Statistics 583, Problem Set 1

Wellner; 3/28/2007

Reading: Chapter 6, sections 6.3 and 6.4;
Lehmann and Romano, TSH, Chapters 6 and 7.
See also Ferguson, MS, Chapter 5, sections 5.6 and 5.7;

Due: Wednesday, April 4, 2007

1. (From Wasserman, *All of Statistics*, page 171; also see Wasserman's pages 161-164 for a brief discussion of permutation tests.) 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 found 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 avoid the use of large - sample methods. What is your conclusion?

2. 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 the conclusion of (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

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

3. 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 $\theta \equiv \lambda/\mu$.
(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 $g_c(x, y) = (cx, cy)$, $c > 0$, and find a UMP invariant test of size α.
(b) Show that the problem of testing $H_0' : \theta = 1$ versus $H_1' : \theta \neq 1$ is invariant *in addition* under the transformation $g(x, y) = (y, x)$, and find a UMP invariant test of size α.
(c) Find UMP invariant tests of the hypotheses in (a) and (b) when $X_1, \ldots, X_m$ are i.i.d. Exponential(λ) and $Y_1, \ldots, Y_n$ are i.i.d. Exponential(μ).

4. Suppose that $X \sim$ Binomial(m, p_1) and $Y \sim$ Binomial(n, p_2) are independent. In 582 lectures last quarter we derived the UMP unbiased (conditional) test of level α for testing $H : p_2 \leq p_1$ versus $K : p_2 > p_1$. It involves rejecting H if $Y > c(t)$ relative to the conditional (hypergeometric) distribution of Y conditional on $T = X + Y = t$, or equivalently if

$$\frac{Y/t - (n/N)}{\sigma_N} > c'(t)$$

where $c'(T) \to_p z_\alpha \equiv \Phi^{-1}(1 - \alpha)$ since

$$\frac{Y/t - (n/N)}{\sigma_N} \to_d Z \sim N(0, 1)$$

conditionally on T if $0 < \liminf(n/N) \leq \limsup(n/N) < 1$. Here $\sigma_N^2 = (1 - (t - 1)/(N - 1))(n/N)(1 - n/N)/t$.
(a) Show that

$$\frac{Y/t - (n/N)}{\sigma_N} = \frac{\sqrt{\frac{mn}{N}}(\frac{Y}{n} - \frac{X}{m})}{\sqrt{\frac{t}{N}(1 - \frac{t-1}{N-1})}}.$$

(b) Use the result of (a) to show that

$$\frac{Y/T - (n/N)}{\sigma_N} \rightarrow_d Z \sim N(0,1)$$

unconditionally under $p_1 = p_2 \equiv p$ (even if $a = \liminf(n/N) < \limsup(n/N) = b$ with possibly $a = 0$ or $b = 1$). [Hint: prove it first under the assumption that $n/N \rightarrow \lambda \in [0,1]$, and then show that the limit is the same even if n/N does not converge by considering subsequences.

(c) What is the (unconditional) limiting behavior of the test statistic $(Y/t - (n/N))/\sigma_N$ under local alternatives of the form $p_2 = p_{2,N} = p_1 + c/\sqrt{N}$ assuming that $n/N \rightarrow \lambda \in (0,1)$?

(d) What does the result of (c) imply about the limiting power of the test under these alternatives?