30100: Mathematical Statistics I **Winter 2022**

Homework 8

Lecturer: Chao Gao

1. For a location family $f(x-\theta)$, show the Fisher information is given by the constant $I_f = \int \frac{(f')^2}{f}$. For a location-scale family $\frac{1}{\tau} f\left(\frac{x-\theta}{\tau}\right)$, show the Fisher information about θ is $\tau^{-2} I_f$.

2. For $X \sim P_\theta$, the Fisher information is given by $I(X;\theta) = \mathbb{E}_\theta \left(\frac{\partial}{\partial\theta} \log p_\theta(X)\right)^2$, where the expectation is taken with respect to the distribution $X \sim P_\theta$. Now consider $(X,Y) \sim P_\theta$, one can similarly define $I(X,Y;\theta)$ and $I(X;\theta)$. Define the conditional Fisher information by $I(Y;\theta|X) = \mathbb{E}_\theta \left(\frac{\partial}{\partial\theta} \log p_\theta(Y|X)\right)^2$, where the expectation is taken with respect to the distribution $(X,Y) \sim P_\theta$.

   (a) Show that Fisher information has chain rule $I(X,Y;\theta) = I(X;\theta) + I(Y;\theta|X)$.

   (b) Consider $X \sim P_\theta$, and $T(X)$ is any statistic. Show data processing inequality: $I(X;\theta) \geq I(T(X);\theta)$. In other words, any manipulation of data will not increase its information about the parameter.

   (c) Show $I(X;\theta) = I(T(X);\theta)$ if $T(X)$ is sufficient.

3. Consider i.i.d. samples $X_1, \cdots, X_n \sim \text{Uniform}(0,\theta)$. Show the MLE $X_{(n)}$ is not asymptotically normal by deriving a limiting distribution for $n(X_{(n)} - \theta)$.

4. Consider i.i.d. samples $X_1, \cdots, X_n \sim P_{\theta^*}$. An M-estimator is defined by

$$\hat{\theta} = \operatorname*{argmax}_\theta \frac{1}{n} \sum_{i=1}^n m(X_i, \theta).$$

   It can be viewed as an extension of the MLE. To guarantee consistency, one requires that $\mathbb{E}_{\theta^*} m(X,\theta)$ is maximized at $\theta = \theta^*$, or $\frac{\partial}{\partial\theta} \mathbb{E}_{\theta^*} m(X,\theta)|_{\theta=\theta^*} = 0$. Let us write $M_n(\theta) = \frac{1}{n}\sum_{i=1}^n m(X_i,\theta)$ and $M(\theta) = \mathbb{E}_{\theta^*} m(X,\theta)$, and use ν_n for the empirical process operator. Assume

   - $|\hat{\theta} - \theta^*| = o_{P_{\theta^*}}(1)$.
   - $m(x, \theta^* + t) = m(x,\theta^*) + t\Delta_{\theta^*}(x) + |t| r(x,t)$, where $\Delta_{\theta^*}(\cdot)$ satisfies $\mathbb{E}_{\theta^*} \Delta_{\theta^*}(X) = 0$ and $r(\cdot,t)$ satisfies $\sup_{t \in U_n} \frac{|\nu_n r(\cdot,t)|}{1+\sqrt{n}|t|} = o_{P_{\theta^*}}(1)$ for any shrinking neighborhood U_n of 0.
   - $M(\theta^* + t) = M(\theta^*) - \frac{1}{2} J_{\theta^*} t^2 + o(t^2)$ for some $J_{\theta^*} > 0$. Note that the first-order term is 0 because of the consistency requirement $\frac{\partial}{\partial\theta} \mathbb{E}_{\theta^*} m(X,\theta)|_{\theta=\theta^*} = 0$.

   Generalize the asymptotic normality of MLE to the M-estimator.

(a) Derive a quadratic expansion of $M_n(\theta)$.

(b) Show $|\hat{\theta} - \theta^*| = O_{P_{\theta^*}}(n^{-1/2})$.

(c) Show $\sqrt{n}(\hat{\theta} - \theta^*) \rightsquigarrow N\left(0, \frac{\mathbb{E}_{\theta^*}\Delta_{\theta^*}(X)^2}{J_{\theta^*}^2}\right)$.

(d) Among all M-estimators, the MLE is the best in the sense that $\frac{\mathbb{E}_{\theta^*}\Delta_{\theta^*}(X)^2}{J_{\theta^*}^2} \geq \frac{1}{I_\theta^*}$. Prove this inequality (you may assume all necessary regularity conditions that allow you to switch the order of integral and derivative as needed). Hint: note that $J_{\theta^*} = -\frac{\partial^2}{\partial\theta^2}\mathbb{E}_{\theta^*}m(X,\theta)|_{\theta=\theta^*}$ and apply Cauchy-Schwarz.

5. Consider i.i.d. samples $X_1, \cdots, X_n \sim N(\theta^*, \sigma^2)$ and $\hat{\theta} = \text{Median}(X_1, \cdots, X_n)$.

   (a) Write $\hat{\theta}$ as an M-estimator.

   (b) Show $M(\theta)$ is maximized at θ^*.

   (c) Find $\Delta_{\theta^*}(\cdot)$ and J_{θ^*} and show $\mathbb{E}_{\theta^*}\Delta_{\theta^*}(X) = 0$.

   (d) Apply the conclusion of Question 4, and find the limiting distribution of $\sqrt{n}(\hat{\theta} - \theta^*)$.