

Statistics 582, Problem Set 9

Wellner; 28 February 2018

Reading: Chapter 6, section 6.3 (pages 25-37 and 6.4 (pages 47 - 51).

Due: Wednesday, March 7, 2018.

Reminder: Final Exam, Monday, March 12, 8:30 - 10:30 (MGH 271)

1. For observations $\underline{X} = (X_1, \dots, X_n)$, let $X_{(1)} \leq \dots \leq X_{(n)}$ denote the *order statistics* of the X_i 's ($X_{(i)} \equiv \mathbb{F}_n^{-1}(i/n)$, $i = 1, \dots, n$) and let $\underline{R} = (R_1, \dots, R_n)$ denote the *ranks* defined by $X_i = X_{(R_i)}$, $i = 1, \dots, n$ (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, \dots, X_n are i.i.d. $F \in \mathcal{F}_{ac}$ (the absolutely continuous df's F on \mathbb{R}) with density f . Show that the order statistics $\underline{X}_{(\cdot)} \equiv (X_{(1)}, \dots, X_{(n)})$ are independent of the ranks \underline{R} and that the order statistics have joint density \bar{p} given by

$$\bar{p}(\underline{x}_{(\cdot)}) = n! \prod_{i=1}^n f(x_{(i)}), \quad -\infty < x_{(1)} < \dots < x_{(n)} < \infty$$

while

$$P(\underline{R} = \underline{r}) = \frac{1}{n!}, \quad \underline{r} \in \Pi \equiv \{ \text{all permutations of } \{1, \dots, n\} \} .$$

(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 $f((x+y)/2)^2 \geq f(x)f(y)$ for all $x, y \in \mathbb{R}$, then $\bar{p}((\underline{x} + \underline{y})/2)^2 \geq \bar{p}(\underline{x})\bar{p}(\underline{y})$ for all $\underline{x}, \underline{y} \in \mathcal{O}_n \equiv \{ \underline{x} \in \mathbb{R}^n : x_1 \leq x_2 \leq \dots \leq x_n \}$.

(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 $\pi \underline{x} \equiv (x_{\pi(1)}, \dots, x_{\pi(n)})$.

(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

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

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 $A \subset \{ \underline{x} \in \mathbb{R}^n : x_1 < x_2 < \dots < x_n \}$ and a fixed permutation \underline{r} of $\{1, \dots, n\}$.

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 , $j = 1, \dots, 7$, and let the growth figures from the C diet be denoted by X_i , $i = 1, \dots, 10$. 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?

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) $\phi(1) = .5, \phi(0) = \phi(2) = \phi(3) = 0$;
- (iii) $\phi(3) = 1, \phi(0) = \phi(1) = \phi(2) = 0$.

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.

5. **Optional bonus problem 1:** Let X and Y be independent exponential random variables with parameters λ and μ respectively: thus $P(X > x) = \exp(-\lambda x)$ and $P(Y > y) = \exp(-\mu y)$ for $x, y \geq 0$. Let $\theta \equiv \lambda/\mu$.
- Show that the problem of testing $H_0 : \theta \leq 1$ versus $H_1 : \theta > 1$ is invariant under the group G of transformations $g_c(x, y) = (cx, cy)$, $c > 0$.
 - Find a maximal invariant.
 - Find a UMP invariant test of size α .
 - Show that the problem of testing $H'_0 : \theta = 1$ versus $H'_1 : \theta \neq 1$ is invariant *in addition* under the transformation $g(x, y) = (y, x)$.
 - Find a UMP invariant test of size α of H'_0 versus H'_1 .
 - Find UMP invariant tests of the hypotheses in (a) and (d) when X_1, \dots, X_m are i.i.d. $\text{Exponential}(\lambda)$ and Y_1, \dots, Y_n are i.i.d. $\text{Exponential}(\mu)$.
6. **Optional bonus problem 2:** (From Wasserman, *All of Statistics*, page 171.) In 1961, 10 essays appeared in the *New Orleans Daily Crescent*. They were signed “Quintus Curtius Snodgrass” and some people suspected they were actually written by Mark Twain. To investigate this, consider the proportion of three letter words founds in an author’s work. From eight Twain essays we have

.225, .262, .217, .240, .230, .229, .235, .217

From 10 Snodgrass essays we have:

.209, .205, .196, .210, .202, .207, .224, .223, .220, .201

- Perform a Wald test for equality of the means. Give a p -value and a 95% confidence interval for the difference of means. What conclusion do you reach?
- Now use a permutation test to test the equality of means. What is your conclusion?