B Math III, Second Semester, March 09, 2007, Midterm

Answer any six questions. All questions carry equal weight.

1. (a). Suppose that $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$. Show that if $f(x, y)$ is a continuous function, then

$$f(X_n, Y_n) \xrightarrow{p} f(X, Y).$$

(b). If in addition $P(Y = 0) = 0$, then show that

$$\frac{X_n}{Y_n} \xrightarrow{p} \frac{X}{Y}.$$

2. Suppose that $\xi_1, \xi_2, ...$are independent and identically distributed random variables such that $E[\xi_1] = \mu$ and $E\left[\xi_1^2\right] < \infty$. Define random variables recursively

$$X_k = \rho X_{k-1} + \xi_k, \quad k \geq 1, \text{ and } X_0 = 0$$

where

$$|\rho| < 1.$$

Let

$$\overline{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k.$$

(a). Show that

$$E\left[\overline{X}_n\right] \to \frac{\mu}{1-\rho}$$

and for any $l \geq 0$,

$$\text{Cov}(X_{k+l}, X_k) = \rho^l \left(1 + \rho^2 + ... + \rho^{2k-2}\right).$$

(b). Show that

$$\overline{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow{p} \frac{\mu}{1-\rho}.$$

3 Let $(X_1, ..., X_n)$ be a random sample with a common mean μ and a common variance σ^2. Suppose that $g(y)$ is differentiable such that

$$g'(\mu) \neq 0,$$

where $g'(\mu)$ is the first derivative of $g(y)$ at $y = \mu$. Recall that this means, if we let $R(\mu, h)$ to be such that $g(\mu + h) - g(\mu) - hg'(\mu) = hR(\mu, h)$, then

$$\sup_{0<|h|\leq\varepsilon} |R(\mu, h)| \to 0 \text{ as } \varepsilon \to 0.$$

(a). Show that $R\left(\mu, \left(\overline{X}_n - \mu\right)\right) \xrightarrow{p} 0$ and $\sqrt{n}\left(\overline{X}_n - \mu\right) R\left(\mu, \left(\overline{X}_n - \mu\right)\right) \xrightarrow{p} 0$.

(b). Using (a), show that $\sqrt{n}\left(g\left(\overline{X}_n\right) - g\left(\mu\right)\right)$ converges in distribution to a normal distribution with mean 0 and variance $\sigma^2 \left(g'\left(\mu\right)\right)^2$.

(If you are applying CLT in (a) and/or (b), explain.)

(c). Let S_n be binomially distributed with parameters n and $p, 0 < p < 1$. Deduce from part (b) that $\sqrt{n}\left(\left(\frac{S_n}{n}\right)^2 - p^2\right)$ converges in distribution to a normal distribution. Give the explicit expression for the variance of the limiting normal distribution.

4. (a). Show that if the random vector $\mathbf{V} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}_k - \boldsymbol{\phi}\boldsymbol{\phi}'\right)$ where $\mathbf{I}_k$ is a unit matrix of order k and $\boldsymbol{\phi}' = \left(\sqrt{\pi_1}, ..., \sqrt{\pi_k}\right)$ with $\pi_j > 0$, $j = 1, ..., k$, and $\pi_1 + ... + \pi_k = 1$, then the quadratic $\mathbf{V}'\mathbf{V}$ has a $\chi^2\left(k-1\right)$ distribution with $k-1$ degrees of freedom.

(b). Consider a multinomial population with k (mutually exclusive) outcomes, with the corresponding probabilities $\pi_1, ..., \pi_k$. We have a random sample of size n from this population. Let N_j = the number of times the jth outcome occurs in the sample, $j = 1, ..., k$. Define

$$\chi_n^2 = \sum_{j=1}^{k} \frac{\left(N_j - n\pi_j\right)^2}{n\pi_j}.$$

Show that, when $\pi_1, ..., \pi_k$ are the true population probabilities, χ_n^2 converges in distribution to a $\chi^2\left(k-1\right)$ distribution with $k-1$ degrees of freedom.

5. The population is as in question 4 (b). For testing the simple null hypothesis

$$\pi_1 = \pi_{10}, \quad ,\pi_{k-1} = \pi_{k-1,0},$$

consider the likelihood ratio statistic

$$\lambda_n = \frac{l_n\left(\boldsymbol{\pi}_0\right)}{\sup_{\boldsymbol{\pi}} l_n\left(\boldsymbol{\pi}\right)},$$

where $\boldsymbol{\pi} = \left(\pi_1, ..., \pi_k\right)'$ and

$$l_n\left(\boldsymbol{\pi}\right) = \pi_1^{N_1} ... \pi_k^{N_k}$$

with N_j as defined in question 4 (b) based on a random sample of size n.

(a). Show that $\sup_{\boldsymbol{\pi}} l_n\left(\boldsymbol{\pi}\right) = l_n\left(\widehat{\boldsymbol{\pi}}_n\right)$ where $\widehat{\boldsymbol{\pi}}_n = \left(\frac{N_1}{n}, ..., \frac{N_k}{n}\right)'$.

(b). Show that, under the null hypothesis

$$-2\log\lambda_n - \sum_{j=1}^{k} \frac{\left(N_j - n\pi_{j0}\right)^2}{n\pi_{j0}} \xrightarrow{p} 0.$$

6. Consider two discrete populations with respective cell probabilities $\boldsymbol{\pi}_1 = (\pi_{11}, ..., \pi_{1k})$ and $\boldsymbol{\pi}_2 = (\pi_{21}, ..., \pi_{2k})$. Define the column vectors

$$\boldsymbol{\phi}_1 = \left(\sqrt{\pi_{11}}, ..., \sqrt{\pi_{1k}}\right)', \qquad \boldsymbol{\phi}_2 = \left(\sqrt{\pi_{21}}, ..., \sqrt{\pi_{2k}}\right)'.$$

Let the vector $\mathbf{V}_1$ be k-variate normal with mean vector $\mathbf{0}$ and variance-covariance matrix $\mathbf{I}_k - \boldsymbol{\phi}_1\boldsymbol{\phi}_1'$. Similarly let $\mathbf{V}_2$ be k-variate normal with mean vector $\mathbf{0}$ and variance-covariance matrix $\mathbf{I}_k - \boldsymbol{\phi}_2\boldsymbol{\phi}_2'$. Assume that $\mathbf{V}_1$ and $\mathbf{V}_2$ are independent.

Let

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \end{bmatrix} = (\mathbf{V}_1', \mathbf{V}_2')',$$

and

$$\boldsymbol{\phi} = \begin{bmatrix} \boldsymbol{\phi}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\phi}_2 \end{bmatrix}_{2k\times 2}.$$

Then

(a).Show that $\mathbf{V}$ is $2k$-variate normal with mean vector $\mathbf{0}$ and variance-covariance matrix $\mathbf{I}_{2k} - \boldsymbol{\phi}\boldsymbol{\phi}'$.

(b). Show that if the quadratic $\mathbf{V}'\mathbf{C}\mathbf{V}$ is such that

$$\mathbf{C}^2 = \mathbf{C} \qquad \text{and} \qquad \mathbf{C}\boldsymbol{\phi} = \boldsymbol{\phi},$$

then $\mathbf{V}'\mathbf{C}\mathbf{V}$ has a $\chi^2(l)$ distribution with the degrees of freedom l given by

$$l = \text{Rank}\,(\mathbf{C}) - 2.$$

(c). Generalize the preceding two population situation to the case of m populations.

7 Consider two multinomial populations, each with the same number of outcomes or cells, with respective cell probabilities $\boldsymbol{\pi}_1 = (\pi_{11}, ..., \pi_{1k})$ and $\boldsymbol{\pi}_2 = (\pi_{21}, ..., \pi_{2k})$. We want to test the null hypothesis

$$(\pi_{11}, ..., \pi_{1k}) = (\pi_{21}, ..., \pi_{2k}) = (\pi_1, ..., \pi_k)\,, \text{ say.}$$

Suppose that we have a sample of size n_1 from the first population, and denote the corresponding sample frequencies by $N_{11}, ..., N_{1k}$. Similarly, let $N_{21}, ..., N_{21k}$ be the sample frequencies corresponding to the sample of size n_2 from the second population.

An appropriate Chi-squared statistic is

$$\chi_n^2 = \sum_{j=1}^{k} \frac{\left(N_{1j} - n_1\widehat{\pi}_j\right)^2}{n_1\widehat{\pi}_j} + \sum_{j=1}^{k} \frac{\left(N_{2j} - n_2\widehat{\pi}_j\right)^2}{n_2\widehat{\pi}_j}$$

where

$$\widehat{\pi}_j = \frac{N_{1j} + N_{2j}}{n_1 + n_2}.$$

(a). Assuming for simplicity that $n_1 = n_2 = n$, show that χ_n^2 has asymptotically a $\chi^2(k-1)$ with $k-1+k-1-k = k-1$ degrees of freedom.

(b). Generalize the preceding parts (a) to the situation of m populations, and show that the resulting χ_n^2 statistic will have asymptotically a $\chi^2(l)$ with $l = (m-1)(k-1)$ degrees of freedom.