Statistics 583, Problem Set 6

Wellner; 5/4/2016

Reading: Chapter 7, sections 7.1- 7.4;
van der Vaart, *Asymptotic Statistics*, chapter 20, pages 291 - 303;
Wasserman, Chapters 2-3, pages 13-41.

Due: Wednesday, May 11, 2016

1. Let $T(F) \equiv \Lambda_F(x_0) = \int_0^{x_0}(1-F_-)^{-1}dF$ where $1-F_-(x) = 1-F(x-) = \lim_{y\nearrow x}(1-F(y)) = P_F(X \ge x)$ and where x_0 is fixed. This is the cumulative hazard function of F at x_0.
(a) Is T a weakly continuous function of F (at continuity points of F)? Is it continuous with respect to the Kolmogorov (i.e. the uniform metric) on distribution functions?
(b) Find the influence function of $T(F)$. **Hint:** see van der Vaart, Lemma 20.10, page 298, and Lemma 20.14, page 300.

2. Let $T(F) \equiv \int(x-\mu(F))^3 dF(y)/\sigma^3(F)$ be the skewness functional where $\mu(F) \equiv \int x dF(x)$ and $\sigma^2(F) = \int(x-\mu(F))^2 dF(x)$.
(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 $T(F)$.
Hint: First calculate the influence functions of $\mu(F)$ and $\sigma^2(F)$; then use the chain rule.
Comment: part (b) is problem 1, Wasserman, page 39; the influence function he gives for T on page 29 does not seem to be correct.

3. Suppose that $\mathcal{F}_+$ is the class of distribution functions F on $\mathbb{R}^+$ with mean $\mu_F = E_F X < \infty$, 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 $T(F)$. (Consider expressing $T(F)$ in terms of two simpler functionals $U(F)$ and $V(F)$ and using the chain rule.)

4. **Optional bonus problem 1:** 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 where F and G are continuous. Let $R_1, \ldots, R_{m+n} \equiv R_N$ be the ranks of $(X_1, \ldots, X_m, Y_1, \ldots, Y_n) \equiv (Z_1, \ldots, Z_N)$: $R_j = N\mathbb{H}_N(Z_j)$ for $j \in \{1, \ldots, N\}$ and where $\mathbb{H}_n = (m/N)\mathbb{F}_m + (n/N)\mathbb{G}_n$ is the pooled empirical distribution function. Thus the $Y-$ranks are given by $(R_{m+1}, \ldots, R_N)$, and $Q_1 < \cdots < Q_n$ be the ordered Y ranks.
(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

$$\begin{aligned} &Var(\sqrt{mn}U_{m,n}) \\ &= (n-1)\int (1-G)^2 dF + (m-1)\int F^2 dG - (N-1)\left(\int F dG\right)^2 + \int F dG \\ &= (n-1)Var[1-G(X)] + (m-1)Var[F(Y)] + \int F dG \left(1 - \int F dG\right) . \end{aligned}$$

(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.)

5. **Optional bonus problem 2:** Consider the Mann-Whitney-Wilcoxon functional $T(F,G)$ as in problem 4.
(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 $T(\mathbb{F}_m, \mathbb{G}_n)$.