BMath.III, Second Semester, 2006. Stat IV. Backup Final

Answer any four of the six questions. All questions carry equal weight.

1. Consider a multinomial population such that its probabilities $\pi_1 = \pi_1(\boldsymbol{\theta}), ..., \pi_k = \pi_k(\boldsymbol{\theta})$ are 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 $\widehat{\boldsymbol{\theta}}$ be an estimator of $\boldsymbol{\theta}$ such that $\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) - \left(-M'_{\boldsymbol{\theta}} M_{\boldsymbol{\theta}}\right)^{-1} M'_{\boldsymbol{\theta}} \mathbf{V}_n \xrightarrow{p} 0$ 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) and

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

Show that (with N_j as in question 1 above), if

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

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

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

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.

2. Suppose that the individuals of a population are described as belonging to one of r categories $A_1, ..., A_r$ with respect to an attribute A, and to one of s categories $B_1, ..., B_r$ with respect to another attribute B. Let π_{ij} be the probability of the occurrence of the joint category $A_i B_j$.

If we have a sample of size n from this population, then we have the respective frequencies N_{ij} corresponding to the categories $A_i B_j$.

We want to test the null hypothesis that the attributes are independent, that is, π_{ij} satisfy the constraints

$$\pi_{ij} = \pi_{i\cdot} \pi_{\cdot j}$$

where $\pi_{i\cdot}$ and $\pi_{\cdot j}$ are marginal probabilities of the events A_i and B_j respectively. For this purpose, the following χ^2_n statistic is used:

$$\chi^2_n = \sum_{j=1}^{s} \sum_{i=1}^{r} \frac{\left(N_{ij} - \frac{N_{i\cdot} N_{\cdot j}}{n}\right)^2}{\frac{N_{i\cdot} N_{\cdot j}}{n}},$$

where

$$N_{i\cdot} = \sum_{j=1}^{s} N_{ij} \quad \text{and} \quad N_{\cdot j} = \sum_{i=1}^{r} N_{ij}.$$

Show that χ_n^2 has asymptotically a $\chi^2(l)$ with l degrees of freedom. Find the degrees of freedom l.

3. Let $(X_1, ..., X_n)$ be a random sample from the population with distribution function F and let $(Y_1, ..., Y_m)$ be a random sample from the population with distribution function G. Assume that the sample $(X_1, ..., X_n)$ is independent of $(Y_1, ..., Y_m)$.

(a). Combine the two samples, and let $R_1, ..., R_n$ be the respective ranks of $X_1, ..., X_n$ and let $R_{n+1}, ..., R_{n+m}$ be the respective ranks of $Y_1, ..., Y_m$ in the combined sample. Show that $\sum_{i=n+1}^{n+m} R_i$, the combined ranks of the second sample, satisfies

$$\sum_{i=n+1}^{n+m} R_i = U_{n,m} + \frac{1}{2} m (m+1)$$

where

$$U_{n,m} = \text{ the number of pairs } (X_i, Y_j) \text{ such that } X_i < Y_j.$$

(b). Find the variance of $U_{n,m}$.

(c). Show that $\frac{U_{n,m}}{nm}$ converges in probability to $p = P[X_1 < Y_1] = E[F(Y_1)]$, as $n, m \to \infty$.

4. Let $(X_1, ..., X_n)$ be a random sample from the population with cumulative distribution function $F(x)$. Let

$$F_n(x) = \sum_{i=1}^{n} I_{(-\infty, x]}(X_i)$$

be the empirical or sample cumulative distribution function based on the sample $(X_1, ..., X_n)$. Assume that $F(x)$ is continuous in x.

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

(a) Show that the process $(\sqrt{n}(F_n(x) - F(x)); -\infty < x < \infty)$ converges in distribution to the Gaussian process $(\eta(F(x)); -\infty < x < \infty)$, where

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

and

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

5. Let $X_1, ..., X_n$ be i.i.d. sample from a population with cumulative distribution function having mean μ and variance σ^2. Show that

$$\frac{1}{\sqrt{n}} \sum \left| X_k - \overline{X}_n \right| - \frac{1}{\sqrt{n}} \sum |X_k - \mu| - \sqrt{n}\left(\overline{X}_n - \mu\right) \lambda_F'(\mu) \xrightarrow{p} 0$$

where $\lambda_F(t) = E_F[|X_1 - t|]$.

Using this deduce the asymptotic distribution of $\sqrt{n}\left(\frac{\sum|X_k - \overline{X}_n|}{n} - \lambda_F(\mu)\right)$.

6. Suppose we want to estimate the mean θ of a normal population with known variance σ^2 on the basis of a sample $X_1, ..., X_n$. Assume that θ has a prior distribution that is normally distributed with mean μ and variance τ^2. Find the Bayes rule for estimating θ under the quadratic loss $l(\theta, a) = (\theta - a)^2$.

Show that the sample mean $\overline{X}_n = \frac{1}{n}(X_1 + ... + X_n)$ is an admissible estimator for θ under the quadratic loss $l(\theta, a) = (\theta - a)^2$. Further state and prove the result upon which you obtain this conclusion.