Statistics 581, Problem Set 3

Wellner; 10/16/2002

Reading: Lehmann & Casella, TPE, pages 54-61 and pages 75-78.
Ferguson, ACILST, pages 1 - 60.
Due: Wednesday, October 23, 2002.

1. A. Ferguson, ACILST, page 24, problem 4. One strategy for evaluating the integral

$$I = \int_1^\infty \frac{1}{x} \sin(2\pi x) dx = .153...$$

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

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

where $Y_1, \ldots, Y_n$ is a sample from the uniform distribution on $[0, 1]$. How well does this approximation work? Does $\hat{I}_n$ 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

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

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

2. Ferguson, ACILST, page 34, problem 1 (modified slightly)
A. Suppose that $X_1, X_2, \ldots$ are i.i.d. in R^2 with distribution giving probability θ_1 to $(1, 0)'$, probability θ_2 to $(0, 1)'$, θ_3 to $(0, 0)'$ and θ_4 to $(-1, -1)$ where $\theta_j \ge 0$ for $j = 1, 2, 3, 4$ and $\theta_1 + \theta_2 + \theta_3 + \theta_4 = 1$. Find the limiting distribution of $\sqrt{n}(\overline{X}_n - E(X_1))$ and describe the resulting approximation to the distributon of $\overline{X}_n$.
B. 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 $Z_n = (1/n)\sum_{i=1}^n 1_{[X_i=1]}$. What is the joint asymptotic distribution of

$$\sqrt{n}((\overline{X}_n, Z_n)' - (\lambda, \lambda e^{-\lambda})')?$$

C. 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 $\hat{\lambda}_n = \overline{X}$?
D. What is the joint asymptotic distribution of $(Z_n, \hat{p}_1)$ (after centering and rescaling)?

3. 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?

4. Ferguson, ACILST, page 34, problem 6. Let $Z_1, Z_2, \ldots$ be i.i.d. continuous random variables. We say a record occurs at k if $Z_k > \max_{i<k} Z_i$. Let $R_k = 1$ if a record occurs at k, and let $R_k = 0$ otherwise. Then $R_1, R_2, \ldots$ are independent Bernoulli random variables with $P(R_k = 1) = 1 - P(R_k = 0) = 1/k$, for $k = 1, 2, \ldots$. Let $S_n = \sum_{k=1}^n R_k$ denote the number of records in the first n observations. Find $E(S_n)$ and $Var(S_n)$, and show that $(S_n - E(S_n))/\sqrt{Var(S_n)} \to_d N(0,1)$.

5. Suppose that $X_1, \ldots, X_n$ are independent $N(0,1)$ random variables, and let $Y_i = X_i^2$, for $i = 1, \ldots, n$. Thus $\sum_1^n Y_i \sim \chi_n^2$.
(a) Show that $\sqrt{n}(\overline{Y}_n - 1) \to_d N(0,$ "*something*"), and find "something".
(b) Show that for each $r > 0$, $\sqrt{n}(\overline{Y}_n^r - 1) \to_d N(0, V^2(r))$ and find $V^2(r)$ as a function of r.
(c) Show that
$$\frac{\sqrt{n}(\overline{Y}_n^{1/3} - (1 - 2/(9n)))}{\sqrt{2/9}} \to_d N(0,1)\,.$$
Does this agree with your result in (b)?
(d) Make normal probablility 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.]

6. **Optional bonus problem:** Ferguson, ACILST, problem 5, page 18: Let $X_{n1}, \ldots, X_{nn}$ be independent, X_{nk} $\sim$Bernoulli(p_{nk}), and let Y_n $\sim$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
$$d_{TV}(P_n, Q_n) \equiv \sup_{A \in \mathcal{B}} |P(S_n \in A) - P(Y_n \in A)| \le \sum_{k=1}^n p_{nk}^2 .$$
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 $T_{nk} \sim$ Poisson(p_{nk}), $k = 1, \ldots, n$ be independent, and let $Z_{nk} \sim$ Bernoulli($1-(1-p_{nk})e^{p_{nk}}$), $k = 1, \ldots, n$ be independent and independent of the T_{nk}'s. Define
$$X_{nk} = 1_{[T_{nk} \ge 1]} + 1_{[T_{nk}=0]} 1_{[Z_{nk}=1]}.$$
Set $S_n = \sum_{k=1}^n X_{nk}$, $Y_n = \sum_{k=1}^n T_{nk}$. Check that $X_{nk} \sim$ Bernoulli(p_{nk}), $Y_n \sim$ Poisson($\sum_1^n p_{nk}$, and
$$\begin{aligned} P(T_{nk} = 0, X_{nk} = 1) &= e^{-p_{nk}} - (1 - p_{nk}) \\ P(T_{nk} \ge 1, X_{nk} = 0) &= 0 \\ P(T_{nk} \ge 2) &= 1 - e^{-p_{nk}} - p_{nk} e^{-p_{nk}} . \end{aligned}$$

Show that

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