M3S3/M4S3 - EXERCISES 3

EFFICIENT ESTIMATION AND TESTING

1. Suppose that Σ is a variance-covariance matrix for k dimensional random variable $\boldsymbol{X} = (X_1, \ldots, X_k)^\mathsf{T}$.

(a) Show that Σ is positive definite.

Hint: let $\boldsymbol{a}$ be an arbitrary real k-vector. You must show that $a^\mathsf{T}\Sigma\, a > 0$. Proceed by computing the variance of the scalar random variable Y formed as the linear combination

$$Y = \sum_{i=1}^{k} a_i X_i = \boldsymbol{a}^\mathsf{T}\boldsymbol{X}.$$

(b) Suppose that Σ is written as a blocked matrix

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

Find the inverse matrix $\Pi = \Sigma^{-1}$ in the form of a blocked matrix

$$\Pi = \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix}$$

by noting that $\Sigma\,\Pi = \mathbf{1}_k$, performing the matrix multiplication, and solving the resulting four matrix equations.

Note: you must obey the rules of matrix multiplication, and remember that, here, only square matrices have inverses.

2. Let $I(\boldsymbol{\theta})$ be the Fisher Information for a two-parameter probability density distribution, $f_{X|\boldsymbol{\theta}}(x|\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\theta_1, \theta_2)^\mathsf{T}$. Denote $I(\boldsymbol{\theta})$ and $I(\boldsymbol{\theta})^{-1}$ by

$$I(\boldsymbol{\theta}) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix} \qquad I(\boldsymbol{\theta})^{-1} = \begin{bmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{bmatrix}.$$

where I_{ij}, I^{ij}, $i, j = 1, 2$ are scalar quantities. Under what conditions does the inequality $(I_{11})^{-1} < I^{11}$ hold, if $I(\boldsymbol{\theta})$ is presumed to be positive definite ?

3. Suppose that probability model $f_{X|\boldsymbol{\theta}}(x|\boldsymbol{\theta})$ is dependent on vector parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^\mathsf{T}$. By using the second-order Taylor *approximation* to the log-likelihood around the MLE, $\widehat{\boldsymbol{\theta}}_n$,

$$\boldsymbol{l}_n(\boldsymbol{\theta}) = \boldsymbol{l}_n(\widehat{\boldsymbol{\theta}}_n) + \dot{\boldsymbol{l}}_n(\widehat{\boldsymbol{\theta}}_n)(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n) + \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)^\mathsf{T}\ddot{\boldsymbol{l}}_n(\widehat{\boldsymbol{\theta}}_n)(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)$$

construct a quadratic approximation to the log-likelihood near to $\widehat{\boldsymbol{\theta}}_n$. Here $\dot{\boldsymbol{l}}_n(\widehat{\boldsymbol{\theta}}_n)$ is a **row** vector.

Derive the quadratic approximation in the following one parameter models

(a) $X \sim Poisson(\lambda)$

(b) $X \sim N(0, \sigma^2)$.

4. The Wald and Rao/Score test statistics derived from a sample of size n, W_n and R_n, for testing

$$\begin{aligned} H_0 &: \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ H_1 &: \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{aligned}$$

when the log density $\log f_{X|\boldsymbol{\theta}}$ admits a finite second derivative with respect to $\boldsymbol{\theta}$ are given by

$$W_n = n(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^T \widehat{I}_n(\widetilde{\boldsymbol{\theta}}_n)(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \tag{1}$$

where $\widetilde{\boldsymbol{\theta}}_n$ is a solution to the likelihood equations,

$$\widehat{I}_n(\widetilde{\boldsymbol{\theta}}_n) = \begin{cases} I(\widetilde{\boldsymbol{\theta}}_n) \\ \\ \dfrac{1}{n}\sum\limits_{i=1}^{n} S(X_i, \widetilde{\boldsymbol{\theta}}_n) S(X_i, \widetilde{\boldsymbol{\theta}}_n)^T \\ \\ -\dfrac{1}{n}\sum\limits_{i=1}^{n} \Psi(\widetilde{\boldsymbol{\theta}}_n, X_i) \end{cases}$$

where is an estimator (with corresponding estimate) of the Fisher Information I derived from the sample,

$$S(X;\boldsymbol{\theta}) = \begin{bmatrix} \dfrac{\partial \log f_{X|\boldsymbol{\theta}}(X|\boldsymbol{\theta})}{\partial \theta_1} \\ \vdots \\ \dfrac{\partial \log f_{X|\boldsymbol{\theta}}(X|\boldsymbol{\theta})}{\partial \theta_d} \end{bmatrix} = \dot{\boldsymbol{l}}(\boldsymbol{\theta}, X)^{\mathsf{T}} \qquad \Psi(\boldsymbol{\theta}, X) = \left[\frac{\partial^2 \log f_{X|\boldsymbol{\theta}}(X|\boldsymbol{\theta})}{\partial \theta_j \partial \theta_l} \right]_{jl} = \ddot{\boldsymbol{l}}(\boldsymbol{\theta}; X)$$

are $d \times 1$ and $d \times d$ quantities respectively, and

$$R_n = Z_n^T \left[I\left(\boldsymbol{\theta}_0\right)\right]^{-1} Z_n \qquad \text{where} \qquad Z_n \equiv Z_n\left(\boldsymbol{\theta}_0\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S\left(X_i, \boldsymbol{\theta}_0\right) \tag{2}$$

(a) Show that, in the one parameter case, the statistics can be expressed as

$$W_n = -(\widetilde{\theta}_n - \theta_0)^2 \ddot{l}_n(\widetilde{\theta}_n) \qquad R_n = -\left\{\dot{l}_n\left(\theta_0\right)\right\}^2 \left\{\ddot{l}_n\left(\theta_0\right)\right\}^{-1}$$

(b) Derive the forms of the Wald, Rao/Score and Likelihood Ratio statistics for testing

$$\begin{aligned} H_0 &: \lambda = \lambda_0 \\ H_1 &: \lambda \neq \lambda_0 \end{aligned}$$

if the data follow a Poisson distribution with parameter $\lambda > 0$.

(c) Derive the form of the Wald statistic when the data are presumed normally distributed with parameters $N\left(\mu, \sigma^2\right)$ in a test of

$$\begin{aligned} H_0 &: \mu = 0 \\ H_1 &: \mu \neq 0 \end{aligned}$$

(i) when σ^2 is **unspecified** under the null and the alternative; the MLE for σ^2 under the null is

$$S_0^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

whereas under the alternative, the joint MLE is $\left(\overline{X}, S^2\right)$ where

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad S^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2.$$

(ii) when $H_0 : (\mu, \sigma) = \boldsymbol{\theta}_0 = \left(0, \sigma_0^2\right)$ and $H_1 : (\mu, \sigma) \neq \boldsymbol{\theta}_0$.