Statistical Inference and Data Analysis: Exam 2019-2020

January 2020

1 Closed Book Part

1.1 Question 1

Suppose we have the following testing problem:

$$H_0 : \theta = \theta_0 \text{ and } H_1 : \theta \neq \theta_0$$

a) Define what a likelihood ratio statistic is, and what is its asymptotic distribution.

b) Suppose we have a sample $X_1, ..., X_n$ of X that follows a Bernoulli distribution with parameter $0 < \theta < 1$, what is the likelihood ratio statistic?

c) Take now -2 times the ln of this statistic, describe the relation between this new statistic and the normal statistic for proportion $\frac{\theta - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}$

1.2 Question 2

a) Explain what a pivotal quantity is.

b) Suppose X has an exponential distribution with parameter θ, find a pivotal quantity and derive an exact $100(1-\alpha)\%$ confidence interval.

1.3 Question 3

Suppose we have a sample $X_1, ..., X_n$ of X which is normal distributed with $N(\mu, \mathbf{\Sigma})$.

a) Derive the maximum likelihood estimator of the samples.

b) Find an exact $100(1-\alpha)\%$ confidence interval for μ.

c) Find an asymptotic $100(1-\alpha)\%$ confidence interval for μ. *(An extra condition was included in this question which i forgot)*

1.4 Question 4

Suppose we have the following variables $(U_1, U_2, U_1 + U_2)^T$

a) Explain what the principal components are.

b) Calculate the proportion of variance of the first principle component.

2 Open Book Part

2.1 Question 1

Suppose we have a random sample $X_1, ..., X_n$ of a normal distribution X with $N(\theta, \theta)$. Hence $E[X^4] = \theta^4 + 6\theta^3 + 3\theta^2$

a) Show that the score function gives the following quadratic equation:

$$\theta^2 + \theta - \frac{1}{n}\sum_i^n X_i^2 = 0$$

and calculate the maximum likelihood estimator, denote it with θ_n^{MLE} (You don't have to check the second derivative)

b) Find the asymptotic distribution of the θ_n^{MLE}.

c) Find the asymptotic relative efficiency $ARE(\bar{X}_n, \theta_n^{MLE})$ and discuss which estimator is better than the other.

d) Suppose we have the following testing problem:

$$H_0 : \theta = \theta_0 \text{ and } H_1 : \theta = \theta_1$$

d1) Show that the credible region is $R = \{X_i^2 > c\}$, you don't have to calculate c.

d2) Suppose now $R = \{\sqrt{n}\frac{\bar{X_n}-\theta_0}{\sqrt{\theta_0}} > c\}$.

i) What is c?

ii) Calculate the probability of a type II error.

iii) Determine the sample size, such that the region is bounded by ... *(I don't remember the exact question)*

2.2 Question 2

A lot of data is given in a table with their eigenvalues and with a 'biplot'.

(Ofcourse, i can't reproduce this data)

a) Calculate the proportion of variance of the first three components.

b) Assign names to the codenames *(=the names of the variables were left out in the plot).*

c) Give a possible interpretation of the first two components.

2.3 Question 3

Suppose we have a linear regression model where the error terms are normal distributed with mean zero. An estimator for this model is $\hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}} = \text{argmin}_\beta \boldsymbol{(Y - X\beta)^T(Y - X\beta)} + \lambda\boldsymbol{\beta^T\beta}$ *(NOTE: argmin means smallest value of this expression)*

a) Show that $\hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}} = (\boldsymbol{X^T X} + \lambda I_n)^{-1}\boldsymbol{X^T Y}$

b) Calculate the bias of this estimator.

c) What will the bias become when $\lambda = 0$, $\lambda \to \infty$