Notes on Q1, Homework 1.

Question

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 $\mu$ 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.

Solution

The question in C&H actually reads, “Write down the likelihood for data $y_0, \ldots, y_n$ ...”; in this correctly worded version the solution goes as follows:

$$f(y_0, \ldots, y_n; \theta) = \frac{1}{(\sqrt{2\pi}\sigma)^{n-1}} \prod_{i=1}^n \exp\left[-\frac{1}{2\sigma^2} \left(y_i - \mu - \rho(y_{i-1} - \mu)\right)^2\right] f(y_0; \theta),$$

writing the joint density as a product of conditionals, and using the Markov property.

The exponent when squared involves the following statistics:

$$\sum_{i=1}^n y_i^2, \sum_{i=1}^n y_{i-1}^2, \sum_{i=1}^n y_i, \sum_{i=1}^n y_{i-1}, \sum_{i=1}^n y_i y_{i-1}. \quad (1)$$

Since $\sum_{i=1}^n y_{i-1} = \sum_{i=0}^{n-1} y_i$, the third and fourth terms can be computed from $\sum_{i=1}^n y_i, y_n, y_0$, and the 1st and 2nd terms can be computed from $\sum_{i=1}^n y_i^2, y_n, y_0$ as well. Thus (1) is equivalent to

$$\left(\sum_{i=1}^n y_i^2, \sum_{i=1}^n y_i, \sum_{i=1}^n y_i y_{i-1}, y_0, y_n\right). \quad (2)$$
In (a), \(f(y_0)\) is a point mass at \(y_0\), so no additional functions of the data are needed to compute the log-likelihood function, and no simplification is available either. In (b) \(f(y_0; \theta)\) depends on \(y_0^2\) and \(y_0\), but again these already appear (2) so no additional functions are needed, nor are there any simplifications. In (c), when \(y_0 = y_n\), so \(f(y_0)\) is a point mass at \(y_n\), then \(\sum_{i=0}^{n-1} y_i = \sum_{i=1}^{n} y_i\), and similarly for \(\sum_{i=1}^{n-1} y_i^2\), so that the endpoint corrections \(y_0, y_n\) are not needed, and we have a 3-dimensional sufficient statistic.

In (a) strictly speaking \(y_0\) is not part of the sufficient statistic, because it is a fixed constant, although the solution given by C&H doesn’t make this distinction.

In the version of the question that I gave, the joint density is

\[
f(y_1, \ldots, y_n; \theta) = \frac{1}{(\sqrt{2\pi}\sigma)^{n-1}} \prod_{i=2}^{n} \exp\left[-\frac{1}{2\sigma^2} \{y_i - \mu - \rho(y_{i-1} - \mu)\}^2\right] f(y_1; \theta),
\]

and \(f(y_1; \theta) = \int f(y_1 \mid y_0; \theta) f(y_0; \theta) dy_0\). The product term now depends on

\[
\sum_{i=2}^{n} y_i, \sum_{i=2}^{n} y_i^2, \sum_{i=2}^{n} y_{i-1}, \sum_{i=2}^{n} y_{i-1}^2, \sum_{i=2}^{n} y_i y_{i-1}. \tag{3}
\]

In the case that \(y_0\) is fixed, \(f(y_1)\) is the density of a \(N(\mu + \rho(y_0 - \mu), \sigma^2)\), so the joint density depends on

\[
\sum_{i=1}^{n} y_i^2, \sum_{i=1}^{n} y_i, \sum_{i=1}^{n} y_i y_{i-1}, y_0, y_n,
\]
as above, although again \(y_0\) isn’t strictly speaking a statistic.

In the case that \(y_0\) follows the stationary distribution given in (b), then so does \(y_1\), in which case the joint density has an exponent that depends on

\[
\sum_{i=1}^{n} y_i^2, \sum_{i=1}^{n} y_i, \sum_{i=1}^{n} y_i y_{i-1}, y_0, y_n, \tag{4}
\]

which is a bit different than for the likelihood based on \((y_0, \ldots, y_n)\).

Finally, in the case that \(y_0 = y_n\), we don’t get the simplification that we do for the likelihood based on \((y_0, \ldots, y_n)\), because the sums start
at 1. (We would if I had written “$y_1 = y_n$”.) David F showed that the marginal distribution of $y_1$ in this case is $N(\mu, \beta)$, where

$$\beta = \sum_k \rho^{2(n-k+1)} \sigma^2.$$  

This then contributes a term as in (b), and no simplification is possible.

I didn’t mark this very rigidly – most people wrote down $L(\theta; y_0, \ldots, y_n)$ without being very specific about it. Let me know if I overlooked anything in your solution.

These statistics are minimal sufficient because they determine the likelihood function.