Statistics 583, Problem Set 2

Wellner; 4/6/2016

Reading: Van der Vaart, Chapter 5, pages 41-67;
JW, 580s Course Notes, Chapter 4, section 4, pages 25-29;
JW, Delft-Cortona-Torgnon Notes, pages 83-93.

Due: Wednesday, April 13, 2016

1. Van der Vaart (1998), problem 25, page 84.
(i) Verify the conditions of Wald's theorem for m_θ the log-likelihood function of the $N(\mu, \sigma^2)$ distribution if the parameter space for $\theta = (\mu, \sigma^2)$ is a compact subset of $\mathbb{R} \times \mathbb{R}^+$.
(ii) Extend m_θ by continuity to the compactification of $\mathbb{R} \times \mathbb{R}^+$. Show that the conditions of Wald's theorem fail at the points $(\mu, 0)$.
(iii) Replace m_θ by the log-likelihood function of a pair of two independent observations from the $N(\mu, \sigma^2)$ distribution, say (X, Y). Show that Wald's theorem now does apply, also with a compactified parameter set.

2. Consider the symmetric location family based on a fixed density f_0 symmetric at 0 defined as follows:
$$\mathcal{P} = \{p_\theta : \ p_\theta(x) = f_0(x - \theta), \quad \theta \in \mathbb{R}\}.$$
Particular cases of interest include:
(a) $f_0 = \phi$, the standard normal density;
(b) $f_0(x) = 2^{-1}\exp(-|x|)$, the Laplace density;
(c) $f_0(x) = \pi^{-1}(1 + x^2)^{-1}$, the standard Cauchy density;
(d) $f_0(x) = e^{-x}(1 + e^{-x})^{-2}$, the logistic density.
(i) For each of these densities compute and plot f_0, $\varphi \equiv -\log f_0$, and the corresponding score functions for location $-f_0'/f_0$. A density f is *log-concave* if $\varphi \equiv -\log f$ is a convex function. Which of the densities (a) - (d) are log-concave?
(ii) Suppose that $X_1, \ldots, X_n$ are i.i.d. P_0. In which of these cases is the MLE explicitly calculable? In which cases is it unique?
(iii) Suppose that $P_0 = P_{\theta_0} \in \mathcal{P}$. In which cases is the MLE consistent?
(iv) Suppose that $P_0 \notin \mathcal{P}$. For which case or cases is it possible that
$$\Theta_0 \equiv \{\theta_0 \in \Theta : \ P_0 m_{\theta_0} = \sup_{\theta \in \Theta} P_0 m_\theta\}$$
contains more than one point? [Hint: see Freedman and Diaconis (1982), *Ann. Statist.* **10**, 454-461, especially page 455.]

3. Suppose that $X_1, \ldots, X_n$ are i.i.d. P on $\mathbb{R}$ with $E|X_1| = \int |x| dP(x) < \infty$. Consider the following measure of dispersion:

$$D_n \equiv n^{-1} \sum_{i=1}^{n} |X_i - \overline{X}_n| = \mathbb{P}_n |X - \overline{X}_n|.$$

The corresponding measure in the population is $d \equiv P|X - PX| = \int |x - \mu(P)| dP(x)$ where $\mu(P) = P(X) = \int x dP(x)$. Use a Glivenko-Cantelli theorem to show that $D_n \to_p d$. [Hint: Let $\delta > 0$. Since $Pr(|\overline{X} - \mu| \le \delta) \to 1$ as $n \to \infty$, it becomes useful to consider the class of functions $\mathcal{F}_\delta \equiv \{f(x,t) \equiv |x - t| : |t - \mu| \le \delta\}$.]

4. Suppose that P_j and Q_j have densities p_j and q_j respectively with respect to μ_j for $j = 1, \ldots, n$, and suppose that X_j has distribution either P_j or Q_j for each $j \le n$. Let $\prod_{j=1}^n P_j$ and $\prod_{j=1}^n Q_j$ denote the product measures.
(i) Show that $\rho(\prod_{j=1}^n P_j, \prod_{j=1}^n Q_j) = \prod_{j=1}^n \rho(P_j, Q_j)$.
(ii) Use the result of (i) to find a formula for $H^2(\prod_{j=1}^n P_j, \prod_{j=1}^n Q_j)$ in terms of $H^2(P_j, Q_j)$, $j = 1, \ldots, n$.
(iii) Specialize the results of (i) and (ii) to the setting in which $P_1 = \cdots = P_n \equiv P$ and $Q_1 = \cdots = Q_n \equiv Q$ where we now write $\prod_{j=1}^n P_j = P^n$ and $\prod_{j=1}^n Q_j = Q^n$.
(iv) Suppose that $\rho(P, Q) < 1$. Using (iii) what can you say about $\lim_n \rho(P^n, Q^n)$ and $\lim_n H(P^n, Q^n)$?

5. (Bonus problem). (i) Suppose that $X_1, \ldots, X_n$ are i.i.d. P on $\mathbb{R}$ where $P(X_1 = x) = 0$ for every fixed $x \in \mathbb{R}$. Let $\mathbb{P}_n$ denote the empirical measure of the X_i's. What is the total variation distance $\sup_{B \in \mathcal{B}} |\mathbb{P}_n(B) - P(B)|$?
(ii) Now suppose that $X_1, \ldots, X_n$ are i.i.d. P on $\mathbb{R}$ where now $P(X_1 \in \mathbb{N}) = 1$ where $\mathbb{N} \equiv \{0, 1, 2, \ldots\}$, the non-negative integers. Let $\mathcal{C}$ denote the collection of all subsets of $\mathbb{N}$. Show that

$$d_{TV}(\mathbb{P}_n, P) = 2^{-1} \sum_{k=0}^{\infty} |\mathbb{P}_n(\{k\}) - P(X_1 = k)| \to_{a.s.} 0,$$

and hence that $\mathcal{C}$ is a $P-$Glivenko-Cantelli class for every P. [Hint: For $k \in \mathbb{N}$, let $\widehat{p}_n(k) \equiv \mathbb{P}_n(\{k\})$, the empirical mass function. Now use Scheffé's theorem.]
(iii) Is there a theorem giving conditions on P under which $\mathcal{C}$ is a $P-$ Donsker class of sets? How would you find it?

6. (Bonus problem). In the same context of problem 3 above, show that $\sqrt{n}(D_n - d) \to_d N(0, V^2)$ for some V^2 if $E(X_1^2) < \infty$.