Statistical Theory - Problem Sheet 2

Spring 2023. Email corrections to Kolyan Ray: kolyan.ray@ic.ac.uk

Instructions: Please attempt the non-starred questions. If you have time, attempt the starred questions (they are not necessarily more difficult).

1. Consider i.i.d. random variables $X_1, \dots, X_n$. For each of the following parametric models of pmf/pdf's, find the MLE of the unknown parameter, the score equation (i.e. $\nabla_\theta \ell_n(\theta) = 0$, where ℓ_n is the log-likelihood based on $X_1, \dots, X_n$) and the Fisher information.

(a) $X_i \sim^{iid}$ Bernoulli(θ), $\theta \in [0, 1]$.

(b) $X_i \sim^{iid} N(\theta, 1)$, $\theta \in \mathbb{R}$.

(c) $X_i \sim^{iid} N(0, \theta)$, $\theta \in (0, \infty)$.

(d) $X_i \sim^{iid} N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)^T \in \mathbb{R} \times (0, \infty)$.

(e) $X_i \sim^{iid}$ Poisson(θ), $\theta \in (0, \infty)$.

(f) $X_i \sim^{iid}$ from model $\{f_\theta : \theta \in (0, \infty)\}$ with pdf $f_\theta(x) = (1/\theta)e^{-x/\theta}$, $x \geq 0$.

(g) $X_i \sim^{iid}$ from model $\{f_\theta : \theta \in (0, \infty)\}$ with pdf $f_\theta(x) = \theta e^{-\theta x}$, $x \geq 0$.

Note: *we already worked out the MLEs for several of these examples in the notes or Problem Sheet 1, so these can just be recalled. We include them all here since we will expand on these examples in future questions.*

2. In which of the examples in Question 1 is the MLE unbiased? When unbiased, deduce whether the variance of the MLE attains the Cramèr-Rao lower bound or not.

3. Let $X_1, \dots, X_n \sim^{iid}$ Poisson(θ) with parameter $\theta > 0$, and let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. Show that $\mathrm{Var}_\theta(\bar{X}_n) \leq \mathrm{Var}_\theta(S_n^2)$ for all $\theta > 0$.

4. Based on n i.i.d. random variable $X_1, \dots, X_n$ from one of the following parametric models, find the maximum likelihood estimator of the parameter ϕ: (i) $\phi = \mathrm{Var}(X_i)$ in a Poisson(θ) model, (ii) $\phi = \mathrm{Var}(X_i)$ in a Bernoulli(θ)-model, (iii) $\phi = (EX_i)^2$ in a $N(\theta, 1)$ model.

5. Let $I_{X_1}(\theta)$ and $I_{(X_1,\dots,X_n)}(\theta)$ be the Fisher information based on $X_1 \sim f_\theta$ and $X_1, \dots, X_n \sim^{iid} f_\theta$, respectively, where $\{f_\theta : \theta \in \Theta\}$ is a 'regular' statistical model (in the sense of the lecture notes). Show that $I_{(X_1,\dots,X_n)}(\theta) = nI_{X_1}(\theta)$.

6. Give an example of functions $g_n, g : \Theta \to \mathbb{R}$, where $\Theta \subset \mathbb{R}$, that have unique maximisers $\hat{\theta}_n, \theta_0$, respectively, such that $g_n(\theta) \to g(\theta)$ for every $\theta \in \Theta$ as $n \to \infty$, but $\hat{\theta}_n \not\to \theta_0$.

This explains why the weak law of large numbers is insufficient to obtain consistency of the MLE when looking at the sketch proof in the lecture notes. We instead require a uniform law of large numbers, which corresponds to uniform convergence of $g_n \to g$.

7. Find the MLE for an i.i.d. sample $X_1, \dots, X_n$ arising from the models (a) $N(\theta, 1)$, where $\theta \in \Theta = [0, \infty)$ and (b) $N(\theta, \theta)$, where $\theta \in \Theta = (0, \infty)$.

8*. Consider $X_1, \dots, X_n$ i.i.d. random variables with probability density

$$f_\theta(x) = \frac{1}{2} e^{-|x-\theta|}, \quad x \in \mathbb{R},\ \theta \in \mathbb{R},$$

of *Laplace distributions*. If n is odd, show that the MLE is equal to the sample median. Discuss what happens when n is even. Can you calculate the Fisher information?

9*. Consider observing an $n \times 1$ random vector $Y \sim N_n(X\theta, I_n)$, where X is a non-random $n \times p$ matrix of full column rank, where $\theta \in \Theta = \mathbb{R}^p$ for $p \leq n$, and where I_n is the $n \times n$ identity matrix. Compute the MLE and find its distribution. Calculate the Fisher information for this model and compare it to the variance of the MLE. Deduce, as a special case, the form of the MLE and Fisher information in the case when $p = n$ and $X = I$.

10. For the examples in Question 1, derive directly (without using the general asymptotic theory for MLEs) the asymptotic distribution of $\sqrt{n}(\hat{\theta}_{MLE} - \theta)$ as $n \to \infty$.

11*. Consider *one* observation of a random vector $X = (X_1, X_2)^T$ from a bivariate normal distribution $N_2(\theta, \Sigma)$, where $\theta = (\theta_1, \theta_2)^T$ and where Σ is an arbitrary but *known* 2×2 positive definite covariance matrix.

(i) Compute the Cramèr-Rao lower bound for estimating the first coefficient θ_1 if (a) θ_2 is known and (b) if θ_2 is unknown.

(ii) Show that the two bounds in (i) coincide when Σ is a diagonal matrix.

(iii) Show that the bound in (i)(a) is always less than or equal to the bound in (i)(b), and give an information-theoretic interpretation of this result.