Statistics 583, Problem Set 4

Wellner; 4/18/2007

Reading: Chapter 7, sections 7.1- 7.4; Wasserman, Chapters 1-2, pages 1-24.
Due: Wednesday, April 25, 2007

1. Suppose that F is a continuous distribution function on $R^+ = (0, \infty)$, and that $X_i \sim F_i$ where $1 - F_i(x) = (1 - F(x))^{\exp(\theta z_i)}$, $i = 1, \ldots, N$ for some numbers $z_1, \ldots, z_N$ and $\theta \in R$.
(a) Find the locally most powerful rank test of $H: \theta \le 0$ versus $K: \theta > 0$.
(b) Describe how you would cary out your test when $N = 5$ and $z_1 = z_2 = z_3 = 0$, $z_4 = z_5 = 1$.
(c) Describe how you would carry out your test when $N = 400$, $z_i = 0$ for $i = 1, \ldots, 150$, $z_i = 1$ for $i = 151, \ldots, 400$.

2. Let $U_{m,n} \equiv T(\mathbb{F}_m, \mathbb{G}_n)$ where $T(F, G) = \int F dG = P(X \le Y)$ is the Mann-Whitney functional and $\mathbb{F}_m$ and $\mathbb{G}_n$ are the empirical df's of $X_1, \ldots, X_m$ i.i.d. with df F, $Y_1, \ldots, Y_n$ i.i.d. with df G.
(a) Show that
$$mnU_{m,n} + n(n+1)/2 = W_{m,n} \equiv \sum_{j=1}^{n} Q_j = \sum_{j=1}^{n} R_{m+j}\,.$$
(b) Show that $EU_{m,n} = P(X \le Y) = \int F dG$ and that
$$\begin{aligned} & Var(\sqrt{mn}U_{m,n}) \\ &= (n-1)\int (1-G)^2 dF + (m-1)\int F^2 dG - (N-1)(\int F dG)^2 + \int F dG \\ &= (n-1)Var[1 - G(X)] + (m-1)Var[F(Y)] + \int F dG(1 - \int F dG)\,. \end{aligned}$$
(c) When $F = G$ use the results of A and B to compute $E_{(F,F)}W_{m,n}$ and $Var_{(F,F)}(W_{m,n})$. (This should agree with calculations for the Wilcoxon rank sum form of the statistic under the null hypothesis via finite sampling calculations.)

3. Suppose that $\mathcal{F}_+$ is the class of distribution functions F on $\mathbb{R}^+$, and consider the functional $T(F)$ defined for a fixed $x_0 \in R^+$ by
$$T(F) \equiv e_F(x_0) \equiv E_F(X - x_0 | X > x_0) = \frac{\int_{x_0}^{\infty}(1 - F(t))dt}{1 - F(x_0)}\,.$$

This functional is the *mean residual life* functional.
(a) For what collection of df's F_0 is T weakly continuous at F_0? For what collection of df's F_0 is T continuous at F_0 with respect to the Kolmogorov metric?
(b) Find the influence function of $T(F)$.

4. Let F be a distribution function on $\mathbb{R}^2$ with finite second moments, and let $\rho(F)$ be the correlation coefficient

$$\rho(F) = \frac{Cov_F(X,Y)}{\sqrt{Var_F(X)Var_F(Y)}}.$$

Assume that $|\rho(F)| < 1$.
(a) Give an example of a sequence of bivariate distributions $\{F_n\}$ satisfying $F_n \to F$, but $\rho(F_n) \to 1$.
(b) Find a collection $\mathcal{F}$ of distribution functions on $\mathbb{R}^2$ so that ρ is weakly continuous on $\mathcal{F}$.

5. Consider the collection $\mathcal{F}_0$ of distribution functions F on R^+ with $0 < E_F X < \infty$ and $E_F X^2 < \infty$. Let $T(F) \equiv \sigma(F)/\mu(F)$ for $F \in \mathcal{F}_0$ where $\sigma^2(F) = Var_F(X)$ and $\mu(F) = E_F(X)$. This is the *coefficient of variation of* F. Find the influence function of $T(F)$.

6. **Optional bonus problem 1:** For distribution functions F on R^+ and $t_0 > 0$, consider the functional $T(F) = \Lambda(t_0) \equiv \int_0^{t_0} \frac{1}{1-F_-} dF$, the cumulative hazard function corresponding to F at t_0. Find the influence function of $T(F)$. What does this mean about asymptotic normality of the natural estimator $T(\mathbb{F}_n)$ of $T(F)$? Can you prove asymptotic normality of $T(\mathbb{F}_n)$ directly?

7. **Optional bonus problem 2:** Consider the Mann-Whitney-Wilcoxon functional $T(F,G)$ as in problem 1.
(a) Show that $T(F,G)$ is continuous at every pair of distributions (F,G) with respect to the Kolmogorov distance $d_K(F,\tilde{F}) \equiv \sup_x |F(x) - \tilde{F}(x)| \equiv \|F - \tilde{F}\|_\infty$: if $\|F_n - F\|_\infty \to 0$ and $\|G_n - G\|_\infty \to 0$, then $T(F_n, G_n) \to T(F,G)$.
(b) Use the result of (a) to prove that $T(\mathbb{F}_n, \mathbb{G}_n) \to_{a.s.} T(F,G)$.
(c) Give an example to show that $T(F,G)$ is *not* weakly continuous at pairs of distribution functions (F,G) with common discontinuity points.
(d) Extend the definition of Gateaux differentiable functions in a natural way to include $T(F,G)$, and then calculate the Gateaux derivative of $T(F,G)$.
(e) Use your calculation in (d) to "guess" the asymptotic variance of $T(\mathbb{F}_m, \mathbb{G}_n)$.