Statistics 583, Problem Set 2 Solutions

Wellner; 4/22/2015

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 $F_n \to_d F$ but $\|F_n - F\|_\infty \not\to 0$.

Solution: (a) If $F_n(x) \to F(x)$ for all x and F is continuous, then for each integer M there exist $x_j \in \mathbb{R}, = 1, \ldots, M+1$, such that $F(x_j) - F(x_{j-1}) = 1/M$, for $j = 1, \ldots, M+1$ where $x_0 \equiv -\infty$ and $x_{M+1} \equiv \infty$. Then we have

$$\begin{aligned}
\|F_n - F_0\|_\infty &= \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = \max_{1\le j\le M+1} \sup_{x_{j-1}\le x\le x_j} |F_n(x) - F(x)| \\
&\le \max_{1\le j\le M+1} (F_n(x_j) - F(x_{j-1}) \vee (F(x_j) - F_n(x_j - 1)) \\
&\le \max_{1\le j\le M} (F_n(x_j) - F(x_j) + 1/M) \vee (F(x_{j-1} - F_n(x_{j-1}) + 1/M) \\
&\le \max_{1\le j\le M} |F_n(x_j) - F(x_j)| + 1/M \\
&\le \epsilon/2 + \epsilon/2 = \epsilon
\end{aligned}$$

by choosing M so that $1/M \le \epsilon/2$ and choosing n so large that $|F_n(x_j) - F(x_j)| \le \epsilon/2$ for $j = 1, \ldots, M$.

(b) Let $C_F \equiv \{x \in \mathbb{R} : \ F \text{ is continuous at } x\}$. Then for $x \in C_F^c = \{x \in \mathbb{R} : F \text{ is discontinuous at } x\}$,

$$F_n(x+) = F_n(x) \to F(x), \quad \text{and} \tag{1}$$

$$\begin{aligned}
F_n(x-) &= F_n(x) - F_n\{x\} = F_n(x) - (F_n(x) - F_n(x-)) \\
&\to F(x) - (F(x) - F(x-)) = F(x-)
\end{aligned} \tag{2}$$

since both $F_n(x) \to F(x)$ and $F_n\{x\} \to F\{x\}$ for every x. Now we proceed much as in (a), but with a bit more attention to left and right limits. For a large integer M and $j \in \{0, \ldots, M\}$ let $x_j \equiv \inf\{x : F(x) \ge j/M\} = F^{-1}(j/M)$. Note that we might have $x_0 = -\infty$ and $x_M = +\infty$. Also note that (1) and (2) both hold with $x = \pm\infty$ since we have the convergence $F_n(\tilde{x}_0) \to F(\tilde{x}_0) \le 1/(2M)$ for $\tilde{x}_0 \equiv F^{-1}(1/(2M))$ and similarly for $\tilde{x}_M \equiv F^{-1}(1 - 1/(2M))$. Then for

$j = 0, 1, \ldots, M-1$ and $x \in [x_j, x_{j+1})$ we have, using the monotonicity of F_n and F,

$$\begin{aligned} F_n(x) - F(x) &\le F_n(x_{j+1}-) - F(x_j) \\ &= F_n(x_{j+1}-) - F(x_{j+1}-) + (F(x_{j+1}-) - F(x_j)) \\ &\le F_n(x_{j+1}-) - F(x_{j+1}-) + 1/M \qquad (3) \end{aligned}$$

if we can show that

$$F(x_{j+1}-) - F(x_j) \le 1/M. \qquad (4)$$

Similarly,

$$\begin{aligned} F_n(x) - F(x) &\ge F_n(x_j) - F(x_{j+1}-) \\ &= F_n(x_j) - F(x_j) - (F(x_{j+1}-) - F(x_j)) \\ &\ge F_n(x_j) - F(x_j) - 1/M \qquad (5) \end{aligned}$$

if (4) holds. But $F(x_j) = F(F^{-1}(j/M)) \ge j/M$ for $j = 1, \ldots, M$, and $F(x_{j+1}-) \le (j+1)/M$ since $x_{j+1} > x$ if and only if $F(x) < (j+1)/M$ implies the claimed inequality by letting $x \nearrow x_{j+1}$. (See Shorack and Wellner (1986), (23) and (24) on page 5.) These two inequalities imply that (4) holds, and hence that both (3) and (5) hold. Now we can argue along the same lines as in (a), with $x_0 \equiv -\infty$ and $x_{M+1} = \infty$,

$$\begin{aligned} \|F_n - F_0\|_\infty &= \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = \max_{0\le j\le M+1} \sup_{x_j \le x < x_{j+1}} |F_n(x) - F(x)| \\ &\le \max_{0\le j\le M-1} (F_n(x_{j+1}-) - F(x_{j+1}-) + 1/M) \vee (F(x_j) - F_n(x_j) + 1/M) \\ &\le \max_{0\le j\le M} (F_n(x_j-) - F(x_j-)) \vee \max_{0\le j\le M} (F(x_j) - F_n(x_j)) + 1/M \\ &\to 1/M < \epsilon \end{aligned}$$

if we choose M so large that $1/M < \epsilon$.

(c) Let $F_n(x) = 1_{[1/n,\infty)}(x)$ and let $F(x) = 1_{[0,\infty)}(x)$. Then $F_n(x) \to F(x)$ for all $x \ne 0$, so $F_n \to_d F$, but $\|F_n - F\|_\infty = |F_n(0) - F(0)| = |0-1| = 1$ for all $n \ge 1$.

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 $EU_{m,n} = P(X \le Y) = \int FdG$ and that

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

(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

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

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.)

Solution: (a) Using empirical distribution function notation, $N\mathbb{H}_N = m\mathbb{F}_m + n\mathbb{G}_n$, so

$$
\begin{aligned}
mnU_{m,n} &= \int m\mathbb{F}_m d(n\mathbb{G}_n) = \int N\mathbb{H}_N d(n\mathbb{G}_n) - \int n\mathbb{G}_n d(n\mathbb{G}_n) \\
&= \sum_{j=1}^{n} N\mathbb{H}_N(Y_j) - \sum_{j=1}^{n} n\mathbb{G}_n(Y_{(j)}) \\
&= \sum_{j=1}^{n} R_{m+j} - \sum_{j=1}^{n} j \\
&= \sum_{j=1}^{n} R_{m+j} - n(n+1)/2\,.
\end{aligned}
$$

(b) The expectation is easy:

$$E(U_{m,n}) = \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} P(X_i \le Y_j) = P(X_1 \le Y_1) = \int F dG\,.$$

For the variance, we first calculate

$$\begin{aligned}
E[mnU_{m,n}]^2 &= \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n} E1_{[X_i \le Y_j, X_k \le Y_l]} \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} E1_{[X_i \le Y_j]} + \sum_{i\ne k}\sum_{j=1}^{n} P(X_i \le Y_j, X_k \le Y_j) \\
&\quad + \sum_{i=1}^{m}\sum_{j\ne l} P(X_i \le Y_j, X_i \le Y_l) + \sum_{i\ne k}\sum_{j\ne l} P(X_i \le Y_j, X_k \le Y_l) \\
&= mnP(X_1 \le Y_1) + m(m-1)nP(X_1 \le Y_1, X_2 \le Y_1) \\
&\quad + mn(n-1)P(X_1 \le Y_1, X_1 \le Y_2) \\
&\quad + m(m-1)n(n-1)P(X_1 \le Y_1, X_2 \le Y_2)\,,
\end{aligned}$$

$$P(X_1 \le Y_1) = \int F dG,$$

$$P(X_1 \le Y_1, X_2 \le Y_1) = EP(X_1 \le Y_1, X_2 \le Y_1 | Y_1) = \int F^2(x) dG(x)\,,$$

$$P(X_1 \le Y_1, X_1 \le Y_2) = EP(X_1 \le Y_1, X_1 \le Y_2 | X_1) = \int (1 - G(x-))^2 dF(x),$$

and

$$P(X_1 \le Y_1, X_2 \le Y_2) = P(X_1 \le Y_1)^2 = (\int F dG)^2\,.$$

It follows by algebra that

$$\begin{aligned}
\mathrm{Var}(mnU_{m,n}) &= E(mnU_{m,n})^2 - \{E(mnU_{m,n})\}^2 \\
&= mn\int FdG + m(m-1)n\int F^2 dG \\
&\quad + mn(n-1)\int (1-G(x-))^2 dF(x) \\
&\quad + m(m-1)n(n-1)\{\int FdG\}^2 - (mn\int FdG)^2 \\
&= m(m-1)n\{\int F^2 dG - (\int FdG)^2\} \\
&\quad + mn(n-1)\{\int (1-G(x-))^2 dF(x) - (\int FdG)^2\} \\
&\quad - mn\int FdG(1-\int FdG)\,.
\end{aligned}$$

By noting that

$$\begin{aligned}
\int FdG = P(X \le Y) &= 1 - P(X > Y) \\
&= 1 - \int G(x-)dF(x) = \int (1-G(x-))dF(x)\,,
\end{aligned}$$

this yields the claimed variance formula (to within a left limit):

$$\begin{aligned}
\mathrm{Var}(\sqrt{mn}U_{m,n}) &= (m-1)Var(F(Y)) + (n-1)Var(G_-(X)) \\
&\quad + \int FdG\left(1-\int FdG\right)\,.
\end{aligned}$$

(c) Dividing both sides in the last display by N yields

$$\begin{aligned}
Var(\sqrt{mn/N}U_{m,n}) &= \frac{n-1}{N}Var(1-G_-(X)) + \frac{m-1}{N}Var(F(Y)) \\
&\quad + N^{-1}\int FdG\left(1-\int FdG\right) \\
&\to (1-\lambda)Var(G_-(X)) + \lambda Var(F(Y)) \\
&= (1-\lambda)\int\int (F(x)\wedge F(y) - F(x)F(y))dG(x)dG(y) \\
&\quad + \lambda\int\int (G(x)\wedge G(y) - G(x)G(y))dF(x)dF(y) \\
&\quad \text{by using (1.4.16) - (1.4.17) of Chapter 1, page 19} \\
&\quad \text{with } X\sim F \text{ and } h=G \\
&\quad \text{and then with } Y\sim G \quad \text{and} \quad h=F \\
&= (1-\lambda)Var\left(\int \mathbb{U}(F)dG\right) + \lambda Var\left(\int \mathbb{V}(G)dF\right)
\end{aligned}$$

since

$$E\left(\int \mathbb{U}(F(x))dG(x)\right) = \int E\mathbb{U}(F(x))dG(x) = \int 0\cdot dG(x) = 0, \quad \text{and}$$

$$\begin{aligned}
E\left(\int \mathbb{U}(F(x))dG(x)\right)^2 &= E\left(\int \mathbb{U}(F(x))dG(x)\int \mathbb{U}(F(y))dG(y)\right) \\
&= E\int\int \mathbb{U}(F(x))\mathbb{U}(F(y))dG(x)dG(y) &(6)\\
&= \int\int E\{\mathbb{U}(F(x))\mathbb{U}(F(y))\}dG(x)dG(y) &(7)\\
&= \int\int \{F(x)\wedge F(y) - F(x)\cdot F(y)\}dG(x)dG(y) &(8)
\end{aligned}$$

and similarly for $Var\left(\int \mathbb{V}(G(x))dF(x)\right)$. Here the interchanges of integration in (6) and (7) are justified by Fubini's theorem since

$$\begin{aligned}
E|\int \mathbb{U}(F(x))dG(x)|^2 &\le E\int \mathbb{U}(F(x))^2dG(x) = \int E\mathbb{U}(F(x))^2dG(x) \\
&= \int F(x)(1-F(x))dG(x) \quad \text{by Tonelli's theorem} \\
&\le (1/4)\int dG(x) < 1/4.
\end{aligned}$$

(d) When $F=G$ continuous we find that

$$E(U_{mn}) = \int FdF = 1/2\,,$$

and, since now $Var[F(Y)] = Var[G(X)] = 1/12$,

$$\begin{aligned} \mathrm{Var}(\sqrt{mn}U_{m,n}) &= (m-1)\frac{1}{12} + (n-1)\frac{1}{12} + \frac{1}{4} \\ &= (N-2)\frac{1}{12} + \frac{1}{4} = (N+1)\frac{1}{12}\,. \end{aligned}$$

Hence from part (a) it follows that

$$E(\sum_{j=1}^{n} Q_j) = n(n+1)/2 + mnE(U_{m,n}) = n(N+1)/2$$

and

$$\mathrm{Var}(\sum_{j=1}^{n} Q_j) = mn\mathrm{Var}(\sqrt{mn}U_{m,n}) = mn(N+1)\frac{1}{12}$$

both of which agree with finite sampling calculations (drawing a sample of size n balls from an urn without replacement where the numbers on the N balls in the urn are $\{1, 2, \ldots, N\}$).

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.

Solution: (a) Van der Vaart, example 12.7, shows (or at least states) that the projection $\hat{U}_{m,n}$ of $U - \theta = \int \mathbb{F}_m d\mathbb{G}_n - \int F dG$ is given by

$$\hat{U}_{m,n} = \frac{1}{n}\sum_{j=1}^{n}\left(F(Y_j) - \int F dG\right) - \frac{1}{m}\sum_{i=1}^{m}\left(G_-(X_i) - \int G_- dF\right).$$

This is easily verified by computing

$$\hat{U}_{m,n} = \frac{1}{m}\sum_{i=1}^{m} E\{(U-\theta)|X_i\} + \frac{1}{n}\sum_{j=1}^{n} E\{(U-\theta)|Y_j\}$$

where we note that

$$
\begin{aligned}
E\{(U-\theta)|X_i\} &= E\left\{\frac{1}{mn}\sum_{i=1}^m\sum_{j=1}^n\left(1_{[X_i\le Y_j]}-\int FdG\right)|X_i\right\}\\
&= \frac{1}{mn}\left(\sum_{i'\ne i}+\sum_{i'=i}\right)\sum_{j=1}^n E\left\{1_{[X_{i'}\le Y_j]}-\int FdG|X_i\right\}\\
&= \frac{1}{mn}\sum_{j=1}^n E\{1_{[X_i\le Y_j]}-\int FdG|X_i\}\\
&= -\frac{1}{mn}\sum_{j=1}^n E\{(1-1_{[X_i\le Y_j]})-(1-\int FdG)|X_i\}\\
&= -\frac{1}{mn}\sum_{j=1}^n E\{1_{[Y_j<X_i]})-\int G_-dF|X_i\}\\
&= -\frac{1}{m}\left(G_-(X_i)-\int G_-dF\right),
\end{aligned}
$$

and, similarly,

$$
\begin{aligned}
E\{(U-\theta)|Y_j\} &= E\left\{\frac{1}{mn}\sum_{i=1}^m\left(\sum_{j'\ne j}+\sum_{j'=j}\right)\left(1_{[X_i\le Y_j']}-\int FdG\right)|Y_j\right\}\\
&= \frac{1}{mn}\sum_{i=1}^m E\left\{1_{[X_i\le Y_j]}-\int FdG|Y_j\right\}\\
&= \frac{1}{n}\left(F(Y_j)-\int FdG\right).
\end{aligned}
$$

This projection argument does not give the exact variance of $U_{m,n}$ as computed in problem 2, but it does give the correct main terms of the approximation, namely

$$
Var(\hat{U}_{m,n}) = \frac{1}{n}Var(F(Y))+\frac{1}{m}Var(G_-(X)).
$$

and hence

$$
Var(\sqrt{mn/N}\hat{U}_{m,n}) = \frac{m}{N}Var(F(Y))+\frac{n}{N}Var(G_-(X)).
$$

(b) VdV, problem 1: With $U\equiv G_n\equiv\sum_{i<j}|X_i-X_j|/\binom{n}{2}$ and $\theta\equiv g\equiv E|X_1-X_2|$ we need to compute $\hat{U}_n=\sum_{i=1}^n E(U-\theta|X_i)=(2/n)\sum_{i=1}^n h_1(X_i)$ where

$$
h_1(x) = Eh(x,X_2)-\theta = E|x-X_2|-E|X_1-X_2| = \int|x-y|dF(y)-E|X_1-X_2|.
$$

Then, since $Eh^2(X_1, X_2) < \infty$ if $E(X_1^2) < \infty$ (since

$$E|X_1 - X_2|^2 \le 2(E(X_1^2) + E(X_2^2)) = 4E(X_1^2)$$

by the c_r-inequality), van der Vaart's Theorem 12.3 yields $\sqrt{n}(U_n - \theta - \hat{U}_n) \to_p 0$, and hence

$$\begin{aligned}\sqrt{n}(G_n - g) &= \frac{2}{\sqrt{n}}\sum_{i=1}^n h_1(X_i) + o_p(1)\\ &\to_d 2N(0, Var(h_1(X))) = N(0, 4Var(h_1(X)))\end{aligned}$$

where

$$Var(h_1(X)) = \int \left(\int (|x-y| - g) dF(y) \right)^2 dF(x).$$

(b) VdV, problem 10: For $\theta = E(X - E(X))^3$, note that for real numbers a, b we have

$$(a-b)^3 = a^3 - 3a^2b + 3ab^2 - b^3.$$

Thus

$$(X - E(X))^3 = X^3 - 3X^2E(X) + 3X(E(X))^2 - (EX)^3. \tag{9}$$

This suggests taking $h(x,y,z) = x^3 - 3x^2y + 3xyz - xyz = x^3 - 3x^2y + 2xyz$. Then we find that if X, Y, and Z are all equal in distribution to X and independent

$$\begin{aligned}Eh(X,Y,Z) &= E(X^3) - 3E(X^2)E(Y) + 2E(X)E(Y)E(Z)\\ &= E(X^3) - 3E(X^2)E(X) + 3E(X)E(X)^2 - (E(X))^3\\ &= E(X - E(X))^3\end{aligned}$$

where the last equality follows from (9). To find a symmetric kernel that works we can take

$$h_s(x_1, x_2, x_3) \equiv \frac{1}{3!}\sum_{\pi} h(x_{\pi(1)}, x_{\pi(2)}, x_{\pi(3)})$$

where the sum is over all 6! permutations of $\{1, 2, 3\}$.

(c) Let $V_n \equiv n^{-2}\sum_{i=1}^n \sum_{j=1}^n h(X_i, X_j)$ and

$$U_n \equiv \frac{1}{\binom{n}{2}}\sum_{i<j} h(X_i, X_j) = \frac{1}{n(n-1)}\sum_{i \ne j} h(X_i, X_j).$$

From van der Vaart's Theorem 12.3 we know that if $Eh^2(X_1, X_2) < \infty$, then $\sqrt{n}(U_n - \theta - \hat{U}_n) \to_p 0$, and hence

$$\sqrt{n}(U_n - \theta) \to_d N(0, 4Var(h_1(X_1)))$$

where $h_1(x) = Eh(x, X_2) - \theta$. But

$$\begin{aligned} V_n - U_n &= n^{-2} \sum_{i=1}^n \sum_{j=1}^n h(X_i, X_j) - \frac{1}{n(n-1)} \sum_{i \ne j} h(X_i, X_j) \\ &= \frac{1}{n^2} \sum_{i=1}^n h(X_i, X_i) + \left\{ \frac{1}{n^2} - \frac{1}{n(n-1)} \right\} \sum_{i \ne j} h(X_i, X_j) \\ &= \frac{1}{n^2} \sum_{i=1}^n h(X_i, X_i) - \frac{1}{n^2(n-1)} \sum_{i \ne j} h(X_i, X_j) \\ &= \frac{1}{n^2} \sum_{i=1}^n h(X_i, X_i) - \frac{1}{n} U_n. \end{aligned} \tag{10}$$

Thus

$$\sqrt{n}(V_n - U_n) \quad = \tfrac{\sqrt{n}}{n^2} \textstyle\sum_{i=1}^n h(X_i, X_i) - \tfrac{1}{\sqrt{n}} U_n.$$

Now $U_n \to_p \theta$, so $n^{-1/2} U_n \to_p 0$, and

$$\begin{aligned} E\{n^{-3/2} \sum_{i=1}^n h(X_i, X_i)\} &= n^{-1/2} Eh(X_1, X_1) \to 0, \quad \text{and} \\ Var\left(n^{-3/2} \sum_{i=1}^n h(X_i, X_i) \right) &= n^{-3} n Var(h(X_1, X_1) \to 0 \end{aligned}$$

since $Eh^2(X_1, X_1) < \infty$ by assumption. Thus the first term in (10) converges to 0 by Chebychev's inequality. It follows that $\sqrt{n}(V_n - \theta) \to_d N(0, 4Var(h_1(X_1)))$.

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

Solution: (a) $T(F,G)$ is continuous at *every* pair (F,G) with respect to the Kolmogorov distance: if $\|F_n - F\|_\infty \to 0$ and $\|G_n - G\|_\infty \to 0$, then

$$T(F_n, G_n) - T(F,G) = \int (F_n - F) dG_n + \int F d(G_n - G)$$

where

$$|\int (F_n - G_n) dG_n| \le \|F_n - G_n\|_\infty \int dG_n = \|F_n - G_n\|_\infty \to 0\,,$$

and, using integration by parts (Proposition 1.4.1, chapter 1, page 17) or Fubini,

$$|\int F d(G_n - G)| = | - \int (G_n(x-) - G(x-)) dF(x)| \le \|G_n - G\|_\infty \to 0\,.$$

(b) Since $\|\mathbb{F}_n - F\|_\infty \to_{a.s.} 0$ and $\|\mathbb{G}_n - G\|_\infty \to_{a.s.} 0$ by the Glivenko-Cantelli theorem, it follows immediately from the continuity proved in (a) that $T(\mathbb{F}_n, \mathbb{G}_n) \to_{a.s.} T(F,G)$.
(c) Here is an example to show that $T(F,G)$ is not weakly continuous at all pairs (F,G): Let $X_n \sim \text{Uniform}(0, 1/n) \equiv F_n$, so that $X_n \to_d 0 \equiv X \sim \delta_0$, and let $Y_n \sim \text{Uniform}(-1/n, 0) \equiv G_n$ so that $Y_n \to_d 0 \equiv Y \sim \delta_0$. Note that $X_n > 0$ a.s. while $Y_n < 0$ a.s.. Then $T(F_n, G_n) = P(X_n \le Y_n) = 0$ for all n, but $T(F,G) = P(X \le Y) = 1$. Hence $T(F,G)$ is *not* weakly continuous at *all* (F,G). However, it is weakly continuous at all pairs (F,G) with *no common discontinuity points*. Here is the proof: suppose that $F_n \to_d F$ and $G_n \to_d G$ where F and G have no common discontinuity points. Then $F_n \times G_n \to_d F \times G$ on $R \times R$: i.e. with $X_n \sim F_n$ and $Y_n \sim G_n$ independent, $(X_n, Y_n) \to_d (X, Y) \sim F \times G$; here (X,Y) are independent with df's F and G respectively. Since F and G have no common discontinuities, the function $g(x,y) \equiv 1_{[x \le y]}$ is continuous a.e. $F \times G$: note that all the mass points of the distribution $F \times G$ on R^2 fall off the diagonal, so $P(X = Y) = \int \{F(x) - F(x-)\} dG(x) = 0$. Hence by the Helly-Bray theorem (Proposition 2.3.7, chapter 2, page 13) it follows that

$$T(F_n, G_n) = E 1_{[X_n \le Y_n]} = E g(X_n, Y_n) \to E g(X, Y) = T(F,G)\,.$$

(d) One simple definition of the Gateaux derivative would be as follows: let $F_t \equiv (1-t)F + tF_1$ and $G_t \equiv (1-t)G + tG_1$ for df's F, F_1, G, G_1. Then

$$\frac{d}{dt} T(F_t, G_t)|_{t=0} = \lim_{t \to 0} \frac{T(F_t, G_t) - T(F,G)}{t} \equiv \dot{T}(F, G, F_1 - F, G_1 - G)\,. \quad (11)$$

We now calculate (11): clearly

$$\begin{aligned} T(F_t, G_t) &= \int \{F + t(F_1 - F)\} d\{G + t(G_1 - G)\} \\ &= \int F dG + t \int (F_1 - F) dG + t \int F d(G_1 - G) \\ &\quad + t^2 \int (F_1 - F) d(G_1 - G)\,, \end{aligned}$$

so

$$\begin{aligned} \frac{d}{dt} T(F_t, G_t)|_{t=0} &= \int (F_1 - F) dG + \int F d(G_1 - G) \\ &= \int (F_1 - F) dG - \int (G_1 - G)_- dF \\ &= \int G_- d(F_1 - F) + \int F d(G_1 - G) \\ &= \int (G_- - \int G_- dF) dF_1 + \int (F - \int F dG) dG_1 \\ &= \dot{T}(F, G; F_1 - F, G_1 - G)\,. \end{aligned}$$

(e) To "guess" the asymptotic variance of $T(\mathbb{F}_m, \mathbb{G}_n)$, write

$$\begin{aligned} &\sqrt{\frac{mn}{N}} (T(\mathbb{F}_m, \mathbb{G}_n) - T(F, G)) \\ &\doteq \sqrt{\frac{mn}{N}} \left\{ \int (G_- \int G_- dF) d(\mathbb{F}_m - F) + \int (F - \int F dG) d(\mathbb{G}_n - G) \right\} \\ &\qquad + \sqrt{\frac{mn}{N}} o(\|\mathbb{F}_m - F\|_\infty \vee \|\mathbb{G}_n - F\|_\infty) \\ &= \sqrt{1 - \lambda_N} \frac{1}{\sqrt{m}} \sum_{i=1}^{m} (G_-(X_i) - \int G_- dF) + \sqrt{\lambda_n} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (F(Y_j) - \int F dG) + o_p(1) \\ &\to_d \sqrt{1 - \lambda} N(0, Var(G_-(X))) + \sqrt{\lambda} N(0, Var(F(Y))) \\ &= N(0, (1 - \lambda) Var(G_-(X)) + \lambda Var(F(Y))) \end{aligned}$$

by using the independence of the X's and Y's to get independence of the two limiting normal distributions in the last line, and assuming that $\lambda_N \to \lambda$. Note that this asymptotic variance agrees with our finite - sample calculations in part (a). The $\doteq$ in the first line above would be rigorous if $T(F, G)$ was Fréchet differentiable with respect to the supremum metric.