Statistics 583, Problem Set 4 Solutions

Wellner; 4/26/2006

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.
(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

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

(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.)

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

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

By noting that

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

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

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

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

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

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 the finite sampling calculations of problem 2.(a) of problem set 3.

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 $\|F_n - F\|_\infty \to 0$ 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

$$\frac{d}{dt}T(F_t, G_t)|_{t=0} = \lim_{t \to 0} \frac{T(F_t, G_t) - T(F,G)}{t} \equiv \dot{T}(F, G, F_1 - F, G_1 - G)\,. \qquad (1)$$

We now calculate (1): 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.

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 $x_0 \in R^+$ 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 $T(F)$.

Solution: First note that with $F_t \equiv (1-t)F + tG$ we have both

$$\frac{d}{dt}(1 - F_t(x_0))\Big|_{t=0} = -(G - F)(x_0)$$

and

$$\frac{d}{dt}\int_{x_0}^\infty (1 - F_t(y))dy\Big|_{t=0} = -\int_{x_0}^\infty (G - F)(y)dy .$$

Thus by the product rule we calculate

$$\begin{aligned} \frac{d}{dt}T(F_t)\Big|_{t=0} &= -\frac{\int_{x_0}^\infty (G-F)(y)dy}{1 - F(x_0)} + \frac{\int_{x_0}^\infty (1 - F(y))dy}{(1 - F(x_0))^2}(G - F)(x_0) \\ &= e_F(x_0)\frac{(G-F)(x_0)}{1 - F(x_0)} - \frac{\int_{x_0}^\infty (G-F)(y)dy}{1 - F(x_0)} . \end{aligned}$$

Taking $G = \delta_x = 1_{[x,\infty)}$ yields the influence function for T at F:

$$\begin{aligned} IC(x;T,F) &= e_F(x_0)\frac{(1_{[x,\infty)}(x_0) - F(x_0))}{1 - F(x_0)} - \frac{\int_{x_0}^\infty (1_{[x,\infty)}(y) - F(y))dy}{1 - F(x_0)} \\ &= e_F(x_0)\frac{(1_{[0,x_0]}(x) - F(x_0))}{1 - F(x_0)} - \frac{\int_{x_0}^\infty (1_{[0,y]}(x) - F(y))dy}{1 - F(x_0)} \\ &= \begin{cases} e_F(x_0) - \frac{\int_{x_0}^\infty 1 - F(y)dy}{1 - F(x_0)} & x \le x_0 \\ \frac{-F(x_0)}{1-F(x_0)}e_F(x_0) - \frac{\int_{x_0}^\infty 1_{[0,y]}(x_0) - F(y)dy}{1-F(x_0)} & x > x_0 \end{cases} \\ &= \begin{cases} 0 & x \le x_0 \\ \frac{-F(x_0)}{1-F(x_0)}e_F(x_0) - \frac{\int_{x_0}^x -F(y)dy}{1-F(x_0)} - \frac{\int_x^\infty 1_{[0,y]}(x_0) - F(y)dy}{1-F(x_0)} & x > x_0 \end{cases} \\ &= \begin{cases} 0 & x \le x_0 \\ \frac{(x - x_0) - e_F(x_0)}{1 - F(x_0)} & x > x_0 \end{cases} \\ &= \frac{[(x - x_0) - e_F(x_0)]1_{(x_0,\infty)(x)}}{1 - F(x_0)} . \end{aligned}$$

Note that

$$E_F[IC^2(X;T,F)] = \frac{Var(X - x_0 | X > x_0)}{1 - F(x_0)} .$$

4. 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. 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?

Solution: To find the influence function of $T(F)$, let $F_t = (1-t)F + t\delta_x$. The distribution function corresponding to $G \equiv \delta_x$ is $1_{(-\infty,y]}(x)$, $y \in R$, so the left limit is $G_-(y) = 1_{[x<y]}$, and the corresponding "at risk" function $1 - G_-(y) = 1_{[x \geq y]} = 1_{[y,\infty)}(x)$. We need to compute

$$\begin{aligned}
&\lim_{t \to 0} \frac{T(F_t) - T(F)}{t} = \frac{d}{dt} T(F_t)|_{t=0} \equiv IC(x;T,F) \equiv \psi_F(x) \\
&= \frac{d}{dt} \left\{ \int_0^{t_0} \frac{1}{1 - (F_t)_-} dF_t \right\} \Big|_{t=0} \\
&= \int_0^{t_0} \frac{1}{1 - F_-} d(\delta_x - F) + \int_0^{t_0} \frac{(\delta_x - F_-)}{(1 - F_-)^2} dF \\
&= \frac{1_{[0,t_0]}(x)}{1 - F_-(x)} - \int_0^{t_0} \frac{1}{1 - F_-} dF + \int_0^{t_0} \frac{1}{1 - F_-} dF - \int_0^{t_0} \frac{(1 - \delta_{x-})}{(1 - F_-)^2} dF \\
&= \frac{1_{[0,t_0]}(x)}{1 - F_-(x)} - \int_0^{t_0} \frac{1_{[x \geq y]}}{(1 - F_-(y))^2} dF(y) \\
&= \frac{1_{[0,t_0]}(x)}{1 - F_-(x)} - \int_0^{x} \frac{1_{[0,t_0]}(y)}{1 - F_-(y)} d\Lambda(y) \\
&= \begin{cases} \frac{1}{1-F_-(x)} - \int_0^x \frac{1}{(1-F_-)^2} dF & \text{if} \quad x \leq t_0 \\ -\int_0^{t_0} \frac{1}{(1-F_-)^2} dF & \text{if} \quad x > t_0 . \end{cases}
\end{aligned}$$

The next to last formula for the influence function of $\Lambda(t_0)$ is natural from a martingale perspective. When F is continuous $F_- = F$, and the influence function computed above reduces to:

$$\begin{aligned}
IC(x;T,F) &= 1_{[x \leq t_0]} - \frac{F(t_0)}{1 - F(t_0)} 1_{[x > t_0]} \\
&= \frac{1_{[x \leq t_0]} - F(t_0)}{1 - F(t_0)} .
\end{aligned}$$

Note that $E_F \psi_F(X) = 0$ and (in the case of a continuous d.f. F)

$$E_F \psi_F^2(X) = \frac{F(t_0)}{1 - F(t_0)} .$$

To prove asymptotic normality of $T(\mathbb{F}_n)$ (assuming that F satisfies $F(t_0) < 1$), write

$$\begin{aligned}
\sqrt{n}(T(\mathbb{F}_n) - T(F)) &= \sqrt{n}\left\{\int_0^{t_0} \frac{1}{1-\mathbb{F}_n(s-)} d\mathbb{F}_n(s) - \int_0^{t_0} \frac{1}{1-F(s-)} dF(s)\right\} \\
&= \int_0^{t_0} \frac{1}{1-\mathbb{F}_n(s-)} d[\sqrt{n}(\mathbb{F}_n(s) - F(s))] \\
&\quad + \int_0^{t_0} \sqrt{n}\left\{\frac{1}{1-\mathbb{F}_n(s-)} - \frac{1}{1-F(s-)}\right\} dF(s) \\
&= \frac{\sqrt{n}(\mathbb{F}_n(t_0) - F(t_0))}{1-\mathbb{F}_n(t_0-)} - \int_0^{t_0} \sqrt{n}(\mathbb{F}_n(s) - F(s))^2 \frac{1}{(1-\mathbb{F}_n(s-))^2} d\mathbb{F}_n(s) \\
&\quad + \int_0^{t_0} \frac{\sqrt{n}(\mathbb{F}_n(s) - F(s))}{(1-\mathbb{F}_n(s-))(1-F(s-))} dF(s) \\
&= \frac{\sqrt{n}(\mathbb{F}_n(t_0) - F(t_0))}{1-\mathbb{F}_n(t_0-)} + o_p(1) \\
&\to_d \frac{\mathbb{U}(F(t_0))}{1-F(t_0-)} \\
&\sim N(0, \frac{F(t_0)}{1-F(t_0)}) \qquad \text{if } F \text{ is continuous}
\end{aligned}$$

since the last two terms can be rewritten as

$$\int_0^{t_0} \frac{\sqrt{n}(\mathbb{F}_n(s) - F(s))}{1-\mathbb{F}_n(s-)} \left\{\frac{d\mathbb{F}_n(s)}{1-\mathbb{F}_n(s-)} - \frac{dF(s)}{1-F(s-)}\right\} = o_p(1)$$

by arguments similar to those we used to deal with the Mann-Whitney Wilcoxon statistic. Alternatively, martingale methods also work.

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 $F_n \to F$, but $\rho(F_n) \to 1$.

(b) Find a collection $\mathcal{F}$ of distribution functions on $\mathbb{R}^2$ so that ρ is weakly continuous on $\mathcal{F}$.

Solution: (a) Without loss of generality we may suppose that F is a bivariate distribution function with zero means, $E_F(X) = E_F(Y) = 0$. Let $F_n = (1 - n^{-1})F + n^{-1}\delta_{(a_n,b_n)}$ with $(a_n, b_n) \in \mathbb{R}^2$. Note that F_n has marginal distribution functions $F_{n,X} = (1-n^{-1})F_X + n^{-1}\delta_{a_n}$, $F_{n,Y} = (1-n^{-1})F_Y + n^{-1}\delta_{b_n}$ respectively where F_X and F_Y are the marginal df's of F. Thus we compute

$$\begin{aligned}
Cov_{F_n}(X,Y) &= E_{F_n}(XY) - E_{F_n}(X)E_{F_n}Y \\
&= (1 - n^{-1}Cov_F(X,Y) + n^{-1}a_n b_n - (n^{-1}a_n)(n^{-1}b_n) \\
&= (1 - n^{-1}Cov_F(X,Y) + n^{-1}(1 - n^{-1})a_n b_n, \\
Var_{F_n}(X) &= E_{F_n}(X^2) - (E_{F_n}(X))^2 = (1 - n^{-1})Var_F(X) + n^{-1}(1 - n^{-1})a_n^2, \\
Var_{F_n}(Y) &= E_{F_n}(Y^2) - (E_{F_n}(Y))^2 = (1 - n^{-1})Var_F(Y) + n^{-1}(1 - n^{-1})b_n^2.
\end{aligned}$$

Choosing $a_n = b_n = n$ yields

$$\begin{aligned}
Cov_{F_n}(X,Y) &= n + o(n) = n(1 + o(1)), \\
Var_{F_n}(X) &= n + o(n) = n(1 + o(1)), \\
Var_{F_n}(Y) &= n + o(n) = n(1 + o(1)).
\end{aligned}$$

Thus we find that

$$\rho(F_n) = \frac{Cov_{F_n}(X,Y)}{\sqrt{Var_{F_n}(X)Var_{F_n}(Y)}} = \frac{n(1+o(1))}{n(1+o(1))} \to 1$$

as $n \to \infty$. Thus ρ is weakly discontinuous at every F.

(b) Consider the following collection of distributions on R^2: for some $r > 2$ and $M < \infty$

$$\mathcal{F}_{r,M} \equiv \{F: \quad E_F|X|^r \le M, \ E_F|Y|^r \le M\}.$$

Then ρ is weakly-continuous on $\mathcal{F}_{r,M}$ at any F with $Var_F(X) > 0$ and $Var_F(Y) > 0$. Here is a proof: let $\{F_n\} \subset \mathcal{F}_{r,M}$ satisfy $F_n \to_d F$. Then with $(X_n, Y_n) \sim F_n$ and $(X,Y) \sim F$ we have $(X_n, Y_n) \to_d (X,Y)$, and by a Skorokhod construction there exist $(X_n^*, Y_n^*) =_d (X_n, Y_n)$ and $(X^*, Y^*) =_d (X,Y)$ defined on a common probability space and satisfying $(X_n^*, Y_n^*) \to_{a.s.} (X^*, Y^*)$. But because $\{F_n \subset \mathcal{F}_{r,M}$, X_n^2, Y_n^2, and $|X_n Y_n|$ are all uniformly integrable: since $r > 2$,

$$EX_n^2 1_{[X_n^2 \ge \lambda]} \ \le \ \frac{1}{\lambda^{r-2}} E|X_n|^r \le \frac{M}{\lambda^{r-2}}$$

so

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} EX_n^2 1_{[X_n^2 \ge \lambda]} \ \le \ \lim_{\lambda\to\infty} \frac{M}{\lambda^{r-2}} = 0$$

and similarly for $\{Y_n^2\}$, so the uniform integrability of $|X_n Y_n|$ follows by Cauchy-Schwarz. The same holds true for the (X_n^*, Y_n^*) pairs since the uniform integrability only depends on the (marginal) distributions. Thus by Vitali's theorem it follows that

$$EX_n^s = EX_n^* s \to EX^* s = EX^s$$

and

$$EY_n^s = EY_n^* s \to EY^* s = EY^s$$

for $s = 1, 2$, while Vitali also yields

$$EX_n Y_n = EX_n^* Y_n^* \to EX^* Y^* = EXY\,.$$

Therefore

$$Var_{F_n}(X_n) \to Var_F(X), Var_{F_n}(Y_n) \to Var_F(Y), \qquad (2)$$

and

$$Cov_{F_n}(X_n, Y_n) \to Cov_F(X, Y). \qquad (3)$$

Since we have assumed that $Var_F(X) > 0$ and $Var_F(Y) > 0$, (2) and (3) yield

$$\rho(F_n) = \frac{Cov_{F_n}(X_n, Y_n)}{\sqrt{Var_{F_n}(X_n) \cdot Var_{F_n}(Y_n)}} \to \frac{Cov_F(X, Y)}{\sqrt{Var_F(X) \cdot Var_F(Y)}} = \rho(F);$$

i.e. ρ is continuous on $\mathcal{F}_{r,M}$ at any F with positive variances.

It is interesting to note that the hypothesis $\{F_n\} \subset \mathcal{F}_{r,M}$ *cannot be weakened to* $\{F_n\} \subset \mathcal{F}_{2,M}$ (and hence it can also not be weakened to the still larger class $\mathcal{F}_{2,\infty}$). Here is a counterexample. Let F be a d.f. on R^2 with $EX = 0 = EY$ and $EX^2 = 1 = EY^2$, and $\rho(F) < 1$ where $(X, Y) \sim F$. Let $M > 1$ be a big number, and consider the class

$$\mathcal{F}_{2,M} = \{F \text{ on } R^2 : \ E_F X^2 \leq M, E_F Y^2 \leq M\}\,.$$

Let $a_n, b_n > 0$; we will specify them in terms of M shortly. Consider the sequence of d.f.'s $\{F_n\} \subset \mathcal{F}_{2,M}$ defined by

$$F_n(x, y) = (1 - \frac{1}{n})F(x, y) + \frac{1}{2n}\delta_{(a_n, b_n)} + \frac{1}{2n}\delta_{(-a_n, -b_n)}.$$

Then for any bounded and continuous function $\psi : R^2 \to R$,

$$\begin{aligned} \int \psi dF_n &= (1 - \frac{1}{n}) \int \psi dF + \frac{1}{2n}\psi(a_n, b_n) + \frac{1}{2n}\psi(-a_n, -b_n) \\ &\to \int \psi dF, \end{aligned}$$

so $F_n \rightarrow_d F$. Furthermore, with $(X_n, Y_n) \sim F_n$,

$$EX_n = (1-1/n)EX = 0, EY_n = 0,$$

$$EX_n^2 \quad = \quad = (1-1/n)EX^2 + \frac{a_n^2}{n} = (1-1/n) + \frac{a_n^2}{n} = M$$

if $a_n^2 = n\{M - (1-1/n)\}$. Similarly,

$$EY_n^2 = (1-1/n) + \frac{b_n^2}{n} = M$$

if $b_n^2 = n\{M - (1-1/n)\}$. With these choices of a_n and b_n,

$$Cov(X_n, Y_n) = (1-1/n)Cov(X,Y) + \frac{a_n b_n}{n},$$

$$\begin{aligned}
\rho(F_n) &= \frac{Cov(X_n, Y_n)}{Var(X_n)Var(Y_n)} \\
&= \frac{(1-1/n)Cov(X,Y) + M - (1-1/n)}{\sqrt{M^2}} \\
&\rightarrow \frac{\rho(F) + M - 1}{M} \neq \rho(F).
\end{aligned}$$

Thus $\rho(F)$ is not continuous on $\mathcal{F}_{2,M}$.