**6/9/97** STATISTICS 583, FINAL EXAM **Wellner**

This exam is **closed book** and **closed notes**. Do **not** refer to the textbooks or any book or notes during the exam.

1. (24 points) **Define** the following terms, providing an appropriate context for your definition.
   (i) A *continuous functional* $T(F)$ with respect to the Kolmogorov metric $d_K$ on distribution functions.
   (ii) A *Gateaux - differentiable function(al)* $T(F)$.
   (iii) The *influence function* corresponding to a *Gateaux - differentiable functional* $T(F)$.
   (iv) The *Prohorov metric* between two probability measures $P$ and $Q$.

2. (24 points) **State three of the following four results**. In each case, provide an appropriate context for your statement.
   (i) Huber's "Master Theorem" for $M-$ estimators.
   (ii) Varadarajan's theorem concerning weak convergence of the empirical measure $\mathbb{P}_n$.
   (iii) An example of a functional $T(F)$ which is *not weakly continuous.*
   (iv) A limit theorem for the bootstrap empirical process $\sqrt{m}(\mathbb{F}_m^* - \mathbb{F}_n)$.

3. (36 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 $Var_G(Y) > 0$. 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 of the pairs.
A. Explain how you would use the jackknife to estimate $n\,Var_H(T_n)$.
B. Explain how you would use the bootstrap to estimate $n\,Var_H(T_n)$. 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?

4. (30 points) Suppose that $X_1, \cdots, X_m$ are i.i.d. exponential($\lambda$), and that $Y_1, \cdots, Y_n$ are i.i.d. exponential($\mu$); thus the density of $X_1$ is $\lambda e^{-\lambda x} 1_{[0,\infty)}(x)$. Consider testing $H : \lambda \le \mu$ versus $K : \lambda > \mu$.

A. Show that this testing problem is invariant with respect to the group of scale changes, $G$ given by $g_c(\underline{x}, \underline{y}) = (c\underline{x}, c\underline{y})$ where $c > 0$.

B. Find the UMP G-invariant test of $H$ versus $K$. [Hint: You may use the fact that the family of distributions $\{\delta^{-1} F_{r,s} : \delta > 0\}$ has monotone decreasing likelihood ratio.]

C. Specify exactly how you would carry out the test derived in B.

5. (46 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* $\equiv P(X > x) = \overline{F}(x)$ versus $1 -$ *specificity* $\equiv 1 - P(Y \le x) = P(Y > x) = \overline{G}(x)$. Thus the ROC curve $R = R_{F,G}$ can be written as

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

A. A good test for a disease has ROC curve with values close to 1 for small $t$ and is everywhere above 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 $G <_s F$.

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 $X_1, \cdots, X_m$ are i.i.d. $F$ and $Y_1, \cdots, Y_n$ 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 $\mathbb{R}_n$ as an estimator of the function $R$ uniformly in $0 \leq t \leq 1$?
(vi) Explain how and why (or why not) you could use the jackknife or bootstrap to estimate the variance of the estimator in (i). Be as detailed and explicit as possible.