Statistics 583, Problem Set 4

Wellner; 4/20/2011

Reading: Chapter 7, sections 7.1- 7.4; Wasserman, Chapters 1-2, pages 1-24.
Due: Wednesday, April 27, 2011
Reminder: Midterm Exam, Friday, 5/6.

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.
   (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
   $$\begin{aligned} &Var(\sqrt{mn}U_{m,n}) \\ &\quad = (n-1)\int (1-G)^2 dF + (m-1)\int F^2 dG - \ (N-1)\left(\int F dG\right)^2 + \int F dG \\ &\quad = (n-1)Var[1-G(X)] + (m-1)Var[F(Y)] + \ \int F dG \left(1 - \int F dG\right)\,. \end{aligned}$$
   (c) Show that if $\lambda_N \equiv m/N \to \lambda \in [0,1]$, then, for independent standard Brownian bridge processes $\mathbb{U}$ and $\mathbb{V}$ it follows that
   $$\begin{aligned} Var(\sqrt{mn/N}U_{m,n}) &= \frac{n-1}{N}Var(1-G(X)) + \frac{m-1}{N}Var(F(Y)) + N^{-1}\int F dG\left(1 - \int F dG\right) \\ &\to (1-\lambda)Var(1-G(X)) + \lambda Var(F(Y)) \\ &= (1-\lambda)\int\int (F(x)\wedge F(y) - F(x)F(y))dG(x)dG(y) \\ &\qquad + \lambda \int\int (G(x)\wedge G(y) - G(x)G(y))dF(x)dF(y) \\ &= (1-\lambda)Var\left(\int \mathbb{U}(F)dG\right) + \lambda Var\left(\int \mathbb{V}(G)dF\right) \end{aligned}$$
   as discussed in class on April 15. [Hint: the variance and covariance formulas in Chapter 1, Section 4, might be useful.]
   (d) 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.)

2. (See also van der Vaart (1998), page 303, problem 3.) 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.
(a) Find the influence function of $T(F)$.
(b) What does this mean about asymptotic normality of the natural estimator $T(\mathbb{F}_n)$ of $T(F)$?
(c) Can you prove asymptotic normality of $T(\mathbb{F}_n)$ directly?

3. 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 $F_n \rightarrow_d F$, but $\rho(F_n) \rightarrow 1 \neq \rho(F)$.
(b) Find a collection $\mathcal{F}$ of distribution functions on $\mathbb{R}^2$ so that ρ is weakly continuous on $\mathcal{F}$.

4. **Optional bonus problem 1:** 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 $T(F)$.

5. **Optional bonus problem 2:** 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 $T(F)$.
(c) Can you prove asymptotic normality of $T(\mathbb{F}_n)$ directly?