Stat 581 Homework 7: Due November 24, 2004

1. (a) Suppose $X_1, ..., X_n$ are i.i.d. $U(\theta - \psi, \theta + \psi)$. Find the minimal sufficient statistic for (θ, ψ) and the MLE of (θ, ψ).

(b) Suppose $X_1, ..., X_n$ are i.i.d. double exponential, each with pdf $f_\theta(x) = \frac{1}{2}\exp(-|x - \theta|)$. Find the minimal sufficient statistic and the MLE of θ.

2. Suppose $X_1, ..., X_n$ are i.i.d. from a distribution symmetric about a location parameter θ and with a strictly positive density over the range of X_i. It is proposed to estimate θ by the average of the pth and $(1-p)$th sample quantiles, $T_n^{(p)} = \frac{1}{2}(F_n^{-1}(p) + F_n^{-1}(1-p))$, where F_n is the empirical distrbution function of $X^{(n)} = (X_1, ..., X_n)$.
(a) Show that the sequence $(T_n^{(p)})$ is consistent for θ, for $0 < p \leq \frac{1}{2}$.
(b) Compare the ARE of sequences of estimators $(T_n^{(p)})$ of θ for varying p – the answer will depend on the density function. How should p be chosen to maximize asymptotic efficiency?
(c) Evaluate your answer to 2(b) for the two examples of 1(a) and 1(b).

3. Suppose that $X_1, ..., X_n$ are i.i.d. from the Poisson distribution with mean $\theta > 0$:

$$P(X_1 = k) \quad = \quad \exp(-\theta)\theta^k/k! \qquad k = 0, 1, 2, 3, ...$$

Let $\widehat{\theta_n} \;=\; \overline{X_n} = n^{-1}\sum_{i=1}^n X_i$.
Two alternative estimators of $P(X = 1) = \theta e^{-\theta}$ are proposed:

$$\widehat{P_{1,n}} \;=\; \widehat{\theta_n}\exp(-\widehat{\theta_n}) \quad \text{and} \quad Z_n = n^{-1}\sum_{i=1}^n I(X_i = 1).$$

(a) Let $W_i \;\equiv\; (X_i, I(X_i = 1))$. Note $(\overline{X_n}, Z_n) \;=\; n^{-1}\sum_{i=1}^n W_i \equiv \overline{W_n}$.
Find the limiting distribution of $n^{\frac{1}{2}}(\overline{W_n} - (\theta, \theta e^{-\theta}))$.

(b) Hence find the limiting distribution of $n^{\frac{1}{2}}((\widehat{P_{1,n}}, Z_n) - (\theta e^{-\theta}, \theta e^{-\theta}))$.

(c) Find the asymptotic relative efficiency (A.R.E.) of the sequence (Z_n) relative to $(\widehat{P_{1,n}})$ as estimators of $\theta e^{-\theta}$. Which sequence of estimators is more efficient?

4. Let $(X_i, Y_i), i = 1, ..., n$ be i.i.d. bivariate normal, with $\mathrm{E}(X_i) = \mathrm{E}(Y_i) = 0$, $\mathrm{var}(X_i) = \mathrm{var}(Y_i) = 1$ and $\mathrm{Cov}(X_i, Y_i) = \rho$. Find the likelihood equation for estimation of ρ. Show that it always (with probability 1) has at least one solution in (-1,1), and that the solution is unique for large enough values of n.

5. Under a random genetic drift model for the changes of frequencies of certain types of genes as a finite-size population evolves, certain functions of these allele frequencies may be assumed to undergo independent Brownian motions. Hence, if initially the values of these frequency functions is $(\theta_1, \ldots, \theta q)$, after a period of time, and after taking a random sample from the current population, the sample values of these functions will be $(X_1, \ldots, X_q)$, where the X_i are independent and $X_i \sim N(\theta_i, \sigma_x^2)$, where σ_x^2 depends on the population size, the sample size, and the time period over which the population evolves.

Thompson (1973: Ann.Hum Genet. 37: 69-80) used this model to estimate the proportions of Norse and Celtic individuals in the founding population of Iceland in about 900 A.D. For simplicity we will here assume that the relative σ^2 values for the three populations are known and equal (not realistic), and so we can scale to take $\sigma^2 = 1$. Thus for the samples from the current Scandinavian (formerly Norse) population we have $X_i \sim N(\theta_i, 1)$; for the current Irish (Celtic) population samples we have $Y_i \sim N(\psi_i, 1)$; and for the Icelandic sample we have $Z_i \sim N(r\theta_i + (1-r)\psi_i, 1)$, where $i = 1, \ldots, q$ and all components of all three q-vectors X, Y, and Z are independent. The parameter r (the proportion of Norse ancestry) is of interest, but of course θ and ψ are unknown.

(a) Define $W(r) = Z - rX - (1-r)Y$, and $h(r) = 1 + r^2 + (1-r)^2 \;=\; \mathrm{var}(W_i(r))$, $i = 1, \ldots, q$. Show that the likelihood equations may be written in the form

$$\begin{aligned} \theta - X &= rW(r)/h(r) \\ \psi - Y &= (1-r)W(r)/h(r) \\ (Z - r\theta - (1-r)\psi)^t(\theta - \psi) &= 0 \end{aligned}$$

(b) Hence show that

$$(Z - r\widehat{\theta} - (1-r)\widehat{\psi}) \;=\; W(r)/h(r)$$

and that maximization of the likelihood is equivalent to maxization of $S(r) = \; -W(r)^t W(r)/h(r)$.

(c) Hence or otherwise show that the likelihood equation for r is

$$g(r) \;=\; r^2(d_{xz}^2 - d_{yz}^2) \;+\; 2r(d_{yz}^2 - d_{xy}^2) \;+\; (d_{xy}^2 - d_{xz}^2) \;=\; 0$$

and that $S'(r)$ has the same sign as $g(r)$.
Here $d_{xy}^2 = (x-y)^t(x-y)$ is Euclidean distance in q-dimensional space, and you are reminded of the cosine formula $2(x-z)^t(y-z) \;=\; d_{xz}^2 + d_{yz}^2 - d_{xy}^2$.

(d) Hence find the MLE of r in $0 \le r \le 1$ in the three cases when $d_{xz} > d_{yz}$
(i) $d_{xz} \;>\; d_{xy} \;>\; d_{yz}$
(ii) $d_{xy} \;>\; d_{xz} \;>\; d_{yz}$
(iii) $d_{xz} \;>\; d_{yz} \;>\; d_{xy}$
(The three cases with $d_{yz} > d_{xz}$ would be given by symmetry, reversing the role of Norse and Celtic populations.)