Statistics 583, Problem Set 2

Wellner; 4/4/2007

Reading: Chapter 6, sections 6.3 and 6.4; Lehmann and Romano, TSH, Chapters 6 and 7. See also Ferguson, MS, Chapter 5, sections 5.6 and 5.7;
Due: Wednesday, April 11, 2006

1. (Problem 10, page 249, Ferguson, MS) Let $\Theta = \{(\Delta, \pi_1, \ldots, \pi_n) : \ \Delta \ge 0, \ \pi = (\pi_1, \ldots, \pi_n)$ is a permutation of $\{1, \ldots, n\}\}$, and let the distribution of $X_1, \ldots, X_n$ given $\theta = (\Delta, \pi_1, \ldots, \pi_n)$ be as independent random variables with gamma distributions, $X_i \sim \text{Gamma}(\alpha, \beta^{-1}\exp(-\Delta b_{\pi_i}))$ where $\alpha > 0$, $\beta > 0$, and $b_1, \ldots, b_n$ are known real numbers with $\sum_1^n b_i = 0$. Consider testing the hypothesis $H : \Delta = 0$ versus the alternative $K : \Delta > 0$. (This is a Gamma-regression model with covariates or predictors b_i in which the relationship between the responses X_i and the covariates b_i have become scrambled or mixed up: we unfortunately don't know the right pairing of X_i and b_i, but we do know that some permutation of the b_i's is correct. Note that problem 11 in Ferguson, MS, gives a more realistic version of the problem in which β is unknown.)
(a) Show that this problem is invariant under the group of permutations of $(X_1, \ldots, X_n)$, and that the distribution of the maximal invariant $(Y_1, \ldots, Y_n) \equiv (X_{(1)}, \ldots, X_{(n)})$(the order statistics) has density

$$f_{\underline{Y}}(\underline{y}|\Delta) = \frac{(\prod_1^n y_i)^{\alpha-1}\exp(-\alpha\Delta\sum_1^n b_i)}{\Gamma(\alpha)^n\beta^{n\alpha}} \sum_{\pi\in\Pi} \exp\left\{-\frac{1}{\beta}\sum_{i=1}^n y_i \exp(-\Delta b_{\pi_i})\right\}$$

for $y_1 < \cdots < y_n$ and zero elsewhere where $sum_{\pi\in\Pi}$ denotes the sum over all permutations π of $\{1, \ldots, n\}$. (b) Show that the locally best invariant test of H versus K (i.e. the test which maximizes the slope of the power function at the null hypothesis) is to reject H when $\sum_{i=1}^n X_i$ is too large.

2. In class in the context of Example 6.3.14 we developed a UMP invariant test under normal hypotheses: "reject H if $T \equiv \sqrt{n}\overline{X}/S > t_{n-1,\alpha}(\delta_0)$ where $\delta_0\sqrt{n}\Phi^{-1}(1 - p_0)$.
(a) Study the limiting power of this test assuming that the Y's (and hence also the X's) have $E(Y^2) < \infty$ and that the Y's are i.i.d. according to the location-scale family $F_{\mu,\sigma}(x) = F_0((x-\mu)/\sigma)$. (You will need to decide on how to specify "local alternatives".)
(b) Now consider the alternative test based on the empirical d.f. of the Y's: "reject $H : p \ge p_0$ (in favor of $K : \ p < p_0$) if $n\mathbb{F}_n(y_0) \le c_{n,\alpha}$ where $c_{n,\alpha}$ is the largest integer satisfying $P(\text{Bin}(n, p_0) \le c_{n,\alpha}) < \alpha$. Study the limiting power of

this test assuming local alternatives of the form $p_n = p_0 - c/\sqrt{n}$ with $c > 0$.
(c) Compare the asymptotic power of the tests in (a) and (b) assuming that F_0 is normal. [Hint: it might be helpful to read example 6.4.2 on page 33 of section 6.4.]

3. Suppose that X_{ijk}, $i = 1, \dots, I$, $j = 1, \dots, J$, $k = 1, \dots, K$ satisfy the general linear model with $\xi_{ijk} = \xi + \mu_i + \eta_j + \delta_{ij}$ where $\sum_i \mu_i = 0$, $\sum_j \eta_j = 0$, $\sum_j \delta_{ij} = 0$ for all i, and $\sum_i \delta_{ij} = 0$ for all j. (δ_{ij} is called the interaction effect of the ith row and the jth column.)
(a) Show that

$$\begin{aligned} S^2 &= \sum\sum\sum(X_{ijk} - \xi - \mu_i - \eta_j - \delta_{ij})^2 \\ &= \sum\sum\sum(X_{ijk} - \overline{X}_{ij.})^2 \\ &\quad + \sum\sum\sum(\overline{X}_{ij.} - \overline{X}_{i..} - \overline{X}_{.j.} + \overline{X}_{...} - \delta_{ij})^2 \\ &\quad + \sum\sum\sum(\overline{X}_{i..} - \overline{X}_{...} - \mu_i)^2 + \sum\sum\sum(\overline{X}_{.j.} - \overline{X}_{...} - \eta_j)^2 \\ &\quad + \sum\sum\sum(\overline{X}_{...} - \xi)^2 \end{aligned}$$

where $\overline{X}_{ij.} = \sum_k X_{ijk}/K$, and so on.
(b) Find the UMP invariant test of the hypothesis of no row effect $H_0 : \mu_1 = \ldots = \mu_I = 0$. What is the distribution of the test statistic under the general linear hypothesis – including the noncentrality parameter?
(c) Find the UMP invariant test of the hypothesis of no interaction effect $H_0 : \delta_{ij} = 0$ for all i, j. What is the distribution of the test statistic under the general linear hypothesis?

4. Consider a two-way classification X_{ij}, $i = 1, \dots, I$, $j = 1, \dots, J$ with the assumptions of the general linear hypothesis for which $EX_{ij} = \alpha + \beta z_i + \eta_j$, where α, β, and η_j are unknown parameters subject to the restriction $\sum \eta_j = 0$, and where z_i are known numbers for which $\sum_1^I z_i = 0$ and $\sum_1^I z_i^2 = 1$.
(a) Find the UMP invariant test of the hypothesis

$$H_0 : \ \eta_1 = \eta_2 = \cdots = \eta_J = 0.$$

(b) What is the distribution of the test statistic under the general linear hypothesis?

5. **Optional bonus problem:** Let $\Theta = \{(\Delta, \nu) : \ \Delta \in \mathbb{R}, \ 1 \le \nu \le n, \ \nu \text{ an integer}\}$ and let the distribution of $(X_1, \dots, X_n)$, given $\theta = (\Delta, \nu)$, be as independent random variables with $X_i \in N(0, 1)$ for $i \ne \nu$, and $X_\nu \sim N(\Delta, 1)$. Test the hypothesis $H_0 : \Delta = 0$ against alternatives $H_1 : \Delta > 0$ or $\overline{H}_1 : \Delta \ne 0$.

(a) Show that this problem is invariant under the group of permutations of $(X_1, \ldots, X_n)$ and that the distribution of the maximal invariant $Y \equiv (Y_1, \ldots, Y_n) = (X_{(1)}, \ldots, X_{(n)})$ (the order statistics) has density

$$f_{\underline{Y}}(y_1, \ldots, y_n|\Delta) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_1^n y_i^2 - \frac{1}{2}\Delta^2\right)(n-1)! \sum_{\nu=1}^n \exp(\Delta y_\nu)$$

for $y_1 < y_2 < \cdots < y_n$ and zero elsewhere.
(b) Show that the locally best invariant test of H_0 versus H_1 is to reject H_0 if $\sum_{i=1}^n X_i$ is too large.
(c) Show that the locally best unbiased invariant test of H_0 versus $\overline{H}_1$ is to reject H_0 if $\sum_1^n X_i^2$ is too large.