Stats 270/370 Homework 6

Due Wednesday, Feb. 17

Problem 1

Sufficient data reduction and exponential families: Let $(\mathcal{X}, \mathcal{B}, \lambda)$ be a probability space, where $\mathcal{X}$ is a region (an open connected set) in $\mathbb{R}^p$ and λ is Lebesgue measure. $\mathcal{P}$ is a family of distributions on $\mathcal{X}$ which are absolutely continuous with respect to λ and f_P denotes the density of $P \in \mathcal{P}$. Let $X_1, \ldots, X_n$ be n independent observations from P.

A family $\mathcal{P}$ is said to be exponential provided there exists a positive integer s, real functions $B, \eta_1, \ldots, \eta_s : \mathcal{P} \to \mathbb{R}$, and real measurable functions $h, \tau_1, \ldots, \tau_s : \mathcal{X} \to \mathbb{R}$, such that

$$f_P(x) = \exp\left[\sum_{i=1}^{s} \eta_i(P)\tau_i(x) - B(P)\right] h(x) \text{ a.e. } [\lambda]. \tag{1}$$

The smallest s which admits a representation of f_P of the form (1) is the *order* of $\mathcal{P}$.

We shall prove:

Theorem 1. *Suppose the probability densities f_P are strictly positive and continuous on the sample space $\mathcal{X} \subset \mathbb{R}^p$ and $n \geq 2$. If T is a continuous, one-dimensional, sufficient statistic on $\mathcal{X}^n$, then $\mathcal{P}$ is exponential of order 1.*

We only focus on the case when $p = 1$, and divide the proof into the following parts. Part (i) and Part (ii) are general statements, and we do not impose the conditions of Theorem 1 for these two parts.

(i) Let C be the space of real continuous functions on $\mathcal{X}$, and P_0 be an arbitrary but fixed element of $\mathcal{P}$. Define $\varphi_P = \log(f_P/f_{P_0})$, and S to be the set of those $\varphi \in C$ for which there exists a function ψ such that

$$\varphi(x_1) + \cdots + \varphi(x_n) = \psi(T(x^{(n)})),\ x^{(n)} \in \mathcal{X}^n.$$

[Note that S depends on T.] Show that T is sufficient if and only if $\varphi_P \in S$, for all $P \in \mathcal{P}$.

(ii) Let V stand for the linear subspace of C spanned by the constant functions and the functions φ_P, $P \in \mathcal{P}$. Show that $\mathcal{P}$ is exponential of order s if and only if $\dim V = s + 1$.

(iii) *Bonus problem.* Assume $p = 1$. Let $\varphi \in S$ and let x_0, y_1, y_2 be points in $\mathcal{X}$, if $\varphi(y_1) = \varphi(y_2)$ and $T(x_0, y_1) < T(x_0, y_2)$ then φ is constant in a neighborhood of x_0.

(iv) Show that $\dim S \leq 2$. Use this to show Theorem 1 holds when $p = 1$. (*hint*: take any φ_1 and $\varphi_2 \in S$ and use part (iii) to show that there exists constants a and b such that $\varphi_2 = a\varphi_1 + b$.)

Remark. (i) The result still holds when $p > 1$.

(ii) If T is a k-dimensional sufficient statistic, then similar results can also be obtained. Rigorously speaking, under the assumptions of Theorem 1 ($n > k$ instead of $n > 2$ here), if the densities f_P have continuous partial derivatives then $\mathcal{P}$ is exponential of order $\leq k$.

(iii) This theorem states that, under certain regularity conditions on the densities f_P, if there exists a sufficient statistic $T = t(X_1, \ldots, X_n)$ which yields a reduction of the data, then the family $\mathcal{P}$ is exponential.

Problem 2

Conditional inference. Suppose an experiment is conducted to measure a parameter θ. Independent unbiased measurements y of θ can be made with either of two instruments, both of which measure with normal errors: for $j = 1, 2$, instrument j produces independent errors with a $N(0, \sigma_j^2)$ distribution. The two error variances σ_1^2 and σ_2^2 are known. When a measurement y_i is made, a record is kept of the instrument used so that after n measurements the data is $(a_1, y_1), , (a_n, y_n)$, where $a_i = j$ if y_i is obtained using instrument j. The choice between instruments is made independently for each observation in such a way that

$$P(a_i = 1) = p = 1 - P(a_i = 2)$$

for some known p. Let n_1 and n_2 denote the number of times instrument 1 and 2 were used, respectively, such that $n_1 = \sum_{i=1}^n 2 - a_i$ and $n_2 = n - n_1$.

(a) Show that the maximum likelihood estimate of θ is given by

$$\hat{\theta} = \frac{\sum_{i=1}^n y_i/\sigma_{a_i}^2}{\sum_{i=1}^n 1/\sigma_{a_i}^2}.$$

(b) Express the expected Fisher information for θ in terms of n, the σ_i^2, and p.

(c) Compute the conditional variance of $\hat{\theta}$ given $(a_1, \ldots, a_n)$. For inference, should one use the unconditional variance of $\hat{\theta}$ or the conditional variance of $\hat{\theta}$ given $(a_1, \ldots, a_n)$? If the unconditional variance is used to form a confidence interval, what do you expect to be true about the coverage probability conditional on n_1 and n_2?