Statistics 581, Problem Set 3 - Revised

Wellner; 10/11/2006

Reading: Lehmann & Casella, TPE, pages 54-61 and pages 75-78.
Ferguson, ACILST, pages 1 - 60.
Due: Wednesday, October 18, 2006.

1. Ferguson, ACILST, page 34, problem 1(b), modified slightly.
Suppose that $X_1, \ldots, X_n$ is a sample from the Poisson distribution with parameter $\lambda > 0$: $P(X_1 = k) = \exp(-\lambda)\lambda^k/k!$, $k = 0, 1, \ldots$. Let $Z_n = (1/n)\sum_{i=1}^n 1_{[X_i=1]}$.
(a) What is the joint asymptotic distribution of
$$\sqrt{n}((\overline{X}_n, Z_n)' - (\lambda, \lambda e^{-\lambda})')?$$
(b) Let $p_1(\lambda) \equiv P_\lambda(X_1 = 1)$. What is the asymptotic distribution of $\hat{p}_1 \equiv p_1(\hat{\lambda}_n)$ where $\hat{\lambda}_n = \overline{X}_n$?
(c) What is the joint asymptotic distribution of $(Z_n, \hat{p}_1)$ (after centering and rescaling)?
(d) Compute the ratio of the asymptotic variances of the two estimators Z_n and $\hat{p}_1$ of $p_1(\lambda)$. Which estimator would you prefer if the Poisson model (assumption) holds? Which estimator would you prefer if the Poisson model (assumption) fails?

2. (From Ferguson, *A Course in Large Sample Theory*, page 65, modified.) In a multinomial experiment with sample size $n = 100$ and 3 cells with null hypothesis $H_0 : \underline{p}_0 = (.2, .6, .2)$, what is the approximate power at the alternative $\underline{p} = (.25, .5, .25)$ 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 $\alpha = .05$? $\alpha = .01$?

3. 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.

4. 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, Poission(λ), ...). How big can γ_2 be? How small can γ_2 be?

5. Ferguson, ACILST, problem 7, page 34.

6. Suppose the same set-up as in the chi-square testing situation considered in lecture in class on 10/8/93 and in the above problem, 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

$$H_n^2 \equiv 4n \sum_{i=1}^{k} (\sqrt{\hat{p}_j} - \sqrt{p_{0,j}})^2 .$$

The statistic H_n^2 is $8n$ times the square of the *Hellinger distance* between $\hat{\underline{p}}$ and $\underline{p}$.
(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 under local alternatives $\underline{p}_n = \underline{p}_0 + \underline{c}/\sqrt{n}$, with $\underline{c}'\underline{1} = 0$, and use this to approximate the power of this test. Compare the (local asymptotic) power of this test to the chi-square test.

7. **Optional bonus problem:** Ferguson, ACILST, problem 5, page 18:
Let $X_{n1}, \ldots, X_{nn}$ 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 $p_{nk} = p_n \to 0$ for all k and $np_n \to \lambda$, 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 $T_{nk} \sim \text{Poisson}(p_{nk})$, $k = 1, \ldots, n$ be independent, and let $Z_{nk} \sim \text{Bernoulli}(1-(1-p_{nk})e^{p_{nk}})$, $k = 1, \ldots, n$ 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 $X_{nk} \sim \text{Bernoulli}(p_{nk})$, $Y_n \sim \text{Poisson}(\sum_1^n p_{nk})$, 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

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