Statistical Inference and Data Analysis

Michaël Liefsoens

June 30th, 2020

Disclaimer: Due to Covid19, this exam was supposed to be finished in three hours, instead of the usual four hours.

1 Closed Book Part

1.1 Question 1

Let $R(T_n, \theta)$ be a risk function.

a) Define what is meant by a Bayes estimator.

b) Let X be a r.v. with density $f(\cdot, \theta)$ and consider a continuous prior $\pi(\theta)$. Provide arguments as to why the Bayes estimator $T_B(\boldsymbol{X})$ is given by

$$T_B(\boldsymbol{X}) = \frac{\int_\Theta \theta \left[\prod_{i=1}^{n} f(x_i;\theta)\right] \pi(\theta)\, d\theta}{\int_\Theta \left[\prod_{i=1}^{n} f(x_i;\theta)\right] \pi(\theta)\, d\theta}.$$

c) Now consider $X \sim \text{Bernoulli}(\theta)$ and suppose that the prior is given by a $\text{Beta}(\alpha, \beta)$ distribution. Give the Bayes estimator for θ.

d) What is its distribution?

e) Give a $(1-\alpha)$ credible region for θ.

1.2 Question 2

Suppose we have $Y_{ij} = \mu_i + \epsilon_{ij}$ for $i = 1, \ldots, k$ and $j = 1, \ldots, n_i$. Suppose $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ and independent. Explain how you would test whether or not the μ_i are equal.

1.3 Question 3

Let $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}$ be $2p$ multivariate normally distributed with

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \sim \mathcal{N}_{2p}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma} \\ \boldsymbol{\Sigma} & \lambda\boldsymbol{\Sigma} \end{pmatrix} \right), \text{ with } \lambda > 1.$$

Define $\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 - \boldsymbol{X}_1 \end{pmatrix}$

a) What is the distribution of $\boldsymbol{Y}$?

b) Show that $\boldsymbol{X}_1$ and $\boldsymbol{X}_2 - \boldsymbol{X}_1$ are independent.

2 Open Book Part

2.1 Question 1

Consider X with cumulative distribution function

$$F_X(x) = P(X \leq x) = (1 - e^{-x})^{\frac{1}{\theta}}, x \geq 0, \theta > 0.$$

a) Find the cumulative distribution of $W = -\log(1 - e^{-X})$. What is its distribution?

b) Find the MLE for θ and state the asymptotic normality result.

c) Find an approximate $(1 - \alpha)$ confidence interval for θ. Limit the use of approximations.

d) Use (a) to give an exact $(1 - \alpha)$ confidence interval.

2.2 Question 2

Define $\boldsymbol{X} = \begin{pmatrix} U_1 & U_2 & U_1 + U_2 \end{pmatrix}^T$ with U_1 and U_2 uniformly distributed over $[0, 1]$.

a) Calculate the variance of $\boldsymbol{X}$.

b) Explain how you would calculate the principal components of $\boldsymbol{X}$, say $\boldsymbol{Z}$.

c) Show that $\text{corr}(\boldsymbol{Z}_i, \boldsymbol{X}_k) = \sqrt{\frac{\lambda_i}{\boldsymbol{\Sigma}_{kk}}}(\boldsymbol{u}_i)_k$, for $i, k = 1, 2, 3$. Here, $\boldsymbol{u}_i$ are the principal directions of $\boldsymbol{X}$ and $\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{u}_1^T\boldsymbol{X} & \boldsymbol{u}_2^T\boldsymbol{X} & \boldsymbol{u}_3^T\boldsymbol{X} \end{pmatrix}^T$, the vector of scores.