Math 445
Midterm Exam – Part I
May 4, 2016

Name: KEY

This exam is closed book and closed notes, except for one standard size sheets of paper (8.5" by 11") which can have notes on both sides. These sheets must be submitted with the exam. No copying, cheating, collaborations, computers, or cell phones are allowed. Show your work and write complete and coherent answers. Be sure to clearly label and justify your solutions to these problems. Illegible or unjustified solutions will not receive full credit. After you have completed the exam, please reaffirm the Lawrence University Honor Code.

1. Where do we expect a bootstrap distribution to be centered?

We expect the bootstrap distribution to be centered at the observed sample statistic.

2. Where will a permutation distribution be centered?

The permutation distribution will be centered at the parameter value specified under the null hypothesis.

3. Let $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$ be random samples drawn from two independent populations and assume that we are interested in the difference between the two population means.

   (a) Describe the algorithm for a two-sample bootstrap used to compare two sample means.

   ① Draw a random sample, with replacement, of size n from $x_1, \ldots, x_n$.
   ② Draw a random sample with replacement, of size m from $y_1, \ldots, y_m$.
   ③ Calculate the difference of sample means between the two bootstrap samples and store this value.
   ④ Repeat steps 1–3 a large number of times, say 10,000.

   (b) Once you have the bootstrap distribution, how would you create a 90% confidence interval for the difference in population means?

   To form a 90% confidence interval from the bootstrap distribution you use the 0.05 quantile (i.e. 5^{th} percentile) as the lower bound and the 0.95 quantile (i.e. 95^{th} percentile) as the upper bound.

4. Let $X_1, \dots, X_n$ be a random sample from a Gamma(α, θ), where $\theta > 0$ and α is a known constant. To answer this problem, use the following parameterization of the gamma distribution:

$$\frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}, \ \mathrm{E}X = \alpha\theta, \ \mathrm{Var}X = \alpha\theta^2$$

(a) Find the method of moments estimator of θ.

Solve $\alpha\theta = \bar{X}_n$ for θ.

$$\Longrightarrow \hat{\theta} = \frac{\bar{X}_n}{\alpha}$$

(b) Is the method of moments estimator a MVUE?

For an estimator to be a MVUE it must (1) be an unbiased estimator and (2) have variance equal to the CRLB.

$$E(\hat{\theta}) = E\left(\frac{\bar{X}_n}{\alpha}\right) = \frac{1}{\alpha}E(\bar{X}_n) = \frac{1}{\alpha}(\alpha\theta) = \theta \Rightarrow \hat{\theta} \text{ is an unbiased estimator of } \theta.$$

$$I_1(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\right]$$

$$= -E\left[\frac{\partial^2}{\partial\theta^2}\left(-\log\Gamma(\alpha) - \alpha\log\theta + (\alpha-1)\log X - \frac{X}{\theta}\right)\right]$$

$$= -E\left[\frac{\partial}{\partial\theta}\left(-\frac{\alpha}{\theta} + \frac{X}{\theta^2}\right)\right]$$

$$= -E\left[\frac{\alpha}{\theta^2} - \frac{2X}{\theta^3}\right] = \frac{-\alpha}{\theta^2} - \frac{2}{\theta^3}E(X) = \frac{-\alpha}{\theta^2} - \frac{2}{\theta^3}(\alpha\theta) = \frac{\alpha}{\theta^2}$$

$$\Longrightarrow \text{CRLB of } Var(\hat{\theta}) = \frac{1}{nI_1(\theta)} = \frac{\theta^2}{n\alpha}$$

$$Var(\hat{\theta}) = Var\left(\frac{\bar{X}_n}{\alpha}\right) = \frac{1}{\alpha^2}Var(\bar{X}_n) = \frac{Var(X)}{n\alpha^2} = \frac{\alpha\theta^2}{n\alpha^2} = \frac{\theta^2}{n\alpha}$$

Since $\hat{\theta}$ is an unbiased estimator of θ and $Var(\hat{\theta})$ achieves the CRLB, $\hat{\theta}$ is a MVUE.

(c) Is the method of moments estimator a consistent estimator of θ?

Since we have already established that $\hat{\theta}$ is unbiased, we must only show that $\lim_{n\to\infty} Var(\hat{\theta}) = 0$ for $\hat{\theta}$ to be consistent.

$$\lim_{n\to\infty} Var(\hat{\theta}) = \lim_{n\to\infty} \frac{\theta^2}{n\alpha} = 0$$

So $\hat{\theta}$ is a consistent estimator of θ.

5. Let $Y_1, \ldots, Y_n$ be a random sample of discrete random variables with the common PDF

$$f(y|\theta) = \theta y^{\theta-1}, \; 0 < y < 1$$

where $\theta > 0$.

(a) Find the maximum likelihood estimator of θ, $\widehat{\theta}_{ML}$.

$$L(\theta | Y_1, \ldots, Y_n) = \prod_{i=1}^{n} \theta Y_i^{\theta-1} = \theta^n \prod_{i=1}^{n} Y_i^{\theta-1}$$

$$l(\theta) = n \log\theta + \sum_{i=1}^{n} (\theta-1) \log Y_i$$

$$= n \log\theta + (\theta-1) \sum_{i=1}^{n} \log Y_i$$

$$\frac{d}{d\theta} l(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \log Y_i \equiv 0$$

Solving for θ we find that $\hat{\theta}_{ML} = \frac{-n}{\sum_{i=1}^{n} \log Y_i}$.

The expected value and variance of the maximum likelihood estimator are given below:

$$E\left(\widehat{\theta}_{ML}\right) = \left(\frac{n}{n-1}\right)\theta \qquad Var\left(\widehat{\theta}_{ML}\right) = \left(\frac{n^2}{(n-1)^2(n-2)}\right)\theta^2$$

(b) Calculate the bias of $\widehat{\theta}_{ML}$. Is the estimator asymptotically unbiased?

$$\text{Bias}(\hat{\theta}_{ML}) = E(\hat{\theta}_{ML}) - \theta$$

$$= \left(\frac{n}{n-1}\right)\theta - \theta$$

$$= \theta \left[\frac{n-(n-1)}{n-1}\right]$$

$$= \left(\frac{1}{n-1}\right)\theta$$

$\lim_{n\to\infty} \text{Bias}(\hat{\theta}_{ML}) = \lim_{n\to\infty} \left(\frac{1}{n-1}\right)\theta = 0 \Rightarrow \hat{\theta}_{ML}$ is asymptotically unbiased.

(c) Calculate the mean square error of $\widehat{\theta}_{ML}$.

$$MSE(\hat{\theta}_{ML}) = Var(\hat{\theta}_{ML}) + [Bias(\hat{\theta}_{ML})]^2$$

$$= \theta^2 \left[\frac{n^2}{(n-1)^2(n-2)} \right] + \theta^2 \left[\frac{1}{(n-1)^2} \right]$$

$$= \theta^2 \left[\frac{n^2+n-2}{(n-1)^2(n-2)} \right]$$

$$= \theta^2 \left[\frac{(n+2)(n-1)}{(n-1)^2(n-2)} \right]$$

$$= \theta^2 \left[\frac{n+2}{(n-1)(n-2)} \right]$$

(d) Define a new estimator $\widehat{\theta}_U$ of θ to be:

$$\widehat{\theta}_U = \left(\frac{n-1}{n} \right) \widehat{\theta}_{ML}$$

Do you prefer this estimator to $\widehat{\theta}_{ML}$? Justify your answer.

$$E(\hat{\theta}_U) = E\left[\left(\frac{n-1}{n}\right)\hat{\theta}_{ML}\right] = \frac{n-1}{n} E(\hat{\theta}_{ML}) = \frac{n-1}{n}\left[\left(\frac{n}{n-1}\right)\theta\right] = \theta$$

$\Longrightarrow$ $\hat{\theta}_U$ is an unbiased estimator of θ.

$$Var(\hat{\theta}_U) = Var\left[\left(\frac{n-1}{n}\right)\hat{\theta}_{ML}\right] = \left(\frac{n-1}{n}\right)^2 Var(\hat{\theta}_{ML})$$

$$= \left(\frac{n-1}{n}\right)^2 \theta^2 \left[\frac{n^2}{(n-1)^2(n-2)}\right]$$

$$= \theta^2 \left(\frac{1}{n-2}\right)$$

$$\Longrightarrow MSE(\hat{\theta}_U) = \frac{\theta^2}{n-2}$$

Since $MSE(\hat{\theta}_U) < MSE(\hat{\theta}_{ML})$, we prefer $\hat{\theta}_U$ to $\hat{\theta}_{ML}$.