

243 Homework 6

Ariel Sim (50% of work) & Samantha Maillie (50% of work)

6/4/2018

Problem 1

This problem was very straightforward particularly given the textbook's walk through with the code. We see an estimated correlation coefficient of .5459189. We later find that the bootstrap produces a smaller standard error than the jackknife. The t-interval approach was a little wider than the others and as a result it crossed over zero and contained negative values. This doesn't give us much insight into the time of correlation between the two variables.

We ran more than just the normal theory confidence intervals. We also ran basic and percentile. Using these three we get confidence intervals that are entirely positive. This makes intuitive sense given we are examining a relationship between LSAT scores and gpa. However, they do drop down to as low as .1 for a lower bound and rise to around .9 for an upper bound. This indicates a positive correlation but gives little insight into the strength of that relationship. See Hw6 01 code for plots and additional information.

Problem 2

a.

The distribution of $\hat{\theta}_{MLE}$ is the distribution of the maximum order statistic. We know the formula for this is $nf(x)F(x)^{n-1}$ plugging in the appropriate values we get that $\hat{\theta}_{MLE}$ follows the distribution $\frac{n}{\theta^n}x_n^{n-1}$.

b.

$$Var(X_n) = E[X_n^2] - E[X_n]^2$$

$$\begin{aligned} E[X_n^2] &= \int_0^3 \frac{n}{3^n} x_n^{n-1} x_n^2 dx_n \\ E[X_n^2] &= \frac{n}{3^n} \int_0^3 x_n^{n+1} dx_n \\ E[X_n^2] &= \frac{n}{3^n} \frac{1}{n+2} 3^{n+2} \\ E[X_n^2] &= \frac{9n}{n+2} \end{aligned}$$

$$\begin{aligned} E[X_n] &= \int_0^3 \frac{n}{3^n} x_n^{n-1} x_n dx_n \\ E[X_n] &= \frac{n}{3^n} \int_0^3 x_n^n dx_n \\ E[X_n] &= \frac{n}{3^n} \frac{1}{n+1} 3^{n+1} \\ E[X_n] &= \frac{3n}{n+1} \end{aligned}$$

$$Var(X_n) = \frac{9n}{n+2} - \left(\frac{3n}{n+1}\right)^2$$

If we let $n = 50$ we find the variance to be approximately equal to .003327.

discussion for c - f

We get a consistently excellent estimation of this using parametric bootstrapping methods. The last run we did gave an estimate of 0.003017126 although we have seen it vary a little bit. The histogram of estimated $\hat{\theta}^*$'s matches very nicely with the true MLE function plot.

We see that the nonparametric estimate fails. Since we sample and recalculate the theta value based on the generated values and then regenerate simulation data we think this is causing the issue. The MLE for theta from the uniform distribution with theta as a parameter is always a biased estimator that underestimates theta. We see quickly that as we recalculate theta it goes to zero and hence the method doesn't work. This distribution tends to have difficulties in many applications that work well for most of the other common distributions.

Plots are contained in the hw6 02 r code file.

Problem 3

Problem 4

a.

$$likelihood = \prod \lambda e^{-\lambda x_i}$$

$$likelihood = \lambda^n e^{-\lambda \sum x_i}$$

$$L = \log - likelihood = n \log(\lambda) - \lambda \sum x_i$$

$$L' = \frac{n}{\lambda} - \sum x_i = 0$$

$$\hat{\lambda}_{MLE} = \frac{n}{\sum x_i} = \frac{1}{\bar{X}}$$

b

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \rightarrow^d N(0, \lambda^2)$$

$$\sqrt{n}(\log(\hat{\lambda}_n) - \log(\lambda)) \rightarrow^d N(0, 1)$$

From the above we know that applying the log function will result in the mean remaining as zero. We also know from the delta method that our new variance will be $(g'(\lambda))^2 Var(\lambda)$ where $g(\lambda) = \log(\lambda)$ and therefore $g'(\lambda) = \frac{1}{\lambda}$. Putting it all together we get the $g'(\lambda)^2$ and $Var(\lambda)$ cancelling each other out so that we have a variance of 1.

c

d See the picture below.

e

References

1. Rizzo, Maria L. Statistical Computing with R. Chapman & Hall/CRC, 2017.

4(d) Given $\lambda \sum X_i \sim \text{Gamma}(n, 1)$ if $X_i \sim \text{Exponential}(\lambda)$
 Let $X \sim \text{Exp}(\frac{1}{2})$, then $2\lambda X = Y \sim \chi^2_2$ since $\chi^2_k \sim G(\frac{k}{2}, \frac{1}{2})$
 Define $h(X_1, \dots, X_n, \lambda) = 2\lambda \sum_{i=1}^n X_i = \sum Y_i$
 Then, $\sum_{i=1}^n Y_i \sim \chi^2_{2n}$.

Let $\chi^2_{2n}(\frac{\alpha}{2})$ be the $100(\frac{\alpha}{2})$ th percentile
 $\chi^2_{2n}(1-\frac{\alpha}{2})$ be the $100(1-\frac{\alpha}{2})$ th percentile.

Then, $P(\chi^2_{2n}(\frac{\alpha}{2}) \leq 2\lambda \sum_{i=1}^n X_i \leq \chi^2_{2n}(1-\frac{\alpha}{2})) = 1-\alpha$

$$P\left(\frac{\chi^2_{2n}(\frac{\alpha}{2})}{2\sum X_i} \leq \lambda \leq \frac{\chi^2_{2n}(1-\frac{\alpha}{2})}{2\sum X_i}\right) = 1-\alpha$$

Note that: $\frac{1}{2} \chi^2_{2n} \sim \text{Gamma}(\frac{2n}{2}, 2(\frac{1}{2})) = G(n, 1)$

Therefore, we have

$$P\left(\frac{G^{-1}(\frac{\alpha}{2})}{\sum X_i} \leq \lambda \leq \frac{G^{-1}(1-\frac{\alpha}{2})}{\sum X_i}\right) = 1-\alpha$$

Since $\hat{\lambda}_n = \frac{n}{\sum X_i}$, we get

$$P\left(\hat{\lambda}_n G^{-1}(\frac{\alpha}{2})/n \leq \lambda \leq \hat{\lambda}_n G^{-1}(1-\frac{\alpha}{2})/n\right) = 1-\alpha$$

and so, the exact C.I. for λ is given by

$$(\hat{\lambda}_n G^{-1}(\frac{\alpha}{2})/n, \hat{\lambda}_n G^{-1}(1-\frac{\alpha}{2})/n)$$

as proved.

Figure 1: Proof for 4d.