

Statistics 581, Problem Set 3

Wellner; 10/8/2008

Reading: Lehmann & Casella, TPE, pages 54-61 and pages 75-78.

Ferguson, ACILST, pages 1 - 60.

Reminder: Make-up lecture, Friday, October 17, 11:30 AM.

Due: Wednesday, October 15, 2008.

1. Ferguson, ACILST, page 34, problem 1(b), modified slightly.

Suppose that X_1, \dots, X_n is a sample from the Poisson distribution with parameter $\lambda > 0$: $P(X_1 = k) = \exp(-\lambda)\lambda^k/k!$, $k = 0, 1, \dots$. Let $Z_n = (1/n) \sum_{i=1}^n 1_{[X_i=2]}$.

- (a) What is the joint asymptotic distribution of

$$\sqrt{n}((\bar{X}_n, Z_n)' - (\lambda, \lambda^2 e^{-\lambda}/2)')$$

- (b) Let $p_2(\lambda) \equiv P_\lambda(X_1 = 2)$. What is the asymptotic distribution of $\hat{p}_2 \equiv p_2(\hat{\lambda}_n)$ where $\hat{\lambda}_n = \bar{X}_n$?

- (c) What is the joint asymptotic distribution of (Z_n, \hat{p}_2) (after centering and rescaling)?

- (d) Compute the ratio of the asymptotic variances of the two estimators Z_n and \hat{p}_2 of $p_2(\lambda)$. Which estimator would you prefer if the Poisson model (assumption) holds? Which estimator would you prefer if the Poisson model (assumption) fails?

2. Suppose that X is a random variable with finite fourth moment; $E|X|^4 < \infty$. Then $\mu_4 = E(X - \mu)^4$ is the fourth central moment of X . The ratio $\mu_4/\sigma^4 \equiv \kappa$ is the *kurtosis* of X (or of the distribution function F of X), and $\gamma_2 \equiv \mu_4/\sigma^4 - 3$ is called the *excess of kurtosis*; note that for any $N(\mu, \sigma^2)$ random variable, $\gamma_2 = 0$. Investigate the value of γ_2 for various classical distributions (t_r , uniform, bernoulli, Poisson(λ), ...). How big can γ_2 be? How small can γ_2 be?

3. (From Ferguson, *A Course in Large Sample Theory*, page 65, modified.) In a multinomial experiment with sample size $n = 100$ and 4 cells with null hypothesis $H_0 : \underline{p}_0 = (.2, .3, .3, .2)$, what is the approximate power at the alternative $\underline{p} = (.1, .4, .4, .1)$ when the level of significance is $\alpha = .05$? $\alpha = .01$? How large a sample size is need to achieve power 0.8 at this alternative when $\alpha = .05$? $\alpha = .01$?

4. Suppose that X_1, \dots, X_n are independent $N(0, 1)$ random variables, and let $Y_i = X_i^2$, for $i = 1, \dots, n$. Thus $\sum_{i=1}^n Y_i \sim \chi_n^2$.

- (a) Show that $\sqrt{n}(\bar{Y}_n - 1) \rightarrow_d N(0, \text{"something"})$, and find "something".

- (b) Show that for each $r > 0$, $\sqrt{n}(\bar{Y}_n^r - 1) \rightarrow_d N(0, V^2(r))$ and find $V^2(r)$ as a function of r .

- (c) Show that

$$\frac{\sqrt{n}(\bar{Y}_n^{1/3} - (1 - 2/(9n)))}{\sqrt{2/9}} \rightarrow_d N(0, 1).$$

Does this agree with your result in (b)?

- (d) Make normal probability plots to compare the approximations in (a) and (c). [The transformation in (c) is called the "Wilson-Hilferty" transformation of a χ^2 random variable.

5. Suppose that X_1, X_2, \dots are i.i.d. positive random variables, and define $\bar{X}_n \equiv n^{-1} \sum_{i=1}^n X_i$, $H_n \equiv 1/(n^{-1} \sum_{i=1}^n (1/X_i))$, and $G_n \equiv \{\prod_{i=1}^n X_i\}^{1/n}$ to be the arithmetic, harmonic, and geometric means respectively. We know that $\bar{X}_n \xrightarrow{a.s.} E(X_1) = \mu$ if and only if $E|X_1| < \infty$.
- (a) Use the SLLN together with appropriate additional hypotheses to show that $H_n \xrightarrow{a.s.} 1/\{E(1/X_1)\} \equiv h$, and $G_n \xrightarrow{a.s.} \exp(E\{\log X_1\}) \equiv g$.
- (c) Use the multivariate CLT and the delta method to find the joint limiting distribution of $\sqrt{n}(\bar{X}_n - \mu, H_n - h, G_n - g)$. You will need to impose or assume additional moment conditions to be able to prove this. Specify these additional assumptions carefully.
6. **Optional bonus problem 1:** A. Ferguson, ACILST, page 24, problem 4. One strategy for evaluating the integral

$$I = \int_1^\infty \frac{1}{x} \sin(2\pi x) dx = .152645\dots$$

by Monte Carlo approximation is as follows. Write the integral, by a change of variable $y = 1/x$, as

$$I = \int_0^1 \frac{1}{y} \sin\left(\frac{2\pi}{y}\right) dy$$

and approximate I by

$$\hat{I}_{1n} = \frac{1}{n} \sum_{i=1}^n \frac{1}{Y_i} \sin\left(\frac{2\pi}{Y_i}\right)$$

where Y_1, \dots, Y_n is a sample from the uniform distribution on $[0, 1]$. How well does this approximation work? Does \hat{I}_{1n} converge to I almost surely?

B. Suppose that I in A is generalized to

$$I_\alpha \equiv \int_1^\infty \frac{1}{x^\alpha} \sin(2\pi x) dx$$

for $\alpha > 0$ (so that the integral I in part A is I_1). Construct the corresponding Monte-Carlo estimator $\hat{I}_{n,\alpha}$ of I_α . For what values of α will the estimator $\hat{I}_{n,\alpha}$ converge to I_α ? (Use the same change of variables as in A.)

C. For what values of α will we have

$$\sqrt{n}(\hat{I}_{n,\alpha} - I_\alpha) \rightarrow_d \text{something?}$$

For those values of α for which this holds, find “something”.

7. **Optional bonus problem 2:** Ferguson, ACILST, problem 5, page 18:

Let X_{n1}, \dots, X_{nn} be independent, $X_{nk} \sim \text{Bernoulli}(p_{nk})$, and let $Y_n \sim \text{Poisson}(\sum_{k=1}^n p_{nk})$. Let P_n be the distribution of $\sum_{k=1}^n X_{nk}$ and let Q_n be the distribution of Y_n . Show that

$$d_{TV}(P_n, Q_n) \equiv \sup_{A \in \mathcal{B}} |P(S_n \in A) - P(Y_n \in A)| \leq \sum_{k=1}^n p_{nk}^2.$$

Note that when $p_{nk} = p_n \rightarrow 0$ for all k and $np_n \rightarrow \lambda$, then $\sum_{k=1}^n p_{nk}^2 = np_n^2 = (np_n)^2/n = O(n^{-1})$.

[Hint: construct S_n and Y_n on a common probability space as follows: let $T_{nk} \sim \text{Poisson}(p_{nk})$, $k = 1, \dots, n$ be independent, and let $Z_{nk} \sim \text{Bernoulli}(1 - (1 - p_{nk})e^{-p_{nk}})$, $k = 1, \dots, n$ be independent and independent of the T_{nk} 's. Define

$$X_{nk} = 1_{[T_{nk} \geq 1]} + 1_{[T_{nk}=0]}1_{[Z_{nk}=1]}.$$

Set $S_n = \sum_{k=1}^n X_{nk}$, $Y_n = \sum_{k=1}^n T_{nk}$. Check that $X_{nk} \sim \text{Bernoulli}(p_{nk})$, $Y_n \sim \text{Poisson}(\sum_{k=1}^n p_{nk})$, and

$$\begin{aligned} P(T_{nk} = 0, X_{nk} = 1) &= e^{-p_{nk}} - (1 - p_{nk}) \\ P(T_{nk} \geq 1, X_{nk} = 0) &= 0 \\ P(T_{nk} \geq 2) &= 1 - e^{-p_{nk}} - p_{nk}e^{-p_{nk}}. \end{aligned}$$

Show that

$$d_{TV}(P_n, Q_n) \leq P(S_n \neq Y_n) \leq \sum_{k=1}^n P(X_{nk} \neq T_{nk}) \leq \sum_{k=1}^n p_{nk}^2.$$