Statistics 583, Final Exam

Wellner; 6/4/2007

Instructions: This is an "in class" and "closed-book" exam. Please do it completely on your own with no books or notes.

1. (40 points) **Define** the following terms.
   (a) A *continuous functional* $T(F)$ with respect to a metric d on distribution functions.
   (b) A *Gateaux - differentiable* functional $T(F)$.
   (c) The *influence function* corresponding to a Gateaux differentiable functional $T(F)$.
   (d) The kernel estimator of a density function f on $\mathbb{R}$.

2. (30 points). Give a complete *statement* of **two** of the following results or theorems:
   (a) An example of a functional $T(F)$ for which the jack-knife *does not* "work."
   (b) A limit theorem for the bootstrap of the sample mean $\overline{X}_n$.
   (c) The Politis-Romano without replacement bootstrap limit theorem.
   (d) Any theorem about asymptotic normality of an estimator via differentiability of the corresponding statistical functional.

3. (40 points). (a) Suppose that $Z_1, \ldots, Z_k$ are i.i.d. with distribution function F_0. Let $V_k \equiv \min_{1 \le j \le k} Z_j$, and suppose that K is a random integer with truncated Poisson distribution

$$P(K = k) = \frac{e^{-\theta}}{1 - e^{-\theta}} \frac{\theta^k}{k!}, \qquad k = 1, 2, \ldots, \quad \theta > 0.$$

Show that V_K has survival function given by

$$P(V_K > v) = \frac{\exp(-\theta F_0(v)) - \exp(-\theta)}{1 - \exp(-\theta)},$$

and hence

$$P(V_K \le v) = \frac{1 - e^{-\theta F_0(v)}}{1 - e^{-\theta}} \equiv \psi_\theta(F_0(v)).$$

(Note that this is valid distribution function for $\theta \in \mathbb{R}$, not just $\theta > 0$.)
(b) Now suppose that $X_1, \ldots, X_m$ are i.i.d. F and $Y_1, \ldots, Y_n$ are i.i.d. with distribution function $G = \psi_\theta(F)$. Find the locally most powerful rank test of

$\theta = 0$ (i.e. $F = G$) versus $\theta > 0$. Recall from calculations similar to those of Homework # 3,

$$\begin{aligned}
\psi'_\theta(u) &= \frac{\theta \exp(-\theta u)}{1 - \exp(-\theta)}, \qquad \text{and} \\
\frac{\partial}{\partial\theta}\psi'_\theta(u) &= \frac{e^{-\theta u}(1 - e^{-\theta} - \theta e^{-\theta})}{(1 - e^{-\theta})^2} - \frac{\theta}{1 - e^{-\theta}} u e^{-\theta u} \\
&\to \frac{1}{2} - u \qquad \text{as } \theta \to 0
\end{aligned}$$

by using L'Hopital's rule twice.

(c) How would you implement the test you found in (b) to achieve a type one error $\alpha = .01$?

4. (40 points). Let $X_1, \ldots, X_n$ be i.i.d. with unknown density function f. Consider the kernel estimator

$$\widehat{f}_n(x) = \int \frac{1}{h_n} K\left(\frac{x - y}{h_n}\right) d\mathbb{F}_n(y)$$

of an unknown density f at a point x with the "box-car" or uniform kernel $K(x) = 2^{-1} 1_{[-1,1]}(x)$. Assume that f' exists at x and is continuous in a neighborhood of x.

(a) Compute $E\widehat{f}_n(x)$ at x explicitly in terms of F and h after taking advantage of the given kernel K.

(b) Use the result of (a) and Taylor expansion to give an expression for $\text{bias}_n(x) = E\widehat{f}_n(x) - f(x)$ in terms of $f'(x)$ assuming that $h_n \to 0$.

(c) Give a formula for $Var(\widehat{f}_n(x))$ in terms of K, F, and h_n, and then use it to give an asymptotic expression for the variance assuming that $h_n \to 0$ and $nh_n \to \infty$.

(d) What is the optimal choice of $h_n = Cn^{-r}$ in this case?

(e) For the optimal choice of r in (d), sketch how you would prove that

$$\sqrt{nh_n}(\widehat{f}_n(x) - f(x)) \to_d N(b(x), \sigma^2(x))$$

and identify the functions $b(x)$ and $\sigma^2(x)$ that will appear here, including their dependence on C.

(f) For the limit distribution identified in (e), find the optimal choice of C (as a function of $f(x)$ and $f'(x)$ and other constants). Explain briefly why this optimal C makes some intuitive sense (as a function of x).

5. (48 points) Consider the functional $T(F) = \int\int |x-y| dF(x) dF(y)$ as a measure of spread or dispersion of the distribution function F. (This functional is sometimes called "Gini's mean difference".)

(a) If $X_1, \ldots, X_n$ are i.i.d. random variables with distribution function F, what is the "principle of substitution" estimator of $T(F)$?

(b) Is the estimator you found in (a) an unbiased estimator of $T(F)$? (Calculate the bias explicitly.)

(c) Use the jackknife to suggest an estimator of $T(F)$ with less bias. Can you find an unbiased estimator of $T(F)$?

(d) Calculate the Gateaux derivative of $T(F)$, and use this to find a formula for the asymptotic variance of $\sqrt{n}(T(\mathbb{F}_n) - T(F))$.

(e) Describe how you would use the bootstrap to estimate $nVar(T(\mathbb{F}_n))$ and

$$H_n(x, F) \equiv P_F(\sqrt{n}(T(\mathbb{F}_n) - T(F)) \leq x)\,,$$

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