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

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

1. Consider a multinomial population with k (mutually exclusive) outcomes, with the corresponding probabilities $\pi_1, ..., \pi_k$ with $\pi_j > 0,\ j = 1, ..., k$, and $\pi_1 + ... + \pi_k = 1$. 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

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

Let

$$\lambda_n = \frac{L(\boldsymbol{\pi}_0)}{\sup_{\boldsymbol{\pi}} L(\boldsymbol{\pi})}$$

where $\boldsymbol{\pi} = (\pi_1, ..., \pi_k)'$ and $L(\boldsymbol{\pi}) = \pi_1^{N_1} ... \pi_k^{N_k}$. Show that when $\boldsymbol{\pi}_0 = (\pi_{10}, ..., \pi_{k0})'$ is the true parameter

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

and hence obtain the asymptotic distribution of $-2\log\lambda_n$.

2. 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

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

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

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

Show, when $\boldsymbol{\theta}$ is the true parameter, that χ_n^2 converges in distribution to the $\chi^2(l)$ 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 i.i.d. with cumulative distribution function $F(x)$. Assume that the distribution of X_1 is symmetric around 0, that is, $F(x) = 1 - F(-x)$ for all x, and $F(x)$ is continuous.

Let $S_1, ..., S_k$ be the ranks of $|X_1|, ..., |X_n|$. Let $R_k = S_k I_k$ where

$$I_k = \operatorname{sign}(X_k).$$

Show that the vectors $(I_1, ..., I_n)$ and $(S_1, ..., S_k)$ are independent.

Show that $a_n^{-1} \sum_{k=1}^n R_k$ converges in distribution to the standard normal distribution, where $a_n^2 = \sum_{k=1}^n k^2$.

Describe the testing problem in a matched pair model and how the use of the statistic $\sum_{k=1}^n R_k$ becomes appropriate in that situation.

5. Let $X_1, ..., X_n$ be i.i.d. with cumulative distribution function $F(x)$. Let $\psi(x, t)$ be monotone in t, that is, either nonincreasing or nondecreasing in t. Further let $\lambda_F(t) = E_F[\psi(X_1, t)]$.

Let t_0 be such that $\lambda_F(t_0) = 0$. Assume that t_0 is isolated, and $\lambda_F(t)$ is differentiable in t in a neighborhood of t_0. Let T_n be an estimator of t_0 such that

$$\sum \psi(X_k, T_n) = 0.$$

Show that $\sqrt{n}(T_n - t_0)$ converges in distribution to a normal distribution with mean 0 and variance

$$\sigma^2 = \frac{E_F\left[\psi^2(x, t_0)\right]}{\left(\lambda_F'(t_0)\right)^2}.$$

By choosing $\psi(x, t) = \phi(x - t)$ with

$$\phi(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ \frac{p}{1-p} & \text{if } x > 0, \end{cases}$$

show that t_0 such that $E_F\left[\psi\left(x, t_0\right)\right] = 0$ is the p-th quantile of F, and hence obtain the asymptotic normality of the sample p-th quantile.

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 $l\left(\theta, a\right) = \left(\theta - a\right)^2$.

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