LTCC/Reid:

Homework Problems for Lecture of 26/11/12; due 03/12/12

1. Suppose $y = y_1, \dots, y_n$ are independent and identically distributed from a distribution with density $f(y;\theta) = \prod_{i=1}^n f_1(y_i;\theta)$, $\theta \in R$. Further let $g(y;\theta) = \sum_{i=1}^n g_1(y_i;\theta)$ be an *unbiased estimating equation* for θ, satisfying $E_\theta\{g(y_i;\theta)\} = 0$ for all θ. The estimate defined by $g(y;\tilde{\theta}_g) = 0$ has asymptotic variance $G^{-1}(\theta) = \{H^{-1}(\theta)J(\theta)H^{-1}(\theta)\}^{-1}$, where $H(\theta) = -\mathrm{E}_\theta\{\nabla_\theta g(y_1;\theta)\}$ and $J(\theta) = \mathrm{var}_\theta\{g(y_1;\theta)\}$. The estimating equation is called *optimal* if it has the largest possible value of $G(\theta)$.

   Show that $G(\theta) \le i_1(\theta)$, where $i_1(\theta)$ is the expected Fisher information in a single observation. This implies that the score equation is the optimum estimating equation.

   Two fun facts that you don't need to prove:

   (a) The multivariate version of this is that $i_1(\underline{\theta}) - G(\underline{\theta})$ is non-negative definite (but you don't need to show this).

   (b) In the autoregressive model

   $$y_i = \theta y_{i-1} + \epsilon_i, \quad i = 1, \dots, n$$

   where y_0 is a constant and ϵ_i are i.i.d. $N(0, \sigma^2)$, show the equation

   $$\Sigma y_i y_{i-1} - n \Sigma y_i^2 = 0$$

   is an unbiased estimating equation obtaining the lower bound.

2. Suppose $Z \sim \sum_{r=1}^m \mu_r X_r^2$, where $X_1, \dots, X_m$ are independent observations from a $N(0,1)$ distribution. If all the μ_r were equal, the distribution of Z would be proportional to a χ^2_m. *Satterthwaite's approximation* (Satterthwaite, 1946) to the distribution of Z is $a\chi^2_b$, where a and b are chosen so that $\mathrm{E}(Z)$ and $\mathrm{var}(Z)$ are equal to the mean and variance of a $a\chi^2_b$ random variable. This idea can also be used to approximate a non-central χ^2 distribution, and arises in the distribution of quadratic forms in unbalanced analysis of variance.

   (a) Find expressions for a and b, in terms of $\mu_1, \dots, \mu_m$.

   (b) Illustrate the approximation numerically in a simple example with, say, $m = 5, 10$. You can choose the values of μ_r in any way you like, but one possibility is to simulate a random vector from $N(0, A)$ for some choice of $A \neq I$; then $X^T X$ will (I think), have the distribution you are looking for. The function `mvrnorm` in the `MASS` library simulates multivariate normal random variables.