Implementation of linear mixed model with AR(1) errors

Rasmus Waagepetersen

May 5, 2022

1 Model with AR(1) errors

Assume
$$Y \sim N(X\beta, V) \tag{1}$$
where V is the covariance matrix of a zero-mean stationary AR(1) process with parameters τ^2 and a (cf. handouts for lecture 2). Then V and V^{-1} have factorizations

$$V = BCB^\mathsf{T} \quad \text{and} \quad V^{-1} = (B^{-1})^\mathsf{T} C^{-1} B^{-1}$$

where C is diagonal and B^{-1} is zero except for the diagonal and the entries just below the diagonal. We can write $C = \tau^2 D$ where τ^2 is the variance of the innovations of the AR(1) process and D only depends on a. Let $V = \tau^2 W$ where $W = BDB^\mathsf{T}$. Let $S = D^{-1/2}B^{-1}$.

1. Show that $\tilde{Y} = SY \sim N(\tilde{X}\beta, \tau^2 I)$ where $\tilde{X} = SX$.
2. Show that the densities f and $\tilde{f}$ of Y and $\tilde{Y}$ are related by
$$f(y) = \tilde{f}(\tilde{y})|S|$$
3. Assume a is known. Argue that estimates of β and τ^2 based on the likelihood of Y coincides with estimates based on the likelihood of $\tilde{Y}$.
4. Write a piece of R code that for a given a produces the maximum likelihood estimates of β and τ^2 and returns the value of the log likelihood function (hint: transform Y and X into $\tilde{Y}$ and $\tilde{X}$ and apply the lm() function in R - see also example code)

5. Simulate a data set from the model (1) and try out your code on this to obtain maximum likelihood estimates of β, τ^2, a.

6. Conduct a simulation study to assess the distribution of the maximum likelihood estimates when $a = 0$, $a = 0.5$ and $a = 0.99$. Try small $n = 20$ and large $n = 1000$.

2 Model extended with independent noise (hidden Markov process)

We now extend the model (1) by adding independent normal errors each with variance σ^2. That is, we consider

$$Y \sim N(X\beta, \tau^2 BDB^\mathsf{T} + \sigma^2 I). \tag{2}$$

This corresponds to the general setting for which we developed maximum likelihood (and restricted maximum likelihood) estimation on the handouts for lecture 2. Note that in this case neither $\tau^2 BDB^\mathsf{T} + \sigma^2 I$ nor its inverse are sparse. Also I do not know how to obtain a square root of the inverse in an efficient manner. However, we can still come up with a computationally efficient implementation of maximum likelihood estimation.

Recall that $Q = (BDB^\mathsf{T})^{-1}$ is a sparse tri-diagonal matrix.

1. let $\phi = \tau^2/\sigma^2$ and show that

$$(\tau^2 BDB^\mathsf{T} + \sigma^2 I)^{-1} = \sigma^{-2}(\phi I + Q)^{-1} Q$$

2. show that the determinant of $|\tau^2 BDB^\mathsf{T} + \sigma^2 I|$ is $\sigma^{n2}|\phi I + Q|/|Q|$ where n is the dimension of Y.

3. one can compute (see accompanying R code) a Cholesky factorization LL^T of $\tilde{Q} = \phi I + Q$. Show that the determinant of $\tilde{Q}$ is the product of squared diagonal elements of L.

4. For a vector z computing Qz is of course straightforward. To compute $x = (\phi I + Q)^{-1} z$ note that this is equivalent to solving the equation $(\phi I + Q)x = z$. How can we apply the Cholesky factorization LL^T of the sparse matrix $(\phi I + Q)$ to solve $(\phi I + Q)x = z$ in an efficient manner (recall L is lower triangular) ?

5. Use the above results to implement maximum likelihood estimation of β and σ^2 given fixed values of a and ϕ. Next use this to implement profile likelihood estimation of a and ϕ (see also example R code).

3 Inference for noise variance

Consider the electricity consumption-temperature data.

1. Fit the models from Section 1-2 with Y equal to the electricity consumption and X the matrix with a first column of ones, a second column given by the temperatures, and a third column given by the binary D variable (weekday vs. weekend).

2. Conduct a likelihood ratio test for $H_0 : \sigma^2 = 0$. Use a parametric bootstrap to approximate the distribution of the likelihood ratio under H_0.