STA261: Assignment 8

Alex Stringer

March, 2018

This assignment is not for credit. Complete the questions as preparation for quizzes and tests.

Suggested reading: Textbook sections 11.1, 11.2.1, 11.3 (including the material on the signed-rank test, which was not covered in lecture)

Do the following questions from textbook section 9.11 (page 362): 1a, 2, 5abdf, 9 (except the part about power),

1. *With thanks to Jerry Brunner*: A random sample of size $n = 150$ yields a sample mean of $\bar{X} = 8.2$ (unless otherwise stated). Use the Normal approximation to the MLE, or the Central Limit Theorem for the sample mean, to give an estimate and a 95% CI for
   (a) θ, if $X_i \sim Bern(\theta_0)$; use $\bar{X} = 0.82$
   (b) λ, if $X_i \sim Poission(\lambda_0)$
   (c) θ, if $X_i \sim Exponential(\theta_0)$, with the parametrization having $E(X) = \theta$
   (d) θ, if $X_i \sim Exponential(\theta_0)$, with the parametrization having $E(X) = 1/\theta$. Can you use your answer to the previous question to avoid having to actually do any calculations?
   (e) θ, if $X_i \sim Unif(0, \theta_0)$. Note that the MLE doesn't depend on $\bar{X}$- do you remember an estimator from earlier in the course that does?
2. *With thanks to Jerry Brunner*: That $Unif(0, \theta_0)$ confidence interval is messy. It also doesn't make sense: we know that θ_0 has to be greater than $max(X_i)$, so we should put the lower bound of our confidence interval at $X_{(n)} = max(X_i)$.
   (a) Show that
   $$X/\theta_0 \sim Unif(0, 1)$$
   (b) Prove that
   $$X_{(n)}/\theta_0 \sim Beta(n, 1)$$
   Hint: compute the CDF directly, $P(X_{(n)}/\theta_0 < x) = P(X_1 < x, \dots, X_n < x)$
   (c) Find a $1 - \alpha$ confidence interval for θ of the form
   $$(X_{(n)}, qX_{(n)})$$
   I.e. find the constant q such that $P(X_{(n)} < \theta_0 < qX_{(n)}) = 1 - \alpha$

(d) For the following random sample from a $Unif(0, \theta_0)$ distribution, evaluate both your confidence interval from this, and the previous question. Which would you prefer to use for making inferences about θ, and why?

$$4.12, 6.42, 10.51, 4.1, 11.57, 13.35, 5.26, 8.5, 0.98, 4.04$$

3. *With thanks to Jerry Brunner*: In question 1, you derived a $1-\alpha$ CI for θ from a $Bern(\theta_0)$ distribution, by using the CLT for the MLE and plugging in the estimate $\hat{\theta}$ for θ_0 in the Fisher Information. Can we improve this confidence interval, by not replacing θ_0 with an estimate?

   (a) Use the CLT to show that

$$P\left(\left(\frac{\sqrt{n}(\bar{X}-\theta_0)}{\sqrt{\theta_0(1-\theta_0)}}\right)^2 < z_{\alpha/2}^2\right) \approx 1-\alpha$$

   (b) Keep going; show that

$$P\left((n+z_{\alpha/2}^2)\theta_0^2 - (2n\bar{X}+z_{\alpha/2}^2)\theta_0 + n\bar{X}^2 < 0\right) \approx 1-\alpha$$

   (c) Because $(n+z_{\alpha/2}^2) > 0$, these random parabolae all open upward. That means that the endpoints of the confidence interval we want are the points at which this parabola (as a function of θ) intersect the x-axis, because between these points, the parabola is less than zero, which is the event that happens with probability $1-\alpha$. Use the quadratic formula and simplify to show that the confidence interval is

$$\frac{2n\bar{X}+z_{\alpha/2}^2}{2(n+z_{\alpha/2}^2)} \pm \frac{\sqrt{4nz_{\alpha/2}^2\bar{X}(1-\bar{X})+z_{\alpha/2}^4}}{2(n+z_{\alpha/2}^2)}$$

   (d) Compare the interval calculated in this way to the interval calculated in question 1. Do you notice much difference for $n=150$? Try it for smaller values of n, say $n=5,10,15,30$. How much data, roughly, is needed for the replacement of θ_0 by $\hat{\theta}$ in the Fisher Information to not change the resulting inferences by much?

4. *Two Sample T-Test.* In lecture, we stated the independent-samples two sample T-Test in the case where both samples have the same size; this doesn't need to be so. Consider the case where we have $X_i, i = 1\ldots n$ and $Y_j, j = 1\ldots m$ for $n \neq m$ as two mutually independent random samples from $N(\mu_x, \sigma^2)$ and $N(\mu_y, \sigma^2)$ distributions.

   (a) Show that

$$\bar{X}-\bar{Y} \sim N\left(\mu_x-\mu_y, \sigma^2\left(\frac{1}{n}+\frac{1}{m}\right)\right)$$

   (b) Show that if σ^2 is known, a $1-\alpha$ confidence interval for $\mu_x-\mu_y$ is given by

$$\left(\bar{X}-\bar{Y}-\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}z_{1-\alpha/2}, \bar{X}-\bar{Y}+\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}z_{1-\alpha/2}\right)$$

(c) Prove Theorem A of section 11.2.1 in the textbook (page 422), which allows you to replace σ^2 by an appropriate estimator, and the z quantiles with t quantiles.

(d) Verify that all the formulas here agree with what was presented in lecture, if $n = m$.

5. Let β_0, β_1 be fixed unknown constants, let $x_i, i = 1 \dots n$ be fixed, known constants, let $\epsilon_i \sim N(0, \sigma^2)$ independently, and let $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. On a previous assignment, you showed that the MLE of $(\beta_0, \beta_1, \sigma^2)$ was

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$.

(a) Show that

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$

(b) Show that

$$\frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \sim \chi^2_{n-2}$$

where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ is the predicted value of y_i in this model.

(c) Using the modified variance estimator

$$M\hat{S}E = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

show that

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{M\hat{S}E}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \sim t_{n-2}$$

(d) Write down a $1 - \alpha$ confidence interval for β_1. In scientific applications, this allows one to answer the question "does the response variable have a linear relationship with x_1?", by checking whether $\beta_1 = 0$ is a plausible value of β_1 given the data and the model.