Topics in statistics III/IV (MATH3361/4071)	Michaelmas term, 2018-2019

Homework 4: One step estimators, Confidence intervals, & hypothesis test

Lecturer: Georgios Karagiannis georgios.karagiannis@durham.ac.uk

Exercise 1

(Log likelihood ratio statistic) Let $x_1, x_2, ..., x_n$ be IID random variables generated from a distribution f_θ labeled by a d-dimensional parameter $\theta \in \Theta \subset \mathbb{R}^d$, and admitting PDF $f(\cdot|\theta)$. Assume the conditions from the Cramér Theorem are satisfied, and that θ_0 is the true value. Prove that

$$W_{\text{LR}}(\theta_0) = -2(\ell_n(\theta_0) - \ell_n(\hat{\theta}_n)) \xrightarrow{D} \chi^2_d$$

it is where $\hat{\theta}_n$ is the MLE of θ.

Hint-1 Expand $\ell_n(\theta_0)$ around $\hat{\theta}_n$ by Taylor expansion

Hint-2 Prove that $W_{\text{LR}}(\theta_0) \xrightarrow{a.s} n(\theta_0 - \hat{\theta}_n)^T \mathcal{I}(\theta_0)(\theta_0 - \hat{\theta}_n)$

Hint-3 Prove that $W_{\text{LR}}(\theta_0) \xrightarrow{D} \chi^2_d$

Exercise 2

Let $x_1, ..., x_n \overset{IID}{\sim} f_\theta$ with unknown parameter $\theta \in (0, \infty)$ and PDF

$$f(x|\theta) = \begin{cases} \theta \exp(-x) + (1-\theta) x \exp(-x) & , x \geq 0 \\ 0 & , x < 0 \end{cases}$$

1. Calculate the moment estimator $\tilde{\theta}_n$ of θ, (I give you a bit of freedom here)
2. Calculate the asymptotic distribution of the $\tilde{\theta}_n$
3. Find the 1-step estimator $\breve{\theta}_n$ of θ such that it can be asymptotically efficient.

Hint: Recall that $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} \mathrm{d}x$, and $\Gamma(a) = (a-1)\Gamma(a-1)$

Exercise 3

Let

$$y_i \overset{\text{ind}}{\sim} \text{Bin}(n, \pi_i)$$

where $i = 1, ..., N$. Consider that the probability of success is modeled such as

$$\text{logit}(\pi_i) = x_i^T \theta \tag{1}$$

where $\text{logit}(\pi_i) = \log(\frac{\pi_i}{1-\pi_i})$. Here $x_i = (x_{i,1}, ..., x_{i,d})^T$ are known vertors containing the values of the d regessions at the i-th observation, and $\theta \in \mathbb{R}^d$.

1. Show that

$$\pi_i = \frac{e^{x_i^T \theta}}{1 + e^{x_i^T \theta}}$$

2. Assume that the MLE $\hat{\theta}$ of θ is known/calculated. Show that the $(1 - a)$ Wald confidence interval for the unknown parameter θ, by using the observed information matrix, is

$$\text{C.I.} : \{ \theta \in \mathbb{R}^d \, : \, (\hat{\theta}_n - \theta)^T X^T (\text{diag}_{\forall i}(n\hat{\pi}_i(1 - \hat{\pi}_i))) X (\hat{\theta}_n - \theta) \leq \chi^2_{d,1-a} \ \}$$

where

$$\hat{\pi}_i = \frac{e^{x_i^T \hat{\theta}}}{1 + e^{x_i^T \hat{\theta}}}$$

X is the so called design matrix from the regression

$$\begin{bmatrix} \text{logit}(\pi_1) \\ \vdots \\ \text{logit}(\pi_N) \end{bmatrix} = \underbrace{\begin{bmatrix} \longleftarrow & x_1^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & x_N^T & \longrightarrow \end{bmatrix}}_{=X} \theta$$

and $\text{diag}_{\forall i}(\heartsuit_i) = \begin{bmatrix} \heartsuit_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \heartsuit_N \end{bmatrix}$.

3. Find the score statistic rejection area for the hypothesis test $\text{H}_0 : \theta = \theta_*$ versus $H_1 : \theta \neq \theta_*$.