Statistics 583, Problem Set 3

Wellner; 4/11/2007

Reading: Chapter 7, sections 7.1- 7.4 (to be handed out on Friday, 4/13); Wasserman, Chapters 1-2, pages 1-24.
Due: Wednesday, April 18, 2007

1. (a) What is the locally best rank test of $F = G$ against $G = (e^{\theta F} - 1)/(e^{\theta} - 1)$, $\theta > 0$?
(b) What is the locally best rank test of $F = G$ against $G = F/(e^{\theta}(1-F)+F)$?
(c) What can you say about the power of these tests (other than the fact that they are locally most powerful)?

2. Suppose that an urn contains N balls with the numbers $z_i = -\log(1-i/(N+1))$, $i = 1, \ldots, N$ and we sample $n < N$ balls from this urn. Let $\overline{Y}_n = n^{-1}\sum_1^n Y_i$ denote the sample mean of the sampled valls.
(a) Calculate the mean $\mu_N = E(\overline{Y}_n)$ and variance $\sigma_N^2 = Var(\overline{Y}_n)$ of $\overline{Y}_n$.
Find the limits of $\overline{z}_N$ and σ_z^2 as $N \to \infty$.
(b) Use the Wald-Wolfowitz-Noether-Hajek finite-sampling CLT to prove that $(\overline{Y}_n - \mu_N)/\sigma_N \to_d N(0, 1)$.
(c) What classical two-sample rank statistic is $\overline{Y}_n$ equivalent to under the null hypothesis (of all $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ equal in distribution with a common continuous distribution function F)?

3. Suppose that $X_1, \ldots, X_n$ are independent Exponential(1) random variables. Let $Y_i \equiv X_{(i)}$, for $i = 1, \ldots, n$, denote the *order statistics* corresponding to $X_1, \ldots, X_n$.
(a) Show that the vector $(Y_1, \ldots, Y_n)$ has the same joint distribution as $(W_1, \ldots, W_n)$ where $W_i \equiv \sum_{j=1}^{i} Z_j/(n-j+1)$ and $Z_1, \ldots, Z_n$ are i.i.d. Exponential(1).
(b) Use the result of (a) to compute $E(Y_i)$, $Var(Y_i)$, and $Cov(Y_i, Y_j)$ for any fixed i, j.

4. Let $X_1, X_2, \ldots, X_N$ be a sample from a distribution with density
$$f_\theta(x) = \theta \exp(\theta x) 1\{x < 0\}, \qquad \theta > 0,$$
and let $V_{(1)} < V_{(2)} < \cdots < V_{(N)}$ denote the order statistics. Show that $Y_1 = V_{(1)} - V_{(2)}, Y_2 = V_{(2)} - V_{(3)}, \cdots, Y_{N-1} = V_{(N-1)} - V_{(N)}, Y_N = V_{(N)}$ are completely independent random variables and that Y_j has density $j\theta e^{j\theta x} 1\{x < 0$ for $j = 1, \ldots, N$.

5. In the context of the two sample problem of testing $H : F = G$ versus $K : F <_s G$, consider an exponential family of distributions

$$f(x; \theta) = c(\theta) \exp(\theta x) h(x)$$

and consider the simple null hypothesis $H_0 : \; f(x) = g(x) = f(x; \theta_0)$ versus the simple alternative $H_1 : \; f(x) = f(x; \theta_0), \; g(x) = f(x; \theta_1)$ with $\theta_0 < \theta_1$. Use the Neyman Pearson lemma to find the best test of H_0 versus H_1 based on the ranks.

6. **Optional bonus problem:** Let $X_1, X_2, \ldots, X_N$ be a sample from a distribution with density $f_\theta(x) = \theta \exp(\theta x) 1\{x < 0\}$, $\theta > 0$, and let $V_{(1)} < V_{(2)} < \cdots < V_{(N)}$ denote the order statistics. Show that $Y_1 = V_{(1)} - V_{(2)}, Y_2 = V_{(2)} - V_{(3)}, \cdots, Y_{N-1} = V_{(N-1)} - V_{(N)}, Y_N = V_{(N)}$ are completely independent random variables and that Y_j has density $j\theta e^{j\theta x} 1\{x < 0$ for $j = 1, \ldots, N$. Use this fact to determine the rejection region of the test you found in problem 4 explicitly when the exponential family $f(x; \theta) = \theta \exp(\theta x) 1\{x < 0\}$; i.e. $c(\theta) = \theta$, $h(x) = 1\{x < 0\}$ in problem 4. Show that the resulting test is a most powerful rank test of $H : \; F = G$ versus $K : \; G = F^2$.