Statistics 581, Final Exam

Wellner; 12/8/2000

1. (32) points) **Define** *each* of the following terms. In each case, provide an appropriate context for your definition.
   (a) The information matrix for θ in a regular parametric model $\mathcal{P} = \{P_\theta : \ \theta \in \Theta \subset R^d\}$.
   (b) The Hellinger distance between two probability measures P and Q.
   (c) The Kullback-Leibler information number $K(P, Q)$ between a probability hfill measure P and a (sub-)probability measure Q.
   (d) The efficient influence function $\tilde{l}_\nu$ for a differentiable parameter $q(\theta) = \nu(P_\theta)$ in a regular parametric model $\mathcal{P}$.

2. (40) **State** *each* of the following results, providing the appropriate (brief) context for your statement:
   (a) A basic result concerning the Kullback-Leibler "distance" $K(P, Q)$.
   (b) The dominated convergence theorem.
   (c) An identity which gives two ways of calculating the information matrix; specify the regularity conditions which are needed for the identity to hold.
   (d) The Lindeberg-Feller central limit theorem.
   (e) The Mann-Wald or continuous mapping theorem.

Do **either** problem 3 or problem 4.

3. (30 points)
   (a) **State** the Glivenko-Cantelli theorem. Then **prove** that it holds if it holds for the case of i.i.d. Uniform$(0, 1)$ random variables.
   (b) **Prove** the Glivenko-Cantelli theorem for i.i.d. Uniform$(0, 1)$ random variables: if $\xi_1, \ldots, \xi_n, \ldots$ are i.i.d. Uniform$(0, 1)$ with empirical distribution function

$$\mathbb{G}_n(t) = \frac{1}{n}\sum_{i=1}^n 1_{[0,t]}(\xi_i)\,, \qquad \text{then} \qquad \sup_{0 \le t \le 1} |\mathbb{G}_n(t) - t| \to_{a.s.} 0\,.$$

4. (30 points)
   Suppose that $\underline{N} \sim \text{Mult}_k(n, \underline{p})$; thus the probability distribution of $\underline{N}$ is given by

$$P_p(\underline{N} = \underline{m}) = \frac{n!}{m_1! \cdots m_k!}\prod_{j=1}^k p_j^{m_j}\,,$$

and the likelihood is

$$L_n(\underline{p}|\underline{N}) = \frac{n!}{N_1! \cdots N_k!}\prod_{j=1}^k p_j^{N_j}\,.$$

Show, without using any calculus, that the MLE of $\underline{p}$ is $\widehat{\underline{p}} = \underline{N}/n$.

5. (54 points) Suppose that $\underline{N}_n = (N_{11}, N_{12}, N_{21}, N_{22}) \sim \text{Mult}_4(n, \underline{p})$ where $\underline{p} = (p_{11}, p_{12}, p_{21}, p_{22})$ where $\sum_{i=1}^2 \sum_{j=1}^2 p_{ij} = 1$. (Thus $\underline{N}_n$ is the sum of n independent $\text{Mult}_4(1, \underline{p})$ random vectors $\{\underline{Y}_i\}_{i=1}^n$.) Since there are really just three independently varying parameters for this problem, it is often useful to re-express the cell probabilities in terms of two marginal probabilities, say $p_{1\cdot} = p_{11} + p_{12}$ and $p_{\cdot 1} = p_{11} + p_{21}$, and ψ, the log of the odds-ratio, defined by

$$\psi \equiv \log \frac{p_{21}/p_{22}}{p_{11}/p_{12}} = \log \frac{p_{12}p_{21}}{p_{11}p_{22}} .$$

You may use the fact that $\psi = 0$ if and only if independence holds for the 2×2 table (i.e. $p_{ij} = p_{i\cdot}p_{\cdot j}$ for $i, j = 1, 2$).
(a) Suggest an estimator of ψ, say $\widehat{\psi}$.
(b) Show that the estimator you proposed in (b) is asymptotically normal and compute the asymptotic variance of your estimator.
(c) One standard test of independence in the 2×2 table is the test based on a Pearson-type chi-square statistic. Write down the chi-square statistic Q_n for this problem, state its asymptotic distribution under the null hypothesis, and explain briefly why the claimed result holds.
(d) Another statistic for testing independence is the likelihood ratio statistic $2 \log \lambda_n$ for testing $H : \psi = 0$ versus $K : \psi \neq 0$. Find this statistic and state its asymptotic distribution under the null hypothesis.
(e) Without doing any additional calculation, what is the asymptotic distribution of the likelihood ratio statistic under local alternatives of the form $\psi_n = tn^{-1/2}$? (Hint: use the result of (b) to find an expression for the (efficient) information for ψ in the presence of the nuisance parameters $p_{1\cdot}$, $p_{\cdot 1}$.)
(f) Suppose that $\psi \neq 0$; i.e. the alternative hypothesis holds. Show that for the statistics Q_n and $2 \log \lambda_n$ from (d) and (e) we have $n^{-1} Q_n \rightarrow_p q$ and $n^{-1} 2 \log \lambda_n \rightarrow_p J$ for some positive constants q and J respectively; you should compute q and J as explicitly as possibly in terms of $\underline{p}$ and/or $(p_{1\cdot}, p_{\cdot 1}, \psi)$.

6. (54 points) (A parametric version of the Cox model). Suppose that $(Y|Z) \sim \text{Exponential}(\lambda e^{\gamma Z})$ where $Z \sim \text{Bernoulli}(\eta)$. Thus the density of $X = (Y, Z)$ is given by

$$p_\theta(y, z) = \lambda e^{\gamma z} \exp(-\lambda e^{\gamma z} y) 1_{[0,\infty)}(y) \eta^z (1 - \eta)^{1-z} 1_{\{0,1\}}(z)$$

where $\theta = (\gamma, \lambda, \eta)$. Suppose that $X_1 = (Y_1, Z_1), \ldots, X_n = (Y_n, Z_n)$ are i.i.d. as X.
(a) Find the scores for $\theta = (\gamma, \lambda, \eta)$ based on one observation.
(b) Find the information matrix for θ.
(c) Compute the information for γ when λ is known (I_{11}) and unknown ($I_{11\cdot 2}$), and explain the difference based on the geometry of the scores.
(d) Write down the score equations for θ and briefly discuss the existence and uniqueness of solutions of these equations.
(e) What does our theory from chapter 4 say about the limiting distribution of $\sqrt{n}(\widehat{\theta} - \theta_0)$ and of $\sqrt{n}(\widehat{\gamma} - \gamma_0)$?
(f) Consider testing $H : \gamma = 0$ versus $K : \gamma \neq 0$. Suggest three test statistics, and briefly discuss the pro's and con's of each. What is the asymptotic distribution of these test statistics under the null hypothesis and local alternatives.

7. (36 points)
(A parametric version of the Cox model, continued). The following problem is in the same context as that of problem 6 above:
Consider estimation of the parameter

$$q(\theta) = \exp(-\lambda e^{\gamma} y_0) = P_\theta(Y \geq y_0 | Z = 1) \equiv \nu(P_\theta)$$

for a fixed number $y_0 > 0$.
(a) Suggest a natural empirical estimator $\widehat{\nu}_n$ of this probability (taking care to note that it is a *conditional probability*).
(b) Show that $\widehat{\nu}_n$ is asymptotically linear, and find its influence function, ψ, explicitly.
(c) Find the efficient influence function $\tilde{l}_\nu$ for estimation of $q(\theta)$ and the related information bound (generically, in terms of the scores).
(d) Describe the relationship between ψ and $\tilde{l}_\nu$ geometrically.