Statistics 583, Midterm Exam Solutions

Wellner; 5/7/2007

1. (30 points) **Define** *any three* of the following terms.
   (a) A maximal invariant with respect to a group G of transformations g on the sample space $\mathcal{X}$).
   (b) A $G-$*invariant test function* ϕ (with respect to a group G.
   (c) The Lévy metric d_L on the set of distribution functions $\mathcal{F}$ on $\mathbb{R}$.
   (d) The Prohorov metric d_{Pr} on the set of all probability measures on a metric space (M, d) with the Borel $\sigma-$field.
   (e) A Hadamard differentiable functional $T : \mathcal{F} \to \mathbb{R}$ with respect to the Kolmogorov metric $\|\cdot\|_\infty$.

   Soluton: See lecture notes, chapters 6 and 7.

2. (30 points) Give a complete **statement** of *any two* 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 formula for the distribution of ranks under the alternative.
   (c) The Wald- Wolfowitz-Noether-Hajek finite sampling central limit theorem.
   (d) Varadarajan's theorem concerning weak convergence of the empirical measure $\mathbb{P}_n$ when $X_1, \ldots, X_n$ are i.i.d. P on a metric space (M, d).

   Soluton: See lecture notes, chapters 6 and 7.

3. (40 points) Consider the collection of functionals $T(F) \equiv T_x(F) = P_F(x \in [X_1 \vee X_2, X_1 \wedge X_2))$ where $X_1, X_2 \sim F$ are independent.
   (a) For fixed $x \in \mathbb{R}$, express this functional explicitly in terms of F and show that $T_x(F) \le 1/2$ for all $x \in \mathbb{R}$. What value of x maximizes $T_x(F)$?
   (b) Under what conditions is $T_x(F)$ weakly continuous? Under what conditions is $T_x(F)$ continuous with respect to the Kolmogorov metric $d_K(F, G) = \|F - G\|_\infty$?
   (c) Find the Gateaux derivative of $T_x(F)$ for fixed x and thereby find the influence function $\psi_{x,F}(y)$ as a function of y for each fixed x. Be careful about confusing x and y here!
   (d) Using any method you want, show that $\sqrt{n}(T_x(\mathbb{F}_n) - T_x(F)) \to_d N(0, V_x^2)$ and find the asymptotic variance V_x^2 in terms of F and x.
   (e) What can you say about the stochastic process $R_n(x) = \sqrt{n}(T_x(\mathbb{F}_n) - T_x(F))$, $x \in \mathbb{R}$?

   Solution: (a) *First approach:* since $X_1 \vee X_2 = X_{(1)}$ and $X_1 \wedge X_2 = X_{(2)}$ are the order statistics of a sample of size $n = 2$ from F with joint density $2! f(y_1) f(y_2)$

on the set $-\infty < y_1 \le y_2 < \infty$ (assuming that F has density f), the probability in question is

$$\begin{aligned} P_F(X_{(1)} \le x < X_{(2)}) &= \int_{y_1 \in (-\infty, x]} \int_{y_2 \in (x, \infty)} 2f(y_1)f(y_2)dy_1dy_2 \\ &= 2F(x)(1 - F(x)). \end{aligned}$$

Thus $T_x(F) = 2F(x)(1 - F(x))$. This approach goes through if F does not have a density by simply writing the integral in the last display as

$$\int_{y_1 \in (-\infty, x]} \int_{y_2 \in (x, \infty)} 2dF(y_1)dF(y_2).$$

Second approach: Note that either $A = \{X_1 = X_1 \vee X_2, X_2 = X_1 \wedge X_2\}$ or $A^c = \{X_2 = X_1 \vee X_2, X_1 = X_1 \wedge X_2\}$ must occur. Thus

$$\begin{aligned} &P_F(X_1 \vee X_2 \le x < X_1 \wedge X_2) \\ &= P_F([X_1 \vee X_2 \le x < X_1 \wedge X_2] \cap A) + P_F([X_1 \vee X_2 \le x < X_1 \wedge X_2] \cap A^c) \\ &= P_F(X_1 \le x < X_2) + P_F(X_2 \le x < X_1) \\ &= F(x)(1 - F(x)) + F(x)(1 - F(x)) \qquad \text{by independence of } \ X_1,\ X_2 \\ &= 2F(x)(1 - F(x)). \end{aligned}$$

Third approach: Letting $A = \{X_1 \wedge X_2 \le x\}$ and $B = \{X_1 \vee X_2 > x\}$, the probability in question is, using

$$\begin{aligned} P(B) &= P(B \cap (A \cup A^c)) \\ &= P((B \cap A) \cup (B \cap A^c)) = P(B \cap A) + P(B \cap A^c), \end{aligned} \tag{1}$$

at the very first step,

$$\begin{aligned} P_F(A \cap B) &= P(B) - P(B \cap A^c) \quad \text{by (1)} \\ &= P(X_1 \vee X_2 > x) - P(X_1 \vee X_2 > x \text{ and } X_1 \wedge X_2 > x) \\ &= 1 - P(X_1 \vee X_2 \le x) - P(X_1 > x \text{ and } X_2 > x) \\ &= 1 - P(X_1 \le x, X_2 \le x) - (1 - F(x))^2 \\ &= 1 - F(x)^2 - (1 - F(x))^2 = 2F(x) - 2(F(x))^2 = 2F(x)(1 - F(x)). \end{aligned}$$

Since $F(x)(1 - F(x)) \le 1/4$ for all x, $T_x(F) = 2F(x)(1 - F(x)) \le 1/2$. Furthermore the maximum is achieved at x such that $F(x) = 1/2$; at $x = F^{-1}(1/2)$ (assuming that $1/2$ is in the range of F).

(b) The functional $T_x(F)$ is weakly-continuous at an distribution function F such that x is a continuity point of F: if $F_n \to_d F$, then $T_x(F_n) = 2F_n(x)(1 - F_n(x)) \to 2F(x)(1 - F(x))$ if $x \in C_F$. On the other hand, $T_x(F)$ is continuous

with respect to the Kolmogorov metric at every F: if $\|F_n - F\|_\infty \to 0$, then $T_x(F_n) = 2F_n(x)(1-F_n(x)) \to 2F(x)(1-F(x)) = T_x(F)$ without any further restrictions, and, indeed,

$$\begin{aligned}
\|T_x(F_n) - T_x(F)\|_\infty &\equiv \sup_x |T_x(F_n) - T_x(F)| = 2\sup_x |F_n(x)(1-F_n(x)) - F(x)(1-F(x))| \\
&\leq 2\sup_x |F_n(x) - F(x)| + 2\sup_x |F_n^2(x) - F^2(x)| \\
&= 2\sup_x |F_n(x) - F(x)| + 2\sup_x |(F_n(x) - F(x))(F_n(x) + F(x))| \\
&\leq 6\|F_n - F\|_\infty \to 0.
\end{aligned}$$

(c) The Gateaux derivative is easily calculated as follows: for $F_t = (1-t)F + tG$

$$\begin{aligned}
\frac{d}{dt} T_x(F_t)\Big|_{t=0} &= 2\frac{d}{dt}\left\{F_t(x)(1-F_t(x))\right\}\Big|_{t=0} \\
&= 2(1-2F(x))\frac{d}{dt}F_t(x)\Big|_{t=0} \\
&= 4(1/2 - F(x))(G(x) - F(x)).
\end{aligned}$$

Taking $G = \delta_y$ so that the distribution function $G(x) = 1_{[y\leq x]}$ gives the influence function

$$\psi_{x,F}(y) = 4(1/2 - F(x))(1_{[y\leq x]} - F(x)).$$

(d) The result of the calculation in (c) suggests that

$$\sqrt{n}(T_x(\mathbb{F}_n) - T_x(F)) \to_d N(0, E_F\psi_{x,F}^2(X))$$

where $E_F\psi_{x,F}^2(X) = 16(1/2-F(x))^2F(x)(1-F(x))$. This follows easily from the delta-method applied to $g(u) = 2u(1-u)$ since $g'(u) = 2(1-2u)$ and $\sqrt{n}(\mathbb{F}_n(x) - F(x)) \to_d Z_x \sim N(0, F(x)(1-F(x))$:

$$\begin{aligned}
\sqrt{n}(T_x(\mathbb{F}_n) - T_x(F)) &= \sqrt{n}(g(\mathbb{F}_n(x)) - g(F(x))) \\
&\to_d g'(F(x))Z_x \\
&= 4(1/2 - F(x))Z_x \\
&\sim N(0, 16(1/2 - F(x))F(x)(1-F(x))).
\end{aligned}$$

(e) To study $\mathbb{R}_n(x) \equiv \sqrt{n}(T_x(\mathbb{F}_n) - T_x(F))$ as a process in x, write

$$\begin{aligned}
\mathbb{R}_n(x) &\equiv \sqrt{n}(T_x(\mathbb{F}_n) - T_x(F)) = 2\sqrt{n}(\mathbb{F}_n(x)(1-\mathbb{F}_n(x)) - F(x)(1-F(x))) \\
&= 2\left\{\sqrt{n}(\mathbb{F}_n(x) - F(x)) - \sqrt{n}(\mathbb{F}_n^2(x) - F^2(x))\right\} \\
&= 2\left\{\sqrt{n}(\mathbb{F}_n(x) - F(x)) - (\mathbb{F}_n(x) + F(x))\sqrt{n}(\mathbb{F}_n(x) - F(x))\right\} \\
&= 2\left\{1 - (\mathbb{F}_n(x) + F(x))\right\}\sqrt{n}(\mathbb{F}_n(x) - F(x)) \\
&\stackrel{d}{=} 2\left\{1 - (\mathbb{G}_n(F(x)) + F(x))\right\}\mathbb{U}_n(F(x)) \\
&\Rightarrow 2(1-2F(x))\mathbb{U}(F(x)) \equiv \mathbb{R}(x)
\end{aligned}$$

where $\mathbb{G}_n(t) = n^{-1}\sum_1^n 1\{\xi_i \le t\}$ is the empirical distribution function of $\xi_1, \ldots, \xi_n$ i.i.d. Uniform$(0,1)$ random variables and $\mathbb{U}_n(t) = \sqrt{n}(\mathbb{G}_n(t) - t)$, by using the Glivenko-Cantelli and Donsker theorems. Note that the process $\mathbb{R}$ is a 0 mean Gaussian process with covariance function

$$Cov(\mathbb{R}(x), \mathbb{R}(y)) = 16(1/2 - F(x))(1/2 - F(y))(F(x) \vee F(y) - F(x)F(y).$$

In particular, $Var(\mathbb{R}(x)) = 16(1/2 - F(x))^2 F(x)(1 - F(x))$ in agreement with the result in (d). Note that this process is 0 with probability 1 at values of x such that $F(x) = 1/2$.

4. (36 points). Suppose that under the null hypothesis H_c, $X_1, \ldots, X_N$ are i.i.d. $F \in \mathcal{F}_c$, while under the alternative hypothesis K_1, the random variables $X_1, \ldots, X_N$ have joint density h given by

$$h(\underline{x}) = \prod_{i=1}^N \lambda_i \exp(-\lambda_i x_i) 1_{(0,\infty)}(x_i)$$

where $\lambda_i = \mu e^{\nu c_i}$ for some fixed, known numbers (covariates) $c_1, \ldots, c_N$, and constants $\mu > 0$ and $\nu \in R$.
(a) Find a most powerful similar test of size α for testing H_c versus K_1. Does your test depend on the values of the constants μ and/or ν?
(b) If $N = 3$, $\lambda_1 = 1$, $\lambda_2 = 2$, and $\lambda_3 = 3$, and we observe $(X_1, X_2, X_3) = (4, 2, 1)$, carry out the test in (a) at level $\alpha = 1/6$.
(c) Find a locally most powerful rank test of $H_0 : \nu = 0$ versus $H_1 : \nu > 0$. Does it matter that μ is unknown in both H_0 and H_1?

Solution: (a) To obtain a similar test of H_c, we condition on the order statistics $\underline{Z} \equiv (X_{(1)}, \ldots, X_{(N)})$ of the sample $X_1, \ldots, X_N$. A most powerful similar test of H_c versus K_1 rejects for those permutations $\underline{z}'$ of $\underline{Z} = \underline{z}$ which lead to large values of $h(\underline{z}')$, or equivalently for large values of

$$\log h(\underline{z}) = \sum_{i=1}^N \{\log \lambda_i - \lambda_i z_i\} = \sum_{i=1}^N \log \lambda_i - \sum_{i=1}^N \lambda_i z_i',$$

or, equivalently, for small values of

$$\sum_{i=1}^N \lambda_i z_i' = \mu \sum_{i=1}^N \exp(\nu c_i) z_i',$$

or, equivalently, for small values of

$$\sum_{i=1}^N \exp(\nu c_i) z_i'.$$

Thus we see that our test does not depend on the value of μ, but it does depend on the value of ν.
(b) When $N = 3$ and $\lambda_i = i$, $i = 1, 2, 3$, and we observe $(X_1, X_2, X_3) = (4, 2, 1)$, we find the following $3! = 6$ values of the test statistic in part (a):

$\underline{z}'/i$	1	2	3	$\sum_1^3 \lambda_i z_i'$
1	1	2	4	17
2	2	1	4	16
3	1	4	2	15
4	2	4	1	13
5	4	1	2	12
6	4	2	1	11

Since the observed value of the test statistic, 11, is the most extreme in the permutation distribution, we would reject H_c at level $\alpha = 1/6$.
(c) To find a locally most powerful rank test in this situation, it is helpful to first consider the density of the observations under the null hypothesis and under the alternatives of interest. In the current case these are

$$h_{\mu,0}(\underline{x}) = \prod_{i=1}^{N} \mu \exp(-\mu x_i) = \mu^N \exp(-\mu \sum_{i=1}^{N} x_i), \quad \text{and,}$$
$$h_{\mu,\nu}(\underline{x}) = \prod_{i=1}^{N} \lambda_i \exp(-\lambda_i x_i) = \mu^N \exp\left(\nu \sum_{1}^{N} c_i - \mu \sum_{1}^{N} e^{\nu c_i} x_i\right)$$

respectively. Thus

$$\frac{h_{\mu,\nu}(\underline{x})}{h_{\mu,0}(\underline{x})} = \exp\left(\nu \sum_{1}^{N} c_i - \mu \sum_{1}^{N} (e^{\nu c_i} - 1) x_i\right).$$

Thus Hoeffding's formula for the distribution of the rank vector $R = (R_1, \ldots, R_N)$ under the alternative hypothesis is

$$P_\nu(R = r) = \frac{1}{N!} E_F \exp\left(\nu \sum_{1}^{N} c_i - \mu \sum_{1}^{N} (e^{\nu c_i} - 1) V_{(r_i)}\right)$$

where $V_{(1)} \le V_{(2)} \le \cdots \le V_{(N)}$ are the order statistics of a sample of size N from $F =$ exponential(μ). To find the locally most powerful rank test of H_0 versus $H_1 : \nu > 0$, we differentiate this probability with respect to ν, evaluate the derivative at $\nu = 0$, and reject for rank vectors $R = r$ with large values of

the resulting statistic. Thus we calculate

$$\begin{aligned}
\frac{\partial}{\partial\nu}P_\nu(R=r)\Big|_{\nu=0} &= \frac{1}{N!}E_F\left\{\frac{\partial}{\partial\nu}\exp\left(\nu\sum_1^N c_i-\mu\sum_1^N(e^{\nu c_i}-1)V_{(r_i)}\right)\Big|_{\nu=0}\right\}\\
&= \frac{1}{N!}E_F\left\{\sum_{i=1}^N c_i-\mu\sum_{i=1}^N c_iE_F(V_{(r_i)})\right\}\\
&= \frac{1}{N!}\sum_{i=1}^N c_i\left(1-\mu E_F(V_{(r_i)})\right) \qquad (2)
\end{aligned}$$

where we note that if $V_1,\ldots,V_N$ are i.i.d. exponential (μ), then $(\mu V_1,\ldots,\mu V_N)\equiv(\tilde V_1,\ldots,\tilde V_N)$ are i.i.d. exponential (1), so that $\mu(V_{(1)},\ldots,V_{(N)})\stackrel{d}{=}(\tilde V_{(1)},\ldots,\tilde V_{(N)})$ are the order statistics of i.i.d exponential(1) random variables. Thus we see that the scores $a_N(i)\equiv 1-\mu E_F(V_{(i)})$ involved in the last display are just the Savage or "log-rank" scores derived in the context of the two-sample exponential scale problem, the two-sample proportional hazards problem, and problem 1 of problem set 4 (where the present ν was called θ, the present c_i's were called z_i's, and we calculated

$$a_N(i)=1-E(\tilde V_{(i)})=1-\sum_{j=1}^i\frac{1}{N-j+1}. \qquad (3)$$

Thus the locally most powerful rank test is "reject H_0 in favor of $H_1:\nu>0$ when $S_N=S_N(R)=\sum_{i=1}^N c_ia_N(R_i)>c_\alpha$ where the $a_N(i)$'s are as in (3).

(d) Beyond the exam: Another possibility here, to get rid of the dependence of the test in (a) on ν, is to look for a test which is a locally most powerful similar test of H_c versus $K:\nu>0$. To find such a test, write $h=h_\nu$ to emphasize the dependence of h on ν. Recall that the conditional distribution of permutations $\underline{z}'$ of $\underline{z}$ under alternatives is

$$p_\nu(\underline{z}';\underline{z})=h_\nu(\underline{z}')/\sum_{\underline{z}''}h(\underline{z}'')\,.$$

Now the conditional power of a test ϕ as a function of ν is

$$\begin{aligned}
\beta_\phi(\nu) &= E_\nu\left\{\phi(\underline{X})\big|\underline{Z}=\underline{z}\right\}\\
&= \sum_{z'}\phi(\underline{z}')p_\nu(\underline{z}')\,.
\end{aligned}$$

Thus the slope of the power function at $\nu=0$ is given by

$$\frac{d}{d\nu}\beta_\phi(\nu)\Big|_{\nu=0} = \sum_{z'}\phi(\underline{z}')\frac{\partial}{\partial\nu}p_\nu(\underline{z}')\Big|_{\nu=0}\,,$$

and by the generalized NP lemma, this leads us to reject H_c for those permutations yielding large values of

$$\frac{\frac{\partial}{\partial\nu}p_\nu(\underline{z}')}{p_\nu(\underline{z}')}\Big|_{\nu=0}$$

Thus we calculate

$$\begin{aligned}
\frac{\frac{\partial}{\partial\nu}p_\nu(\underline{z}')}{p_\nu(\underline{z}')} &= \frac{\partial}{\partial\nu}\log p_\nu(\underline{z}') \\
&= \frac{\partial}{\partial\nu}\left\{\log h_\nu(\underline{z}') - \log\sum_{z''} h_\nu(\underline{z}'')\right\} \\
&= \frac{\partial}{\partial\nu}\left\{\sum_1^N (\log\lambda_i - \lambda_i z_i') - C_\nu(\underline{z})\right\} \\
&= \sum_{i=1}^N \{c_i - c_i\lambda_i z_i'\} - \frac{\partial}{\partial\nu}C_\nu(\underline{z})\,.
\end{aligned}$$

When this is evaluated at $\nu = 0$, we find that

$$\frac{\frac{\partial}{\partial\nu}p_\nu(\underline{z}')}{p_\nu(\underline{z}')}\Big|_{\nu=0} = \sum_{i=1}^N \{c_i - \mu c_i z_i'\} + C_0'(\underline{z})\,,$$

and hence the locally most powerful similar test rejects H for small values of $\sum_1^N c_i z_i'$, which does not involve knowledge of ν, in contrast to the most powerful test in (a) above – which did rely on the value of ν fixed in the specification of a fixed alternative.