Statistics 583, Final Exam Solutions

Wellner; 6/8/2011

1. (36 points) **Define any three of** the following terms.
   (a) A Gateaux - differentiable functional $T : \mathcal{F} \to \mathbb{R}$ and the corresponding influence function.
   (b) A metric d between distribution functions which is *compatible with respect to the empirical distribution function*. Give one example of such a metric.
   (c) The Nadaraya-Watson estimator $\hat{r}_n$ of a regression function r on $[a, b] \subset \mathbb{R}$.
   (d) The *bounded Lipschitz metric* d_{BL} between two probability measures P and Q.
   (e) The U- and V- statistics and natural $V-$functional corresponding to a symmetric kernel $h(x_1, \ldots, x_r)$ (and d.f.'s F with $E_F|h(X_1, \ldots, X_r)| < \infty$).
   (f) A "rule-of-thumb" band-width h_n based on a normality assumption.

   Solution: See course notes.

2. (36 points). Give a complete *statement* of **three** of the following results or theorems:
   (a) An example of a functional $T(F)$ which is always weakly lower-semicontinuous.
   (b) A limit theorem for the the bootstrap empirical process $\sqrt{m}(\mathbb{F}_m^* - \mathbb{F}_n)$ when $m \wedge n \to \infty$.
   (c) Any theorem about the "bootstrap working" for a differentiable statistical functional.
   (d) A general result concerning the the jackknife estimator of variance of the estimator $T_n \equiv T(\mathbb{F}_n)$ of a functional $T(F)$.
   (e) Hoeffding's exponential inequality for a sum of independent bounded random variables, with $a_i \le X_i \le b_i$ for $i = 1, \ldots, n$
   (f) Your favorite theorem about the bias and variance of a kernel density estimator leading to a (theoretical) optimal bandwidth

   Solution: See course notes.

   Do **both** problem 3 **and** problem 4.

3. (40 points). Suppose that H is a bivariate distribution function of a pair of positive random variables (X, Y) with marginal distribution functions F and G (i.e $F(x) = H(x, \infty)$ and $G(y) = H(\infty, y)$), and with $EX^4 < \infty$, $EY^4 < \infty$, $\mu(F) > 0$, and $\sigma^2(G) = Var_G(Y) > 0$. Consider the functional

$$T(H) = \frac{\sigma(F)/\mu(F)}{\sigma(G)/\mu(G)}$$

the ratio of the marginal *coefficients of variation* $cv(F) \equiv \sigma(F)/\mu(F)$ and $cv(G) \equiv \sigma(G)/\mu(G)$; here $\mu(F) = E_F(X)$, $\sigma^2(F) = Var_F(X)$ and similarly for G. Suppose that we observe i.i.d. pairs (X_i, Y_i) from the distribution H and estimate $T(H)$ by $T_n \equiv T(\mathbb{H}_n)$ where $\mathbb{H}_n$ is the empirical distribution function (or empirical measure) of the pairs.

(a) Explain how you would use the jackknife to estimate $nVar_H(T_n)$.

(b) Explain how you would use the bootstrap to estimate $nVar_H(T_n)$. Discuss both the ideal bootstrap estimator and the Monte-Carlo implementation thereof.

(c) Do you believe that $\sqrt{n}(T_n - T(H)) \to_d N(0, V^2)$ for some V^2 under the above hypotheses? Why or why not?

(d) Will the jackknife estimator of variance "work" in this situation? Will the bootstrap estimator of variance "work" in this situation?

(e) Describe the bootstrap estimator of

$$Q_n(x, H) \equiv Pr_H(\sqrt{n}(T(\mathbb{H}_n) - T(H)) \le x)$$

distinguishing clearly in your description between the "ideal bootstrap" and the Monte-Carlo implementation thereof.

Solution: (a) Let $\mathbb{H}_{n,i}$ denote the empirical distribution of the data with the ith pair (X_i, Y_i) omitted. Let $T_{n,i} \equiv T(\mathbb{H}_{n,i})$, $T_{n,\cdot} \equiv n^{-1}\sum_{i=1}^n T_{n,i}$ and let $T^*_{n,i} \equiv nT_n - (n-1)T_{n,i}$. Then the Jackknife estimator of $nVar_H(T_n)$ is

$$\frac{1}{n-1}\sum_{i=1}^n \{T^*_{n,i} - \overline{T}^*_n\}^2.$$

(b) The ideal bootstrap estimator of $nVar_H(T_n)$ is $nVar_{\mathbb{H}_n}(T_n)$. To implement this, we would draw B bootstrap samples

$$(X^*_{j1}, Y^*_{j1}), \ \ldots, \ (X^*_{jn}, Y^*_{jn}), \qquad j = 1, \ldots, B,$$

let $\mathbb{H}^*_{j,n}(x, y) \equiv n^{-1}\sum_{i=1}^n 1_{[X^*_{ji} \le x, \ Y^*_{ji} \le y]}$ be the empirical distribution function of the jth bootstrap sample, and compute $T^*_{j,n} \equiv T(\mathbb{H}^*_{j,n})$, $j = 1, \ldots, B$ Then the bootstrap estimator of $nVar_H(T_n)$ is just

$$n\frac{1}{B}\sum_{j=1}^B \{T^*_{j,n} - \overline{T^*_n}\}^2.$$

(c) Because $T(H)$ is a smooth functional of the marginal first and second moments, and the natural substitution estimators of these parameters are jointly asysmptotically normal under the hypotheses $EX^4 < \infty$, $EY^4 < \infty$, it is clear

that $\sqrt{n}(T(\mathbb{H}_n)-T(H)) \to_d N(0,V^2)$ for some V^2 by the delta method. Because $T(H)$ is a ratio of moments, and because the central limit theorem does a better job of approximating sums rather than moments, it seems likely that use of the logarithmic transformation $g(x)=\log x$ might be helpful: $\sqrt{n}(\log(T(\mathbb{H}_n)) - \log(T(H))$ will probably converge to normality faster than $\sqrt{n}(T(\mathbb{H}_n)-T(H))$.
(d) Because $T(H)$ is a smooth functional of moments, and because the bootstrap works for moments, it follows by preservation of bootstrap convergence under differentiable mappings that the bootstrap will "work" in this situation. Similarly, the jackknife will also estimate $nVar_H(T_n)$ consistently in this problem.
(e) Much as in (b), the ideal bootstrap estimator of $Q_n(x,H)$ is

$$Q_n(x,\mathbb{H}_n) \equiv Pr_{\mathbb{H}_n^*}(\sqrt{n}(T(\mathbb{H}_n^*)-T(\mathbb{H}_n)) \le x).$$

To implement this, we would draw B bootstrap samples

$$(X_{j1}^*,Y_{j1}^*),\ \ldots,\ (X_{jn}^*,Y_{jn}^*), \qquad j=1,\ldots,B,$$

let $\mathbb{H}_{j,n}^*(x,y) \equiv n^{-1}\sum_{i=1}^n 1_{[X_{ji}^*\le x,\ Y_{ji}^*\le y]}$ be the empirical distribution function of the jth bootstrap sample, and compute $T_{j,n}^* \equiv T(\mathbb{H}_{j,n}^*)$, $j=1,\ldots,B$. Then the (Monte-Carlo implementation of the) bootstrap estimator $Q_n(x,\mathbb{H}_n)$ of $Q_n(x,H)$ is just

$$\frac{1}{B}\sum_{j=1}^B 1_{\{\sqrt{n}(T_{j,n}^*-\overline{T_n^*})\le x\}}\ .$$

4. (40 points). Suppose that $\{S_n\}$ and $\{T_n\}$ are arbitrary sequences of statistics (with $E(S_n^2)<\infty$ and $E(T_n^2)<\infty$ for each n) satisfying

$$\frac{Var(S_n-T_n)}{Var(T_n)} \to 0. \tag{1}$$

(a) Show that (1) implies that $Var(S_n)/Var(T_n) \to 1$.
[Hint: Write $Var(S_n)=Var(T_n+(S_n-T_n))$, compute, and then use the Cauchy-Schwarz inequality.]
(b) Show that $2Cov(S_n,T_n) = Var(S_n)+Var(T_n)-Var(S_n-T_n)$.
(c) Let

$$R_n \equiv \frac{S_n-E(S_n)}{\sqrt{Var(S_n)}} - \frac{T_n-E(T_n)}{\sqrt{Var(T_n)}}.$$

Use the identity in (b) to show that

$$\begin{aligned} Var(R_n) &= 2-2\frac{Cov(S_n,T_n)}{\sqrt{Var(S_n)Var(T_n)}} \\ &= 2-\sqrt{\frac{Var(S_n)}{Var(T_n)}}-\sqrt{\frac{Var(T_n)}{Var(S_n)}}+\sqrt{\frac{Var(T_n)}{Var(S_n)}}\cdot\frac{Var(S_n-T_n)}{Var(T_n)}. \end{aligned}$$

(d) Use the result of (a) together with the identity you proved in (c) to show that (1) implies $Var(R_n) \to 0$.

(e) Does the conclusion of (d) imply $R_n \to_p 0$?

Solution:

(a) Now

$$\frac{Var(S_n)}{Var(T_n)} = 1 + \frac{Var(S_n - T_n) + 2Cov(S_n - T_n, T_n)}{Var(T_n)}$$

so

$$\begin{aligned}
\left|\frac{Var(S_n)}{Var(T_n)} - 1\right| &\leq \frac{Var(S_n - T_n)}{Var(T_n)} + 2\frac{|Cov(S_n - T_n, T_n)|}{Var(T_n)} \\
&\leq \frac{Var(S_n - T_n)}{Var(T_n)} + 2\frac{\sqrt{Var(S_n - T_n) \cdot Var(T_n)}}{Var(T_n)} \\
&= \frac{Var(S_n - T_n)}{Var(T_n)} + 2\sqrt{\frac{Var(S_n - T_n)}{Var(T_n)}} \\
&\to 0 + 0 = 0
\end{aligned}$$

if (1) holds.

(b) Now $Var(S_n - T_n) = Var(S_n) + Var(T_n) - 2Cov(S_n, T_n)$, so rearranging yields

$$2Cov(S_n, T_n) = Var(S_n) + Var(T_n) - Var(S_n - T_n)$$

as claimed.

(c) Note that

$$\begin{aligned}
Var(R_n) &= 2 - 2\frac{Cov(S_n, T_n)}{\sqrt{Var(S_n)Var(T_n)}} \\
&= 2 - \sqrt{\frac{Var(S_n)}{Var(T_n)}} - \sqrt{\frac{Var(T_n)}{Var(S_n)}} + \sqrt{\frac{Var(T_n)}{Var(S_n)}} \cdot \frac{Var(S_n - T_n)}{Var(T_n)}.
\end{aligned}$$

(d) If (1) holds, then we have $Var(S_n)/Var(T_n) \to 1$ by (a), and hence it follow from the identity in the the last display that

$$Var(R_n) \;\to\; 2 - 1 - 1 + 1 \cdot 0 = 0$$

using both (1) and (a) .

(e) Since $E(R_n) = 0$ and $Var(R_n) \to 0$ by (b), $R_n \to_p 0$ follows by Chebychev's inequality (or Markov's inequality).

Do **either** problem 5 **or** problem 6 (but **not both**).

5. (40 points). Let F be a bivariate d.f. with marginal d.f.'s F_1 and F_2 respectively. Let $a : [0,1]^2 \mapsto \mathbb{R}$ be differentiable, and define
$$T(F) = \int\int a(F_1(x), F_2(y))dF(x,y).$$
Consider $\sqrt{n}(T(\mathbb{F}_n) - T(F))$ where $\mathbb{F}_n$ is the (bivariate) empirical d.f. of $(X_1, Y_1), \ldots, (X_n, Y_n)$ i.i.d. F on $\mathbb{R}^2$.
(a) Find the influence function of $T(F)$.
(b) What asymptotic variance do you expect to be able to demonstrate for $\sqrt{n}(T(\mathbb{F}_n) - T(F))$ when $a(u,v) = u \cdot v$?
(c) Will the bootstrap work for estimation of $H_n(x, F) \equiv Pr_F(\sqrt{n}(T(\mathbb{F}_n) - T(F)) \le x)$ when a is differentiable and $\int\int a^2(F_1(x), F_2(y))dF(x,y) < \infty$?

Solution: Let $F_\epsilon = (1-\epsilon)F + \epsilon G$. Then
$$\begin{aligned}
\lim_{\epsilon \searrow 0} \frac{T(F_\epsilon) - T(F)}{\epsilon} &= \frac{\partial}{\partial \epsilon} T(F_\epsilon)\Big|_{\epsilon=0} \\
&= \int\int a(F_1(x), F_2(y))d(G-F)(x,y) \\
&\quad + \int\int \frac{\partial}{\partial u} a(F_1(x), F_2(y))(G_1(x) - F_1(x))dF(x,y) \\
&\quad + \int\int \frac{\partial}{\partial v} a(F_1(x), F_2(y))(G_2(y) - F_2(y))dF(x,y).
\end{aligned}$$
Taking $G = \delta_{x_0,y_0}$ gives the influence function $IC((x_0,y_0); T, F)$:
$$\begin{aligned}
IC((x_0,y_0); T, F) &= a(F_1(x_0), F_2(y_0)) - T(F) \\
&\quad + \int\int \frac{\partial}{\partial u} a(F_1(x), F_2(y))(1_{[x_0,\infty)}(x) - F_1(x))dF(x,y) \\
&\quad + \int\int \frac{\partial}{\partial v} a(F_1(x), F_2(y))(1_{[y_0,\infty)}(y) - F_2(y))dF(x,y) \\
&\equiv \psi_F(x_0, y_0).
\end{aligned}$$
(b) When $a(u,v) = uv$, we have $(\partial/\partial u)a(u,v) = v$, $(\partial/\partial v)a(u,v) = u$, so
$$\begin{aligned}
\psi_F(x_0,y_0) &= F_1(x_0)F_2(y_0) - T(F) \\
&\quad + \int\int F_2(y)(1_{[x_0,\infty)}(x) - F_1(x))dF(x,y) \\
&\quad + \int\int F_1(x)(1_{[y_0,\infty)}(y) - F_2(y))dF(x,y) \\
&\equiv b(x_0,y_0) - T(F) + b_1(x_0) + b_2(y_0),
\end{aligned}$$

where $b(x_0, y_0) \equiv F_1(x_0)F_2(y_0)$, and hence we would expect the asymptotic variance to be

$$V^2 = E_F\psi_F^2(X,Y) = Var_F(b(X,Y) + b_1(X) + b_2(Y)).$$

(c) If a is differentiable, then it can be shown that $T(F)$ is Hadamard differentiable with respect to the Kolmogorov metric on bivariate distribution functions, and hence then it follows from our theory of the bootstrap for differentiable functionals that the bootstrap of $H_n(x; F) \equiv Pr(\sqrt{n}(T(\mathbb{F}_n) - T(F)) \leq x)$ will work.

6. (40 points). Let $T(F) = \int_0^1 F^{-1}(u)k(u)du$ for some function k.
   (a) Show that $T(F) = 0$ for every F symmetric about 0 if k is symmetric about 1/2. (Note: F is symmetric about 0 if and only if $F^{-1}(u)$ is odd about 1/2, and k is symmetric about 1/2 if and only if k is even about 1/2.)
   (b) Show that $T(F)$ is location equivariant (i.e. $T(F(\cdot - a)) = a + T(F)$) if and only if $\int_0^1 k(u)du = 1$.
   (c) Find the Gateaux derivative of T? What does this lead you to conjecture about $\sqrt{n}(T_n - T(F))$ where $T_n \equiv T(\mathbb{F}_n)$?
   (d) Would the bootstrap work for estimating $nVar_F(T_n)$?

Solution: (a) When F is symmetric about 0 so that F^{-1} is odd about 1/2, we have

$$T(F) = \int_0^1 F^{-1}(u)k(u)du = \int_{-1/2}^{1/2} F^{-1}(1/2+s)k(1/2+s)ds = 0$$

since $F^{-1}(1/2+s)$ is odd and $k(1/2+s)$ is even (on $[-1/2, 1/2]$).
(b) If $\int_0^1 k(u)du = 1$, then since $F(\cdot - a))^{-1} = F^{-1}(\cdot) + a$, we have

$$\begin{aligned} T(F(\cdot - a)) &= \int_0^1 (a + F^{-1}(u))k(u)du \\ &= a\int_0^1 k(u)du + \int_0^1 F^{-1}(u)k(u)du \\ &= a + T(F); \end{aligned}$$

i.e. $T(F)$ is location equivariant. On the other hand, if T is location equivariant then equality in the last line above forces $\int_0^1 k(u)du = 1$.
(c) Recall that the Gateaux derivative of $Q_u(F) \equiv F^{-1}(u)$ is, assuming that F has positive density f at $F^{-1}(u)$,

$$\dot{Q}_u(F; G-F) = -\frac{-(G-F)(F^{-1}(u))}{f(F^{-1}(u))}.$$

Thus the Gateaux derivative of $T(F) = \int_0^1 F^{-1}(u)k(u)du$ should be, under some modest regularity condition,

$$\begin{aligned}
\lim_{\epsilon \searrow 0} \frac{T(F_\epsilon) - T(F)}{\epsilon} &= \frac{\partial}{\partial \epsilon} T(F_\epsilon)\Big|_{\epsilon=0} \\
&= \int_0^1 k(u) \frac{-(G-F)(F^{-1}(u))}{f(F^{-1}(u))} du.
\end{aligned}$$

(d) I would expect that the bootstrap would work for $nVar_F(T_n)$ if the weight function k is not too heavy near zero and one (say e.g. that it is bounded) and that F has some corresponding moments (say $E_F X^2 < \infty$).

Do **either** problem 7 **or** problem 8 (but **not both**).

7. (40 points). Let $V_{n,1}$ and $V_{n,2}$ be two $V-$statistics based on two kernels $h_1(x_1, x_2)$ and $h_2(x_1, x_2)$ of order 2. Let $V_1(F)$ and $V_2(F)$ denote the corresponding (population) $V-$functionals. Assume that $E_F h_j^2(X_1, X_2) < \infty$ for $j = 1, 2$.
   (a) Find the (marginal) asymptotic distributions of $\sqrt{n}(V_{n,1} - V_1(F))$ and $\sqrt{n}(V_{n,2} - V_2(F))$. Are these $V-$functionals asymptotically linear?
   (b) Find the joint asymptotic distribution of $(\sqrt{n}(V_{n,1} - V_1(F)), \sqrt{n}(V_{n,2} - V_2(F)))$.
   (c) How does your result in (b) change if the two kernels are $h_1(x_1, \ldots, x_{r_1})$ and $h_2(x_1, \ldots, x_{r_2})$ with possibly different orders r_1 and r_2?

Solution: (a) For a single symmetric kernel h of order 2 and corresponding functional $T(F) = \iint h(x,y)dF(x)dF(y)$ we calculated the influence function in class: with $F_\epsilon \equiv (1-\epsilon)F + \epsilon G$ we have

$$\begin{aligned}
T(F_\epsilon) &= T(F) + \epsilon \iint h(x,y)dF(x)d(G-F)(y) \\
&\quad + \epsilon \iint h(x,y)d(G-F)(x)dF(y) \\
&\quad + \epsilon^2 \iint h(x,y)d(G-F)(x)d(G-F)(y),
\end{aligned}$$

and hence the Gateaux derivative is given by

$$\begin{aligned}
\lim_{\epsilon \searrow 0} \frac{T(F_\epsilon - T(F)}{\epsilon} &= 2 \iint h(x,y)d(G-F)(x)dF(y) \\
&= 2\left(\int h(x,y)dF(y) - T(F)\right) \quad \text{when} \quad G = \delta_x.
\end{aligned}$$

In class we proceeded to show that the corresponding $U-$ statistic had a linear - representation via Hájek projection, and that the U and V statistics are asymptotically equivalent, both with the same linear representation. Applying this twice (to $V_{1,n}$ with kernel h_1 and to $V_{2,n}$ with kernel h_2) yields

$$\sqrt{n}(V_{1,n} - V_1(F)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_F^{(1)}(X_i) + o_p(1),$$
$$\sqrt{n}(V_{2,n} - V_2(F)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_F^{(2)}(X_i) + o_p(1),$$

where

$$\psi_F^{(1)}(x) = 2\left(\int h_1(x,y)dF(y) - V_1(F)\right),$$
$$\psi_F^{(2)}(x) = 2\left(\int h_2(x,y)dF(y) - V_2(F)\right).$$

Since $E_F\psi_F^{(j)}(X_1) = 0$ for $j = 1,2$ and the assumption $E_F h_j^2(X_1, X_2) < \infty$ implies $v_j^2 \equiv E_F\psi_F^{(j)}(X_1) < \infty$ for $j = 1,2$, the central limit theorem yields

$$\sqrt{n}(V_{1,n} - V_1(F)) \to_d N(0, v_1^2), \quad \text{and}$$
$$\sqrt{n}(V_{2,n} - V_2(F)) \to_d N(0, v_2^2).$$

(b) Joint asymptotic normality follows easily from the asymptotic linearity in (2) and (2): by the multivariate CLT,

$$\sqrt{n}\begin{pmatrix} V_{1,n} - V_1(F) \\ V_{2,n} - V_2(F) \end{pmatrix} \to_d N_2(0, \Sigma)$$

where

$$\Sigma \equiv \begin{pmatrix} v_1^2 & c \\ c & v_2^2 \end{pmatrix}$$

and $c \equiv E\left(\psi_F^{(1)}(X_1)\psi_F^{(2)}(X_1)\right)$.

(c) When the kernels h_1 and h_2 are of possibly different orders r_1 and r_2, the asymptotic linearity statement in (a) still holds, but with the influence functions there replaced by $\psi_F^{(j)}$ given by

$$\psi_F^{(1)}(x) = r_1\left(\int h_1(x, x_2, \ldots, x_{r_1})dF(x_2)dF(x_3)\cdots dF(x_{r_1}) - V_1(F)\right),$$
$$\psi_F^{(2)}(x) = r_2\left(\int h_2(x, x_2, \ldots, x_{r_2})dF(x_2)dF(x_3)\cdots dF(x_{r_2}) - V_2(F)\right).$$

Thus the joint asymptotic normality statement in (b) continues to hold, but with the variance-covariance matrix Σ given in terms of $\psi_F^{(j)}$, $j = 1,2$.

8. (40 points). Suppose that $T(F) = Cov_F(X,Y) = E_F(X - E(X))(Y - E(Y))$ where $(X,Y) \sim F$ on $\mathbb{R}^2$.
(a) Show that $T(F)$ can be written as a $V-$functional:

$$T(F) = \int\int h((x_1,y_1),(x_2,y_2))dF(x_1,y_1)dF(x_2,y_2)$$

for some kernel h.
(b) What are the corresponding $U-$ and $V-$functionals U_n and V_n yielding natural estimators of $T(F)$?
(c) Find the limiting distribution of the $U-$ estimator of $T(F)$.

Solution: (a) Let $(\tilde{X},\tilde{Y}) \sim F$ be independent of (X,Y). Then, since $E\tilde{X} = EX$ and $E\tilde{Y} = EY$,

$$\begin{aligned} E_F(X-\tilde{X})(Y-\tilde{Y}) &= E_F(X - EX - (\tilde{X} - E(\tilde{X})))(Y - EY - (\tilde{Y} - E(\tilde{Y}))) \\ &= E_F(XEX)(Y - EY) + 0 + 0 + E(\tilde{X} - E(\tilde{X}))(\tilde{Y} - E(\tilde{Y})) \\ &= Cov_F(X,Y) + Cov_F(\tilde{X},\tilde{Y}) = 2Cov_F(X,Y). \end{aligned}$$

Thus

$$Cov_F(X,Y) = E_F h((X,Y),(\tilde{X},\tilde{Y}))$$

where $h((x_1,y_1),(x_2,y_2)) \equiv \frac{1}{2}(x_1 - x_2)(y_1 - y_2)$.
(b) If $(X_1,Y_1),\ldots,(X_n,Y_n)$ are i.i.d. F with empirical distribution $\mathbb{P}_n \equiv n^{-1}\sum_{i=1}^n \delta_{(X_i,Y_i)}$, then natural $V-$functional is just

$$V_n = T(\mathbb{P}_n) = \int\int h((x_1,y_1),(x_2,y_2))d\mathbb{P}_n(x_1,y_1)d\mathbb{P}_n(x_2,y_2)$$

where h is as in (a) The natural unbiased estimator is

$$U_n \equiv \frac{1}{\binom{n}{2}} \sum_c h((X_{i_1},Y_{i_1}),(X_{i_2},Y_{i_2}))$$

with h also as in (a).
(c) For U_n as in (b) we have, by the asymptotic linearity of $U-$statistics proved in class,

$$\sqrt{n}(U_n - T(F)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_F(X_i,Y_i) + o_p(1)$$

where

$$\begin{aligned}
\psi_F(x,y) &= 2\left(\int h((x,y),(u,v))dF(u,v) - T(F)\right) \\
&= E_F(x-\tilde{X})(y-\tilde{Y}) - 2Cov_F(X,Y) \\
&= xy - yE_FX - xE_FY + E(XY) - 2Cov_F(X,Y) \\
&= (x-EX)(y-EY) - EX\cdot EY + E(XY) - 2Cov_F(X,Y) \\
&= (x-EX)(y-EY) - Cov_F(X,Y).
\end{aligned}$$

Thus, assuming that $E_F(X^2Y^2) < \infty$,

$$\sqrt{n}(U_n - T(F)) \to_d N(0, V_F^2)$$

where

$$V_F^2 = Var_F((X-EX)(Y-EY)).$$