STA 3000F (Fall, 2013)

Homework 1.(September 27; due October 11)

1. CH 2.2 Suppose that random variables Y_r follow the first order autoregressive process
$$Y_r = \mu + \rho(Y_{r-1} - \mu) + \epsilon_r,$$
where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. $N(0, \sigma^2)$ and $|\rho| < 1$. Write down the likelihood for data $y_1, \ldots, y_n$ in the cases where the initial value y_0 is

   (a) a given constant;

   (b) normally distributed with mean μ and variance $\sigma^2/(1-\rho^2)$;

   (c) assumed equal to y_n.

   Find the minimal sufficient statistic for $\theta = (\mu, \rho, \sigma^2)$ in each case.

2. SM 3.5.6 & 4.9.17

   (a) Suppose Z_1 and Z_2 are independent standard normal random variables. Let $Y = Z_1$ and $W = Z_2 - \lambda Z_1$. Show that the conditional density of Y, given that $W < 0$ is
$$f(y; \lambda) = 2\phi(y)\Phi(\lambda y),$$
where $\phi(\cdot), \Phi(\cdot)$ are the density and distribution function of a standard normal, respectively. This is called the 'skew-normal' density function, with skewness parameter λ. It may be helpful to sketch the density for a few values of λ. There is an `R` package called `sn` that makes this easy.

   (b) If $y_1, \ldots, y_n$ is a random sample with density $\sigma^{-1} f_0\{(y-\mu)/\sigma; \lambda\}$, where $f_0(\cdot; \lambda)$ is the skew-normal density function of part (a), write down the log-likelihood function for $\theta = (\mu, \sigma, \lambda)$. What is the minimal sufficient statistic for θ?

3. CH 2.11 Suppose that $Y_1, \ldots, Y_n$ are i.i.d. with probability density $f(y; \theta), y \in \mathbb{R}, \theta \in \mathbb{R}$, and that $T = t(Y)$ is a one-dimensional sufficient

statistic for θ for all values of n. If θ_1 and θ_2 are any two fixed values of θ, show that for any θ

$$\frac{\frac{\partial}{\partial y_i}\log\{f(y_i;\theta)/f(y_i;\theta_1)\}}{\frac{\partial}{\partial y_i}\log\{f(y_i;\theta_2)/f(y_i;\theta_1)\}}$$

is independent of y_i and hence is a function of θ alone. Since θ_1 and θ_2 are fixed, deduce from this that $f(y;\theta)$ is of exponential family form.

This result generalizes to samples of size n from a density with d-dimensional parameter θ: if the support of the density is independent of θ, and if there is a d-dimensional sufficient statistic, $d < n$, then the model is an exponential family. From this point of view sufficiency is somewhat specialized. This result is attributed to Darmois (1935), Pitman (1936) and Koopman (1937) in TSH (§2.7), but to Fisher (1934) and Dynkin (1961) in Pace, L. and Salvan, A. (1997). *Principles of Statistical Inference: a neo-Fisherian perspective.* World Scientific, Singapore. This book has lots of good explanations of parametric statistical inference.

4. Let $y_1, \ldots, y_n$ be independent observations from the regression-scale model
$$y_i = x_i^T\beta + \sigma\epsilon_i, \quad i = 1, \ldots, n$$
where x_i and β are $p \times 1$ vectors, and the density of ϵ_i is $f_0(\cdot)$. The usual estimates of β and σ are
$$\hat\beta = (X^TX)^{-1}X^Ty, \quad \hat\sigma = (y - X\hat\beta)^T(y - X\hat\beta)/(n-p),$$
where X is the matrix with ith row x_i^T, and $y = (y_1, \ldots, y_n)^T$.
   (a) Show that $a = (a_1, \ldots, a_n)$, with $a_i = (y_i - x_i^T\hat\beta)/\hat\sigma, i = 1, \ldots, n$ forms an $n-p-1$-dimensional ancillary statistic.
   (b) Show that $s = \hat\beta, s$ forms a $p+1$-dimensional sufficient statistic
   (c) If the ϵ_i follow a standard normal distribution, show that a and s are independently distributed.