Statistics 583, Problem Set 1 Solutions

Wellner; 4/6/2016

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 $(1/2)\rho(P,Q)^2 \le \eta(P,Q) \le \rho(P,Q)$.

Solution: (a) This identity follows by direct calculation:

$$\begin{aligned} H^2(P,Q) &= \frac{1}{2}\int \{\sqrt{p}-\sqrt{q}\}^2 d\mu = \frac{1}{2}\int (p - 2\sqrt{pq} + q) d\mu \\ &= 1 - \int \sqrt{pq}\, d\mu \quad \text{since} \quad \int p\, d\mu = 1 = \int q\, d\mu. \end{aligned}$$

(b) Let $\delta \equiv p - q = \delta^+ - \delta^-$. Since

$$0 = 1 - 1 = \int (p-q) d\mu = \int \delta\, d\mu = \int \delta^+ d\mu - \int \delta^- d\mu$$

implies that $\int \delta^+ = \int \delta^-$, we find that

$$\begin{aligned} V(P,Q) &= \frac{1}{2}\int |\delta| d\mu = \frac{1}{2}\int (\delta^+ + \delta^-) d\mu = \int \delta^+ d\mu \\ &= \int_{p\ge q} p\, d\mu - \int_{p \ge q} q\, d\mu = \int_{p\ge q} p\, d\mu + \int_{p<q} p\, d\mu - \int_{p<q} p\, d\mu - \int_{p\ge q} q\, d\mu \\ &= 1 - \int (p \wedge q) d\mu. \end{aligned}$$

(c) The inequality on the right side follows easily by writing

$$\begin{aligned}\rho(P,Q) &= \int \sqrt{pq}d\mu = \int_{p\geq q} \sqrt{pq}d\mu + \int_{q>p} \sqrt{pq}d\mu \\ &\geq \int_{p\geq q} qd\mu + \int_{q>p} pd\mu = \int (p\wedge q)d\mu.\end{aligned}$$

To prove the inequality on the left side, first note that

$$2 = \int (p+q)d\mu = \int \{(p\wedge q) + (p\vee q)\}d\mu.$$

Thus

$$\begin{aligned}\rho(P,Q)^2 &\leq \left(\int \sqrt{pq}d\mu\right)^2 = \left(\int \sqrt{(p\wedge q)(p\vee q)}d\mu\right)^2 \\ &\leq \int (p\wedge q)d\mu \cdot \int (p\vee q)d\mu \quad \text{by the Cauchy-Schwarz inequality} \\ &= \eta(P,Q)(2-\eta(P,Q) \leq 2\eta(P,Q).\end{aligned}$$

This yields the first inequality in (c).

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

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

Solution: The first inequality on the left side follows easily from 1(c) above: since $\rho(P,Q) \geq \eta(P,Q)$, problems 1(b) and 1(a) yield

$$V(P,Q) = 1 - \int (p\wedge q)d\mu \geq 1 - \int \sqrt{pq}d\mu = H^2(P,Q).$$

On the other hand,

$$\begin{aligned}V(P,Q) &= \frac{1}{2}\int |p-q|d\mu = \int |(\sqrt{p}-\sqrt{q})(\sqrt{p}+\sqrt{q})|d\mu \\ &\leq \frac{1}{2}\left\{\int (\sqrt{p}-\sqrt{q})^2 d\mu \cdot \int (\sqrt{p}+\sqrt{q})^2 d\mu\right\}^{1/2} \\ &\qquad\qquad \text{by the Cauchy-Schwarz inequality} \\ &= \frac{1}{2}\sqrt{2}H(P,Q)\left(2+2\int \sqrt{pq}d\mu\right)^{1/2} \\ &= H(P,Q)\left(1+1-H^2(P,Q)\right)^{1/2} = \sqrt{2}H(P,Q)\left(1-2^{-1}H^2(P,Q)\right)^{1/2}.\end{aligned}$$

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.)

   Solution: Here we use the same notation as in the solution of problem 1(b): $\delta \equiv (p - q) = \delta^+ - \delta^-$. Now for any set $A \in \mathcal{A}$ we have

$$0 = \int_{\mathcal{X}} \delta d\mu = \int_A \delta d\mu + \int_{A^c} \delta d\mu.$$

   It follows that

$$\begin{aligned} 2|P(Q) - Q(A)| &= 2\left|\int_A (p - q) d\mu\right| = \left|\int_A \delta d\mu\right| + \left|\int_{A^c} \delta d\mu\right| \\ &\leq \int_A |\delta| d\mu + \int_{A^c} |\delta| d\mu = \int |\delta| d\mu \end{aligned}$$

   and equality holds if $A = \{p \geq q\}$. This yields the claimed equality.

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 function ϕ which minimizes $a + Db$ where D is a fixed positive 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. $P + Q$).

   Solution: (a) We can write

$$a + Db = E_Q(1 - \phi) + E_P \phi = 1 - \int \phi(q - Dp) d\mu$$

   and we want to choose ϕ to minimize it. This is clearly minimized by taking $\phi = \phi_0$ where

$$\phi_0(x) = \begin{cases} 1, & \text{if } q - Dp > 0, \\ \text{anything}, & \text{if } q - Dp = 0, \\ 0, & \text{if } q - Dp < 0. \end{cases}$$

   Now note that the Bayes risk of any test ϕ with respect to the prior $\Lambda = (\lambda, 1 - \lambda)$ on P and Q can be written as

$$\mathcal{R}(\phi, \Lambda) = (1 - \lambda) E_Q(1 - \phi) + \lambda E_P \phi = \frac{1}{1 + D} E_Q(1 - \phi) + \frac{D}{1 + D} E_P \phi = \frac{1}{1 + D}(a + Db)$$

with $\lambda = D/(1+D)$ and hence $1-\lambda = 1/(1+D)$. Hence the test ϕ_0 is the Bayes rule with respect to the prior $(\lambda, 1-\lambda) = (D/(1+D), 1/(1+D))$.
(b) By (a), when $D = 1$ the rule $\phi_0(x) \equiv 1\{q \ge p\}$ is Bayes with respect to the prior $(\lambda, 1-\lambda) = (1/2, 1/2)$. Twice the Bayes risk, or equivalently the sum of the ordinary risks, is (using the notation in 1(b))

$$2 \min_{\phi}(a+b)/2 = 1 - \int_{q \ge p} (q-p) d\mu = 1 - \int \delta^{-} d\mu = 1 - V(P,Q) = \eta(P,Q).$$

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.

Solution: Note that

$$V(P_n, P_0) = V(P_0, P_n) = \int \delta_n^{+} d\mu$$

where $0 \le \delta_n^{+} \equiv (p_0 - p_n) 1_{[p_0 \ge p_n]} \le p_0$, $\int p_0 d\mu = 1 < \infty$,
and $\delta_n^{+} \to_{a.e.} 0$. Thus $V(P_0, P_n) = \int \delta_n^{+} \to 0$ by the dominated convergence theorem.
(b) *Example 1.* Suppose that $X_1, \ldots, X_n$ are i.i.d. exp(1) and $M_n \equiv \max_{1 \le i \le n} X_i$. Then

$$\begin{aligned} P(M_n - \log n \le x) &= P(M_n \le x + \log n) = P(X_i \le x + \log n \text{ for all } n) \\ &= F(x + \log n)^n = \left(1 - \frac{e^{-x}}{n}\right)^n \to e^{-e^{-x}}. \end{aligned}$$

Thus $M_n - \log n \to_d M_\infty$ where M_0 has the Gumbel distribution $F_0(x) = \exp(-\exp(-x))$. Moreover, $M_n - \log$ has density function f_n given by

$$f_n(x) = e^{-x} \left(1 - \frac{e^{-x}}{n}\right)^{n-1} \to e^{-x} e^{-e^{-x}} \equiv f_0(x)$$

where f_0 is the Gumbel density function. It then follows from Problem 4 that $V(P_n, P_0) \to 0$, which is considerably stronger than convergence in distribution. [Question: at what rate does $V(P_n, P_0) \to 0$.]
Example 2. Suppose that $X_n \sim$ Bin(n, p_n) where $np_n \to \lambda$. Then X_n has density

$\{p_n(k) : 0 \le k \le n\}$ with respect to counting measure on $\{0, 1, 2, \ldots\}$ given by

$$\begin{aligned} p_n(k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{k! n^k} (np_n)^k \left(1 - \frac{np_n}{n}\right)^{n-k} \\ &\to \frac{1}{k!} \lambda^k \exp(-\lambda) \equiv P_0(X_0 = k) \end{aligned}$$

for each $k = 0, 1, 2, \ldots$; here P_0 is the Poisson(λ) distribution on $\{0, 1, 2, \ldots\}$. Thus by Problem 4 it follows that

$$V(P_n, P_0) = \frac{1}{2} \sum_{k=1}^{\infty} |P(X_n = k) - P_0(X_0 = k)| \to 0.$$

A generalization and improvement of this result due to Hodges and Le Cam (1960) is as follows: Let $X_i \sim \text{Bern}(p_i)$ for $i = 1, \ldots, n$ be independent and let $S_n = \sum_{i=1}^n X_i$. Let $\lambda \equiv \sum_{i=1}^n p_i = E(S_n)$, and suppose that $Z \sim \text{Poisson}(\lambda)$. Write P_n for the probability distribution of S_n and P_0 for the Poisson(λ) probability distribution. Then $V(P_n, P_0) \le \sum_{i=1}^n p_i^2$. [See Hodges and Le Cam (1960), *Ann. Math. Stat.* **31**, 737-740; Ferguson, ACILST, problem 3.5, page 18; and the book *Poisson Approximation* (1992) by Barbour, Holst, and Janson, for further developments.]

Example 3. Suppose that $X_1, X_2, \ldots$ are i.i.d. random variables with finite second moment $EX_1^2 < \infty$ and characteristic function $\phi(t) = E\exp(itX_1)$ satisfying $\int |\phi(t)|^\nu dt < \infty$ for some $\nu \ge 1$. Then $\sqrt{n}\overline{X}_n - \mu)/\sigma$ satisfies the central limit theorem in total variation:

$$\sup_{B \in \mathcal{B}} |P(\sqrt{n}(\overline{X}_n - \mu) \in B) - P(Z \in B)| \to 0 \quad \text{as} \quad n \to \infty$$

where $Z \sim N(0, 1)$ and hence $P(Z \in B) = \Phi(B) = \int_B \phi(z)dz$ for the standard normal density ϕ. See e.g. van der Vaart (1998), pages 22 - 23.