Stat 581 Homework 9: Due December 10, 2003

1. Suppose that $X_1, ..., X_n$ are i.i.d. with Weibull density $c\theta^{-c}x^{c-1}\exp(-(x/\theta)^c)I_{(0,\infty)}(x)$.

(a) Show that $(X/\theta)^c$ is a standard exponential r.v.

(b) Find the Information, $I(\theta)$, when c is known.

(c) When (c, θ) are both unknown, find the Rao (Score) test of the hypothesis $c = 1$. (You may here **assume** the information matrix given in JAW notes P. 3.13).

(d) Why would you use the Rao (Score) test in this example?

2. In a sample from a trinomial distribution, the cell counts are (n_1, n_2, n_3), $n_1 + n_2 + n_3 = n$, and the three cell probabilities are $(\theta, \theta^2 + \theta^4, 1 - \theta - \theta^2 - \theta^4)$, where $0 < \theta < \frac{1}{2}$.

(a) Show that n_1/n is a consistent asymptotically Normal estimator of θ.

(b) Find the information matrix for θ, as a (messy) function of θ.

(c) Show that n_1/n is not an asymptotically efficient estimator of θ.

(d) Find the one-step estimator of θ based of the initial estimator n_1/n.

3. Individuals can be one of four blood types A, B, O and AB. These types are determined by the types of the two alleles carried by an individual– known as the genotype of the individual. There are three types of alleles, A, B and O, with frequencies p, q and r, $p + q + r = 1$.

Blood type	Genotype	Frequency	Blood types	Genotype	Frequency
A	AA	p^2	A	AO	$2pr$
B	BB	q^2	B	BO	$2qr$
O	OO	r^2	AB	AB	$2pq$

Thus the frquencies of the four observable blood types are $p^2 + 2pr$, $q^2 + 2qr$, r^2 and $2pq$.

A sample of N individuals is observed, with n_A, n_B, n_O and n_{AB} of the four types, and (p, q, r) is to be estimated.
(a) By considering the data as an incomplete observation from the sample counts of the 6 genotypes, derive an EM-algorithm for obtaining the MLE of (p, q, r).
(b) Estimate (p, q, r) when $n_A = 182$, $n_B = 60$, $n_O = 176$ and $n_{AB} = 17$.
(c) Estimate the covariance matrix of the MLE. Hint: Use parameter (p, q) with $r = 1 - p - q$.

4. Suppose X_i, $i = 1, \ldots, n$, are i.i.d. from the mixture distribution

$$\theta P(\lambda) \; + \; (1-\theta)P(\mu) \;\; 0 \le \theta \le 1, \lambda \; > \; 0 \; , \quad \mu \; > \; 0 \; .$$

where $P(\lambda)$ denotes the Poisson distribution mean λ.

(a) Let $Z_i = e_1 = (1, 0)'$ if X_i is from the first component $P(\lambda)$ and let $Z_i = e_2 = (0, 1)'$ if X_i is from the second component $P(\mu)$. Let $S_j = \{i : \; Z_i \; = \; e_j\}$, $j \; = \; 1, 2$. Show that the natural sufficient statistics for (θ, λ, μ) for the family of distributions of the "complete data" $\{(X_i, \; Z_i); \; i = 1, \ldots, n\}$ are $(\sum_{S_1} X_i, \; \sum_{S_2} \; X_i, \; \sum_{S_1} \; 1)$

(b) Hence construct the EM algorithm equations for obtaining the MLE of (θ, λ, μ) from a sample $x_1, ..., x_n$.

(c) Investigate briefly the performance of your algorithm when
(i) $n = 7$, $\mathbf{x} = (0, 1, 2, 6, 7, 9, 10)$
(ii) $n = 7$, $\mathbf{x} = (3, 3, 5, 5, 6, 6, 7)$.

5. A simple model for a quantitative trait (e.g. height) on a set of k related individuals gives rise to a multivariate Normal distribution for the observed trait values: $\mathbf{y}$ is distributed as $N_k(\mathbf{0}, \boldsymbol{\sigma^2}\mathbf{G} + \boldsymbol{\tau^2}\mathbf{I})$ where $\mathbf{G}$ is a known positive definite matrix, and $\mathbf{I}$ is the identity matrix. An EM-type formulation (which actually has a "real" interpretation for the latent variables $\mathbf{z}$) is $\mathbf{y} = \mathbf{z} + \mathbf{e}$ where $\mathbf{z}$ is $N_k(\mathbf{0}, \sigma^2\mathbf{G})$, and $\mathbf{e}$ is $N_k(\mathbf{0}, \tau^2\mathbf{I})$.

(a) Show that if $\mathbf{z}$ and $\mathbf{y}$ could be observed, then the MLEs of σ^2 and τ^2 would be $k^{-1}\mathbf{z}'\mathbf{G}^{-1}\mathbf{z}$ and $k^{-1}(\mathbf{y}-\mathbf{z})'(\mathbf{y}-\mathbf{z})$

(b) Let $\mathbf{V} = \mathrm{var}(\mathbf{y}) = \boldsymbol{\sigma^2}\mathbf{G} + \boldsymbol{\tau^2}\mathbf{I}$. Show

$$\begin{aligned}
\mathrm{var}(\mathbf{z} \mid \mathbf{y}) &= (\sigma^{-2}\mathbf{G}^{-1} + \tau^{-2}\mathbf{I})^{-1} = \sigma^2\tau^2\mathbf{V}^{-1}\mathbf{G} \\
\mathbf{a} &= \mathrm{E}(\mathbf{z} \mid \mathbf{y}) = \sigma^2\mathbf{V}^{-1}\mathbf{G}\mathbf{y}
\end{aligned}$$

(c) Hence show

$$\begin{aligned}
\mathrm{E}(\mathbf{z}'\mathbf{G}^{-1}\mathbf{z} \mid \mathbf{y}) &= \mathbf{a}'\mathbf{G}^{-1}\mathbf{a} + \sigma^2\tau^2\mathrm{tr}(\mathbf{V}^{-1}) \\
\mathrm{E}((\mathbf{y}-\mathbf{z})'(\mathbf{y}-\mathbf{z}) \mid \mathbf{y}) &= (\mathbf{y}-\mathbf{a})'(\mathbf{y}-\mathbf{a}) + \sigma^2\tau^2\mathrm{tr}(\mathbf{V}^{-1}\mathbf{G})
\end{aligned}$$

and write down the EM equations to find the MLE of (σ^2, τ^2) given data $\mathbf{y}$.