Statistics 582, Problem Set 9 Solutions

Wellner; 9 March 2018

1. For observations $\underline{X} = (X_1, \ldots, X_n)$, let $X_{(1)} \le \ldots \le X_{(n)}$ denote the *order statistics* of the X_i's ($X_{(i)} \equiv \mathbb{F}_n^{-1}(i/n)$, $i = 1, \ldots, n$) and let $\underline{R} = (R_1, \ldots, R_n)$ denote the *ranks* defined by $X_i = X_{(R_i)}$, $i = 1, \ldots, n$ (if $X_i = X_j$ for some $i < j$, define the ranks by $R_i < R_j$ and $X_i = X_{(R_i)}$).
(a) Suppose that $X_1, \ldots, X_n$ are i.i.d. $F \in \mathcal{F}_{ac}$ (the absolutely continuous df's F on R) with density f. Show that the order statistics $\underline{X}_{(\cdot)} \equiv (X_{(1)}, \ldots, X_{(n)})$ are independent of the ranks $\underline{R}$ and that the order statistics have joint density $\bar{p}$ given by
$$\bar{p}(\underline{x}_{(\cdot)}) = n! \prod_{i=1}^{n} f(x_{(i)}), \qquad -\infty < x_{(1)} < \ldots < x_{(n)} < \infty$$
while
$$P(\underline{R} = \underline{r}) = \frac{1}{n!}, \qquad \underline{r} \in \Pi \equiv \{ \text{ all permutations of } \{1, \ldots, n\}\} .$$
(b) Show that if the density f of the X_i's is log-concave, then the joint density $\bar{p}$ of the order statistics $\underline{X}_{(\cdot)}$ is log-concave; i.e. show that if $f((x+y)/2)^2 \ge f(x)f(y)$ for all $x, y \in \mathbb{R}$, then $\bar{p}((\underline{x}+\underline{y})/2)^2 \ge \bar{p}(\underline{x})\bar{p}(\underline{y})$ for all $\underline{x}, \underline{y} \in \mathcal{O}_n \equiv \{\underline{x} \in \mathbb{R}^n : x_1 \le x_2 \le \cdots \le x_n\}$.
(c) Show that (a) continues to hold for any joint distribution p of the $\underline{X}$ which is symmetric with respect to permutation of its coordinates: $p(\pi\underline{x}) = p(\underline{x})$ for all $\underline{x}$ and $\pi \in \Pi$ where $\pi x \equiv (x_{\pi(1)}, \ldots, x_{\pi(n)})$.
(d) If the joint density p of $\underline{X}$ is general (not permutation symmetric), show that the joint density $\bar{p}$ of the order statistics is given by
$$\bar{p}(\underline{x}_{(\cdot)}) = \sum_{\pi \in \Pi} p(\pi \underline{x}_{(\cdot)}) ,$$
and
$$P(\underline{R} = \underline{r} | \underline{X}_{(\cdot)} = \underline{x}_{(\cdot)}) = \frac{p(r\underline{x}_{(\cdot)})}{\bar{p}(\underline{x}_{(\cdot)})} .$$
Hint: Do (d) first by computing $P(\underline{X}_{(\cdot)} \in A)$ and $P(\underline{R} = \underline{r}, \underline{X}_{(\cdot)} \in A)$ for a Borel set $A \subset \{\underline{x} \in \mathbb{R}^n : x_1 < x_2 < \cdots < x_n\}$ and a fixed permutation $\underline{r}$ of $\{1, \ldots, n\}$.

Solution: I will prove (d) first; then (a) and (c) follow as corollaries:
(d) Suppose that $\underline{X}$ has joint density p. Then for any set Borel set $A \subset \{\underline{x} \in \mathbb{R}^n : \ x_1 < x_2 < \cdots < x_n\}$

$$\begin{aligned} P(\underline{X}_{(\cdot)} \in A) &= \int_{[\underline{x}_{(\cdot)} \in A]} p(x_1, \ldots, x_n) \ dx_1 \ldots dx_n \\ &= \sum_{r \in \Pi} \int_{[R(\underline{x})=r, \ \underline{x}_{(\cdot)} \in A]} p(x_1, \ldots, x_n) \ dx_1 \ldots dx_n \\ &= \sum_{r \in \Pi} \int_A p(x_{(r_1)}, \ldots, x_{(r_n)}) \ dx_{(1)} \ldots dx_{(n)} \\ &= \int_A \overline{p}(x_{(1)}, \ldots, x_{(n)}) \ dx_{(1)} \ldots dx_{(n)} \end{aligned}$$

where we have used the fact that the correspondence between $(x_1, ..., x_n)$ and $(x_{(1)}, \ldots, x_{(n)})$ is one-to-one and linear with Jacobian $= 1$ on each subset $[R = r]$, $r \in \Pi$. This proves that

$$\overline{p}(\underline{x}_{(\cdot)}) = \sum_{\pi \in \Pi} p(\pi \underline{x}_{(\cdot)}) \ .$$

Similarly,

$$\begin{aligned} P(R = r, \ \underline{X}_{(\cdot)} \in A) &= \int_{[R=r, \ \underline{x}_{(\cdot)} \in A]} p(x_1, \ldots, x_n) \ dx_1 \cdots dx_n \\ &= \int_A p(x_{(r_1)}, \ldots, x_{(r_n)}) \ dx_{(1)} \ldots dx_{(n)} \\ &= \int_A \frac{p(x_{(r_1)}, \ldots, x_{(r_n)})}{\overline{p}(x_{(1)}, \ldots, x_{(n)})} \overline{p}(x_{(1)}, \ldots, x_{(n)}) \ dx_{(1)} \ldots dx_{(n)} \end{aligned}$$

since $\overline{p}(x_{(1)}, \ldots, x_{(n)}) = 0$ implies $p(x_{r(1)}, \ldots, x_{r(n)}) = 0$ for each $r \in \Pi$. This implies that

$$P(\underline{R} = \underline{r} | \underline{X}_{(\cdot)} = \underline{x}_{(\cdot)}) = \frac{p(r\underline{x}_{(\cdot)})}{\overline{p}(\underline{x}_{(\cdot)})} \ .$$

(c) When $p(\underline{x}) = p(\pi \underline{x})$ for all $\pi \in \Pi$, then

$$\overline{p}(\underline{x}_{(\cdot)}) = n! p(\underline{x}_{(\cdot)}),$$

and

$$P(\underline{R} = \underline{r} | \underline{X}_{(\cdot)} = \underline{x}_{(\cdot)}) = \frac{p(r\underline{x}_{(\cdot)})}{\overline{p}(\underline{x}_{(\cdot)})} = \frac{p(r\underline{x}_{(\cdot)})}{n! p(\underline{x}_{(\cdot)})} = \frac{1}{n!}.$$

Hence R is independent of $\underline{X}_{(\cdot)}$, and $P(R = r) = 1/n!$ for each $r \in \Pi$.
(a) This follows easily from (c) since, in this case, for any permutation π

$$p(\pi \underline{x}) = \prod_{i=1}^{n} f(x_{\pi(i)}) = \prod_{i=1}^{n} f(x_i) = p(\underline{x}).$$

(b) If the marginal density f of the X_i's is log-concave, then

$$\begin{aligned}
\overline{p}((\underline{x}_{(\cdot)} + \underline{y}_{(\cdot)})/2)^2 &= n!^2 \prod_{i=1}^{n} f((\underline{x}_{(i)} + \underline{y}_{(i)})/2)^2 \\
&\geq n!^2 \prod_{i=1}^{n} f(\underline{x}_{(i)}) f(\underline{y}_{(i)}) \\
&= \left(n! \prod_{i=1}^{n} f(\underline{x}_{(i)}) \right) \left(n! \prod_{i=1}^{n} f(\underline{y}_{(i)}) \right) \\
&= \overline{p}(\underline{x}_{(\cdot)}) \cdot \overline{p}(\underline{y}_{(\cdot)}).
\end{aligned}$$

Thus the joint density of the order statistics is also log-concave.

2. In a comparison of the effect on growth of two diets B and C, a number of growing rats were placed on these two diets, and the following growth figures were observed after 7 weeks:

$$B : 156, 183, 120, 113, 138, 145, 142$$
$$C : 109, 107, 119, 162, 121, 123, 76, 111, 130, 115.$$

(This data is from Lehmann (1975), *Nonparametrics: Statistical Methods based on Ranks*, problem 20, page 108. Lehmann references the original article in a footnote on the same page.)
(a) Let the growth figures from the B diet be denoted by Y_j, $j = 1, \ldots, 7$, and let the growth figures from the C diet be denoted by X_i, $i = 1, \ldots, 10$. Assuming that the X_i's are i.i.d. with mean μ and that the Y_j's are i.i.d. with mean ν and that the X's and Y's have a common variance, use a two-sample t-test to test $H : \mu \geq \nu$ versus $K : \mu < \nu$. What is the p-value for your test?
(b) Now use a two-sample permutation test (or approximation thereof via sampling) to test the same hypotheses as in (a): what is the p-value (or approximate p-value) for your permutation test? Is this in (approximate) agreement with the limit theorem we proved in class?

Solution: (1) I compute, with $m = 10$, $n = 7$,

$$\tau_{m,n} = \frac{\sqrt{\frac{mn}{N}}(\overline{Y}_n - \overline{X}_m)}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}}} = 2.302,$$

and the $p-$value is $P(t_{10,7} \geq 2.302) = .018$.
(b) Here the number of elements in the permutation distribution of $\tau_{m,n} = \tau_{10,7}$ is $\binom{17}{7} = 19448$. By sampling 3000 permutations (or samples without replacement from the pooled data), I find an estimated $p-$value of 0.0183, which is in close agreement with the $p-$value of 0.018 for the two-sample (normal theory) $t-$test we obtain in (a). Figure 1 gives the histogram of values of $\overline{Y}$ for the 3000 sampled labelling of the pooled data as X's and Y's. By enumerating the entire list of all 19448 values of $\tau_{m,n}$ obtained from all possible relabelling of the data as X's and Y's I get a $p-$value of 0.0186.

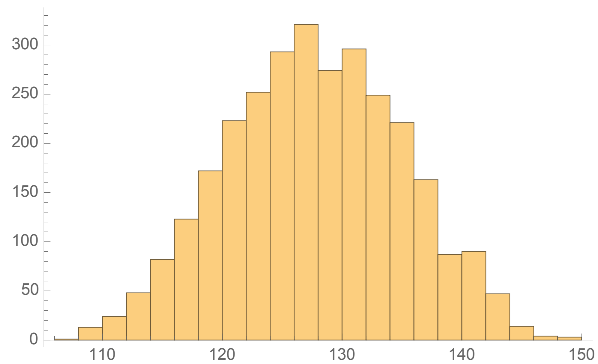

Figure 1: Sampled permutation distribution, two-sample $t-$statistic $\tau_{m,n}$, 3000 "chooses" or "re-labellings"

3. (Problem 3.7, Lehmann and Romano, TSH, page 94.) Suppose that the distribution of X is given by

x	0	1	2	3
$p_\theta(x)$	θ	2θ	$.9 - 2\theta$	$.1 - \theta$

where $0 < \theta < .1$. For testing $H : \theta = .05$ against $\theta > .05$ at level $\alpha = .05$, determine which of the following tests (if any) is UMP:
(i) $\phi(0) = 1$, $\phi(1) = \phi(2) = \phi(3) = 0$;
(ii) $\phi(1) = .5$, $\phi(0) = \phi(2) = \phi(3) = 0$;
(ii) $\phi(3) = 1$, $\phi(0) = \phi(1) = \phi(2) = 0$.

Solution: The likelihood ratios $P_{\theta'}(X = x)/P_\theta(X = x)$

x	0	1	2	3
$P_\theta(X = x)$	$\theta/2$	θ	$.9 - \theta$	$.1 - \theta/2$
$\frac{P_{\theta'}(X=x)}{P_\theta(X=x)}$	$\frac{\theta'}{\theta}$	$\frac{\theta'}{\theta}$	$\frac{9-10\theta'}{9-10\theta}$	$\frac{1-5\theta}{1-5\theta'}$

It is easy to check that

$$\frac{\theta'}{\theta} > \frac{9-10\theta'}{9-10\theta} > \frac{1-5\theta'}{1-5\theta}$$

Hence this family has monotone decreasing likelihood ratio in x (though not strictly), and strictly decreasing likelihood ratio in

$$\begin{aligned} T(x) &= 1\{x=0\} + 1\{x=1\} + 2\cdot 1\{x=2\} + 3\cdot 1\{x=3\} \\ &= x1\{x>0\} + 1\{x=0\}. \end{aligned}$$

It follows from the Karlin - Rubin theorem that a UMP test of $H : \theta \le \theta_0 = .05$ (of its level) is given by

$$\phi(X) = 1_{[T(X)<k]} + \gamma(X)1_{[T(X)=k]}. \tag{1}$$

(i) Note that the test $\phi_1(X) = 1\{X=0\}$ is of the form (1) with $k=1$ and $\gamma(X) = 1\{X=0\}$ and it has level $\alpha = .05$; hence it is a UMP test of H versus K. The power of ϕ_1 is given by $\beta_1(\theta) \equiv E_\theta\phi_1(X) = \theta/2$.
(ii) The test $\phi_2(X) = (.5)1\{X=1\}$ is also of the form (1) with $k=1$ and $\gamma(X) = .5\cdot 1\{X=1\}$ and it has level $\alpha = .05$. Hence it is also a UMP test of H versus K. The power of ϕ_2 is given by $\beta_2(\theta) \equiv E_\theta\phi_2(X) = \theta/2$.
(iii) The test $\phi_3(X) = 1\{X=3\}$ is clearly not of the form (1). It has power function $\beta_3(\theta) = E_\theta\phi_3(X) = .1-\theta/2$, so $\beta_3(.05) = .05$, but $\beta_3(\theta) > .05$ for $\theta < .10$ while $\beta_3(\theta) < .05$ for $\theta > \theta_0 = .10$. In fact, this is a UMP test of $\tilde{H} : \theta \ge \theta_0$ versus $\tilde{K} : \theta < \theta_0$.

4. Read TPE (Lehmann and Casella) pages 160 - 162 concerning the notion of *equivariance* of an estimator $\delta = \delta(X)$ under a group of transformations G. Relate this to *invariance* of a (test) function ϕ under a group of transformations G. Illustrate equivariance with two examples.

   Solution: Let $X \sim P_\theta$ for $\theta \in \Theta$, and suppose that we want to estimate some function $h(\theta) \in \mathcal{H}$. For a group of transformations G on the sample space $\mathcal{X}$, we typically induce a group $\overline{G}$ on the parameter space Θ via the correspondence $gX \sim P_{\overline{g}\theta}$. Suppose, moreover, that $\overline{G}$ induces a group G^* on $\mathcal{H}$ via $h(\overline{g}\theta) = g^*h(\theta)$. If $\delta : \mathcal{X} \to \mathcal{H}$ yields an estimator $\delta(X)$ of $h(\theta)$, then we expect to use $g^* \circ \delta(X)$ or $\delta(gX)$ to estimate $h(\overline{g}\theta)$. Thus equivariance is just the requirement that $g^* \circ \delta(X) = \delta(gX)$.

   It is fairly straightforward to relate this to the testing situation, in which case $h(\theta) = 1_{\Theta_K}(\theta)$ and the induced group G^* reduces to the trivial group $G = \{e\}$.

   Here are two examples of equivariance in estimation:

Example 1. Location Suppose that $X = (X_1, \ldots, X_n)$ where the X_i's are i.i.d. $N(\theta, \sigma^2)$ where $\sigma^2 > 0$ is known. We want to estimate $h(\theta) = \theta$. If $G = \{g_c : g_c(x) = x + c\underline{1},\ c \in R\}$, then the induced group on the parameter space is $\overline{G} = \{\overline{g}_c : g_c(\theta) = \theta + c,\ c \in R\}$, and this is also the group G^* in the discussion above. Note that for the usual estimator $\delta(X) = \overline{X} = n^{-1} \sum_1^n X_i$ we have

$$\delta(g_c X) = \overline{X} + c = g_c^*(\overline{X}),$$

i.e. $\delta = \overline{X}$ is (location) equivariant.

Example 2. Scale Suppose that $X = (X_1, \ldots, X_n)$ where the X_i's are i.i.d. $N(0, \theta^2)$ where $\theta > 0$ is unknown. We want to estimate $h(\theta) = \theta^2$. If $G = \{g_c : \ g_c(x) = cx,\ x \in R^n,\ c > 0\}$, then the induced group on the parameter space $\Theta = \{\theta : \ \theta > 0\}$ is $\overline{G} = \{\overline{g}_c : \ \overline{g}_c(\theta) = c\theta\}$, and in this case the group $G^* = \{g_c^* : \ g_c^*(h) = c^2 h : \ c > 0\}$ since $h(\overline{g}\theta) = (c\theta)^2 = c^2 h(\theta)$. For the natural (consistent) estimator $\delta(X) = S_X^2 = n^{-1} \sum_1^n (X_i - \overline{X})^2$ of $h(\theta) = \theta^2$, we have

$$\delta(g_c(X)) = c^2 S_X^2 = g_c^*(\delta(X)),$$

i.e. $\delta = S_X^2$ is (scale)-equivariant.