BMath III, Stat IV, Dec. 04, 2008, Final.

Answer any five of the six questions. All questions carry equal weight. Read the questions carefully.

1. Consider a multinomial population such that its probabilities $\pi_1 = \pi_1(\boldsymbol{\theta}), ..., \pi_k = \pi_k(\boldsymbol{\theta})$ are known functions of the parameter $\boldsymbol{\theta}$ where $\boldsymbol{\theta} = (\theta_1, ..., \theta_q)$ is a q-vector, $q < k-1$. Assume that we have sample of size n.

Let $\widehat{\boldsymbol{\theta}}_n$ be an estimator of $\boldsymbol{\theta}$ such that $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) - \left(M'_{\boldsymbol{\theta}} M_{\boldsymbol{\theta}}\right)^{-1} M'_{\boldsymbol{\theta}} \mathbf{V}_n \xrightarrow{p} 0$

where $M_{\boldsymbol{\theta}} = \left[\frac{1}{\sqrt{\pi_j(\boldsymbol{\theta})}} \frac{\partial \pi_j(\boldsymbol{\theta})}{\partial \theta_s}\right]_{k \times q}$ (assumed to be of rank q), where

$$\mathbf{V}'_n = \left(\frac{N_1 - n\pi_1(\boldsymbol{\theta})}{\sqrt{n\pi_1(\boldsymbol{\theta})}}, ..., \frac{N_k - n\pi_k(\boldsymbol{\theta})}{\sqrt{n\pi_k(\boldsymbol{\theta})}}\right)$$

with N_j = the number of times the jth outcome occurs in the sample, $j = 1, ..., k$..Let

$$\chi^2_n = \sum_{j=1}^{k} \frac{\left(N_j - n\pi_j\left(\widehat{\boldsymbol{\theta}}_n\right)\right)^2}{n\pi_j\left(\widehat{\boldsymbol{\theta}}_n\right)}.$$

Show that when $\boldsymbol{\theta}$ is the true parameter,

$$\chi^2_n \quad - \quad \mathbf{V}'_n \left(\mathbf{I}_k - M_{\boldsymbol{\theta}} \left(M'_{\boldsymbol{\theta}} M_{\boldsymbol{\theta}}\right)^{-1} M'_{\boldsymbol{\theta}}\right) \mathbf{V}_n \xrightarrow{p} 0.$$

Show, when $\boldsymbol{\theta}$ is the true parameter, that χ^2_n converges in distribution to the $\chi^2(l)$ with l degrees of freedom. Find the degrees of freedom l.

(Here you may use the fact that when $\mathbf{V} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_k - \boldsymbol{\phi}\boldsymbol{\phi}')$, the quadratic $\mathbf{V}'\mathbf{C}\mathbf{V}$ under certain restrictions on $\mathbf{C}$ has a χ^2 distribution with a certain degrees of freedom)

2. 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 $\boldsymbol{\pi}_2 = (\pi_{21}, ..., \pi_{2k})$. Suppose that it is known that $(\pi_{11}, ..., \pi_{1k}) = (\pi_1(\boldsymbol{\theta}_a), ..., \pi_k(\boldsymbol{\theta}_a))$ where $\boldsymbol{\theta}_a = (\theta_{a1}, ..., \theta_{aq})$ and $(\pi_{21}, ..., \pi_{2k}) = (\pi_1(\boldsymbol{\theta}_b), ..., \pi_k(\boldsymbol{\theta}_b))$ where $\boldsymbol{\theta}_b = (\theta_{b1}, ..., \theta_{bq})$. It is assumed that the functional forms of π_j's are known, but $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_b$ are unknown parameters. Thus the two populations are the same except for the difference in parameters $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_b$.

We want to test the null hypothesis

$$(\theta_{a1}, ..., \theta_{aq}) = (\theta_{b1}, ..., \theta_{bq}) = (\theta_1, ..., \theta_q) = \boldsymbol{\theta}.$$

Suppose we have a sample of size n from the first population and a sample of the same size n from the second population. Assume that the samples are independent. Let N_{1j} be the number of times the j-th outcome occurs in the sample from the first population and let N_{2j} be the number of times the j-th outcome occurs in the sample from the second population , $j = 1, ..., k$. Then an appropriate statistic is

$$\chi_n^2 = \sum_{j=1}^{k} \frac{n\left(\pi_j\left(\widehat{\theta}_{na}\right) - \pi_j\left(\widehat{\theta}_n\right)\right)^2}{\pi_j\left(\widehat{\theta}_n\right)} + \sum_{j=1}^{k} \frac{n\left(\pi_j\left(\widehat{\theta}_{nb}\right) - \pi_j\left(\widehat{\theta}_n\right)\right)^2}{\pi_j\left(\widehat{\theta}_n\right)}$$

where $\widehat{\boldsymbol{\theta}}_{na}$ is an asymptotic MLE of $\boldsymbol{\theta}_a$ (based on $N_{11}, ..., N_{1k}$), $\widehat{\boldsymbol{\theta}}_{nb}$ is an asymptotic MLE of $\boldsymbol{\theta}_b$ (based on $N_{21}, ..., N_{2k}$) and $\widehat{\boldsymbol{\theta}}_n$ is an asymptotic MLE of $\boldsymbol{\theta}$ under the null hypothesis (which will be based both on $N_{11}, ..., N_{1k}$ and $N_{21}, ..., N_{2k}$.)

Show that the χ_n^2 has the asymptotic $\chi^2(l)$ distribution with $l = q+q-q = q$ degrees of freedom.

(Here you may use the fact that if $\mathbf{V}_1 \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}_{k_1} - \phi_1\phi_1'\right)$ and $\mathbf{V}_2 \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}_{k_2} - \phi_2\phi_2'\right)$ with $\mathbf{V}_1$ and $\mathbf{V}_2$ independent, where $\phi_1 = \left(\sqrt{\pi_{11}}, ..., \sqrt{\pi_{1k}}\right)'$, $\phi_2 = \left(\sqrt{\pi_{21}}, ..., \sqrt{\pi_{2k}}\right)'$, then, letting $\mathbf{V} = \left(\mathbf{V}_1' \vdots \mathbf{V}_2'\right)'$, the quadratic $\mathbf{V}'\mathbf{C}\mathbf{V} \sim \chi^2(l)$ under suitable conditions on $\mathbf{C}$.)

3. Let $(X_1, ..., X_n)$ be a random sample from the population with cumulative distribution function $F(x)$. Assume that $F(x)$ is continuous in x. Let

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{(-\infty,x]}(X_i)$$

be the empirical cumulative distribution function based on the sample $(X_1, ..., X_n)$.

(a) Show that $\sup_{-\infty<x<\infty} |F_n(x) - F(x)| \xrightarrow{P} 0$.

(b) Let $(\eta(t); 0 \le t \le 1)$ be a Brownian Bridge, that is, a Gaussian process such that

$$E[\eta(t)] = 0 \quad \text{for all } 0 \le t \le 1$$

and

$$E[\eta(s)\eta(t)] = \min\{t, s\} - ts \quad \text{for all } 0 \le s, t \le 1.$$

Show that the process $(\sqrt{n}(F_n(x) - F(x)); -\infty < x < \infty)$ converges in distribution to the Gaussian process $(\eta(F(x)); -\infty < x < \infty)$, in the sense that for every finite $-\infty < x_1 < ... < x_k < \infty$, the random vector

$$\left(\sqrt{n}(F_n(x_i) - F(x_i));\ \ i = 1, ..., k\right) \Longrightarrow (\eta(F(x_i));\ \ i = 1, ..., k).$$

4. Let

$$S = \left\{\frac{j}{2^l}, j = 0, 1, ..., 2^l,\ l = 0, 1, ...\right\}.$$

For a given random process $(H(t), 0 \le t \le 1)$, suppose you are given that, for every positive integers $m < m_0$,

$$\sup_{|t-s|\le \frac{1}{2^m},\ t,s\in S} |H(t) - H(s)|$$

$$\le\ 2\sum_{l=m+1}^{m_0} \sup_{0\le q<2^l} \left| H\left(\frac{q+1}{2^l}\right) - H\left(\frac{q}{2^l}\right) \right| + 2 \sup_{|u-v|\le \frac{1}{2^{m_0}},\ t,s\in S} |H(u) - H(v)|.$$

Then show that

$$\lim_{\delta\to 0}\ \limsup_{n\to\infty} P\left[\sup_{|t-s|\le\delta} |\eta_n(t) - \eta_n(s)| > \tau\right] = 0 \quad \text{for all } \tau > 0,$$

where

$$\eta_n(t) = \sqrt{n}\left(F_n^*(t) - t\right), \quad ,0 \le t \le 1,$$

with $F_n^*(t)$ the empirical distribution function corresponding to a sample from the uniform distribution over the interval $(0, 1)$.

5. Let $X_1, ..., X_n$ be i.i.d. with the probability density function $f(x)$. Assume that the distribution of X_1 is symmetric around 0, that is, $f(x) = f(-x)$ for all x.

Let $R_1^+, ..., R_n^+$ be the ranks of $|X_1|, ..., |X_n|$.

Show that the vectors $(\text{sign}(X_1), ..., \text{sign}(X_n))$ and $\left(R_1^+, ..., R_n^+\right)$ are independent.

Show that the Wilcoxon statistic $a_n^{-1}\sum_{i=0}^n R_i^+ \text{ sign}(X_i)$ converges in distribution to the standard normal distribution, where $a_n^2 = \sum_{k=1}^n k^2$.

6. Let $(X_1, ..., X_n)$ be an i.i.d. sample from a population with cumulative distribution function $F(x)$. Let ξ_p be the population p-th quantile and let $\widehat{\xi}_{np}$ be the corresponding sample p-th quantile, $0 < p < 1$. Assume that $F(x)$ is continuously differentiable in the neighborhood of ξ_p.

Show that

$$\sqrt{n}\left(\widehat{\xi}_{np} - \xi_p\right) - \frac{1}{f(\xi_p)}\sqrt{n}\left(p - F_n(\xi_p)\right) \xrightarrow{P} 0,$$

where $f(x) = \frac{dF(x)}{dx}$. (If you use any general result from which you obtain this result, then you will have to state and prove that result.)