Statistics 581, Problem Set 4

Wellner; 10/18/2017

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.

Due: Wednesday, October 25, 2017

1. Suppose that $X_{n,1}, \ldots, X_{n,n}$ are independent Bernoulli$(p_{n,1}), \ldots,$ Bernoulli$(p_{n,n})$ respectively. Let $T_n = X_{n,1} + \cdots + X_{n,n}$, $\mu_n = \sum_{i=1}^n p_{n,i}$, 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 $\sigma_n^2 \to \infty$, then $(T_n - \mu_n)/\sigma_n \to_d N(0,1)$.
(b) Suppose that $X_{n,i}$ and T_n are as in (a) with $p_{n,i} = p_n \equiv p_0 + cn^{-1/2}$ for $1 \le i \le n$ and $n \ge c^2/\min\{p_0^2, (1-p_0)^2\}$. Show that $\sqrt{n}(\widehat{p}_n - p_n) \to_d N(0, p_0(1-p_0))$.
(c) Show that the key condition of the Liapunov CLT implies the Lindeberg condition (and hence the Lindeberg-Feller CLT also holds).

2. (a) Suppose that $\underline{N}_n \sim$ 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 $\underline{p}_n$.)
(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 $H_0 : \underline{p}_0 = (1/3, 1/3, 1/3)$, 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. 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 i.i.d. random variables $X_1, \ldots, X_n$ with $EX_1 > 0$ and $E|X_1|^4 < \infty$; i.e. show that $\sqrt{n}(S_n^2/\overline{X}_n - \sigma^2/\mu) \to_d$ "something" and find "something".
(b) Find the asymptotic distribution of $S_n^2/\overline{X}_n$ under H_0 and show that it is independent of λ.

4. (Continuation of problem 3 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 $x = 0, 1, \ldots$.
(b) Show that $E(X) = r/b$ and $Var(X) = (r/b) + r/b^2 > (r/b) = E(X)$, and hence if $b \equiv b_n = \sqrt{n}/\lambda_0$ and $r \equiv r_n = \sqrt{n}$, we have, letting E_n and Var_n denote expectation and variance under (r_n, b_n), $E_n(X) \to \lambda_0$ and $Var_n(X) \to \lambda_0$, while $\sqrt{n}(Var_n(X) - \lambda_0) \to \lambda_0^2$. (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 $X_n \to_d X_0 \sim \text{Poisson}(\lambda_0)$.
(d) Now suppose that $X_{ni} \sim \text{Negative Binomial}(b_n/(1+b_n), r_n)$ for $i = 1, \ldots, n$ are independent with b_n and r_n as in (b). Let $\overline{X}_n = n^{-1}\sum_{i=1}^n X_{ni}$ and $S_n^2 = (n-1)^{-1}\sum_1^n (X_{ni} - \overline{X}_n)^2$ as in problem 1. Use the results of (b) to show that under this family of local alternatives to the Poisson distribution we have

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

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.

5. **Optional bonus problem 1.** 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 $\overline{X}_n \to_{a.s.} E(X_1) = \mu$ if and only if $E|X_1| < \infty$.
(a) Show that $H_n \le G_n \le \overline{X}_n$ and that $h \le g \le \mu$ (where h and g are defined in (b)).
(b) Use the SLLN together with appropriate additional hypotheses to show that $H_n \to_{a.s.} 1/\{E(1/X_1)\} \equiv h$, and $G_n \to_{a.s.} \exp(E\{\log X_1\}) \equiv g$.
(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).