Statistics 582, Problem Set 9

Wellner; 3/4/2009

Reading: Chapter 6, sections 6.1 and 6.2 (through page 19); section 6.5 (pages 42-46).
Due: Wednesday, March 11, 2009.

1. Let X and Y be independent random variables with geometric distributions

$$p_{X,Y}(x,y|\theta_1,\theta_2) = (1-\theta_1)(1-\theta_2)\theta_1^x\theta_2^y, \qquad x,y \in \{0,1,\ldots\}.$$

where $0 < \theta_j < 1$, $j = 1,2$. Find a UMP unbiased test of size $\alpha = .20$ for testing
(a) $H_0 : \theta_1 \le \theta_2$ versus $H_1 : \theta_1 > \theta_2$.
(b) $H_0 : \theta_1 = \theta_2$ versus $H_1 : \ \theta_1 \ne \theta_2$.
(c) For what functions $\varphi(\theta_1,\theta_2)$ do our methods guarantee existence of a UMP unbiased test of $H_0 : \varphi(\theta_1,\theta_2) = 0$ versus $H_1 : \varphi(\theta_1,\theta_2) \ne 0$?

2. (From Wasserman, *All of Statistics*, page 171). In 1861, 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, we will 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 (which avoids the use of large - sample methods). What is your conclusion?

3. 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 (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 $\pi x \equiv (x_{\pi(1)}, \ldots, x_{\pi(n)})$.
(c) If the joint distribution 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

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

4. **Optional bonus problem:** (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_μ 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 μ.
(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

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

Show that k_n can be chosen so that $E_{P_0^n} \phi_n(\underline{X}) \to 0$ and $E_{Q_n}^n (1 - \phi_n(\underline{X}_n)) \to 0$.