UNIVERSITY OF MASSACHUSETTS
Department of Mathematics and Statistics
ADVANCED EXAM - Mathematical Statistics and Probability
Monday, August 28, 2006

Work all problems. 70 points are required to pass.

1. (25 points) Let $\mathbf{X}_1, \ldots \mathbf{X}_n$ be i.i.d. where $\mathbf{X}_i$ has a pmf or pdf in a regular k parameter exponential family in canonical form; that is with density $f_{\mathbf{X}_i}(\mathbf{x}_i|\boldsymbol{\theta}) = h(\mathbf{x}_i)d(\boldsymbol{\theta})exp(\sum_{j=1}^k \theta_j t_j(\mathbf{x}_i))$ (where the θ_j are functionally independent so the parameter space is of dimension k). Let $\mathbf{X}$ denote the collection of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with $\mathbf{x}$ defined similarly, and let $f_{\mathbf{X}}(\mathbf{x}, \boldsymbol{\theta}) = \prod_i f_{\mathbf{X}_i}(\mathbf{x}_i|\boldsymbol{\theta})$ denote the joint pdf/pmf of $\mathbf{X}$.

   (a) Find the score vector $S(\mathbf{x}, \boldsymbol{\theta}) = \partial f_{\mathbf{X}}(\mathbf{x}, \boldsymbol{\theta})/\partial \boldsymbol{\theta}$.

   (b) Show that $E(S(\mathbf{X}, \boldsymbol{\theta})) = \mathbf{0}$. Note that we are now considering the score vector as a random vector as a function of $\mathbf{X}$.

   (c) Use the previous part to find $E(t_j(\mathbf{X}))$ where $t_j(\mathbf{X}) = \sum_{i=1}^n t_j(\mathbf{X}_i))$.

   (d) Determine the $k \times k$ Hessian matrix $H(\mathbf{x}, \boldsymbol{\theta}) = \partial S(\mathbf{x}, \boldsymbol{\theta})/\partial \boldsymbol{\theta}$.

   (e) The information matrix for β, based on $\mathbf{X}$ is defined to be $I_n(\boldsymbol{\theta}) = E(S(\mathbf{x}, \boldsymbol{\theta})S(\mathbf{x}, \boldsymbol{\theta})')$.
   i) Show that here $I_n(\boldsymbol{\theta}) = -E(H(\mathbf{X}, \boldsymbol{\theta}))$.
   ii) Show that $I_n(\boldsymbol{\theta}) = n \cdot I_1(\boldsymbol{\theta})$ and give an expression for the components of $I_1(\boldsymbol{\theta})$ that involves the function $d(\boldsymbol{\theta})$ as well as its first and second derivatives.

2. (15 points) Let $X_1, \ldots, X_n$ be i.i.d. from $N(\mu, \sigma^2)$.

   (a) Suppose σ^2 is a known constant, and let $\pi(\mu) = c$ be an improper prior for μ. Find the posterior p.d.f. for μ. Find the Bayes estimate for μ.

   (b) Suppose σ^2 is unknown, and let $\pi(\mu, \sigma^2) = \sigma^{-2} I_{(0,\infty)}(\sigma^2)$ be an improper prior for (μ, σ^2). Show that the posterior p.d.f. of (μ, σ^2) given $x = (x_1, \ldots, x_n)$ is $\pi(\mu, \sigma^2|x) = \pi_1(\mu|\sigma^2, x)\pi_2(\sigma^2|x)$, where $\pi(\mu|\sigma^2, x)$ is $N(\overline{x}, \sigma^2/n)$ and $\pi_2(\sigma^2|x)$ is inverse gamma. (The inverse gamma p.d.f. is $f(z; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{-(\alpha+1)} \exp\left(\frac{-\beta}{z}\right), z > 0$.)

3. (30 points) Let (X_i, Y_i) be iid pairs for $i = 1, \ldots, n$.

   (a) A predictor of Y based on X is some function of X, say $g(X)$. The best unbiased predictor of X is the predictor $g(X)$ which i) has $E[g(X)] = E(Y)$ and ii) minimizes $E[(Y - g(X))^2]$. Prove that $g(X) = E(Y|X)$ is the best unbiased predictor of Y. (Note that this question is completely divorced from the remaining parts.)

   (b) Define the 5×1 vector $\mathbf{Z}_i$ with $\mathbf{Z}_i' = [X_i, Y_i, X_i^2, Y_i^2, X_iY_i]$. State the multivariate central limit theorem and use it to give $\boldsymbol{\mu}_Z$ and Σ_Z such that

   $$n^{1/2}(\bar{\mathbf{Z}}_n - \mu_Z) \Rightarrow N(\mathbf{0}, \Sigma_Z),$$

   where $\Rightarrow$ denotes convergence in distribution and $\bar{\mathbf{Z}}_n = \sum_{i=1}^n \mathbf{Z}_i/n$.

   (c) Consider the random variable $W_n = g(\bar{\mathbf{Z}}_n)$, where g is continuous function, differentiable at $\boldsymbol{\mu}$.

i. Use the limiting distribution result given in b) to argue that W_n is a consistent estimator of $g(\boldsymbol{\mu})$.

ii. Argue that $n^{1/2}(W_n - g(\boldsymbol{\mu}_Z)) \Rightarrow N(\mathbf{0}, \mathbf{d}'\Sigma_Z\mathbf{d}$ (Hint: Use the multivariate version of Taylor's theorem). Describe the components of $\mathbf{d}$.

iii. Use the result of part ii) to find the approximate large sample distribution of the sample correlation coefficient

$$r = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\left[\sum_i (X_i - \bar{X})^2 \sum_i (Y_i - \bar{Y})^2\right]^{1/2}}.$$

You don't need to write our or simplify $\mathbf{d}'\Sigma_Z\mathbf{d}$, but be sure to explain clearly how the components of $\mathbf{d}$ are obtained in this case.

4. (15 points)(a) Calculate the characteristic function of a Poisson random variable ξ, where

$$P(\xi = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots,$$

and λ is a positive constant.

(b) For each integer $n \geq 1$ let the independent random variables $\xi_{n1}, \ldots, \xi_{nn}$ be such that

$$P(\xi_{nk} = 1) = p_{nk} \text{ and } P(\xi_{nk} = 0) = q_{nk}, \quad p_{nk} + q_{nk} = 1.$$

Assume that as $n \to \infty$,

$$\max_{1 \leq k \leq n} p_{nk} \to 0 \text{ and } \sum_{k=1}^{n} p_{nk} \to \lambda > 0.$$

Prove that for each non-negative integer m we have that

$$P(\xi_{n1} + \xi_{n2} + \ldots + \xi_{nn} = m) \to \frac{e^{-\lambda}\lambda^m}{m!}, \quad n \to \infty.$$

5. (15 points) (a) Assume ξ_n, $n = 1, 2, \ldots$ is a sequence of Gaussian random variables, $\xi_n \sim \mathcal{N}$ (μ_n, σ_n) such that

$$|\mu_n| \leq C < \infty, \quad |\sigma_n| \leq C < \infty.$$

Show the sequence of corresponding probability measures is tight.

(b) Show that the limiting measure from part (a) is still Gaussian.

Hint: Recall the density of the Gaussian distribution is $(2\pi\sigma_n^2)^{-1/2} \exp\left(-(x - \mu_n)^2 / 2\sigma_n^2\right)$ and its characteristic function is $\phi_{\xi_n}(t) = \exp\left(it\mu_n - \frac{1}{2}t^2\sigma_n^2\right)$.