Statistics 583, Final Exam Solutions

Wellner; 6/7/2006

1. (40 points) **Define** the following terms.
   (a) A Fréchet - differentiable functional $T : \mathcal{F} \to \mathbb{R}$ with respect to a metric d_* on $\mathcal{F}$.
   (b) A Hadamard - differentiable functional $T : \mathcal{F} \to \mathbb{R}$ with respect to a metric d_* on $\mathcal{F}$.
   (c) A metric d between distribution functions which is compatible with respect to the empirical distribution function.
   (d) The kernel estimator of a density function f on $\mathbb{R}$.

   Solution: See course notes and Wasserman.

2. (30 points). Give a complete *statement* of **two** of the following results or theorems:
   (a) An example of a functional $T(F)$ which is *not* weakly continuous.
   (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) Some version of Hoeffding's theorem about the distribution of the rank vector $\underline{R}$ under alternatives.
   (d) Any theorem about asymptotic normality of an estimator via differentiability of the corresponding statistical functional.

   Solution: See course notes and Wasserman.

3. (40 points). Suppose that $X_1, \ldots, X_m$ are i.i.d. with continuous d.f. F on $\mathbb{R}^+$ and $Y_1, \ldots, Y_n$ are i.i.d. with continuous d.f. G where G and F are related by $G = F^{\Delta}$ where $\Delta > 1$. Let $\underline{R} = (R_1, \ldots, R_n)$ denote the ranks of $\underline{Z} = (Z_1, \ldots, Z_N) = (X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ where $N = m + n$, and let $\underline{Q} = (Q_1, \ldots, Q_n)$ denote the ordered Y ranks.
   (a) If $m = 1$, $n = 2$, and $\Delta = 1$, calculate $P_{\Delta}(\underline{Q} = \underline{q})$ for all possible values of $\underline{q}$.
   (b) If $m = 1$, $n = 2$, and $\Delta = 2$, calculate $P_{\Delta}(\underline{Q} = \underline{q})$ for all possible values of $\underline{q}$.
   (c) What is the locally most powerful rank test S_N for testing $H: F = G$ versus $K: G = F^{\Delta}$, $\Delta > 1$ (for general m and n)?
   (d) If you observe $\underline{Q} = (3, 4)$, what is the p-value of the locally most powerful test in (c)?
   (e) What is the asymptotic distribution of S_N under the null hypothesis?

   Solution: (a) Under the hypothesis $\Delta = 1$ we have $G = F$ and all three possible values of $\underline{Q}$, namely $(1, 2), (1, 3), (2, 3)$, are equally likely with probability 1/3:

$$P_{\Delta=1}(\underline{Q} = \underline{q}) = 1/3, \qquad \underline{q} \in \{(1, 2), (1, 3), (2, 3)\}.$$

(b) Under the hypothesis $\Delta = 2$ we can calculate $P_{\Delta=2}(\underline{Q} = \underline{q})$ via Hoeffding's formula:

$$P_{\Delta=2}(\underline{Q} = \underline{q}) = \frac{1}{\binom{N}{n}} E \prod_{j=1}^{n} \psi'(U_{(q_j)}) = \frac{1}{\binom{3}{2}} E \prod_{j=1}^{2} \psi'(U_{(q_j)})$$

where $\psi(u) = G(F^{-1}(u)) = u^2$ and $(U_{(1)}, \ldots, U_{(3)})$ are the order statistics of $U_1, \ldots, U_3$ i.i.d. as Uniform$(0,1)$. Since $\psi'(u) = 2u$, this yields

$$P_{\Delta=2}(\underline{Q} = \underline{q}) = \frac{4}{3} E(U_{(q_1)} U_{(q_2)}) = \frac{4}{3} E(U_{(q_1)} U_{(q_2)}).$$

Thus we calculate

$$\begin{aligned} P_{\Delta=2}(\underline{Q} = (1,2)) &= (4/3)3! \int_0^1 \int_0^{u_3} \int_0^{u_2} u_1 u_2 du_1 du_2 du_3 \\ &= \frac{8}{2} \int_0^1 \int_0^{u_3} u_2^3 du_2 du_3 = \frac{8}{2 \cdot 4} \int_0^1 u_3^4 du_3 \\ &= \frac{8}{2 \cdot 4 \cdot 5} = \frac{8}{40} = \frac{1}{5} = \frac{3}{15}; \end{aligned}$$

$$\begin{aligned} P_{\Delta=2}(\underline{Q} = (1,3)) &= (4/3)3! \int_0^1 \int_0^{u_3} \int_0^{u_2} u_1 u_3 du_1 du_2 du_3 \\ &= \frac{8}{2} \int_0^1 \int_0^{u_3} u_2^2 u_3 du_2 du_3 = \frac{8}{2 \cdot 3} \int_0^1 u_3^4 du_3 \\ &= \frac{8}{2 \cdot 3 \cdot 5} = \frac{8}{30} = \frac{4}{15}; \end{aligned}$$

$$\begin{aligned} P_{\Delta=2}(\underline{Q} = (2,3)) &= (4/3)3! \int_0^1 \int_0^{u_3} \int_0^{u_2} u_2 u_3 du_1 du_2 du_3 = \frac{8}{1} \int_0^1 \int_0^{u_3} u_2^2 u_3 du_2 du_3 \\ &= \frac{8}{1 \cdot 3} \int_0^1 u_3^4 du_3 = \frac{8}{1 \cdot 3 \cdot 5} = \frac{8}{15}. \end{aligned}$$

Note that $3/15 + 4/15 + 8/15 = 15/15 = 1$.
(c) To find the locally most powerful rank test of $H: F = G$ versus $K: G = F^{\Delta}$, We start with Hoeffding's formula as in (b) for a general $\Delta > 0$:

$$\begin{aligned} P_{\Delta}(\underline{Q} = \underline{q}) &= \frac{1}{\binom{N}{n}} E \prod_{j=1}^{n} \psi'_{\Delta}(U_{(q_j)}) = \frac{1}{\binom{N}{n}} E\{\prod_{j=1}^{n} \Delta U_{(q_j)}^{\Delta-1}\} \\ &= \frac{\Delta^n}{\binom{N}{n}} E\{\prod_{j=1}^{n} U_{(q_j)}^{\Delta-1}\} \end{aligned}$$

Since we want to maximize the slope of the power function at $\Delta = 1$, we are lead to rejecting for $\underline{q}$ with large values of

$$\begin{aligned}\frac{d}{d\Delta} P_{\Delta}(\underline{Q} = \underline{q})\Big|_{\Delta=1} &= \left\{ \frac{n\Delta^{n-1}}{\binom{N}{n}} E\{\prod_{j=1}^{n} U_{(q_j)}^{\Delta-1}\} \right. \\ &\qquad \left. + \frac{\Delta^n}{\binom{N}{n}} \frac{\partial}{\partial\Delta} E\{\exp((\Delta-1)\sum_{j=1}^{n} \log(U_{(q_j)}))\} \right\} \Big|_{\Delta=1} \\ &= \frac{1}{\binom{N}{n}} \left\{ n + \sum_{j=1}^{n} E\{\log U_{(q_j)}\} \right\}.\end{aligned}$$

Now if $U \sim \text{Uniform}(0,1)$, then $\log U = -\log(1/U) \stackrel{d}{=} -Y$ where $Y \sim \text{exponential}(1)$. Thus, using the fact that the order statistics $Y_{(i)}$ of N i.i.d. exponential(1) random variables satisfy

$$Y_{(i)} \stackrel{d}{=} \sum_{j=1}^{i} \frac{V_j}{N-j+1}$$

where $V_1, \ldots, V_N$ are i.i.d. exponential (1), and hence (because of the $-$ sign), $\log U_{(i)} \stackrel{d}{=} -\log Y_{(N-i+1)}$, it follows that

$$\begin{aligned} a_N(i) &\equiv E\{\log U_{(i)}\} = -EY_{(N-i+1)} \\ &= -\sum_{j=1}^{N-i+1} \frac{1}{N-j+1} = -\sum_{k=i}^{N} \frac{1}{k} \end{aligned}$$

$i = 1, \ldots, N$. Thus the locally most powerful rank test rejects for large values of $S_N \equiv \sum_{j=1}^{n} a_N(Q_j)$. [For example, if $m = 1$ and $n = 2$, $a_N(i)$ is given by $a_3(3) = -1/3$, $a_3(2) = -(1/3) - (1/2) = -5/6$, $a_3(1) = -(1/3) - (1/2) - 1 = -11/6$; and the value of the statistic when $\underline{Q} = (2,3)$ is $S_3(\underline{Q}) = a_3(2) + a_3(3) = -5/6 - 1/3 = -7/6$, while $S_3((1,3)) = a_3(1) + a_3(3) = -11/6 - 1/3 = -13/3$, $S_3((1,2)) = a_3(1) + a_3(2) = -11/6 - 5/6 = -16/6 = -8/3$. Hence we reject at level $\alpha = 1/3$ if $\underline{Q} = (2,3)$.]

(d) From (c) it follows immediately that the p value for the locally most powerful test if $Q = (2,3)$ is observed is 1/3.

(e) Now under the null hypothesis $S_N = \sum_{j=1}^{N} a_N(Q_j) \stackrel{d}{=} \sum_{i=1}^{n} \tilde{Y}_i$ where $\tilde{Y}_1, \ldots, \tilde{Y}_n$ are the result of n draws without replacement from the urn containing the balls labeled $a_N(1), \ldots, a_N(N)$. Thus it follows from the WWNH finite sampling CLT that

$$\frac{S_N - ES_N}{\sqrt{Var(S_N)}} = \frac{\overline{\tilde{Y}}_n - \overline{a}_N}{\sigma_N} \to_d N(0,1)$$

where $\overline{a}_N \equiv N^{-1} \sum_{i=1}^N a_N(i)$ and

$$\sigma_N^2 = \left(1 - \frac{n-1}{N-1}\right) \frac{\sigma_a^2}{n}$$

where $\sigma_a^2 = N^{-1} \sum_{i=1}^N (a_N(i) - \overline{a}_N)^2$.

4. (40 points). Suppose that $X_1, \ldots, X_n$ are i.i.d. F with density function f having $p > 2$ continuous derivatives at a point x. Suppose that we use a "kernel estimator" $\widehat{f}_n(x)$ based on the bandwidth $h = h_n$ and kernel k *of order* p: i.e. k satisfies

$$\int k(z)dz = 1, \qquad \int zk(z)dz = 0, \ldots, \int z^{p-1}k(z)dz = 0,$$
$$\int |z|^p k(z)dz < \infty, \qquad \int k^2(z)dz < \infty.$$

Such a kernel cannot be a probability density function since the condition $\int z^2 k(z)dz = 0$ forces k to take negative values; thus such kernels are sometimes called "higher-order kernels". For example, $k(x) = 8^{-1}(9 - 15x^2)1_{[-1,1]}(x)$ is a kernel of order $p = 4$.

(a) Use the same method as in class and homework to show that the resulting estimator $\widehat{f}_n(x)$ has bias given by

$$E\{\widehat{f}_n(x)\} - f(x) = \frac{h_n^p}{p!}(-1)^p \int z^p k(z) f^{(p)}(z - h_n z^*)dz$$

(b) Use the same methods as in class and homework to show that $\widehat{f}_n(x)$ has variance

$$Var(\widehat{f}_n(x)) = \frac{f(x)}{nh_n} \int k^2(z)dz + o((nh_n)^{-1}).$$

(c) Combine (a) and (b) to find an asymptotic expression for $E(f(x) - \widehat{f}_n(x))^2$, and hence show that $\widehat{f}_n(x)$ achieves the optimal rate of convergence $n^{p/(2p+1)}$ with optimal bandwidth choice $h_{n,opt} = n^{-1/(2p+1)}$.

Solution: (a) As in the case developed in class,

$$\begin{aligned} E\{\widehat{f}_n(x)\} &= \frac{1}{h_h} \int k\left(\frac{x-y}{h_n}\right) f(y)dy \\ &= \int k(z) f(x - h_n z)dz. \end{aligned}$$

Now instead of expanding to second order as in class, we expand f to order p:

$$f(y) = f(x) + f'(x)(y-x) + \frac{1}{2!}f''(x)(y-x)^2 + \cdots + \frac{1}{p!}f^{(p)}(x^*)(y-x)^p$$

where y^* satisfies $|y^* - x| \le |y - x|$. With $y = x - h_n z$ so that $y - x = -h_n z$, this yields

$$\begin{aligned} E\{\widehat{f}_n(x)\} &= \int k(z) f(x - h_n z) dz \\ &= \int k(z) \left\{ f(x) - f'(x) h_n z + \frac{1}{2!} f''(x)(h_n z)^2 + \cdots \right. \\ &\qquad \left. + (-1)^p \frac{1}{p!} f^{(p)}(x - h_n z^*)(h_n z)^p \right\} dz \\ &= f(x) - 0 + 0 + \cdots + (-1)^p \frac{1}{p!} h_n^p \int z^p k(z) f^{(p)}(x - h_n z^*) dz \end{aligned}$$

where z^* satisfies $|(x - h_n z^*) - x| \le |(x - h_n z) - x| = h_n |z|$. Thus we deduce that

$$E\{\widehat{f}_n(x)\} - f(x) = (-1)^p \frac{1}{p!} h_n^p f^{(p)}(x) \int z^p k(z) dz + o(h_n^p)$$

(b) As in the case $p = 2$ developed in class,

$$\begin{aligned} Var(\widehat{f}_n(x)) &= \frac{1}{nh_n^2} \left\{ \int k\left(\frac{x-y}{h_n}\right)^2 f(y) dy - \left(\int k\left(\frac{x-y}{h_n}\right) f(y) dy \right)^2 \right\} \\ &= \frac{1}{nh_n} \int k^2(z) f(x - h_n z) dz - O(n^{-1}) \\ &= \frac{1}{nh_n} f(x) \left\{ \int k^2(z) dz + o(1) \right\} + O(n^{-1}). \end{aligned}$$

(a) Combining the results of (a) and (b) yields

$$\begin{aligned} E(f(x) - \widehat{f}_n(x))^2 &= \frac{1}{nh_n} f(x) \left\{ \int k^2(z) dz + o(1) \right\} + O(n^{-1}) \\ &\qquad + \frac{[f^{(p)}(x)]^2}{(p!)^2} h_n^{2p} \left(\int z^p k(z) dz \right)^2 + o(h_n^{2p}) \\ &\equiv A_x h_n^{2p} + \frac{B_x}{nh_n} + O(n^{-1}) + o(h_n^{2p}) \\ &\equiv R(h_n) + O(n^{-1}) + o(h_n^{2p}). \end{aligned}$$

Differentiating $R(h)$ with respect to h yields $R'(h) = A_x 2p h^{2p-1} - \frac{B_x}{n}\frac{1}{h^2}$, and solving for $h \equiv h_{n,opt}$ yields

$$h_{n,opt} = \left(\frac{B_x}{n2pA_x}\right)^{1/(2p+1)}.$$

The resulting approximation of the mean square error is

$$\begin{aligned} R(h_{n,opt}) &= A_x \left(\frac{B_x}{n2pA_x}\right)^{2p/(2p+1)} + \frac{B_x}{n\left(\frac{B_x}{n2pA_x}\right)^{1/(2p+1)}} \\ &= A_x^{1/(2p+1)} B_x^{2p/(2p+1)} \left\{(2p)^{-(2p)/(2p+1)} + (2p)^{1/(2p+1)}\right\} n^{-2p/(2p+1)}. \end{aligned}$$

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

5. (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, 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? What transformation g of T_n might lead to a better approximation using this asymptotic distribution?

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

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.

6. (40 points). In the context of testing for a disease, let $X \sim F$ denote the outcome of the test for a diseased individual and let $Y \sim G$ denote the outcome of the test for a non-diseased individual. Assuming that $X > x$ (or $Y > x$) leads to classifying the individual as "diseased", the *Receiver Operating Characteristic* or ROC curve R is a plot of *sensitivity* $\equiv P(X > x) = 1 - F(x) \equiv \overline{F}(x)$ versus $1-$*specificity*$\equiv 1 - P(Y \le x) = P(Y > x) = 1 - G(x) \equiv \overline{G}(x)$. Thus the ROC curve $R = R_{F,G}$ can be written as

$$R(t) = \overline{F}(\overline{G}^{-1}(t)) = 1 - F(G^{-1}(1-t)), \qquad 0 < t < 1.$$

(a) A good test for a disease has ROC curve with values close to 1 for small t and is everywhere above the line $I(t) = t$. Show that $R(t) \ge t$ for $0 \le t \le 1$ with strict inequality for some t if and only if $G <_s F$ (i.e. G is stochastically smaller than F).
(b) Consider $A \equiv \int_0^1 R(t)dt$ as a measure of the quality of the disease test (values close to 1 indicating an excellent test). Show that A can be expressed in terms of the Mann-Whitney-Wilcoxon functional $\int F dG$.

(c) Suppose that $X_1, \ldots, X_m$ are i.i.d. F and $Y_1, \ldots, Y_n$ are i.i.d. G, and consider estimation of the ROC curve $R \equiv R_{F,G}$ on the basis of the data.

(i) Propose a nonparametric estimator $\mathbb{R}_{m,n}(t)$ of $R(t) = R_{F,G}(t)$.

(ii) Give conditions on F and G which imply that your estimator in (i) is consistent for a fixed $t \in (0,1)$.

(iii) Compute the Gateaux derivatives of the functionals $T_1(F) \equiv R_{F,G}(t)$ for fixed G and t and $T_2(G) \equiv R_{F,G}(t)$ for fixed F and t.

(iv) Give conditions on F and G which imply that your estimator in (i) is asymptotically normal (for a fixed $t \in (0,1)$). Find the influence function of your estimator (with help from (iii)).

(v) What can you say about your estimator $\mathbb{R}_{m,n}$ as an estimator of the function R uniformly in $0 \le t \le 1$?

(vi) Explain how and why (or why not) you could use the jackknife or bootstrap to estimate the variance of the estimator in (i).

Solution: (a) Now $R(t) = 1 - F(G^{-1}(1-t)) \ge t$ for all $0 \le t \le 1$ if and only if $1 - t \ge F(G^{-1}(1-t))$ for all $0 \le t \le 1$, if and only if $u \ge F(G^{-1}(u))$ for all $0 \le u \le 1$, if and only if $F^{-1}(u) \ge G^{-1}(u)$ for all $0 \le u \le 1$, if and only if $F(x) \le G(x)$ for all $-\infty < x < \infty$, and the latter means $G \le_s F$. If strict inequality holds for some t in the first inequality ($R(t) > t$, then strict inequality holds in $F(x) < G(x)$ for some x, and hence $G <_s F$.

(b) Note that

$$
\begin{aligned}
A &\equiv \int_0^1 R(t)dt = 1 - \int_0^1 F(G^{-1}(1-t))dt \\
&= 1 - \int_0^1 F \circ G^{-1}(u)du \\
&= 1 - EF \circ G^{-1}(U) \qquad \text{where } \ U \sim \text{Uniform}(0,1), \\
&= 1 - EF(Y) \qquad \text{since } \ Y \equiv G^{-1}(U) \ \text{ has distribution } \ G \\
&= 1 - \int F dG = 1 - P(X \le Y) = P(Y < X).
\end{aligned}
$$

(c) (i) A natural estimator is $\mathbb{R}_{m,n}(t) \equiv R_{\mathbb{F}_m, \mathbb{G}_n}(t) = 1 - \mathbb{F}_m(\mathbb{G}_n(1-t))$.

(ii) Now $\mathbb{F}_m(x) \rightarrow_{a.s.} F(x)$ uniformly in x by the Glivenko-Cantelli theorem, and similarly, $\mathbb{G}_n(y) \rightarrow_{a.s.} G(y)$ uniformly in Y. Furthermore, $\mathbb{G}_n^{-1}(t) \rightarrow_{a.s.} G^{-1}(t)$ for any t for which G^{-1} is continuous at t. It follows that $\mathbb{R}_{m,n}(t) \rightarrow_{a.s.} R(t)$ for any t which is a point of continuity of G^{-1}.

(iii) The Gateaux derivatives of the functionals $T_1(F) = R_{F,G}(t)$ for fixed G and t can be easily computed as follows: Let F_1 be a fixed distribution function *not* $= F$, and set $F_\epsilon \equiv (1-\epsilon)F + \epsilon F_1$. Then

$$\begin{aligned} \frac{d}{d\epsilon}T_1(F_\epsilon)\Big|_{\epsilon=0} &= \frac{d}{d\epsilon}\{1 - F_\epsilon(G^{-1}(1-t))\}\Big|_{\epsilon=0} \\ &= -(F_1 - F)(G^{-1}(1-t)) \equiv \dot{T}_1(F; F_1 - F). \end{aligned}$$

Similarly, let G_1 be a fixed distribution function $\neq G$, and set $G_\epsilon \equiv (1-\epsilon)G + \epsilon G_1$. Then

$$\begin{aligned} \frac{d}{d\epsilon}T_2(G_\epsilon)\Big|_{\epsilon=0} &= \frac{d}{d\epsilon}\{1 - F(G_\epsilon^{-1}(1-t))\Big|_{\epsilon=0} \\ &= -f(G^{-1}(1-t))\frac{d}{d\epsilon}G_\epsilon^{-1}(1-t)\Big|_{\epsilon=0} \\ &= \frac{f(G^{-1}(1-t))}{g(G^{-1}(1-t))}\{G_1(G^{-1}(1-t)) - (1-t)\} \end{aligned}$$

because

$$\begin{aligned} 0 &= \frac{d}{d\epsilon}(1-t)\Big|_{\epsilon=0} = \frac{d}{d\epsilon}G_\epsilon(G_\epsilon^{-1}(1-t))\Big|_{\epsilon=0} \\ &= \frac{d}{d\epsilon}\{G(G_\epsilon^{-1}(1-t)) + \epsilon(G_1 - G)(G_\epsilon^{-1}(1-t))\}\Big|_{\epsilon=0} \\ &= g(G^{-1}(1-t))\frac{d}{d\epsilon}G_\epsilon^{-1}(1-t)\Big|_{\epsilon=0} + (G_1 - G)(G^{-1}(1-t)). \end{aligned}$$

(iv) Suppose that f and G are differentiable at $G^{-1}(1-t)$ with derivatives f and g respectively and suppose that $g(G^{-1}(1-t)) > 0$. Then, for the case $m = n$

$$\begin{aligned} &\sqrt{n}(\mathbb{R}_{n,n}(t) - R(t)) = -\sqrt{n}(\mathbb{F}_n(\mathbb{G}_n^{-1}(1-t)) - F(G^{-1}(t))) \\ &= -\sqrt{n}(\mathbb{F}_n(\mathbb{G}_n^{-1}(1-t)) - F(\mathbb{G}_n^{-1}(1-t))) - \sqrt{n}(F(\mathbb{G}_n^{-1}(1-t)) - F(G^{-1}(1-t))) \\ &= -\sqrt{n}(\mathbb{F}_n(G^{-1}(1-t)) - F(G^{-1}(1-t))) + o_p(1) \\ &\qquad -\frac{F(\mathbb{G}_n^{-1}(1-t)) - F(G^{-1}(1-t))}{\mathbb{G}_n^{-1}(1-t) - G^{-1}(1-t)}\sqrt{n}(\mathbb{G}_n^{-1}(1-t) - G(1-t)) \\ &= -\sqrt{n}(\mathbb{F}_n(G^{-1}(1-t)) - F(G^{-1}(1-t))) + o_p(1) \\ &\qquad -\frac{f}{g}(G^{-1}(1-t))\sqrt{n}(\mathbb{G}_n^{-1}(1-t) - G(1-t)) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\{-(1_{(-\infty,G^{-1}(1-t)]}(X_i) - F(G^{-1}(1-t))) \\ &\qquad + \frac{f}{g}(G^{-1}(1-t))(1_{(-\infty,G^{-1}(1-t)]}(Y_i) - (1-t))\} + o_p(1), \end{aligned}$$

so the influence function in this case is

$$\begin{aligned} \psi_{F,G}(x,y) &= -(1_{(-\infty,G^{-1}(1-t)]}(x) - F(G^{-1}(1-t))) \\ &\qquad + \frac{f}{g}(G^{-1}(1-t))(1_{(-\infty,G^{-1}(1-t)]}(y) - (1-t))\}. \end{aligned}$$

When $m \neq n$ it is natural to normalize by $\sqrt{mn/N}$ (as we did in the case of the Mann-Whitney-Wilcoxon statistic), and then, assuming that $\lambda_N \to \lambda$, the influence function for $\mathbb{R}_{m,n}(t)$ is given by the pair of functions

$$\begin{pmatrix} -\sqrt{1-\lambda}(1_{(-\infty, G^{-1}(1-t)]}(x) - F(G^{-1}(1-t))) \\ +\sqrt{\lambda}\frac{f}{g}(G^{-1}(1-t))(1_{(-\infty, G^{-1}(1-t)]}(y) - (1-t)) \end{pmatrix}.$$

(v) Concerning our estimator $\mathbb{R}_{m,n}$ as an estimator of the function R uniformly in $0 \le t \le 1$, first, note that

$$\begin{aligned}
\sup_{0\le t\le 1} |\mathbb{R}_{m,n}(t) - R(t)| &= \sup_{0\le t\le 1} |\mathbb{F}_m(\mathbb{G}_n^{-1}(1-t)) - F(G^{-1}(1-t)))| \\
&\le \sup_{0\le t\le 1} |\mathbb{F}_m(\mathbb{G}_n^{-1}(1-t)) - F(\mathbb{G}_n^{-1}(1-t)))| \\
&\qquad \sup_{0\le t\le 1} |F(\mathbb{G}_n^{-1}(1-t)) - F(G^{-1}(1-t))| \\
&\le \sup_{-\infty<x<\infty} |\mathbb{F}_m(x) - F(x)| \\
&\qquad + \sup_{0\le t\le 1} |F(G^{-1}(\Gamma_n^{-1}(1-t))) - F(G^{-1}(1-t))|
\end{aligned}$$

where, in the last line, we have replaced $\mathbb{G}_n^{-1}$ by something equal in distribution (jointly in n), namely $G^{-1}(\Gamma_n^{-1})$ where $\Gamma_n(t) = n^{-1}\sum_{i=1}^n 1_{[0,t]}(\xi_i)$, the empirical distribution function of n i.i.d. Uniform$(0,1)$ random variables. By the Glivenko-Cantelli theorem we have $\|\mathbb{F}_m - F\|_\infty \to_{a.s.} 0$, so the first term converges a.s. to 0. Also by the Glivenko-Cantelli theorem and symmetry about the identity, $\|\Gamma_n - I\|_\infty \to_{a.s.} 0$. Thus we see that the second term will converge a.s. to 0 if the function $F(G^{-1}(1-t))$ is uniformly continuous on $[0,1]$; i.e. if $F \circ G^{-1}$ is continuous on the closed interval $[0,1]$.
(vi) Concerning use of the jackknife or bootstrap to estimate the variance of the estimator in (i): because the jackknife fails for the median and other quantiles, and because $R_{F,G}$ involves $G^{-1}(1-t)$, it seems likely that the jackknife will fail for $R_{F,G}$. On the other hand, the bootstrap "works" for quantiles and for linear statistics (and differentiable functionals more generally), and hence it will work for $R_{F,G}(t)$.