Statistics 583, Midterm Exam Solutions

Wellner; 5/6/2009

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) A metric d_* compatible with the empirical distribution function or the empirical measure.
   (e) A Fréchet-differentiable functional $T : \mathcal{F} \to \mathbb{R}$ with respect to a metric d_*.

   Solution: See course notes, Chapters 6 and 7.

2. (30 points) Give a complete **statement** of *any two* of the following results:
   (a) Hoeffding's formula for the distribution of ranks under the alternative.
   (b) The Wald- Wolfowitz-Noether-Hájek finite sampling central limit theorem.
   (c) 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).
   (d) 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.

   Solution: See course notes, Chapters 6 and 7.

   Do either problem 3 or problem 4.

3. (36 points) Suppose that $X_1, \ldots, X_m$ are i.i.d. exponential(λ) and that $Y_1, \ldots, Y_n$ are i.i.d. exponential(μ). Thus the density of X_1 is $\lambda \exp(-\lambda x) 1_{[0,\infty)}(x)$. 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 likelihood ratio.
   (c) Specify as exactly as possible how you would carry out the test derived in (b).

   Solution: (a) If $X \sim$ exponential(λ), then

$$\begin{aligned} P_\lambda(cX > t) &= P_\lambda(X > t/c) = \exp(-\lambda t/c) \\ &= \exp(-(\lambda/c)t) = P_{\lambda/c}(X > t), \end{aligned}$$

and similarly for $Y \sim$ exponential(μ). Hence the induced group on the parameter space is $\overline{g}(\lambda, \mu) = (\lambda/c, \mu/c)$. Note that for any $\overline{g} \in \overline{G}$ we have $\overline{g}\Theta_0 = \{(\lambda/c, \mu/c) :$

$\lambda \le \mu\} = \{(\lambda, \mu) : \lambda \le \mu\} = \Theta_0$, and $\overline{g}\Theta = \{(\lambda/c, \mu/c) : (\lambda, \mu) \in R^+ \times R^+\} = \Theta$. Hence the testing problem is invariant under the group G.
(b) By sufficiency we may reduce to consideration of $(S, T) = (\sum_{i=1}^m X_i, \sum_{j=1}^n Y_j)$. The induced group G^* on the space of the sufficent statistic is given by $G^* = \{g^*(s,t) = (cs, ct) : c > 0\}$, and the maximal invariant for the group G^* is $V = S/T$; the corresponding $\overline{G}-$maximal invariant is $\delta = \lambda/\mu$. Now $2\lambda X_i \sim \chi_2^2$, and similarly $2\mu Y_j \sim \chi_2^2$. Hence $2\lambda S \sim \chi_{2m}^2$ and $2\mu T \sim \chi_{2n}^2$. Hence

$$\frac{n}{m}V = \frac{\mu}{\lambda} \cdot \frac{2\lambda S/2m}{2\mu T/2n} = \delta^{-1} F_{2m,2n}$$

where $F_{2m,2n}$ has an $F-$distribution with degrees of freedom $2m, 2n$. Since the family $\delta^{-1}F_{r,s}$ has monotone decreasing monotone likelihood ratio, we conclude that the UMP G-invariant test of H versus K is given by "reject H if $nV/m < F_{2m,2n,\alpha}$ where $P(F_{2m,2n} \le F_{2m,2n,\alpha}) = \alpha$. (Alternatively, "reject H if $m/(nV) = (n^{-1}T/m^{-1}S) > F_{2n,2m,1-\alpha}$" where $P(F_{2n,2m} \ge F_{2n,2m,1-\alpha}) = \alpha$.
(c) See above.

4. (36 points) Suppose that X_1 has continuous distribution function F and Y_1, Y_2 are independent of X_1 and themselves independent with distribution function $G = F^2$. Let $\underline{Q} = (Q_1, Q_2)$ denote the ordered Y ranks.
(a) Is $F <_s G$? (Explain why or why not.)
(b) Compute the probabilities $P_{F,G}(\underline{Q} = \underline{q})$ for $\underline{q} \in \{(1,2), (1,3), (2,3)\}$ under the alternative $G = F^2$.
(c) Use (b) to find the most powerful rank test of $F = G$ versus $G = F^2$ at level $\alpha = 1/3$.

Solution: (a) Here $F <_s G$ holds since $G(x) = F(x)^2 < F(x)$ for all x such that $0 < F(x) < 1$.
(b) Now $G = \psi(F)$ with $\psi(u) = u^2$. Hence $\psi'(u) = 2u$, and by Hoeffding's formula

$$P(\underline{Q} = \underline{q}) \;=\; \frac{1}{\binom{3}{2}} E \prod_{j=1}^{2} \psi'(U_{(q_j)}) = \frac{1}{3}E[2U_{(q_1)} 2U_{(q_2)}] = \frac{4}{3}E[U_{(q_1)}U_{(q_2)}]$$

where $(U_{(1)}, U_{(2)}, U_{(3)})$ are the order statistics of a sample of 3 Uniform$(0,1)$ random variables. Thus we compute

$$\begin{aligned} E(U_{(1)}U_{(2)}) &= 3! \iiint_{0 \le u_1 \le u_2 \le u_3 \le 1} u_1 u_2 du_1 du_2 du_3 \\ &= 3! \frac{1}{2} \int_0^1 \left(\int_0^{u_3} u_2^3 du_2 \right) du_3 \\ &= \frac{3!}{8} \int_0^1 u_3^4 du_3 = \frac{3}{20}. \end{aligned}$$

Hence $P(\underline{Q} = (1,2)) = 4/20 = 1/5 = 3/15$. Similarly,

$$\begin{aligned} E(U_{(1)}U_{(3)}) &= 3! \iiint_{0 \le u_1 \le u_2 \le u_3 \le 1} u_1 u_3 du_1 du_2 du_3 \\ &= \frac{3!}{2} \int_0^1 \left(\int_0^{u_3} u_2^2 du_2 \right) u_3 du_3 \\ &= \int_0^1 u_3^4 du_3 = \frac{1}{5}. \end{aligned}$$

Hence $P(\underline{Q} = (1,3)) = 4/15$. Finally, for $\underline{q} = (2,3)$,

$$\begin{aligned} E(U_{(2)}U_{(3)}) &= 3! \iiint_{0 \le u_1 \le u_2 \le u_3 \le 1} u_2 u_3 du_1 du_2 du_3 \\ &= 3! \int_0^1 \left(\int_0^{u_3} u_2^2 du_2 \right) u_3 du_3 \\ &= 2 \int_0^1 u_3^4 du_3 = \frac{2}{5}, \end{aligned}$$

and this yields $P(\underline{Q} = (2,3)) = 8/15$. Note that

$$P(\underline{Q} = (1,2)) + P(\underline{Q} = (1,3)) + P(\underline{Q} = (2,3)) = \frac{3}{15} + \frac{4}{15} + \frac{8}{15} = 1.$$

Alternatively,

$$\begin{aligned} P(\underline{Q} = (1,2)) &= P(Y_1 \le X_1, Y_2 \le X_1) = \int G^2 dF \\ &= \int F^4 dF = \int_0^1 u^5 du = 1/5 = 3/15, \\ P(\underline{Q} = (2,3)) &= P(X_1 \le Y_1, X_1 \le Y_2) = \int (1-G)^2 dF \\ &= \int (1-F^2)^2 dF = \int_0^1 (1-u^2)^2 du = 8/15, \\ P(\underline{Q} = (1,3)) &= 2P(Y_1 \le X_1 \le Y_2) = 2 \int G(1-G) dF \\ &= 2 \int F^2(1-F^2) dF = 2 \int_0^1 u^2(1-u^2) du = 4/15. \end{aligned}$$

(c) Since $P(\underline{Q} = (2/3)) = 8/15 > 4/15 = P(\underline{Q} = (1/3)) > P(\underline{Q} = (1,2)) = 3/15$, we conclude from the Neyman-Pearson lemma that the most powerful rank test of $F = G$ versus $G = F^2$ at level $\alpha = 1/3$ is given by "reject H if $\underline{Q} = (2,3)$".

Do either problem 5 or problem 6.

5. (36 points) (a) State the general linear model in its "canonical form", specifying the parameter space Θ of the general model, and the canonical null hypothesis Θ_0.
(b) Under what group $G = G_1 \oplus G_2 \oplus G_3 \oplus G_4$ of transformations on the sample space are both Θ and Θ_0 invariant under the induced group $\overline{G}$ on the parameter space?
(c) What is the usual form of the general linear model? Explain the relationship of the "usual form" to the "canonical form", including a transformation which relates them explicitly.

Solution: See Chapter 6 course notes, pages 25-28.

6. (36 points) Suppose that F is a distribution function on $(0, \infty)$ given in terms of a distribution function H on $(0, \infty)$ by

$$F(x) = \frac{1}{\mu} \int_0^x y dH(y)$$

where $\mu \equiv \int_0^\infty y dH(y) < \infty$. Then F is the *length-biased distribution* corresponding to H. Suppose also that $\int_0^\infty y^{-1} dH(y) < \infty$
(a) Show that the distribution function H can be expressed in terms of F by

$$H(x) \equiv H_F(x) \equiv \frac{\int_0^x y^{-1} dF(y)}{\int_0^\infty y^{-1} dF(y)}.$$

(b) Fix $x_0 \in (0, \infty)$, and let $T(F) \equiv H_F(x_0)$ where H_F is as in (a). Find the Gateaux derivative of $T(F)$ at F and the influence curve $IC(x; T, F)$.
(c) Use the calculation of (b) to "guess" the asymptotic variance of the limiting distribution of $\sqrt{n}(T(\mathbb{F}_n) - T(F))$.
(d) How would you proceed to prove the result suggested by (b) and (c)?

Solution: (a) Since $F(x) = \mu^{-1} \int_0^x y dH(y)$, we have $\Delta F(0) = 0$ since $\Delta H(0) = 0$, and hence $dF(x) = \mu^{-1} x dH(x)$ implies

$$dH(x) = \mu \frac{1}{x} dF(x), \qquad \text{or} \qquad H(x) = \mu \int_0^x \frac{1}{y} F(y).$$

Since $1 = H(\infty) = \mu \int_0^\infty y^{-1} dF(y)$, we deduce that

$$\mu^{-1} = \int_0^\infty y^{-1} dF(y),$$

and hence that

$$H(x) = \frac{\int_0^x y^{-1} dF(y)}{\int_0^\infty y^{-1} dF(y)} \equiv H_F(y).$$

(b) Let $F_t = (1-t)F + tG$. Then

$$\begin{aligned} T(F_t) &= H_{F_t}(x_0) = \frac{\int_0^{x_0} y^{-1} dF_t(y)}{\int_0^{\infty} y^{-1} dF_t(y)} \\ &= \frac{\int_0^{x_0} y^{-1} dF(y) + t\int_0^{x_0} y^{-1} d(G-F)(y)}{\int_0^{\infty} y^{-1} dF(y) + t\int_0^{x_0} y^{-1} d(G-F)(y)}. \end{aligned}$$

Hence

$$\begin{aligned} \dot{T}(F, G-F) &= \frac{d}{dt} T(F_t)\big|_{t=0} \\ &= \frac{\int_0^{x_0} y^{-1} d(G-F)(y)}{\int_0^{\infty} y^{-1} dF(y)} \\ &\quad + \frac{\int_0^{x_0} y^{-1} dF(y)}{\left(\int_0^{\infty} y^{-1} dF(y)\right)^2} (-1) \int_0^{\infty} y^{-1} d(G-F)(y) \\ &= \mu \int_0^{x_0} y^{-1} d(G-F)(y) - \mu H(x_0) \int_0^{\infty} y^{-1} d(G-F)(y) \\ &= \mu \int_0^{\infty} \frac{1}{y} (1_{[0,x_0]}(y) - H(x_0)) d(G-F)(y) \\ &= \mu \int_0^{\infty} \frac{1}{y} (1_{[0,x_0]}(y) - H(x_0)) dG(y). \end{aligned}$$

Thus when $G = \delta_x$, we find that

$$IC(x; T, F) = \mu \frac{1}{x} (1_{[0,x_0]}(x) - H(x_0)).$$

It follows that

$$\begin{aligned} V^2 &= V^2(T,F) = E(IC^2(X;T,F)) \\ &= \mu^2 \int_0^{\infty} \frac{1}{y^2} (1_{[0,x_0]}(y) - H(x_0))^2 dF(y) \\ &= \mu \int_0^{\infty} \frac{1}{y} (1_{[0,x_0]}(y) - H(x_0))^2 dH(y), \end{aligned}$$

and we expect that with $T(\mathbb{F}_n) \equiv G_{\mathbb{F}_n}(x_0)$ we will have

$$\sqrt{n}(T(\mathbb{F}_n) - T(F)) \to_d N(0, V^2(T,F)).$$

(d) We could proceed to show that $H_F(x_0)$ is Fréchet or Hadamard differentiable with respect to the Kolmogorov metric or some other metric that is "compatible" with the empirical distribution function or empirical measure. Then the asymptotic normality indicated in (c) will hold via one of the (functional) delta-method theorems 7.4.2, 7.4.3, or 7.4.4