Statistics 581, Problem Set 4

Wellner; 10/17/2018

Reading: Course Notes, Chapter 2, pages 26-40;
Ferguson, ACILST pages 60-66, 22-23, 87-93;
vdVaart, Asym. Stat., sections 3.1-3.5, 4.1, pages 25-37; sections 21.1-21.2, pages 304-310.
Reminder: Midterm exam, Friday 2 November 2018.

Due: Wednesday, October 24, 2018

1. 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.

2. 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 $n^{-1} Q_n \rightarrow_p$ 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 $n = 40$, $\alpha = .05$, 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.

3. Suppose that $X_1, X_2, \ldots$ are i.i.d. positive random variables, and define $\overline{X}_n \equiv n^{-1} \sum_{i=1}^n X_i$, $H_n \equiv 1/(n^{-1} \sum_{i=1}^n (1/X_i))$, and $G_n \equiv \{\prod_{i=1}^n X_i\}^{1/n}$ to be the *arithmetic, harmonic, and geometric* means respectively. We know that $\overline{X}_n \rightarrow_{a.s.} E(X_1) = \mu$ if and only if $E|X_1| < \infty$.
(a) Use the SLLN together with appropriate additional hypotheses to show that $H_n \rightarrow_{a.s.} 1/\{E(1/X_1)\} \equiv h$, and $G_n \rightarrow_{a.s.} \exp(E\{\log X_1\}) \equiv g$.

(c) Use the multivariate CLT and the delta method to find the joint limiting distribution of $\sqrt{n}(\overline{X}_n - \mu, H_n - h, G_n - g)$. You will need to impose or assume additional moment conditions to be able to prove this. Specify these additional assumptions carefully.
(d) Suppose that $X_i \sim \text{Gamma}(r, \lambda)$ with $r > 0$. For what values of r are the hypotheses you imposed in (c) satisfied? Compute the covariance of the limiting distribution in (c) as explicitly as you can in this case.
(e) Use the result in (c) to show that $\sqrt{n}(G_n/\overline{X}_n - g/\mu) \to_d N(0, V^2)$ and compute V^2 explicitly when $X_i \sim \text{Gamma}(r, \lambda)$ with r satisfying the conditions you found in (d).

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

6. **Optional bonus problem 1:** Ferguson, ACILST, problem 5, page 50. Ferguson, ACILST, problem 5, page 50: (The Poisson dispersion test). A standard test of the hypothesis H_0 that a distribution is Poisson(λ) for some λ is to reject H_0 if the ratio of the sample variance to the sample mean, $S_n^2/\overline{X}_n$, is too large. This test is good against alternatives whose variance is greater than the mean, such as the negative binomial distribution or any other mixture of Poisson distributions.
(a) Find the asymptotic distribution of $S_n^2/\overline{X}_n$ for general distributions.
(b) Find the asymptotic distribution of $S_n^2/\overline{X}_n$ under H_0 and show that it is independent of λ.

7. **Optional bonus problem 2:** Ferguson, ACILST, problem 4, page 55. Suppose that $(X_i - \mu)/\sigma$, $i = 1, \ldots, m$ and $(Y_j - \nu)/\tau$, $j = 1, \ldots, n$ are i.i.d. $(0, 1, \mu_4) < \infty$; thus γ_2 is the same for the two populations. 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 $N \equiv m + n$, find the true asymptotic size of this test for non-normal X's and Y's as above.