Midterm Examination
Statistics 200C

Ferguson — Friday, May 9, 2008

1. Suppose that $X_1, X_2, \ldots$ are i.i.d. random variables with the uniform distribution on the interval $[0,+1]$ (density $f(x) = \mathrm{I}(0 < x < +1)$. Let $\overline{X}_n = (1/n)\sum_1^n X_j$ and $T_n = (1/n)\sum_1^n \mathrm{I}(X_j < 1/2)$. Thus, T_n is the proportion of X_j's that are less than 1/2.

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

(b) What is the asymptotic distribution of $\overline{X}_n/T_n$?

2. Let $X_1, X_2, \ldots$ be i.i.d. exponential random variables with mean 1 and variance 1 (density $f(x) = e^{-x}\mathrm{I}(x > 0)$). Let $Y_j = \sqrt{i}(X_i - 1)$ for all i.

(a) Assuming the Lindeberg condition is satisfied for $S_n = \sum_1^n Y_i$, what is the asymptotic distribution of $\overline{Y}_n$?

(b) Show that the Lindeberg condition is satisfied for S_n.

3. Consider a multinomial distribution with c cells, sample size n and probability vector $\boldsymbol{p} = (p_1, \ldots, p_c)$. Let n_j denote the number of observations that fall in cell j for $j = 1, \ldots, c$.

(a) What is Pearson's chi-square for testing the hypothesis that the probability vector of the multinomial distribution is a fixed vector $\boldsymbol{p}$? What is its asymptotic distribution under this hypothesis?

(b) Find the transformed chi-square that replaces each cell frequency by its logarithm.

(c) What is the approximate large sample distribution of this transformed chi-square if the true value of $\boldsymbol{p}$ is $\boldsymbol{q}$?

4. (a) Give an example of a sequence $X_1, X_2, \ldots$ that is m-dependent (for some m) but not stationary.

(b) Give an example of a sequence $X_1, X_2, \ldots$ that is stationary but not m-dependent (for any m).

(c) If $Y_1, Y_2, \ldots$ are i.i.d. random variables with mean 0 and variance σ^2, and if $X_i = Y_i \cdot Y_{i+1} + Y_{i+2}$, what is the asymptotic distribution of $\overline{X}_n$?

5. Let $X_1, \ldots, X_n$ be i.i.d. from a Pareto distribution on $(0, \infty)$, with $F(x) = x/(x+\theta)$ for $x > 0$, where $\theta > 0$ is a scale parameter.

(a) Find the asymptotic distribution of the median, $X_{(\lceil n/2 \rceil)}$.

(b) Find the asymptotic joint distribution of $X_{(\lceil n/2 \rceil)}$ and $X_{(\lceil 3n/4 \rceil)}$, the median and the third quartile.

Solutions to the Midterm Examination Statistics 200C

Ferguson Spring 2008

1. (a) $\mathrm{E}X_j = 1/2$, $\mathrm{EI}(X_j < 1/2) = 1/2$, $\mathrm{Var}(X_j) = 1/12$, $\mathrm{Var}(\mathrm{I}(X_j < 1/2)) = 1/4$, and $\mathrm{Cov}(X_j, \mathrm{I}(X_j < 1/2)) = \int_0^{1/2} x\,dx - (1/4) = (1/8) - (1/4) = -1/8$. So,

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - 1/2 \\ T_n - 1/2 \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1/12 & -1/8 \\ -1/8 & 1/4 \end{pmatrix}\right)$$

(b) $g(x,y) = x/y$, so $\dot{g}(x,y) = (1/y, -x/y^2)$, and $\dot{g}(1/2,1/2) = (2,-2)$. So

$$\sqrt{n}(\frac{\overline{X}_n}{T_n} - 1) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \frac{1}{3} + 1 + 1) = \mathcal{N}(0, \frac{7}{3}).$$

2. (a) $\mathrm{E}Y_i = 0$ and $\mathrm{Var}(Y_i) = i$, so $B_n^2 = \sum_1^n i = n(n+1)/2$. If the Lindeberg condition holds, then $S_n/B_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$, where $S_n = \sum Y_i$. Since $B_n^2 \sim n^2/2$, we have $\overline{Y}_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/2)$.

(b)
$$\begin{aligned} \frac{1}{B_n^2}\sum_{i=1}^n \mathrm{E}(Y_i^2 \mathrm{I}(|Y_i| > B_n\epsilon)) &= \frac{1}{B_n^2}\sum_{i=1}^n i\mathrm{E}((X_i-1)^2\mathrm{I}((X_i-1)^2 > \frac{B_n^2\epsilon^2}{i})) \\ &\le \frac{1}{B_n^2}\sum_{i=1}^n i\mathrm{E}((X_i-1)^2\mathrm{I}((X_i-1)^2 > \frac{B_n^2\epsilon^2}{n})) \\ &= \mathrm{E}((X_1-1)^2\mathrm{I}((X_1-1)^2 > \frac{(n+1)\epsilon^2}{2})). \end{aligned}$$

since the X_i are identically distributed. This tends to zero as $n \to \infty$ since the second moment of X_1 is finite. Thus, the Lindeberg condition holds.

3. (a) $\chi^2_{\mathrm{P}} = n\sum_1^c (\hat{p}_j - p_j)^2/p_j$, where $\hat{p}_j = n_j/n$. $\chi^2_{\mathrm{P}} \xrightarrow{\mathcal{L}} \chi^2_{c-1}$, the chi-square distribution with $c-1$ degrees of freedom.

(b) With $g(p) = \log(p)$ and $g'(p) = 1/p$, the transformed chi-square becomes

$$\chi^2_{\mathrm{T}} = n\sum_1^c \frac{(\log(\hat{p}_j) - \log(p_j))^2}{g'(p_j)^2 p_j} = n\sum_1^c p_j(\log(\hat{p}_j) - \log(p_j))^2.$$

(c) $\chi^2_{\mathrm{T}} \sim \chi^2_{c-1}(\gamma^2)$, the non-central chi-square distribution with $c-1$ degrees of freedom and noncentrality parameter $\gamma^2 = n\sum_1^c p_j(\log(q_j) - \log(p_j))^2$ (or $\gamma^2 = n\sum_1^c (q_j - p_j)^2/p_j$).

4. (a) If $X_j \in \mathcal{N}(j, \sigma^2)$ are independent, then $X_1, X_2, \ldots$ are 0-dependent, but not stationary.

(b) Let Y_i be i.i.d. and let $X_j = \sum_{i=0}^\infty \beta^i Y_{j-i}$. Then $X_1, X_2, \ldots$ are stationary but not m-dependent for any m.

(c) The X_i are 2-dependent and stationary, with $\mathrm{E}X_i = 0$, $\sigma_{00} = \mathrm{Var}(X_1) = \mathrm{Var}(Y_1Y_2)$ $+\mathrm{Var}(Y_3) = \sigma^4 + \sigma^2$, $\sigma_{01} = \mathrm{Cov}(X_1, X_2) = 0$, and $\sigma_{02} = \mathrm{Cov}(X_1, X_3) = 0$. Therefore,

$$\sqrt{n}(\overline{X}_n - 0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_{00} + 2\sigma_{01} + 2\sigma_{02}) = \mathcal{N}(0, \sigma^4 + \sigma^2).$$

5. (a) The median satisfies $x/(x+\theta) = 1/2$ and so is $x_{1/2} = \theta$. The density is $f(x) = F'(x) = \theta/(x+\theta)^2$, so $f(x_{1/2}) = f(\theta) = 1/(4\theta)$. Therefore, $\sqrt{n}(X_{\lceil n/2 \rceil} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, (1/4)/(1/(4\theta))^2) = \mathcal{N}(0, 4\theta^2)$.

(b) The third quartile satisfies $x/(x+\theta) = 3/4$ and so is $x_{3/4} = 3\theta$. Since $f(3\theta) = 1/(16\theta)$, we have

$$\sqrt{n}\left(\begin{pmatrix} X_{\lceil n/2 \rceil} - \theta \\ X_{\lceil 3n/4 \rceil} - 3\theta \end{pmatrix}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4\theta^2 & 8\theta^2 \\ 8\theta^2 & 48\theta^2 \end{pmatrix}\right)$$