Statistics 583, Problem Set 2

Wellner; 4/8/2015

Reading: Chapter 7, sections 7.1- 7.4;
Wasserman, Chapters 1-2, pages 1-24;
Van der Vaart, Chapter 20, pages 291-303.
Due: Wednesday, April 22, 2015
Reminder: Midterm Exam, Wednesday, May 6.

1. Let F_n and F be distribution functions on the real line. Show that:
   (a) If $F_n(x) \to F(x)$ for all x and F is continuous, then $\|F_n - F\|_\infty \to 0$.
   (b) If $F_n(x) \to F(x)$ and $F_n\{x\} \to F\{x\}$ for every x, then $\|F_n - F\|_\infty \to 0$.
   (c) Give an example of a sequence of distribution functions $\{F_n\}$ such that $\|F_n - F\|_\infty \not\to 0$.

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 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

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

   (c) 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

$$\begin{aligned}
&Var(\sqrt{mn/N}U_{m,n}) \\
&= \frac{n-1}{N}Var(1-G(X)) + \frac{m-1}{N}Var(F(Y)) + N^{-1}\int FdG\left(1-\int FdG\right) \\
&\to (1-\lambda)Var(1-G(X)) + \lambda Var(F(Y)) \\
&= (1-\lambda)\int\int (F(x)\wedge F(y) - F(x)F(y))dG(x)dG(y) \\
&\qquad + \lambda\int\int (G(x)\wedge G(y) - G(x)G(y))dF(x)dF(y) \\
&= (1-\lambda)Var\left(\int \mathbb{U}(F)dG\right) + \lambda Var\left(\int \mathbb{V}(G)dF\right)
\end{aligned}$$

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.)

3. Read van der Vaart, *Asymptotic Statistics*, Chapter 12, pages 161 - 172.
(a) Briefly compare the treatment of the Mann-Whitney statistic in VdV's example 12.7, page 166 to the results obtained in problem 2 above.
(b) Do problems 1 and 3, VdV page 171.
(c) Do Problem 10, VdV page 172.

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