Statistics 3858b Assignment 2

Handout February 12, 2018; Due date: Feb 28, 2018 (after the reading week)

These problems are all from the course text unless otherwise stated.

1. 8.10.16 a-c; part (c) means to calculate the variance for the limiting or normal approximation of the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$, which involves $I(\theta)$; see theorem 8.5B.

2. 8.10.18. For (d) find a sufficient statistic of dimension 1, that is a minimal sufficient statistic; it is okay for this problem to leave derivatives of $\log(\Gamma(\cdot))$ as derivatives without simplifying as they are in terms of trigamma which must be evaluated numerically.

3. 8.10.55 (change the parts as follows). (a) and find Fisher's information;

   (b)

   (c) - use the parametric bootstrap to approximate the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$; Use $M = 1000$ replicates in your bootstrap. How should this be related to your part (a).

   (d) give the 95% CI for θ based on the result in (c).

   Aside : For a given value of θ in part (c) you will need to simulate from 4-nomial distribution. You can make use of the R function rmultinom; see the R help for multinomial distribution.

4. 9.11.4

5. 9.11.17

Suggested problems

1. 8.10.19

2. 8.10.48 (recall chapter 4, approximation of moments).

   Also write a R program to compare, with discussion, these two estimators using the Monte Carlo simulation method, at $\lambda = .1$ and 1. Do these for data sample $n = 50$ and 100, and $M = 1000$ Monte Carlo replicates. Within the R code, calculate the efficiency (ratio of variances) and comment on how well these compare with the propogation of errors method above. Recall Section 4.6, and page 162 where *propogation of errors is discussed*.

   Remark : This is yet a third example of an estimator. It is neither the MLE or method of moments estimator.

3. 8.10.52

4. 8.10.46. Bootstrap here (part e and f) means the parametric bootstrap as discussed in class. For part (g) obtain the confidence intervals marginally, that is for each parameter separately; for this you will need to obtain the observed information matrix $I(\hat{\theta})$.

5. 8.10.57

6. 8.10.59

7. 8.10.60 (g)

   Hint : Show that $Y = \tau X$, where X has the distribution in the problem, then $Y \sim$ exp, mean 1.

   Next find the distribution of $\tau\bar{X}$ (it is a Gamma distribution, but this distribution does not involve the parameter τ). Then using quantiles of the distribution of Y construct a confidence interval for τ by solving for τ from the two inequalities

   $$q_L \leq \tau\hat{\tau} \leq q_U$$

   where q_L, q_U are lower and upper quantiles of the distribution of Y. If $n = 10$ what are these quantiles so that the confidence interval is a 95% confidence interval.

8. 8.10 problems 68-73.

9. This problem shows the student that the estimation methods studied in this course can be used for stochastic processes such as Markov chains. Here we only consider a simple two state Markov chain. For students who have not yet studied Markov chains, you should skip this problem.

   A Markov chain with state space $\{0, 1\}$ has transition matrix

$$P = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix} .$$

   Suppose the starting value for the Markov chain is $X_0 = 0$, and that data given this starting position is $X_1, X_2, \ldots, X_n$. The initial state is $x_0 = 0$.

   In this problem you will work through the (conditional upon initial value 0) MLE for this Markov chain. The purpose of this problem is to show the student that estimation can be done for Markov chains where one gets to observe the transitions.

   For students who have not studied Markov chains, at least the part which allows one to find the probability of $X_1, X_2, \ldots, X_n$, such students can skip this problem. An R program to simulate a Markov chain will be provided.

   (a) Give the log likelihood function. Obtain a formula for the MLE of α and β.

   (b) A data set of length $n + 1$, with $n = 200$ is obtained and has the following summary statistics

$$n_{0,0} = 95, n_{0,1} = 32, n_{1,0} = 32, n_{1,1} = 41$$

   where $n_{l,k}$ counts the number of transitions from state l to state k in the data.

   Obtain the maximum likelihood estimators of α and β.

   (c) Using always the starting value or initial value of $X_0 = 0$, write R code to obtain the parametric bootstrap simulation approximation for the sampling distribution of the MLE.

   You may make use of some relevant R functions given on the assignments page in your R code. These allow you to simulate the two state Markov chain and calculate the sample statistics needed.

   (d) Using your code and the estimates above give the 95% confidence interval for α based on the parametric bootstrap approximation, with $M = 500$ bootstrap replicates.