Statistics 581, Problem Set 4

Wellner; 10/20/2010

Reading: Course Notes, Chapter 2, sections 3-6; Ferguson, ACILST pages 44 - 66.
Due: Wednesday, October 27, 2010.
Reminder: Make-up lecture on Monday, 25 October, Savery 132, 11:30 - 12:20.

1. Suppose that $X_1, X_2, \ldots$ are i.i.d. (μ, σ^2) with $\mu_4 < \infty$. Let $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ and $S_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ be the sample mean and sample variance respectively.
(a) Show that
$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ S_n^2 - \sigma^2 \end{pmatrix} \rightarrow_d \underline{Z} \sim N_2(0, \Sigma)$$
where
$$\begin{pmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{pmatrix}.$$
(b) Suppose $\mu \neq 0$. Use (a) to find the limiting distribution of the sample *coefficient of variation* $C_n \equiv S_n/\overline{X}_n$; i.e. show that $\sqrt{n}(C_n - c) \rightarrow_d N(0, V^2)$ with $c \equiv \sigma/\mu$ and find V^2.

2. (a) 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 $\underline{Z}_n \rightarrow_d \underline{Z}$ 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 $Q_n \rightarrow_d \chi^2_{k-1}(\delta)$ with $\delta = \sum_{j=1}^k c_j^2/p_{0,j}$ under the local alternative $\underline{p}_n$.)
(b) (Ferguson, *A Course in Large Sample Theory*, page 65.) In a multinomial experiment with sample size $n = 100$ and 3 cells with null hypothesis $H_0 : \underline{p}_0 = (.25, .5, .25)$, what is the approximate power at the alternative $\underline{p} = (.2, .6, .2)$ when the level of significance is $\alpha = .05$? $\alpha = .01$? How large a sample size is need to achieve power 0.9 at this alternative when $\alpha = .05$? $\alpha = .01$?

3. Suppose the same set-up as in the chi-square testing situation considered in lecture in class but now, for testing $H_0 : \underline{p} = \underline{p}_0$ versus $K_0 : \underline{p} \neq \underline{p}_0$, instead of the chi-square statistic Q_n, consider the test statistic given by
$$H_n^2 \equiv 4n\sum_{i=1}^k (\sqrt{\hat{p}_i} - \sqrt{p_{i0}})^2.$$
The statistic H_n^2 is $4n$ times the square of the *Hellinger distance* between $\underline{\hat{p}}$ and $\underline{p}_0$.
(a) Find the limiting distribution of H_n^2 under the null hypothesis H_0.

(b) Find the limit of $n^{-1}H_n^2$ under fixed alternatives $\underline{p} \neq \underline{p}_0$ in K_0, and use this to show that the test based on H_n^2 is consistent against K_0.
(c) Find the limiting distribution of H_n^2 under local alternatives $\underline{p}_n = \underline{p}_0 + \underline{c}/\sqrt{n}$, and use this to approximate the power of this test. Compare the (local asymptotic) power of this test to the chi-square test.

4. 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 $\underline{\hat{\beta}}$ be the least squares estimator of $\underline{\beta} = (\alpha, \theta')'$; i.e. $\underline{\hat{\beta}} = (X^T X)^{-1} X^T \underline{Y}$. Suppose that $n^{-1}(X^T X) \to D$ 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).

5. 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

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

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.

6. **Optional bonus problem 1:** This is a continuation of problem 5 above. 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 $\psi_n = t n^{-1/2}$; 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.

7. **Optional bonus problem 2:** Suppose that $(X_i - \mu)/\sigma$, $i = 1, \ldots, m$ and $(Y_j - \nu)/\tau$, $j = 1, \ldots, n$ are iid $(0, 1, \mu_4 < \infty)$ (thus γ_2 is the same for the two populations), and let S_X^2 and S_Y^2 denote the sample variances of the X's and Y's respectively. The classical $F-$test based on the assumption that all the standardized X's and Y's are $N(0,1)$ rejects $H_0 : \tau \le \sigma$ in favor of $H_1 : \tau > \sigma$ if $F \equiv S_Y^2 / S_X^2 > F_{n-1, m-1, \alpha}$. Assuming that $m/N \to \lambda \in [0, 1]$ as $m \wedge n \to \infty$ where $N = m + n$, find the true asymptotic size of this test for non-normal X's and Y's as above.