Statistics 583, Midterm Exam

Wellner; 5/5/2006

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

1. (30 points) **Define** *any three* of the following terms.
   (a) A $G-$*invariant function* ϕ of the data (with respect to a group G of transformations g on the sample space $\mathcal{X}$).
   (b) A *maximal invariant* function T of the data (with respect to a group G of transformations g on the sample space $\mathcal{X}$).
   (c) A continuous functional $T : \mathcal{F} \to \mathbb{R}$ with respect to a metric d_* on $\mathcal{F}$.
   (d) A Gateaux - differentiable functional $T : \mathcal{F} \to \mathbb{R}$ with respect to a metric d_* on $\mathcal{F}$.

2. (30 points) Give a complete **statement** of *any one* of the following results:
   (a) A theorem about the existence of a UMP $G-$invariant test in the case that both the $G-$maximal invariant and the $\overline{G}-$maximal invariant are real-valued.
   (b) Hoeffding's theorem about the distribution of the rank vector $\underline{R}$ under the alternative hypothesis in the context of the two- sample problem of testing $H : F = G$ versus $K : F <_s G$ with $F, G \in \mathcal{F}_c$.
   (c) Any theorem about consistency of an estimator via continuity of statistical functionals.
   (d) Any theorem about asymptotic normality of an estimator via differentiability of the corresponding statistical functional.

   Do either problem 3 or problem 4.

3. (50 points) Suppose that $X_1, \ldots, X_m$ are i.i.d. $F \in \mathcal{F}_c$, the set of all continuous d.f.'s on R, and that $Y_1, \ldots, Y_n$ are i.i.d. G where, for some $\theta \in R$,

$$\frac{1 - G(x)}{G(x)} = e^{\theta} \frac{1 - F(x)}{F(x)}$$

   for all x. Consider testing $H : \theta = 0$ versus $K : \theta > 0$.
   (a) Under what group of transformations G is this testing problem invariant?
   (b) What is the maximal invariant $T(\underline{X}, \underline{Y})$ for the group G?
   (c) What is the $\overline{G}-$maximal invariant on the parameter space?
   (d) What does Hoeffding's formula say about the distribution of the maximal invariant under the alternative K?
   (e) Use (d) to find the locally most powerful rank test of H versus K. What is the name of this test statistic?

4. (50 points) Suppose that an urn contains N balls with the numbers $a_N(1), \ldots, a_N(N)$ written on the balls. Suppose that a sample of n balls is drawn from the urn without replacement: let the numbers on the sampled balls be $Y_1, \ldots, Y_n$, and let $\overline{Y}_n = n^{-1} \sum_{i=1}^n Y_i$.
   (a) What is the mean of $\overline{Y}_n$?
   (b) What is the variance of $\overline{Y}_n$?
   (c) If $\underline{R} = (R_1, \ldots, R_N)$ is a random permutation of $\{1, \ldots, N\}$, what is the relationship between $n\overline{Y}_n$ and $\sum_{j=1}^n a_N(R_j)$?
   (d) Under some condition on the numbers $a_N(i)$, a CLT holds for an appropriately standardized version of $\overline{Y}_n$, and hence also for $\sum_{j=1}^n a_N(R_j)$. State this condition and the theorem.
   (e) If $a_N(j) = j$, $j = 1, \ldots, N$, compute the mean and variance in (a) and (b), and make the conclusion of the CLT in (d) explicit. What is the name of the corresponding rank test and when is it a locally most powerful test based on the ranks?

5. (40 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) Show that the estimator you found in (a) is a biased estimator of $T(F)$ and calculate the bias explicitly.
   (c) Compute the influence function of $T(F)$.
   (d) Use the result of (c) to guess the asymptotic distribution of $\sqrt{n}(T(\mathbb{F}_n) - T(F))$.