=\mathbb{E}\left[\frac{1}{\frac{1}{m-1}\sum_{i=1}^{m}(\exp(Y_i)-\hat{\mu})^2}\right]>\left[\frac{1}{\mathbb{E}\left(\frac{1}{m-1}\sum_{i=1}^{m}(\exp(Y_i)-\hat{\mu})^2\right)}\right]=\frac{1}{1/n}=n$$

hence, the method of moments estimator for n is biased upwards.

e) Show that the method of moments estimator for μ derived in part (c) is consistent.

Answer: $\hat{\mu}$ is an average of iid random variables with mean μ, so WLLN applies: $\hat{\mu}\to_p\mu$.

f) Is the method of moments estimator for n consistent? No need to provide a proof. Instead, provide a valid argument that uses results seen in class.

Answer: In other words, sketch how one would show: $\hat{n}\to_p n$. There is a long way and a short way:

- (short) In class, we established that both M_2 and S^2 are consistent estimators of $\sigma_Y^2=1/n$ – these are the two options for $1/\hat{n}$ given in part (c). So $1/\hat{n}\to_p 1/n$. By the continuous mapping theorem, then, $\hat{n}\to_p n$.

- (long) To establish consistency, we usually try to apply a LLN, but this is not an option in this case because $\hat{n}$ cannot be written as an average of iid random variables. Thus we would use an alternative method taking advantage of Chebychev's inequality and the definition of convergence in probability. To do this, first calculate $\mathbb{E}[(\hat{n}-n^*)^2]$, where n^* is a constant representing the true value of n. Then using Chebychev:

$$P(|\hat{n}-n^*|\geq\epsilon)\leq\frac{\mathbb{E}[(\hat{n}-n^*)^2]}{\epsilon^2}$$

  Taking the limit of the RHS as $n\to\infty$ should result in 0, which is then sufficient to satisfy the definition of convergence in probability.

8. (15) **Taken from lecture's example.** A researcher is interested in learning about the distribution of opportunity costs of preserving hectares of forest in the Amazon. He proposes a model for preservation costs given by

$$c(Q; W) = a + \frac{W}{2}q^2, \tag{1}$$

where q are the number of hectares preserved and $a > 0$.

Note that the marginal cost of preserving q hectares of forest is proportional to the random variable W. A Payments for Ecosystem Services (PES) program pays p for each hectare of land preserved, where p is a known constant. Farmers decide how many hectares to submit by setting the marginal cost of preservation equal to the per-hectare compensation. Hence,

$$Q^* = \frac{p}{W} \tag{2}$$

The researcher observes the number of hectares submitted to the program for a random sample of farmers, $Q_1^*, ..., Q_n^*$.

(a) Derive the method of moments estimator for the mean marginal cost determinant W, μ_W, that is consistent with the cost model in (1) and (2).

Answer: Each agent has cost individual cost parameter W_i equal to:

$$W_i = \frac{p}{Q_i^*}$$

and thus the MoM estimation is:

$$\mathbb{E}(W_i) = \mathbb{E}\left(\frac{p}{Q_i^*}\right)$$

$$\hat{\mu}_W = \frac{1}{n}\sum_{i=1}^{n}\frac{p}{Q_i^*}$$

(b) Derive an unbiased method of moments estimator for the variance of W, σ_W^2.

Answer: In order for this to be unbiased, we need to correct for the df:

$$\hat{\sigma}_W^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{p}{Q_i^*} - \hat{\mu}_W\right)^2$$

For questions (c) through (f) assume that $W \sim \text{exponential}(\lambda)$.

(c) Write the pdf of Q^* under the new assumption for the distribution of W.

Answer: By the rule for mapping transformations of RV's, we have

$$f_{Q^*}(Q^*) = f_W(w^{-1}(Q^*))\left|\frac{d}{dQ^*}w^{-1}(Q^*)\right|$$

And so with $w^{-1}(Q^*) = p/Q^*$, we have:

$$f_{Q^*}(Q^*) = \lambda \exp\left(-\lambda\frac{p}{Q^*}\right)\frac{p}{(Q^*)^2} \text{ for } Q^* > 0$$

(d) What is the MLE estimator for λ?

Answer: Given the distribution in the previous section, this is straightforward. The joint distribution function is:

$$f(q_1^*, ..., q_n^*) = \prod_{i=1}^{n} \lambda \exp\left(-\lambda\frac{p}{q_i^*}\right)\frac{p}{q_i^{*2}}$$

and thus the log-likelihood function is:

$$\log f(q_1^*, ..., q_n^*) = n\log(\lambda) - \lambda\sum_{i=1}^{n}\left(\frac{p}{q_i^*}\right) + \sum_{i=1}\log\frac{p}{q_i^{*2}}$$

This is minimized for:

$$\frac{\partial}{\partial\lambda}\log f(q_1^*, ..., q_n^*) = \frac{n}{\hat{\lambda}} - \sum_{i=1}^{n}\left(\frac{p}{q_i^*}\right) = 0$$

$$\hat{\lambda}_{ML} = \frac{n}{\sum_{i=1}^{n}\left(\frac{p}{q_i^*}\right)}$$

(e) What is the MLE estimator for the variance of W?

Answer: That is, what is the MLE estimator for λ^{-2}. There is a trick for this questions (that, there's a short way and a long way):

- (short) Recall that MLE possesses the invariance property: if $\hat{\theta}$ is the ML estimator for θ, then $\tau(\hat{\theta})$ is the ML estimator for $\tau(\theta)$. Thus, because $\lambda^{-2} = (\lambda)^{-2}$, we have

$$\widehat{\lambda^{-2}}_{ML} = \left(\frac{\sum_{i=1}^{n} \left(\frac{p}{q_i^*} \right)}{n} \right)^2$$

- (long) Let's set $\xi = \lambda^{-2}$, and thus $W \sim exp(\xi^{-1/2})$

$$f(w_1, ..., w_n) = \prod_{i=1}^{n} \xi^{-1/2} \exp\left(-\xi^{-1/2} \cdot w_i\right)$$

  and thus the log-likelihood function is:

$$\log f(w_1, ..., w_n) = -\frac{n}{2} \log(\xi) - \xi^{-1/2} \sum_{i=1}^{n} w_i$$

  This is minimized for:

$$\frac{\partial}{\partial \xi} \log f(w_1, ..., w_n) = \frac{n}{2\hat{\xi}} + \frac{1}{2\xi^{-3/2}} \sum_{i=1}^{n} w_i = 0$$

  and so finally:

$$\hat{\xi} = \widehat{\lambda^{-2}}_{ML} = \left(\frac{\sum_{i=1}^{n} w_i}{n} \right)^2 = \bar{w}^{-2}$$

(f) Provided that the distributional assumption on W holds, which estimator for the variance of W is more efficient? Choose between the one you derived in part (b) and the one you derived in part (e). Explain.

Answer: That is, compare $\hat{\sigma}_W^2$ and $\hat{\xi}$. Because ML estimation is efficient, $\hat{\xi}$ is more efficient. Note that this is an asymptotic result; with finite samples, the bias induced in the ML estimation may not be sufficiently small to warrant ML over MM.

Bernoulli

$P(X = x|p) = p^x(1-p)^{(1-x)}$; $x = 0, 1$; $0 \le p \le 1$

Binomial

$P(X = x|n, p) = \begin{pmatrix} n \\ x \end{pmatrix} p^x(1-p)^{(n-x)}$; $x = 0, 1, 2, ..., n$; $0 \le p \le 1$

Chi-squared

$f(x|k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$

Discrete uniform

$P(X = x|N) = \frac{1}{N}$; $x = 1, 2, ..., N$; $N = 1, 2, ...$

Poisson

$P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$; $x = 0, 1, 2, ...$; $0 \le \lambda < \infty$

Uniform

$f(x|a, b) = \frac{1}{b-a}$; $x \in [a, b]$

Exponential

$f(x|\beta) = \lambda e^{-\lambda x}$; $0 \le x < \infty$, $\lambda > 0$

Logistic

$f(x|\mu, \beta) = \frac{1}{\beta} \frac{\exp(-(x-\mu)/\beta)}{[1+\exp(-(x-\mu)/\beta)]^2}$; $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\beta > 0$

Normal

$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$; $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$