Statistics 581, Problem Set 3

Wellner; 10/17/2001

Reading: Lehmann & Casella, TPE, pages 54-61 and pages 75-78.
Ferguson, ACILST, pages 1 - 60.
Due: Wednesday, October 25, 2001.

1. Ferguson, ACILST, page 18, problem 3: Suppose that X_n is a sequence of random variables with densities f_n and X is a random variable with density f with respect to a common dominating measure μ, and that $f_n(x) \to f(x)$ as $n \to \infty$ for all x. Show that
$$Eg(X_n) \to Eg(X)$$
for all bounded measurable functions g.

2. Suppose that Y is a random variable with $E(Y^2) < \infty$.
(a) Show that
$$Var(Y) = E\{Var(Y|X)\} + Var\{E(Y|X)\}\,.$$
(b) Suppose that X and Y are random variables defined on the probability space $(\Omega, \mathcal{A}, P)$, and let $\mathcal{D}$ be the $\sigma-$field generated by X; $\mathcal{D} = \sigma[X] \equiv X^{-1}(\mathcal{B})$ where $\mathcal{B}$ is the Borel sigma-field for the real line R. Let $L_2(\Omega, \mathcal{D}, P)$ be the space of all random variables which are $\mathcal{D}$ measurable and Show that $Y - E(Y|X) \perp L_2(\Omega, \mathcal{D}, P)$ in the sense that
$$E\{(Y - E(Y|X))Z\} = 0$$
for all $Z \in L_2(\Omega, \mathcal{D}, P)$.
[Hint: You may use the following fact: if $Z \in L_2(\Omega, \mathcal{D}, P)$, then $Z = g(X)$ for some measurable function g from R to R with $E\{g^2(X)\} < \infty$.]
(c) Interpret (a) and (b) geometrically.
(d) Suppose that $Y \sim \chi^2_n(\delta)$. Compute $E(Y)$ and $Var(Y)$.

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. Suppose that $X, X_1, \ldots, X_n$ are i.i.d. with mean μ, variance σ^2, and $E|X|^4 < \infty$.
(a) Show that the sample variance $S_n^2 = \sum_{i=1}^n (X_i - \overline{X}_n)^2/(n-1)$ satisfies
$$\sqrt{n}(S_n^2 - \sigma^2)/\sqrt{2}\sigma^2 \to_d N(0, 1 + \gamma_2/2).$$
where $\mu_4 \equiv E(X-\mu)^4$ and $\gamma_2 \equiv \mu_4/\sigma^4 - 3$ is called the *excess of kurtosis*.
(b) Suppose that you want to test $H : \sigma \le \sigma_0^2$ versus $K : \sigma^2 > \sigma_0^2$ for σ_0 a fixed number, and you base your test on normal theory, but in fact the X's are *not normal* with $\gamma_2 \neq 0$. What effect does this have on the level (or size or actual type one error) of the normal theory test?

5. Suppose that $X_1, \ldots, X_n$ are independent Poisson(λ) random variables (so $P(X_1 = k) = e^{-\lambda}\lambda^k/k!$, $k = 0, 1, \ldots$).
(a) Show that $\sqrt{n}(\overline{X}_n - \lambda) \to_d N(0, \text{“}something\text{''})$.
(b) Show that the sequence $\{\sqrt{n}|\overline{X}_n - \lambda|\}$ is uniformly integrable and find $\lim_{n\to\infty} E(\sqrt{n}|\overline{X}_n - \lambda|)$.
(c) Let $g(x) = x^\gamma$ for $x \ge 0$ and $0 < \gamma < \infty$. Show that $\sqrt{n}(g(\overline{X}_n) - g(\lambda)) \to_d N(0, V^2)$ and compute V^2 explicitly in terms of λ and γ. For what γ is V^2 constant in λ? Is this the value of γ that makes $g(\overline{X}_n)$ “most nearly normal”?

6. **Optional bonus problem:** Ferguson, ACILST, problem 5, page 18:
Let $X_{n1}, \ldots, X_{nn}$ be independent, X_{nk} ~Bernoulli(p_{nk}), and let Y_n ~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}) 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

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