

Statistics 583, Problem Set 4

Wellner; 4/19/2006

Reading: Chapter 7, sections 7.1- 7.4; Wasserman, Chapters 1-2, pages 1-24.

Due: Wednesday, April 26, 2006

Reminder: Second make-up lecture, Monday, 4/24, 9:30 AM.

1. Let $U_{m,n} \equiv T(\mathbb{F}_m, \mathbb{G}_n)$ where $T(F, G) = \int FdG = P(X \leq Y)$ is the Mann-Whitney functional and \mathbb{F}_m and \mathbb{G}_n are the empirical df's of X_1, \dots, X_m i.i.d. with df F , Y_1, \dots, 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 $EU_{m,n} = P(X \leq Y) = \int FdG$ and that

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

(c) When $F = G$ use the results of A and B to compute $E_{(F,F)}W_{m,n}$ and $\text{Var}_{(F,F)}(W_{m,n})$. (This should agree with calculations for the Wilcoxon rank sum form of the statistic under the null hypothesis via finite sampling calculations.)

2. 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 $\|F_n - F\|_\infty \rightarrow 0$ and $\|G_n - G\|_\infty \rightarrow 0$, then $T(F_n, G_n) \rightarrow T(F, G)$.

(b) Use the result of (a) to prove that $T(\mathbb{F}_n, \mathbb{G}_n) \rightarrow_{a.s.} T(F, G)$.

(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 $T(\mathbb{F}_m, \mathbb{G}_n)$.

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 $x_0 \in \mathbb{R}^+$ 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. Find the influence function of $T(F)$.

4. For distribution functions F on \mathbb{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?

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