Statistics 581, Problem Set 4

Wellner; 10/15/2008

Reading: Course Notes, Chapter 2, sections 3-6; Ferguson, ACILST pages 44 - 66.
Due: Wednesday, October 22, 2008.
Reminder: Make-up lecture on Friday, 17 October, EEB 042, 11:30 - 12:20.

1. Suppose that $\underline{N}_n \sim \text{Mult}_k(n, \underline{p})$ and $\widehat{p} = \underline{N}_n/n$. Suppose that the true $\underline{p}$ is $\underline{p}_n = \underline{p}_0 + n^{-1/2}\underline{c}$ where $\underline{1}^T\underline{c} = 0$. Use the Cramér - Wold device together with either the Liapunov or the Lindeberg-Feller CLT to show that

$$\underline{Z}_n = \left(\frac{N_{n,1} - np_{n,1}}{\sqrt{np_{0,1}}}, \ldots, \frac{N_{nk} - np_{n,k}}{\sqrt{np_{0,k}}} \right)$$

satisfies $\underline{Z}_n \to_d \underline{Z}$ where $\underline{Z} \sim N_k(0, I - \sqrt{p_0}\sqrt{p_0}^T)$. (It therefore follows, as outlined in class, that the chi-square statistic $Q_n \to_d \chi^2_{k-1}(\delta)$ with $\delta = \sum_{j=1}^k c_j^2/p_{0,j}$ under the local alternative $\underline{p}_n$.)

2. Suppose that $\underline{N}_n \sim \text{Mult}_k(n, \underline{p})$ and $\widehat{p} = \underline{N}_n/n$. Define a family of functions ϕ_s for $-1 \le s \le 2$ by

$$\phi_s(x) = \frac{1 - s + sx - x^s}{s(1-s)}, \qquad x \in \mathbb{R}^+, \quad s \neq 0, 1,$$

and define $\phi_1(x) = x(\log x - 1) + 1$, $\phi_0(x) = \log(1/x) + x - 1$. Now define a family of statistics for testing $H : \underline{p} = \underline{p}_0$ versus $K : \underline{p} \neq \underline{p}_0$ by

$$T_n(s) \equiv 2n \sum_{j=1}^k p_{0j} \phi_s \left(\frac{\hat{p}_j}{p_{0j}} \right).$$

(a) Show that $T_n(s)$ reduces to the following statistics discussed in class:
(i) $T_n(2)$ is the Pearson chi-square statistic Q_n; (ii) $T_n(1)$ is $2 \log \lambda_n$ where λ_n is the likelihood ratio statistic; (iii) $T_n(-1)$ is Neyman's version of the chi-square statistic, Q_n^{Neyman}; and (iv) $T_n(1/2)$ is the Hellinger statistic H_n^2.
(b) Show that $n^{-1}T_n(s)$ converges in probability to a deterministic limit $t(s)$ under a general $\underline{p}$, and identify the limit explicitly in cases (i) - (iv) of part (a). Do any of these limiting parameters have names?
(c) Find the limiting distribution of $T_n(1/2)$ under the null hypothesis H.
Notes: This problem is related to the statistics treated in Cressie and Read, JRSS B 46 (1984), 440 - 464, and also to the "transformed" chi-square statistics discussed in Ferguson, ACILST, pages 59 and 66. See also Jager and Wellner, Ann. Statist. 35 (2007), 2018-2053 for related material in a different vein.

3. Ferguson, ACILST, problem 4, page 55, modified slightly: suppose that the sample sizes (of X's and Y's) are m and n respectively, and that $m/(m+n) \to \lambda \in (0, 1)$.

4. Suppose that $\underline{N}_n = (N_{11}, N_{12}, N_{21}, N_{22}) \sim \text{Mult}_4(n, \underline{p})$ where $\underline{p} = (p_{11}, p_{12}, p_{21}, p_{22})$ where $\sum_{i=1}^2 \sum_{j=1}^2 p_{ij} = 1$. (Thus $\underline{N}_n$ is the sum of n independent $\text{Mult}_4(1, \underline{p})$ random vectors $\{\underline{Y}_i\}_{i=1}^n$.) Since there are really just three independently varying parameters for this problem, it is often useful to re-express the cell probabilities in terms of two marginal probabilities, say $p_{1\cdot} = p_{11} + p_{12}$ and $p_{\cdot 1} = p_{11} + p_{21}$, and ψ, the log of the odds-ratio, defined by

$$\psi \equiv \log \frac{p_{21}/p_{22}}{p_{11}/p_{12}} = \log \frac{p_{12}p_{21}}{p_{11}p_{22}}\,. \tag{1}$$

You may use the fact that $\psi = 0$ if and only if independence holds for the 2×2 table (i.e. $p_{ij} = p_{i\cdot}p_{\cdot j}$ for $i, j = 1, 2$).
(a) Suggest an estimator of ψ, say $\widehat{\psi}$.
(b) Show that the estimator you proposed in (a) is asymptotically normal and compute the asymptotic variance of your estimator.

5. This is a continuation of problem 3. One standard test of independence in the 2×2 table is the test based on a Pearson-type chi-square statistic.
(a) Write down the chi-square statistic Q_n for this problem, state its asymptotic distribution under the null hypothesis, and explain briefly why the claimed result holds.
(b) Suppose that the alternative hypothesis holds. Show that under the alternative hypothesis $n^{-1}Q_n \to_p$ some constant q and compute q as explicitly as possible.
(c) Find the asymptotic distribution of Q_n under local alternatives of the form $\psi_n = tn^{-1/2}$; i.e. $\underline{p}_n \equiv (p_{11,n}, p_{12,n}, p_{21,n}, p_{22,n}) = \underline{p}_0 + \underline{c}n^{-1/2}$ where

$$\psi_0 \equiv \log\left(\frac{p_{21,0}p_{12,0}}{p_{11,0}p_{22,0}}\right) = 0$$

and $\underline{1}'\underline{c} = 0$.
(d) Suppose that $n = 30$, $\alpha = .02$, and the true $\underline{p}$ is $\underline{p} = (.3, .2, .1, .4)$. Give an approximation to the power of the chi-square test at this particular alternative.

6. **Optional bonus problem.**

Suppose that $Y_i = \alpha + \theta'(x_i - \overline{x}) + \epsilon_i$, $i = 1, \ldots, n$, where $\epsilon_i \sim (0, \sigma^2)$ are i.i.d. and the x_i's are known vectors in R^k. Equivalently, $\underline{Y} = X\underline{\beta} + \underline{\epsilon}$ where

$$X^T = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 - \overline{x} & x_2 - \overline{x} & \cdots & x_n - \overline{x} \end{pmatrix}$$

so that X is an $n \times (k+1)$ matrix. Let $\hat{\underline{\beta}}$ be the least squares estimator of $\underline{\beta} = (\alpha, \theta')'$; i.e. $\hat{\underline{\beta}} = (X^TX)^{-1}X^T\underline{Y}$. Suppose that $n^{-1}(X^TX) \to D$ where D is positive definite.
(a) What additional condition(s) do you need to impose to prove that

$$\sqrt{n}(\hat{\beta}_n - \beta) \to_d N_{k+1}(0, \text{“}\textit{something}\text{”})\,?$$

(b) Find “something” in part (a).