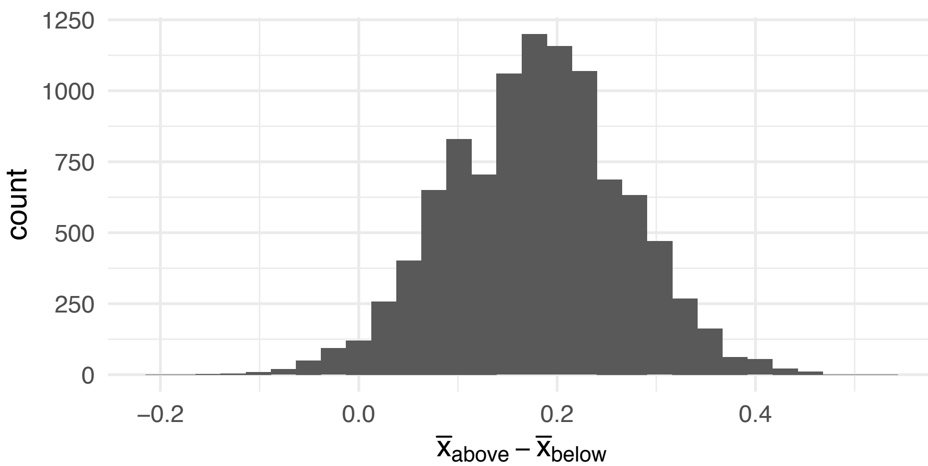
Math 445 Name: KEY
Exam 2
May 16, 2017

1. A pollution-control inspector suspected that a riverside community was releasing semi-treated sewage into a river and was consequently changing the level of dissolved oxygen of the river. To this, the engineer drew 15 randomly selected specimens of river water at a location above the town and another 15 specimens below. The dissolved oxygen readings in part per million (ppm) were recorded from each specimen. Summary statistics from the samples are displayed below.

	min	Q1	median	Q3	max	mean	SD	n
Above	4.7	4.8	4.9	5	5.2	4.92	0.16	15
Below	4.2	4.6	4.7	4.9	5.5	4.74	0.32	15

A two-sample bootstrap was run in order to compare the average dissolved oxygen concentrations above and below town. Below is the resulting histogram of the bootstrap distribution along with select summary statistics.



min	Q1	median	Q3	max	mean	SD	n
-0.21	0.12	0.18	0.24	0.52	0.18	0.09	10000

(a) Calculate the bootstrap estimate of the bias.

$\bar{X}_{above} - \bar{X}_{below} = 4.92 - 4.74 = 0.18$ ppm, which is also the avg. of the bootstrap stats $\Rightarrow$ $Bias_{boot} = 0$

(b) Provide a bootstrap estimate of the standard error of the difference in sample means.

$SE_{boot} = 0.09$

(c) Percentiles of the bootstrap distribution are provided below. Obtain a 99% bootstrap confidence interval for the difference in average dissolved oxygen concentrations above and below town.

0.5%	1%	2.5%	5%	10%	20%	80%	90%	95%	97.5%	99%	~~0.995%~~ 99.5%
-0.06	-0.03	0.00	0.03	0.07	0.11	0.25	0.29	0.32	0.35	0.39	0.41

−0.06 to 0.41 ppm

(d) Interpret the confidence interval you obtained in part (b) in the context of the problem.

We are 99% confident that the average dissolved oxygen is between 0.06 ppm lower to 0.41 ppm higher above town than it is below town.

(e) Outline the algorithm for the two-sample bootstrap that was used in this problem.

1. Draw a random sample, with replacement, of size $n = 15$ from the above town sample.

2. Draw a random sample, with replacement, of size $m = 15$ from the below town sample.

3. Calculate the difference in means, $\bar{X}^*_{above} - \bar{X}^*_{below}$, between the two bootstrap samples (and store this value).

4. Repeat Steps 1-3 a large number of times. (10,000 here)

2. Let $X_1, \ldots, X_n$ be i.i.d. random variables with PDF

$$f(x|\theta) = (\theta+1)x^{\theta}, \quad 0 \le x \le 1$$

(a) Find the maximum likelihood estimator of θ.

$$L(\theta|x_1, \ldots, x_n) = \prod_{i=1}^{n} (\theta+1) X_i^{\theta} = (\theta+1)^n \prod_{i=1}^{n} X_i^{\theta}$$

$$\ell(\theta) = n \log(\theta+1) + \theta \sum_{i=1}^{n} \log X_i$$

$$\frac{d}{d\theta} \log(\theta) = \frac{n}{\theta+1} + \sum_{i=1}^{n} \log X_i \overset{set}{=} 0$$ and solve for θ.

$$\hat{\theta} = \frac{-n}{\sum_{i=1}^{n} \log X_i} - 1$$

Checking the 2^{nd} deriv. of $\ell(\theta)$: $\frac{d^2}{d\theta^2} \ell(\theta) = \frac{-n}{(\theta+1)^{2}} < 0$ since $n > 0$,

so we have found a max.

(b) Find the asymptotic variance of the maximum likelihood estimator you found in part (a).

We know that $Var(\hat{\theta}) \approx \frac{1}{I_n(\theta)}$ as $n \to \infty$.

$$I_1(\theta) = -E\left[\frac{d^2}{d\theta^2} \log f(x|\theta)\right] = -E\left[\frac{d^2}{d\theta^2} \log(\theta+1) + \theta \log X\right]$$

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

So $$Var(\hat{\theta}) \approx \frac{1}{I_n(\theta)} = \frac{1}{n/(\theta+1)^2} = \frac{(\theta+1)^2}{n}$$

3. If Y has a binomial distribution with n trials and success probability p, show that Y/n is a consistent estimator of p.

Let $Y = \sum_{i=1}^{n} X_i$ where $X_i \overset{iid}{\sim} Bern(p)$, so $\frac{Y}{n} = \bar{X}$.

The WLLN then gives that $\lim_{n\to\infty} P(|\bar{X} - p| < \epsilon) = 1$

$\forall\, \epsilon > 0$, since $E(X_i) = p$.

4. Let X_1, X_2, X_3 be a random sample from a distribution with PDF

$$f(x|\theta) = 2x\theta^2, \quad 0 \le x \le 1/\theta$$

Let $T = \frac{1}{9}X_1 + \frac{1}{9}X_2 + \frac{1}{3}X_3$ be an estimator of $\frac{1}{\theta}$.

(a) Find the bias of T.

$$E(X) = \int_0^{1/\theta} x 2\theta^2 x\,dx = 2\theta^2 \int_0^{1/\theta} x^2\,dx = 2\theta^2 \cdot \frac{1}{3}x^3\Big|_0^{1/\theta} = \frac{2}{3}\theta^2\left(\frac{1}{\theta}\right)^3 = \frac{2}{3\theta}$$

$$E(T) = \frac{1}{9}E(X_1) + \frac{1}{9}E(X_2) + \frac{1}{3}E(X_3) = \frac{2}{3\theta}\left[\frac{1}{9}+\frac{1}{9}+\frac{1}{3}\right] = \frac{10}{27\theta}$$

$$Bias(T) = E(T) - \frac{1}{\theta} = \frac{10}{27\theta} - \frac{1}{\theta} = \frac{1}{\theta}\left(\frac{10}{27}-1\right) = \frac{-17}{27\theta}$$

(b) Find the MSE of T.

$$E(X^2) = \int_0^{1/\theta} x^2 2\theta^2 x\,dx = 2\theta^2 \int_0^{1/\theta} x^3\,dx = 2\theta^2 \cdot \frac{1}{4}x^4\Big/_0^{1/\theta} = \frac{1}{2}\theta^2 \cdot \frac{1}{\theta^4} = \frac{1}{2\theta^2}$$

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{1}{2\theta^2} - \left(\frac{2}{3\theta}\right)^2 = \frac{1}{2\theta^2} - \frac{4}{9\theta} = \frac{1}{18\theta^2}$$

$$\text{So } Var(T) = \frac{1}{9^2}Var(X_1) + \frac{1}{9^2}Var(X_2) + \frac{1}{3^2}Var(X_3)$$
$$= Var(X)\left[\frac{1}{9^2}+\frac{1}{9^2}+\frac{1}{3^2}\right]$$
$$= \frac{1}{18\theta^2}\left[\frac{11}{9^2}\right] = \frac{11}{1458\theta^2}$$

$$MSE(T) = Var(T) + [Bias(T)]^2 = \frac{11}{1458\theta^2} + \frac{1}{\theta^2}\left(\frac{17}{27}\right)^2$$
$$= \frac{11+17^2(2)}{1458\theta^2}$$

(c) If possible, use T to get an unbiased estimator for $\frac{1}{\theta}$.

Notice that $E(T) = \frac{10}{27} \cdot \frac{1}{\theta}$, so

$E\left(\frac{27}{10}T\right) = \frac{1}{\theta}$. Thus $\frac{27}{10}T$ is an unbiased estimator of $\frac{1}{\theta}$.

5. Let $Y_1, \ldots, Y_n$ denote a random sample from a power family distribution with PDF

$$f(y|\alpha) = \frac{\alpha}{3^\alpha} y^{\alpha-1}, \quad 0 \le y \le 3$$

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

$$E(Y) = \int_0^3 y \cdot \frac{\alpha}{3^\alpha} y^{\alpha-1}\, dy = \frac{\alpha}{3^\alpha} \int_0^3 y^\alpha dy = \frac{\alpha}{3^\alpha} \cdot \frac{1}{\alpha+1} y^{\alpha+1} \Big|_0^3 = \frac{\alpha}{3^\alpha} \cdot \frac{1}{\alpha+1} 3^{\alpha+1} = \frac{3\alpha}{\alpha+1}$$

So the MoM eq'n is: $\frac{3\alpha}{\alpha+1} = \bar{Y}$

$$\Rightarrow \hat{\alpha} = \frac{\bar{Y}}{3-\bar{Y}}$$

(b) The maximum likelihood estimator of α is $\frac{n}{\sum \log(X_i) - n\log 3}$. Describe how you would determine which estimator (MoM or ML) that you prefer. (You do not need to actually determine this, I simply want you to tell me *how* you would approach this comparison.)

Answers will vary, but should focus on comparing bias and variance. The MSE is a good metric to compare.

6. For each of the following state whether it is TRUE or FALSE.

(a) Maximum Likelihood estimators are always unbiased.

FALSE

(b) Maximum Likelihood estimators are always consistent.

TRUE

(c) Maximum Likelihood estimators are always asymptotically normal.

TRUE

7. Let $\widehat{\theta}_{ML}$ be an unbiased maximum likelihood estimator for θ, and let $\widehat{\theta}_{MoM}$ be an unbiased method of moments estimator for θ. Let $Var(\widehat{\theta}_{ML}) = \eta$. If $\widehat{\theta}_{ML} \neq \widehat{\theta}_{MoM}$, which of the following is a possible value for $Var(\widehat{\theta}_{MoM})$? Select the best response and justify your answer.

(a) 0.75η

(b) 1.00η

(c) 1.33η

(d) Either (b) or (c).

ML estimates are asymptotically efficient, so it's variance will be close to the CRLB. Since the estimators are to equal, the variance of the MoM estimator will be larger.