Statistics 583, Final Exam

Wellner; 6/8/98

Instructions: This is a "closed-book" exam. You should do this exam without using any books or notes.

1. (32 points) **Define** *any four* of the following terms. In each case, provide an appropriate context for your definition.
   (a) An *unbiased test* ϕ.
   (b) A *uniformly most powerful unbiased (UMPU) test* ϕ.
   (c) A *Similar On the Boundary* (SOB) test ϕ.
   (d) A maximal invariant with respect to a group G.
   (e) An *invariant test* ϕ with respect to a group G.
   (f) A *continuous functional* $T(F)$ with respect to the Kolmogorov metric d_K on distribution functions.
   (g) A *Fréchet differentiable functional* $T(F)$ with respect to a metric d.

2. (32 points) **State** *any four* of the following results:
   (a) Varadarajan's theorem concerning weak convergence of the empirical measure $\mathbb{P}_n$.
   (b) An example of a functional $T(F)$ which is *not weakly continuous*.
   (c) The Wald-Wolfowitz-Noether-Hajek finite sampling central limit theorem.
   (d) Hoeffding's formula for the distribution of ranks (under the alternative).
   (e) A Central Limit Theorem for a functional $T(F)$ which is Fréchet - differentiable with respect to a metric which is compatible with the empirical distribution function (or empirical measure).
   (f) Any large sample theorem for Efron's nonparametric bootstrap.

3. (48 points) Suppose that $X \sim$ Binomial(m, p_1) and $Y \sim$ Binomial(n, p_2) are independent. Show how to find the UMP unbiased test of size $\alpha \in (0,1)$ of $H: \ p_1 \geq p_2$ versus $K: \ p_1 < p_2$. Describe how you would carry out the test, give a name for a key parameter in the relevant conditional distribution, name the conditional distribution obtained under the null hypothesis, and describe this distribution in terms of an urn model.

Do *either* problem 4 *or* problem 5:

4. (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?

5. (50 points) Suppose that an urn contains N balls with the numbers $z_N(1), \ldots, z_N(N)$ 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 $\sum_{j=}^n z_N(R_j)$?
(d) Under some condition on the numbers $z_N(i)$, a CLT holds for an appropriately standardized version of $\overline{Y}_n$. State this condition and the theorem.
(e) Does the condition of the theorem you stated in (c) hold if the numbers $z_N(1), \ldots, z_N(N)$ are in fact $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ where $N = m + n$ and $X_1, \ldots, X_m$ are i.i.d. F with $E_F X_1^2 < \infty$ and $Y_1, \ldots, Y_n$ are i.i.d. G with $E_G Y_1^2 < \infty$?
(f) Briefly describe the relevance of the CLT in (c) for a permutation test of the difference in means of two populations. Briefly describe the relevance of the CLT in (c) for two-sample linear rank statistics.

6. (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.