

## Statistics 581, Problem Set 4

Wellner; 10/21/2009

**Reading:** Course Notes, Chapter 2, sections 3-6; Ferguson, ACILST pages 44 - 66.

**Due:** Wednesday, October 28, 2009.

1. Ferguson, ACILST, problem 5, page 50: (The Poisson dispersion test). A standard test of the hypothesis  $H_0$  that a distribution is  $\text{Poisson}(\lambda)$  for some  $\lambda$  is to reject  $H_0$  if the ratio of the sample variance to the sample mean,  $S_n^2/\bar{X}_n$ , 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  $S_n^2/\bar{X}_n$  for general distributions.
  - (b) Find the asymptotic distribution of  $S_n^2/\bar{X}_n$  under  $H_0$  and show that it is independent of  $\lambda$ .

2. (Continuation of problem 1 above.) Suppose that  $(X|\Lambda) \sim \text{Poisson}(\Lambda)$  where  $\Lambda \sim \Gamma(r, b)$  with density  $b^r \lambda^{r-1} \exp(-b\lambda)/\Gamma(r)$  for some  $r > 0$  and  $b > 0$ .
  - (a) Show that the marginal distribution of  $X$  is Negative Binomial  $(b/(1+b), r)$  with density (probability mass function)

$$P_{r,b}(X = x) = \frac{\Gamma(r+x)}{x!\Gamma(r)} \left(\frac{b}{1+b}\right)^r \frac{1}{(1+b)^x}$$

for  $x = 0, 1, \dots$

(b) Show that  $E(X) = r/b$  and  $\text{Var}(X) = (r/b) + r/b^2 > (r/b) = E(X)$ , and hence if  $b \equiv b_n = \sqrt{n}/\lambda_0$  and  $r \equiv r_n = \sqrt{n}$ , we have, letting  $E_n$  and  $\text{Var}_n$  denote expectation and variance under  $(r_n, b_n)$ ,  $E_n(X) \rightarrow \lambda_0$  and  $\text{Var}_n(X) \rightarrow \lambda_0$ , while  $\sqrt{n}(\text{Var}_n(X) - \lambda_0) \rightarrow \lambda_0^2$ . (Hint: Use our results for computing the mean and variance conditionally on another random variable.)

(c) Show that if  $X_n \sim \text{Negative Binomial}(b_n/(1+b_n), r_n)$  with  $b_n$  and  $r_n$  as in (b), then  $X_n \rightarrow_d X_0 \sim \text{Poisson}(\lambda_0)$ .

(d) Now suppose that  $X_{ni} \sim \text{Negative Binomial}(b_n/(1+b_n), r_n)$  for  $i = 1, \dots, n$  are independent with  $b_n$  and  $r_n$  as in (b). Let  $\bar{X}_n = n^{-1} \sum_{i=1}^n X_{ni}$  and  $S_n^2 = (n-1)^{-1} \sum_{i=1}^n (X_{ni} - \bar{X}_n)^2$  as in problem 1. Use the results of (b) to show that under this family of local alternatives to the Poisson distribution we have

$$\sqrt{n}(S_n^2/\bar{X}_n - 1) \rightarrow_d N(c, 2)$$

for some  $c \neq 0$  and find  $c$ . Use this to approximate the power of the test in problem 1 for this particular sequence of alternatives.

3. Suppose that  $\underline{N}_n \sim \text{Mult}_k(n, \underline{p})$  and  $\hat{\underline{p}} = \underline{N}_n/n$ . 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

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

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  $Q_n \rightarrow_d \chi_{k-1}^2(\delta)$  with  $\delta = \sum_{j=1}^k c_j^2/p_{0,j}$  under the local alternative  $\underline{p}_n$ .)

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  $\psi$ , the log of the odds-ratio, defined by

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

You may use the fact that  $\psi = 0$  if and only if independence holds for the  $2 \times 2$  table (i.e.  $p_{ij} = p_{i\cdot}p_{\cdot j}$  for  $i, j = 1, 2$ ).

(a) Suggest an estimator of  $\psi$ , say  $\hat{\psi}$ .

(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 4. One standard test of independence in the  $2 \times 2$  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  $\psi_n = tn^{-1/2}$ ; 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  $n = 50$ ,  $\alpha = .05$ , 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.

6. Suppose that  $X_{n,1}, \dots, X_{n,n}$  are independent Bernoulli( $p_{n,1}$ ),  $\dots$ , Bernoulli( $p_{n,n}$ ) respectively. Let  $T_n = X_{n,1} + \dots + X_{n,n}$ ,  $\mu_n = \sum_{i=1}^n p_{n,i}$ , and  $\sigma_n^2 \equiv \sum_{i=1}^n p_{n,i}(1-p_{n,i})$ .

(a) Use the Liapunov central limit theorem to show that if  $\sigma_n^2 \rightarrow \infty$ , then  $(T_n - \mu_n)/\sigma_n \rightarrow_d N(0, 1)$ .

(b) Show that the key condition of the Liapunov CLT implies the Lindeberg condition.

7. **Optional bonus problem 1:** Ferguson, ACILST, problem 1, page 54: Find the asymptotic distribution of the estimate of the regression coefficient  $\hat{\beta} = S_{XY}/S_X^2$  when sampling from a bivariate distribution (with finite fourth moments). What is its asymptotic variance when sampling from a bivariate normal distribution?

8. **Optional bonus problem 2:** Suppose that  $\underline{N}_n \sim \text{Mult}_k(n, \underline{p})$  and  $\hat{p} = \underline{N}_n/n$ . Define a family of functions  $\phi_s$  for  $-1 \leq s \leq 2$  by

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

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  $\lambda_n$  is the likelihood ratio statistic; (iii)  $T_n(-1)$  is Neyman's version of the chi-square statistic,  $Q_n^{\text{Neyman}}$ ; 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.