Statistics 583, Problem Set 2 Solutions

Wellner; 4/12/2000

1. Suppose that $\mathcal{P} \equiv \{p(x,\theta) : x \in R, \theta \in \Theta \subset R\}$ is a family of density functions with respect to Lebesgue measure such that the mixed second derivative

$$\frac{\partial^2}{\partial\theta\partial x} \log p(x,\theta)$$

exists for all x and θ. Show that if this second partial derivative is ≥ 0 for all x and θ, then $\mathcal{P}$ has monotone likelihood ratio in x.

Solution: If the second partial derivative satisfies

$$\frac{\partial^2}{\partial\theta\partial x} \log p(x,\theta) \geq 0 \tag{1}$$

for all θ and all x, then

$$\frac{\partial}{\partial x} \log \frac{p(x,\theta')}{p(x,\theta)} \geq 0 \tag{2}$$

for all $\theta' > \theta$ and all x, and conversely. To see this, note that

$$\begin{aligned}
\frac{\partial}{\partial\theta} \log p(x,\theta) &= \frac{\frac{\partial}{\partial\theta} p(x,\theta)}{p(x,\theta)} \\
&= \lim_{\theta' \downarrow \theta} \frac{p(x,\theta') - p(x,\theta)}{\theta' - \theta} \frac{1}{p(x,\theta)} \\
&= \lim_{\theta' \downarrow \theta} \left\{ \frac{p(x,\theta')/p(x,\theta) - 1}{\theta' - \theta} \right\} \\
&= \lim_{\theta' \downarrow \theta} \left\{ \frac{\log[p(x,\theta')/p(x,\theta)]}{\theta' - \theta} \right\}
\end{aligned}$$

since $\log(z) - (z-1) = o(z-1)$. But (2) holds if and only if $\log[p(x,\theta')/p(x,\theta)]$ is a nondecreasing function of x for each $\theta' > \theta$; i.e. $\mathcal{P}$ has monotone likelihood ratio in x.

2. Suppose that $p(x,\theta) = g(x-\theta)$, $x \in R$, $\theta \in R$, where g is a fixed density function on R. Show that this family has monotone likelihood ratio in x if and only if $-\log g$ is convex. [Hint: see Lehmann, TSH, section 9.2, page 509.]

Solution: This is solved in Lehmann, TSH, page 509. The following is almost verbatim from that page. The condition of monotone likelihood ratio in x means

$$\frac{g(x-\theta')}{g(x-\theta)} < \frac{g(x'-\theta')}{g(x'-\theta)} \qquad \text{for all } \; x < x', \quad \theta < \theta' \, .$$

By taking the logarithm of both sides and re-arranging, it is equivalent to

$$\log g(x'-\theta) + \log g(x-\theta') \leq \log g(x-\theta) + \log g(x'-\theta') \, .$$

Since we can write

$$x - \theta = t(x - \theta') + (1 - t)(x' - \theta)x' - \theta' = (1 - t)(x - \theta') + t(x' - \theta)$$

where $t = (x' - x)/(x' - x + \theta' - \theta)$, the previous display can be written, with $f(y) \equiv -\log g(y)$, as

$$f(t(x - \theta') + (1 - t)(x' - \theta)) + f((1 - t)(x - \theta') + t(x' - \theta)) \leq f(x - \theta') + f(x' - \theta).$$

A sufficient condition for (3) is that f is convex: f is convex if and only if

$$f(ta + (1 - t)b) \leq tf(a) + (1 - t)f(b)$$

for all $t \in [0, 1]$ and a, b in the domain of f. Note that this implies that

$$f(ta+(1-t)b)f((1-t)a+tb) \leq tf(a)+(1-t)f(b)+(1-t)f(a)+tf(b) = f(a)+f(b)\,,$$

and this is exactly the inequality in (3) with $a = x - \theta'$, $b = x' - \theta$. In the other direction, let $a < b$ be any fixed real numbers. Choose x, x', θ, θ' so that $x - \theta' = a$, $x' - \theta = b$, $x' - \theta' = x - \theta$. Then $x - \theta = [(x - \theta') + (x' - \theta)]/2 = (a + b)/2$, and (3) implies

$$2f((a + b)/2) \leq f(a) + f(b)\,, \text{or} \;\; f((a + b)/2) \leq \frac{1}{2}(f(a) + f(b))\,.$$

This implies that $f = -\log(g)$ is convex.

If g is twice differentiable, then this also follows from the result of problem 1: Note that the mixed second derivative as in problem 1 is given by :

$$\begin{aligned}
\frac{\partial^2}{\partial\theta\partial x}\log p(x, \theta) &= \frac{\partial^2}{\partial\theta\partial x}\log g(x - \theta) \\
&= \frac{\partial}{\partial\theta}\frac{g'(x - \theta)}{g(x - \theta)} \\
&= -\frac{\partial}{\partial y}\frac{g'(y)}{g(y)}\Big|_{y=x-\theta} \\
&= \frac{\partial^2}{\partial y^2}\left(-\log g(y)\right)\Big|_{y=x-\theta} \\
&\geq 0
\end{aligned}$$

if and only if $-\log g(y)$ is convex (when g is twice differentiable).

3. (i) Suppose that $X \sim P$ on a measurable space $(\mathcal{X}, \mathcal{A})$. Consider testing the simple hypothesis $H_0 : P = P_0$ versus the simple alternative $H_1 : P = P_1$ based on one observation X from either P_0 or P_1. Let ϕ be a test of H_0 versus H_1, and let $a \equiv R(0, \phi) \equiv E_0\phi(X)$, $b \equiv R(1, \phi) \equiv E_1(1 - \phi(X))$ be the risks associated with a $0 - 1$ loss function. Find a test ϕ which minimizes the weighted sum of the risks $c_0R(0, \phi) + c_1R(1, \phi) = c_0a + c_1b$.
(ii) Show that when $c_0 = c_1$ (both positive), the minimum weighted risk can be expressed in terms of $\int p_0 \wedge p_1 d\mu$ (where p_0, p_1 are densities of P_0, P_1 respectively

with respect to a common dominating measure μ), and hence to the total variation distance $d_{TV}(P_0, P_1)$ between P_0 and P_1.

Solution: (i) The weighted sum of risks is

$$\begin{aligned} c_0 R(0,\phi) + c_1 R(1,\phi) &= c_0 E_0 \phi(X) + c_1 E_1 (1 - \phi(X)) \\ &= c_1 + \int \phi(x) \{c_0 p_0(x) - c_1 p_1(x)\} \, d\mu(x) \end{aligned}$$

where p_0, p_1 are densities of P_0, P_1 with respect to a common dominating measure μ. This is minimized by any rule of the form

$$\phi(x) = \begin{cases} 1 & \text{if } c_0 p_0(x) < c_1 p_1(x) \\ 0 & \text{if } c_0 p_0(x) > c_1 p_1(x) \ . \end{cases}$$

The minimum weighted risk is

$$\begin{aligned} & c_1 + \int (c_0 p_0(x) - c_1 p_1(x)) 1_{[x: c_0 p_0(x) < c_1 p_1(x)]} d\mu(x) \\ &= c_1 + c_0 P_0(c_0 p_0(X) < c_1 p_1(X)) - c_1 P_1(c_0 p_0(X) < c_1 p_1(X)) \, . \end{aligned}$$

When $c_0 = c_1 > 0$, any rule of the form

$$\phi(x) = \begin{cases} 1 & \text{if } p_0(x) < p_1(x) \\ 0 & \text{if } p_0(x) > p_1(x) \end{cases}$$

minimizes the weighted risk, and the minimum weighted risk is

$$\begin{aligned} & c_0 \left\{ 1 + \int_{[x: p_1(x) > p_0(x)]} (p_0(x) - p_1(x)) d\mu(x) \right\} \\ &= c_0 \left\{ \int_{[p_1 > p_0]} p_0 d\mu + \int_{[p_1 \le p_0]} p_1 d\mu \right\} \\ &= c_0 \int p_0 \wedge p_1 d\mu \\ &= c_0 \{1 - d_{TV}(P_0, P_1)\} \end{aligned}$$

since

$$\begin{aligned} d_{TV}(P_0, P_1) &\equiv \sup_{B \in \mathcal{B}} |P_0(B) - P_1(B)| \\ &= \frac{1}{2} \int |p_0 - p_1| d\mu \\ &= 1 - \int p_0 \wedge p_1 d\mu \, . \qquad (3) \end{aligned}$$

Proof of (3): Let $\delta \equiv p_0 - p_1$. For any fixed set B, since $\int \delta d\mu = 0$,

$$\int_B \delta d\mu = - \int_{B^c} \delta d\mu \, ,$$

and hence for any fixed set B we have

$$\begin{aligned}
2|P_0(B) - P_1(B)| &= 2|\int_B \delta d\mu| \\
&= |\int_B \delta d\mu| + |\int_{B^c} \delta d\mu| \\
&\leq \int_B |\delta|d\mu| + \int_{B^c} |\delta|d\mu| \\
&= \int |\delta| d\mu\,,
\end{aligned}$$

and equality holds if $B = [x : \delta(x) \geq 0]$. Thus

$$2d_{TV}(P_0, P_1) \equiv 2 \sup_{B \in \mathcal{B}} |P_0(B) - P_1(B)| = \int |p_0 - p_1| d\mu\,.$$

Also from the above proof,

$$\begin{aligned}
2d_{TV}(P_0, P_1) &= \int |p_0 - p_1| d\mu = 2 \int_{[p_0 \geq p_1]} (p_0 - p_1) d\mu \\
&= 2\left\{\int_{p_0 \geq p_1} p_0 d\mu + \int_{p_0 < p_1} p_0 d\mu - \int_{p_0 < p_1} p_0 d\mu - \int_{p_0 \geq p_1} p_1 d\mu\right\} \\
&= 2\left\{1 - \int p_0 \wedge p_1 d\mu\right\}\,.
\end{aligned}$$

4. A random variable X takes on the values $1, 2, 3, 4$ with probability distribution P_0 or P_1 with mass functions p_0 and p_1 given in the following table:

x	1	2	3	4
$p_0(x)$	.1	.2	.3	.4
$p_1(x)$	.4	.2	.2	.2

(a) Find a most powerful test of size $\alpha = .2$ for testing P_0 versus P_1 and determine its power.
(b) Find a test ϕ which minimizes the sum of risks $a + b$ where a, b are as defined in the previous problem.
Compute the minimum risk and the total variation distance between P_0 and P_1.

Solution: (a) If we add a row with p_1/p_0 to the above table, we find:

x	1	2	3	4
$p_0(x)$	.1	.2	.3	.4
$p_1(x)$	.4	.2	.2	.2
$p_1(x)/p_0(x)$	4	1	2/3	1/2

Thus the most powerful test of P_0 versus P_1 at level $\alpha = .2$ is the test $\phi(X) = 1_{[X=1]} + (.5)1_{[X=2]}$. For this test we have

$$E_0\phi(X) = P_0(X=1) + .5P_0(X=2) = .1 + (.5)(.2) = .2$$

and the power is

$$E_1\phi(X) = P_1(X=1) + .5P_1(X=2) = .4 + (.5)(.2) = .5\,.$$

(b) As we showed in the previous problem, any test ϕ of the form

$$\phi(x) = 1_{[x:\, p_1(x) > p_0(x)]} + \gamma 1_{[x:\, p_1(x) = p_0(x)]}$$

minimizes the sum of the risks. The minimum total risk is, from the preceding problem,

$$\int p_0 \wedge p_1 d\mu = .1 + .2 + .2 + .2 = .7\,.$$

Note that the most powerful test (the NP test) of part (a) is of this form and has $E_0\phi(X) + E_1(1-\phi(X)) = .2 + .5 = .7$. The total variation distance between P_0 and P_1 is $d_{TV}(P_0, P_1) = 1 - .7 = .3 = (1/2)(.3 + 0 + .1 + .2)$. The square of the Hellinger distance is $1 - \int \sqrt{p_0 p_1} d\mu = 1 - 0.927792 = 0.07221$, so $H(P_0, P_1) \doteq 0.26871$. Note that we always have

$$H^2(P_0, P_1) \le d_{TV}(P_0, P_1) \le \sqrt{2} H(P_0, P_1)\,. \tag{4}$$

Here is the proof of (4):

$$\begin{aligned}
H^2(P_0, P_1) &= 1 - \int \sqrt{p_0 p_1} d\mu \le 1 - \int p_0 \wedge p_1 d\mu \\
&\qquad \text{since } \int p_0 \wedge p_1 d\mu \le \int \sqrt{p_0 p_1} d\mu \\
&= d_{TV}(P_0, P_1) = \frac{1}{2} \int |\sqrt{p_0} - \sqrt{p_1}||\sqrt{p_0} + \sqrt{p_1}| d\mu \\
&\le H(P_0, P_1) \left\{ \frac{1}{2} \int |\sqrt{p_0} + \sqrt{p_1}|^2 d\mu \right\}^{1/2} \\
&= H(P_0, P_1) \left\{ \frac{1}{2} \left(2 + 2 \int \sqrt{p_0 p_1} d\mu \right) \right\}^{1/2} \\
&= H(P_0, P_1) \left\{ 2 - H^2(P_0, P_1) \right\}^{1/2} \\
&= \sqrt{2} H(P_0, P_1)\,.
\end{aligned}$$