Midterm Examination
Statistics 200C

Ferguson — Monday, May 14, 2007

1. Suppose that $X_1, X_2, \ldots$ are two-valued random variables with $\mathrm{P}(X_n = n) = p_n$ and $\mathrm{P}(X_n = 0) = 1 - p_n$. Under what conditions on the p_n is it true that

(a) $X_n \xrightarrow{P} 0$?

(b) $X_n \xrightarrow{q.m.} 0$?

(c) $X_n \xrightarrow{a.s.} 0$?

2. Suppose $X_1, X_2, \ldots$ are independent with $\mathrm{P}(X_j = a_j) = \mathrm{P}(X_j = -a_j) = 1/2$ for all j, where $a_1, a_2, \ldots$ is a bounded sequence of numbers satisfying $\sum_1^n a_j^2 \to \infty$ as $n \to \infty$. Note that $\mathrm{E}X_j = 0$ for all j. Let $S_n = \sum_1^n X_j$ and let $B_n^2 = \mathrm{Var}(S_n)$.

(a) What is the Lindeberg condition?

(b) Show using the Lindeberg condition that $S_n/B_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$.

3. Suppose we observe i.i.d. random variables, $X_1, \ldots, X_n$ from a population with mean μ_x and variance σ_x^2. Suppose for each i we also observe $Y_i = \beta X_i + e_i$, where β is an unknown constant, and $e_1, \ldots, e_n$ are i.i.d. random variables with mean 0 and variance σ_e^2, independent of $X_1, \ldots, X_n$.

(a) Using the multivariate central limit theorem, find the asymptotic joint distribution of $\overline{X}_n$ and $\overline{Y}_n$, including the asymptotic covariance matrix.

(b) Suppose that $\mu_x > 0$, and consider the estimate of β given by $\hat{\beta}_n = \overline{Y}_n/\overline{X}_n$. Find the asymptotic distribution of $\hat{\beta}_n$.

4. Let $X_1, X_2, \ldots$, be i.i.d, taking values "Red", "White", and "Blue" with probability 1/3 each. Let Y_i be 1 if $X_i = \text{red}, X_{i+1} = \text{white}, X_{i+2} = \text{blue}$, and let $Y_i = 0$ otherwise.

(a) Explain why the Y_i form an m-dependent stationary sequence (for what m?).

(b) Find the asymptotic distribution (suitably normalized) of $\overline{Y}_n$.

5. In sampling from a population of $N = 3n$ objects having values $z_1, z_2, \ldots, z_N$, first a sample of size n is taken without replacement. Later a second sample of size n is taken from the remaining $N - n$ objects without replacement. The difference of the means of the two samples is used to compare the samples. This leads to a rank statistic of the form $S_N = \sum_1^N z_j a(R_j)$, where $a(i) = 1$ for $i = 1, \ldots, n$, $a(i) = -1$ for $i = n+1, \ldots, 2n$, and $a(i) = 0$ for $i = 2n+1, \ldots, N = 3n$.

(a) Give the mean and the variance of S_N.

(b) Under what condition on the z_i is it true that $(S_N - \mathrm{E}S_N)/\sqrt{\mathrm{Var}(S_N)} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$?

Solutions to the Midterm Examination
Statistics 200C

Ferguson
Spring 2007

1. (a) $X_n \xrightarrow{P} 0$ if and only if $p_n \to 0$ as $n \to \infty$.
(b) $X_n \xrightarrow{q.m.} 0$ if and only if $\mathrm{E}X_n^2 = n^2 p_n \to 0$ as $n \to \infty$.
(c) $X_n \xrightarrow{a.s.} 0$ if (and only if, provided the X_n are independent) $\sum_1^\infty p_n < \infty$.

2. (a) The Lindeberg condition for independent X_{nj} with mean zero and variance σ_{nj}^2 is for every $\epsilon > 0$,

$$\frac{1}{B_n^2}\sum_{j=1}^n \mathrm{E}(X_{nj}^2 \mathrm{I}(|X_{nj}| > \epsilon B_n) \to 0$$

(b) We apply the Lindeberg condition with $X_{nj} = X_j$. We are given $\sum_1^n a_i^2 \to \infty$ as $n \to \infty$, and $a_1, a_2, \ldots$ is bounded, say $|a_j| < C$ for all j. We have $\mathrm{E}X_j = 0$ and $\mathrm{Var}X_j = a_j^2$. Hence for $S_n = \sum_1^n X_j$, we have $B_n^2 = \mathrm{Var}(S_n) = \sum_1^n a_j^2$. The Lindeberg condition holds:

$$B_n^{-2}\sum_1^n \mathrm{E}(X_j^2\mathrm{I}(|X_j| > \epsilon B_n)) = B_n^{-2}\sum_1^n a_j^2 \mathrm{I}(|a_j| > \epsilon B_n)) \quad \text{since } X_j = |a_j| \text{ w.p. } 1$$

$$\leq B_n^{-2}\sum_1^n a_j^2 \mathrm{I}(C > \epsilon B_n)) = \mathrm{I}(C > \epsilon B_n) \to 0. \quad \text{since } |a_j| < C \text{ and } B_n \to \infty.$$

We may conclude $S_n/B_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$.

3. (a) $(X_1, Y_1), (X_2, Y_2), \ldots$ are i.i.d. with mean μ and covariance matrix Σ where

$$\mu = \begin{pmatrix} \mu_x \\ \beta\mu_x \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_x^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \beta^2\sigma_x^2 + \sigma_e^2 \end{pmatrix}$$

Therefore, $\sqrt{n}((\overline{X}_n, \overline{Y}_n)^T - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma)$.

(b) With $g(x,y) = y/x$, we have $g(\mu) = \beta$, $\dot{g}(x,y) = (-y/x^2, 1/x)$, and $\dot{g}(\mu) = (-\beta/\mu_x, 1/\mu_x)$. So,

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \dot{g}(\mu)\Sigma\dot{g}(\mu)) = \mathcal{N}(0, \sigma_e^2/\mu_x^2).$$

4. (a) The distribution of $(Y_1, \ldots, Y_n)$ and the distribution of $(Y_{t+1}, \ldots, Y_{t+n})$ are the same for every t and n, so the sequence is stationary. The sets of variables $\{Y_1, \ldots, Y_n\}$ and $(Y_{n+3}, Y_{n+4}, \ldots\}$ are independent since the former involves only $X_1, \ldots, X_{n+2}$ and the latter only involve $X_{n+3}, \ldots$ which are independent, so the sequence is 2-stationary.

(b) We have $EY_1 = 1/27$, $\mathrm{Var}(Y_1) = 1/27(1 - 1/27)$, $\mathrm{Cov}(Y_1, Y_2) = -1/27^2$ and $\mathrm{Cov}(Y_1, Y_3) = -1/27^2$. Hence,

$$\sqrt{n}(\frac{1}{n}S_n - \frac{1}{27}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \frac{1}{27}(1 - \frac{1}{27}) - 2\frac{1}{27^2} - 2\frac{1}{27^2}) = \mathcal{N}(0, \frac{22}{27^2})$$

This uses the theorem that states: for a stationary m-dependent sequence, $Y_0, Y_1, Y_2, \ldots$ with finite variance,

$$\sqrt{n}(\overline{Y}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_{00} + 2\sigma_{01} + \cdots + 2\sigma_{0m}),$$

where $\mu = \mathrm{E}Y_0$, and $\sigma_{0j} = \mathrm{Cov}(Y_0, Y_j)$ for $j = 0, 1, \ldots, m$.

5. (a) Since $\bar{a}_N = 0$, we have $\mathrm{E}S_N = 0$. The variance of S_N is $(N^2/(N-1))\sigma_z^2\sigma_a^2$, and since $\sigma_a^2 = (1/N)\sum_1^N a(i)^2 = 2/3$, we have $\mathrm{Var}(S_N) = (2/3)(N^2/(N-1))\sigma_z^2$.
(b) For asymptotic normality of S_N, we need

$$\frac{\max_j (z_j - \bar{z}_N)^2 \max(a(j) - \bar{a}_N)^2}{N\sigma_z^2\sigma_a^2} \to 0.$$

We have $\max_j (a(j) - \bar{a}_N)^2 = 1$, and $\sigma_a^2 = 2/3$. The above condition becomes

$$\frac{\max_j (z_j - \bar{z}_N)^2}{\sum_1^n (z_i - \bar{z}_i)^2} \to 0.$$