# Statistics 582, Problem Set 9

Wellner; 3/4/2015

**Reading:** Chapter 6, section 6.3 (pages 25-37 and 6.4 (pages 47 - 51).
**Due:** Wednesday, March 11, 2015.
**Reminder:** Final Exam, Monday, March 16, 8:30 - 10:30 (MGH 271)

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 $\{1, \ldots, n\}$.

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 $\mu$ and that the $Y_j$'s are i.i.d. with mean $\nu$ 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. 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 $\phi$ under a group of transformations $G$. Illustrate equivariance with two examples.

4. Let $X$ and $Y$ be independent exponential random variables with parameters $\lambda$ and $\mu$ respectively: thus $P(X > x) = \exp(-\lambda x)$ and $P(Y > y) = \exp(-\mu y)$ for $x, y \geq 0$. Let $\theta \equiv \lambda/\mu$.
(a) Show that the problem of testing $H_0 : \theta \leq 1$ versus $H_1 : \theta > 1$ is invariant under the group $G$ of transformations $g_c(x, y) = (cx, cy)$, $c > 0$.
(b) Find a maximal invariant.
(c) Find a UMP invariant test of size $\alpha$.
(d) Show that the problem of testing $H_0' : \theta = 1$ versus $H_1' : \theta \neq 1$ is invariant *in addition* under the transformation $g(x, y) = (y, x)$.
(e) Find a UMP invariant test of size $\alpha$ of $H_0'$ versus $H_1'$.
(f) Find UMP invariant tests of the hypotheses in (a) and (d) when $X_1, \ldots, X_m$ are i.i.d. Exponential($\lambda$) and $Y_1, \ldots, Y_n$ are i.i.d. Exponential($\mu$).

5. **Optional bonus problem 1:** (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

(a) 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?
(b) Now use a permutation test to test the equality of means. What is your conclusion?

6. **Optional bonus problem 2:** (From the material on consistency of Neyman Pearson tests, section 6.1; and Donoho and Jin, *Ann. Statist.* **32** (2004), 962-994)
Let $p_\mu(x) \equiv \phi(x-\mu)$ denote the density of $X \sim N(\mu, 1)$, and let $P_\mu$ denote the corresponding measure on $\mathbb{R}$.
(a) Consider testing $H: \ P = P_0$ versus $K: \ Q = P_\mu$ with $\mu > 0$ fixed. Compute $\rho(P, Q) = \rho(P_0, P_\mu) = \int \sqrt{p_0(x) p_\mu(x)} dx$ explicitly as a function of $\mu$.
(b) Compare the power functions of the following two tests of $H$ versus $K$ when $\mu = 1$:
(i) The Neyman - Pearson test with $\alpha = .05$; (ii) The Neyman - Pearson type test with $k = k_n = 1$ (in the notation of Theorem 6.1.4, page 8, Chapter 6).
(c) Now suppose $\mu = \mu_n \equiv t/\sqrt{n}$ with $t > 0$ and consider testing $H: P = P_0$ versus $K_n: \ P = P_{\mu_n}$.
(i) Find the limit of $\rho(P_0, P_{\mu_n})^n$ as a function of $t$.
(ii) Compare the limiting power function of the two tests in (b).
(d) Now suppose that $Q = Q_n$ is given by the mixture

$$q(x) = q(x; \mu, \epsilon) = (1-\epsilon) p_0(x) + \epsilon p_\mu(x)$$

where $\epsilon = \epsilon_n = n^{-\beta}$ with $1/2 < \beta < 1$ and $\mu = \mu_n = \sqrt{2r \log n}$ with $r > \rho^*(\beta)$ and where the function

$$\rho^*(\beta) = \begin{cases} \beta - 1/2, & 1/2 \le \beta < 3/4, \\ (1-\sqrt{1-\beta})^2, & 3/4 \le \beta < 1. \end{cases}$$

This is a model for "sparse normal means". Show that $k_n$ can be chosen so that $E_{P_0^n} \phi_n(\underline{X}) \to 0$ and $E_{Q_n}^n (1 - \phi_n(\underline{X}_n)) \to 0$.