Math 10B with Professor Stankova
Worksheet, Discussion #21; Thursday, 4/11/2019
GSI name: Roy Zhao

Statistics and Estimators

Concepts

1. A **statistic** is a function of random variables and we use them to estimate values that we aren't always given. The **estimator of the mean** is

$$\hat{\mu} = \frac{x_1 + \cdots + x_n}{n}.$$

The **biased estimator of the standard deviation** is

$$s_*^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2.$$

The **unbiased estimator of the standard deviation**, also known as the **sample standard deviation** is

$$s^2 = \frac{n}{n-1}s_*^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2.$$

The **95% confidence interval** of the *population* mean is

$$(\hat{\mu} - 2\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\mu} + 2\frac{\hat{\sigma}}{\sqrt{n}}).$$

The PDF of a normal distribution with mean μ and standard deviation σ is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}.$$

Example

2. The number of rainy days in Honolulu in a year is Poisson distributed. Suppose that last ten years have had $4, 3, 2, 6, 5, 4, 1, 3, 4, 3$ rainy days. What is the 95% confidence interval for λ?

Solution: First we calculate $\hat{\mu} = \bar{x} = \frac{4+3+2+6+5+4+1+3+4+3}{10} = 3.5$. Then for a Poisson distribution, we have $\hat{\lambda} = \hat{\mu} = 3.5$. We also have $\hat{\sigma} = \sqrt{\hat{\lambda}} = \sqrt{3.5}$. So the 95% confidence interval is

$$(\hat{\mu} - 2\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\mu} + 2\frac{\hat{\sigma}}{\sqrt{n}}) = (3.5 - \frac{2\sqrt{3.5}}{\sqrt{10}}, 3.5 + \frac{2\sqrt{3.5}}{\sqrt{10}}).$$

3. Now suppose that we did not know that the number of rainy days was Poisson distributed. What is the 95% confidence interval for the average number of rainy days per year?

Solution: If we don't know that the distribution is Poisson, then we need another estimate for the standard deviation. We use the formula

$$s^2 = \frac{1}{10-1}[(4-3.5)^2 + (3-3.5)^2 + \cdots + (4-3.5)^2 + (3-3.5)^2] = \frac{18.5}{9} = \frac{37}{18}.$$

So now $\hat{\sigma} = s = \sqrt{37/18}$. So our 95% confidence interval is

$$(\hat{\mu} - 2\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\mu} + 2\frac{\hat{\sigma}}{\sqrt{n}}) = (3.5 - \frac{2\sqrt{37/18}}{\sqrt{10}}, 3.5 + \frac{2\sqrt{37/18}}{\sqrt{10}})$$

Problems

4. **TRUE** False If $f(x) = \frac{1}{a}e^{-(x-2019)^2/b}$ is the PDF of a normal distribution. Then $\pi = \frac{a^2}{b}$.

Solution: Under our formula, we have $a = \sigma\sqrt{2\pi}$ and $b = 2\sigma^2$ so $\frac{a^2}{b} = \frac{\sigma^2 \cdot 2\pi}{2\sigma^2} = \pi$.

5. **TRUE** False If we know both the biased estimator s_* and the unbiased estimator s, we can find out the sample size n.

Solution: We have $s_*^2/s^2 = \frac{n-1}{n} = 1 - \frac{1}{n}$ so we can solve for n.

6. **TRUE** False For a geometric distribution, our estimate for the probability p is $\hat{p} = \frac{1}{\bar{x}+1}$.

Solution: We have $\bar{x} = \hat{\mu} = \frac{1-\hat{p}}{\hat{p}} = \frac{1}{\hat{p}} - 1$ so solving gives $\hat{p} = \frac{1}{\bar{x}+1}$.

7. True **FALSE** The smaller the 95% confidence interval is, the lower our confidence is that the true parameter is in that interval.

Solution: We are always just 95% confident that the true parameter is in the interval.

8. True **FALSE** The smaller the 95% confidence interval is, the higher our confidence is that the true parameter is in that interval.

Solution: We are always just 95% confident that the true parameter is in the interval.

9. True **FALSE** The 95% confidence interval means that there is 95% chance that the parameter is in the interval.

Solution: The parameter is a fixed number so it is either in the interval or not in it.

10. True **FALSE** Chebyshev's inequality says that 95% of the sample data much lie within 2 standard deviations of the mean.

Solution: Chebyshev's only gives a bound of $1 - \frac{1}{2^2} = 75\%$.

11. I flip a biased coin 100 times and get 64 heads. What is the 95% confidence interval for p?

Solution: This is a sum of Bernoulli trials. Our best estimate for $\hat{p} = \hat{\mu} = \frac{64}{100} = \frac{16}{25}$. So then our standard deviation is $\sqrt{\hat{p}(1-\hat{p})} = \sqrt{\frac{16}{25}\frac{9}{25}} = \frac{12}{25}$. Then our 95% confidence interval for $\hat{p} = \hat{\mu}$ is

$$(\hat{\mu} - 2\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\mu} + 2\frac{\hat{\sigma}}{\sqrt{n}}) = (0.64 - \frac{2(12/25)}{\sqrt{100}}, 0.64 + \frac{2(12/25)}{\sqrt{100}}) = (0.64 - \frac{12}{125}, 0.64 + \frac{12}{125}).$$

12. For the upcoming ASUC elections, you ask 400 people if they support the basic needs referendum, and 256 of them do. What is the 95% confidence interval for the percentage of all students who support the referendum?

Solution: This is again a sum of Bernoulli trials. Our best estimate for $\hat{p} = \hat{\mu} = \frac{256}{400} = \frac{16}{25}$. So then our standard deviation is $\sqrt{\hat{p}(1-\hat{p})} = \sqrt{\frac{16}{25}\frac{9}{25}} = \frac{12}{25}$. Then our 95% confidence interval for $\hat{p} = \hat{\mu}$ is

$$(\hat{\mu} - 2\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\mu} + 2\frac{\hat{\sigma}}{\sqrt{n}}) = (0.64 - \frac{2(12/25)}{\sqrt{400}}, 0.64 + \frac{2(12/25)}{\sqrt{400}}) = (0.64 - \frac{6}{125}, 0.64 + \frac{6}{125}).$$

13. Suppose I keep asking students if they are voting in the ASUC elections until I find someone who is. I do this multiple times and suppose that the number of students I have to ask before finding someone who is voting is $2, 2, 1, 8, 3, 5, 6, 3, 3, 7$. What is the 95% confidence interval for the average number of times we need to ask before we ask someone who is voting?

Solution: This is a geometric distribution. We have $\hat{\mu} = \bar{x} = \frac{2+2+1+8+3+5+6+3+3+7}{10} = 4$. Then because this is a geometric distribution and we include the person who is voting, the success, we have $\hat{\mu} = \frac{1-\hat{p}}{\hat{p}}$ so $\hat{p} = \frac{1}{5}$. The standard deviation is $\hat{\sigma}^2 = \frac{1-\hat{p}}{\hat{p}^2} = 20$. Then, the 95% confidence interval for $\hat{\mu} = \frac{1-\hat{p}}{\hat{p}}$ is

$$(\hat{\mu} - 2\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\mu} + 2\frac{\hat{\sigma}}{\sqrt{n}}) = (4 - \frac{2\sqrt{20}}{\sqrt{10}}, 4 + \frac{2\sqrt{20}}{\sqrt{10}}) = (4 - 2\sqrt{2}, 4 + 2\sqrt{2}).$$

14. Do the previous problem if we use the sample standard deviation to calculate our confidence interval.

Solution: Now we set $\hat{\sigma} = s$ where

$$s^2 = \frac{1}{10-1}[(2-4)^2 + (2-4)^2 + \cdots + (3-4)^2 + (7-4)^2] = \frac{50}{9}.$$

So the 95% confidence interval is

$$(\hat{\mu} - 2\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\mu} + 2\frac{\hat{\sigma}}{\sqrt{n}}) = (4 - \frac{2\sqrt{50/9}}{\sqrt{10}}, 5 + \frac{2\sqrt{50/9}}{\sqrt{10}}).$$

15. Every morning for 10 days I try to do the water bottle flip 25 times. I am successful $7, 8, 3, 4, 8, 5, 4, 5, 2, 4$ times. What is the 95% confidence interval for the number of times I am successful tomorrow morning?

Solution: This is multiple binomial distributions with $n = 25$. We want to find the 95% confidence interval $\hat{\mu} = n\hat{p} = 25\hat{p}$ and $\hat{\mu} = \bar{x} = \frac{7+8+3+4+8+5+4+5+2+4}{10} = 5$. So $\hat{p} = \frac{5}{25} = \frac{1}{5}$ and the standard deviation is $\hat{\sigma} = \sqrt{n\hat{p}(1-\hat{p})} = \sqrt{25(1/5)(4/5)} = 2$. So the 95% confidence interval is

$$(\hat{\mu} - 2\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\mu} + 2\frac{\hat{\sigma}}{\sqrt{n}}) = (5 - \frac{2 \cdot 2}{\sqrt{10}}, 5 + \frac{2 \cdot 2}{\sqrt{10}}).$$

16. A Pareto distribution is given by the PDF $f(x) = \frac{p}{x^{p+1}}$ for $x \geq 1$ and 0 for $x < 1$ for some parameter p. Suppose I draw from this distribution 4 times and get the values $1, 2, 1, 1, 1$. What is the 95% confidence interval for μ?

Solution: The mean of the Pareto distribution is $\frac{p}{p-1}$. The standard deviation is $\sqrt{\frac{p}{(p-1)^2(p-2)}}$. So again we calculate $\hat{\mu} = \frac{\hat{p}}{\hat{p}-1} = \frac{1+2+1+1+1}{5} = \frac{6}{5}$ so we see $\hat{p} = 6$. Then $\hat{\sigma} = \sqrt{\frac{6}{5^2 \cdot 4}} = \frac{\sqrt{6}}{10}$. So the 95% confidence interval for μ is

$$(\hat{\mu} - 2\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\mu} + 2\frac{\hat{\sigma}}{\sqrt{n}}) = (\frac{6}{5} - \frac{2 \cdot \sqrt{6}/10}{\sqrt{5}}, \frac{6}{5} + \frac{2 \cdot \sqrt{6}/10}{\sqrt{5}}).$$

17. An exponential distribution is given by the PDF $f(x) = ce^{-cx}$ for $x \geq 0$ and 0 for $x < 0$. I draw from this distribution 5 times and get the values $\frac{1}{6}, 0, \frac{1}{3}, 1, \frac{1}{6}$. What is the 95% confidence interval for μ?

Solution: For this distribution, we need to first calculate the mean and standard deviation and how they depend on c by taking integrals. If we do this, we get $\mu = \frac{1}{c}$ and $\sigma = \frac{1}{c}$. So in this case, we have $\hat{\mu} = \frac{1}{\hat{c}} = \bar{x} = \frac{1/6+0+1/3+1+1/6}{5} = \frac{1}{3}$ so $\hat{c} = 3$. So $\hat{\sigma} = \frac{1}{\hat{c}} = \frac{1}{3}$. Now the 95% confidence interval for μ is

$$(\hat{\mu} - 2\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\mu} + 2\frac{\hat{\sigma}}{\sqrt{n}}) = (\frac{1}{3} - \frac{2 \cdot 1/3}{\sqrt{5}}, \frac{1}{3} + \frac{2 \cdot 1/3}{\sqrt{5}}).$$