Statistics 583, Problem Set 1

Wellner; 3/30/2016

Reading: Van der Vaart, Asymp Statist, pages 211-212;
Ferguson, ACLST, Chapter 2, pages 8 - 9 and problems 6 - 8, page 12.
(See also Tsybakov INS, section 2.4, page 83 ff.)

Due: Wednesday, April 6, 2016

1. For two probability measures on a measurable space $(\mathcal{X}, \mathcal{A})$, let $H(P,Q)$ and $V(P,Q)$ be the Hellinger and total variation distances between P and Q defined by

$$H^2(P,Q) = \frac{1}{2}\int \{\sqrt{p} - \sqrt{q}\}^2 d\mu \quad \text{and} \quad V(P,Q) = \frac{1}{2}\int |p-q| d\mu$$

respectively. Here $p = dP/d\mu$ and $q = dQ/d\mu$ where μ is any measure dominating both P and Q (e.g. $\mu = P + Q$). [Note that some authors do not include the constant factor 1/2 in the definitions of $H^2(P,Q)$ or $V(P,Q)$.]
(a) Show that $H^2(P,Q) = 1 - \int \sqrt{pq}\, d\mu \equiv 1 - \rho(P,Q)$; here $\rho(P,Q) = \int \sqrt{pq}\, d\mu \le 1$ is known as the *Hellinger affinity*.
(b) Show that $V(P,Q) = 1 - \int (p \wedge q) d\mu \equiv 1 - \eta(P,Q)$; here $p \wedge q \equiv \min\{p,q\}$ and $\eta(P,Q) = \int (p \wedge q) d\mu \le 1$ is called the *total variation affinity*.
(c) Use (a) and (b) and problem 1 to show that $(1/2)\rho(P,Q)^2 \le \eta(P,Q) \le \rho(P,Q)$.

2. With the same notation as in problem 1 show that

$$H^2(P,Q) \le V(P,Q) \le \sqrt{2} H(P,Q) \left(1 - 2^{-1} H^2(P,Q)\right)^{1/2}.$$

3. Using the notation in problem 1, show that $V(P,Q) = \sup_{A \in \mathcal{A}} |P(Q) - Q(A)|$. (This justifies the terminology "total variation distance", and is sometime known as Scheffé's theorem.)

4. Consider testing the simple hypothesis $H: X \sim P$ versus the simple alternative $K: X \sim Q$. Let ϕ be a test of H versus K, and let $a \equiv E_Q(1-\phi)$, $b \equiv E_P \phi$.
(a) Find a test ϕ which minimizes $a + Db$ where $D > 0$ is a fixed number. Relate the test you find to the Bayes rule for some prior $\Lambda = (\lambda, 1-\lambda)$ on $\{P, Q\}$.
(b) When $D = 1$, relate the minimized total $a + b$ to the Bayes risk and to the total variation distance $V(P,Q)$ between P and Q (or $\int (p \wedge q) d\mu$ for some dominating measure μ, e.g. $P + Q$).

5. (a) Suppose that $\{p_n\}$ is a sequence of densities with respect to a dominating measure μ that satisfies $p_n(x) \to p_0(x)$ for all $x \in \mathcal{X}$ where p_0 is the density corresponding to a probability measure P_0. Show that $V(P_n, P_0) \to 0$.
(b) Give two examples of such a sequence p_n, including one in which the dominating measure μ is Lebesgue measure on $\mathbb{R}$ and one in which the dominating measure is counting measure on the non-negative integers.

6. **Optional bonus problem 1:** Suppose that $X_1, \ldots, X_n$ are i.i.d. P on $(\mathcal{X}, \mathcal{A})$, let let $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ denote the empirical measure, and let $\mathbb{G}_n \equiv \sqrt{n}(\mathbb{P}_n - P)$ be the empirical process. Let $\mathcal{F}$ be a class of real valued measurable functions from $\mathcal{X}$ to $\mathbb{R}$ with $\mathcal{F} \subset L_2(P)$, and let $f_j \in \mathcal{F}$ for $j = 1, \ldots, k$. Thus $\mathbb{G}_n(f_j) = \sqrt{n}(\mathbb{P}_n(f_j) - P(f_j))$ for $j = 1, \ldots, k$. Use the multivariate CLT to show that

$$(\mathbb{G}_n(f_1), \ldots, \mathbb{G}_n(f_k)) \to (\mathbb{G}(f_1), \ldots, \mathbb{G}(f_k)) \sim N_k(0, \Sigma_k(f, P))$$

where $\Sigma_k(f, P)$ is the $k \times k$ matrix with i, jth entry given by

$$Cov(f_i(X_1), f_j(X_1)) = P(f_i f_j) - P(f_i)P(f_j)$$

and $\mathbb{G}$ denotes a mean zero Gaussian process with the same covariance structure as $\mathbb{G}_n$.