Stat 300A Theory of Statistics

Homework 6

Andrea Montanari *Due on November 14, 2018*

- Solutions should be complete and concisely written. Please, use a separate sheet (or set of sheets) for each problem.
- We will be using Gradescope (`https://www.gradescope.com`) for homework submission (you should have received an invitation) - no paper homework will be accepted. Handwritten solutions are still fine though, just make a good quality scan and upload it to Gradescope.
- You are welcome to discuss problems with your colleagues, but should write and submit your own solution.

1: Sufficient statistics (TPE 1.6.32,1.6.33))

(a) Consider two statistical models (i.e. two classes of distributions) $\mathscr{P}_0$, $\mathscr{P}_1$ on the same sample space $\mathcal{X}$, such that $\mathscr{P}_0 \subseteq \mathscr{P}_1$. Let $\boldsymbol{T}$ be a sufficient statistics for $\mathscr{P}_1$. Show that it is sufficient for $\mathscr{P}_0$ as well.

(b) Continuing from the previous point, assume that, for any $\mathsf{P} \in \mathscr{P}_1$ there exists $\mathsf{P}' \in \mathscr{P}_0$ such that, for any $N \subseteq \mathcal{X}$ with $\mathsf{P}'(N) = 0$, we have $\mathsf{P}(N) = 0$. Show that, if $\boldsymbol{T}$ is sufficient for $\mathscr{P}_0, \mathscr{P}_1$, and is complete for $\mathscr{P}_0$, then it is complete for $\mathscr{P}_1$.

(c) Let $\mathscr{P}$ be the family of distributions of n i.i.d. random variables $X_1, \dots, X_n$, with some common density $\mathsf{p}(\,\cdot\,)$ on $\mathbb{R}$. (In other words $\mathscr{P} = \{\mathsf{p}^{\otimes n} : \ \mathsf{p} \ \text{ is a density on } \ \mathbb{R}\}$ is a class of probability distributions on $\mathbb{R}^n$.) Prove that the order statistics $X_{(1)}, \dots, X_{(n)}$ is complete.

[Hint: you can use the submodel $\mathscr{P}_0$ formed by density of the form $\exp\{\theta_1 \sum_{i=1}^n x_i + \cdots + \theta_n \sum_{i=1}^n x_i^n - \sum_{i=1}^n x_i^{2n}$. You can also use the fact that if $\sum_{i=1}^n a_i^k = \sum_{i=1}^n b_i^k$ for all $k \le n$, then $(a_i)_{i\le n}$ is a permutation of $(b_i)_{i\le n}$.]

(d) Continuing from the previous point, determine an UMVU estimator of $\mathsf{P}(X_1 \le x) = \int_{-\infty}^x f(t)\,\mathrm{d}t$ in the class $\mathscr{P}$.

2: Unbiased estimation from Binomials (TPE 2.1.17)

Let $\mathsf{P}_\theta = \mathrm{Binom}(n, \theta)$, $\theta \in \Theta = [0,1]$. We consider unbiased estimation of $g(\theta) = \theta^3$.

(a) Show that, for $n \le 2$, no unbiased estimator exists. What happens for $n = 3$?

(b) Use the orthogonality condition to construct an UMVU estimator for $n > 3$.

3: Logistic regression

Consider a logistic regression model, where we are given i.i.d. pairs $(Y_i, \boldsymbol{X}_i)$, $i \le n$, with $\boldsymbol{X}_i \in \mathbb{R}^d$ a feature vector, and $Y_i \in \{0,1\}$ a label (or response variable). The distribution of the $\boldsymbol{X}_i$ given by $\mathsf{p}_{\boldsymbol{X}}$, and the distribution of Y_i given $\boldsymbol{X}_i$ is given by

$$\mathsf{P}_{\boldsymbol{\theta}}\big(Y_i = y_i \big| \boldsymbol{X}_i = \boldsymbol{x}_i\big) = \frac{e^{y_i \langle \boldsymbol{\theta}, \boldsymbol{x}_i\rangle}}{1 + e^{\langle \boldsymbol{\theta}, \boldsymbol{x}_i\rangle}}\,. \tag{1}$$

(a) Derive an expression for the Fisher information matrix $\mathbf{I}_{\mathrm{F}}(\boldsymbol{\theta})$.

(b) Assume $\mathsf{p}_{\boldsymbol{X}} \sim \mathsf{N}(\mathbf{0}, \mathrm{I}_d)$. Show that

$$\mathbf{I}_{\mathrm{F}}(\boldsymbol{\theta}) = c_0 \mathrm{I}_d + c_1 \boldsymbol{\theta}\boldsymbol{\theta}^{\mathsf{T}}\,, \tag{2}$$

where $c_0 = c_0(\|\boldsymbol{\theta}\|)$ and $c_1 = c_1(\|\boldsymbol{\theta}\|_2)$ are scalars that depend on the norm of $\boldsymbol{\theta}$. Provide expressions for c_0, c_1 in terms of one-dimensional integrals.

(c) Generalize the previous formula for $\mathsf{p}_{\boldsymbol{X}} \sim \mathsf{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

[Hint: In solving this problems, it might be useful to remember Gaussian integration by parts. If $\boldsymbol{X} \sim \mathsf{N}(\mathbf{0}, \boldsymbol{\Sigma})$ takes values in $\mathbb{R}^d$, and $f : \mathbb{R}^d \to \mathbb{R}$ is differentiable and such that the expectations below make sense, then

$$\mathbb{E}\{X_i\, f(\boldsymbol{X})\} = \sum_{j=1}^{d} \Sigma_{ij}\; \mathbb{E}\Big\{\frac{\partial f}{\partial x_j}(\boldsymbol{X})\Big\}\,. \tag{3}$$

]