Inference in multivariate autoregressive process and its extensions

Pierre Gloaguen

2020-06-19

Contents

Chapter 1

Introduction

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 1. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter **??**.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

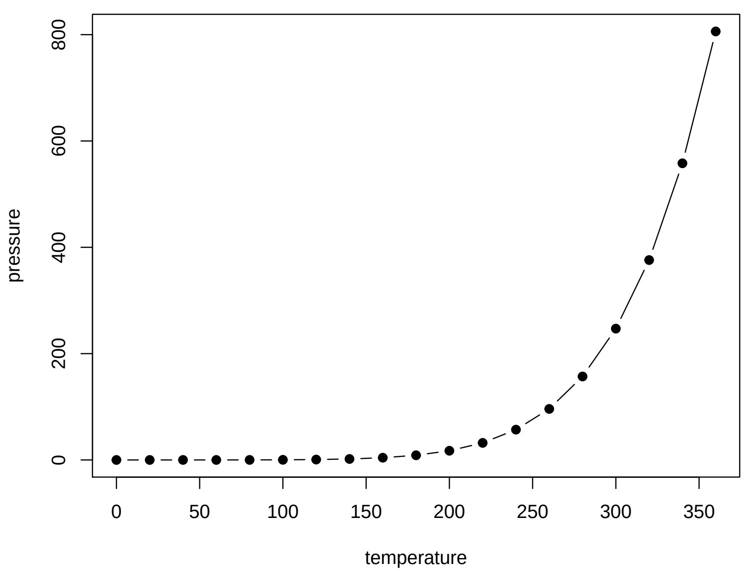

Figure 1.1: Here is a nice figure!

Table 1.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 1.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 1.1.

```r
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2020) in this sample book, which was built on top of R Markdown and **knitr** (**?**).

Chapter 2

Multivariate autoregressive model

In this chapter, we focus on the case where observations consist in a multivariate time series $y_0, \dots, y_n$ such that for any $0 \leq t \leq n$, $y_t \in \mathbb{R}^d$, we denote:

$$y_t = \begin{pmatrix} y_{t,1} \\ \vdots \\ y_{t,d} \end{pmatrix}$$

2.1 Model

We assume that these observations are realisations of random variables $Y_0, \dots, Y_n$ such that:

$$\begin{aligned} &Y_0 \sim \chi_0(\cdot), \\ &Y_t = m + \mathbf{A} Y_{t-1} + E_t,\ 1 \leq t \leq n \end{aligned} \tag{2.1}$$

where χ_0 is some probability distribution over $\mathbb{R}^d$, $m \in \mathbb{R}^d$ and $\mathbf{A} \in \mathcal{M}_{d\times d}$ are parameters, E_t is a $d-$dimensionnal vector such that:

$$E_t \overset{ind.}{\sim} \mathcal{N}\left(0, \mathbf{\Sigma}\right).$$

2.2 Inference

In this simple context, inference consists in finding the maximum likelihood estimates of unknown parameters[1] $\hat{m}$, $\hat{\mathbf{A}}$ and $\mathbf{\Sigma}$.

Inference is straightforward here as we can recognize in (2.1) a multivariate linear model:

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E},$$

where

$$\mathbf{Y} = \begin{pmatrix} Y_1' \\ \vdots \\ Y_n' \end{pmatrix} \in \mathcal{M}_{n\times d},$$

$$\mathbf{X} = \begin{pmatrix} 1 & Y_0' \\ \vdots & \\ 1 & Y_{n-1}' \end{pmatrix} \in \mathcal{M}_{n\times (d+1)},$$

$$\mathbf{B} = \begin{pmatrix} m' \\ A' \end{pmatrix} \in \mathcal{M}_{(d+1)\times d},$$

$$\mathbf{E} = \begin{pmatrix} E_1' \\ \vdots \\ E_n' \end{pmatrix} \in \mathcal{M}_{n\times d}.$$

Thus, $\hat{m}$ and $\hat{\mathbf{A}}$ can be obtained using the classical estimate

$$\hat{\mathbf{B}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}, \tag{2.2}$$

and $\hat{\mathbf{\Sigma}}$ is then obtained classicaly as:

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{m} - \hat{\mathbf{A}}Y_{t-1}\right)\left(Y_t - \hat{m} - \hat{\mathbf{A}}Y_{t-1}\right)' \tag{2.3}$$

[1] In the case where multiple time series are observed, unknown parameters for the initial distribution $\chi_0(\cdot)$ could be considered

Chapter 3

Switching autoregressive system

In this chapter, we focus on a more complex system involving autoregressive structure. We still focus on a time series $y_0, \dots, y_n$ of values in $\mathbb{R}^d$. However, it is now supposed that the time series dynamics could change through time, according to an unobserved stochastic process in a discrete space. This unobserved process might model different regimes of the dynamics (see Rabiner (1989) for selected applications).

3.1 Model

Taking the same notations and dimensions as in equation (2.1) We assume that these observations are realisations of random variables $Y_0, \dots, Y_n$ such that:

$$\begin{aligned} Z_0 &\sim \chi_{0,Z}(\cdot) \\ Z_t|Z_{t-1} &\sim p(z_t|Z_{t-1}) \\ Y_0 &\sim \chi_{0,Y}(\cdot|Z_0), \\ Y_t|Z_t &= m(Z_t) + \mathbf{A}(Z_t)Y_{t-1} + E_t,\ 1 \leq t \leq n \end{aligned}$$

where

- $\{Z_t\}_{0\leq t\leq n}$ is an homogeneous Markov chain taking value on the finite space $\mathbb{K} = \{1, \dots, K\}$, and of transition matrix denoted by $\mathbf{P}$,
- $\{m(k) \in \mathbb{R}^d,\ \mathbf{A}(k) \in \mathcal{M}_{d\times d}\}_{k=1,\dots,K}$ are unknown parameters.
- $\chi_{0,Z}(\cdot)$ and $\chi_{0,Y}(\cdot|Z_0)$ are some probability distributions over $\mathbb{K}$ and $\mathbb{R}^d$

- $p(z_t|Z_{t-1})$ is the law of Z_t condtionnally to Z_{t-1} (here the line of $\mathbf{P}$ given by Z_{t-1}).
- E_t is a random vector such that:

$$E_t \overset{ind.}{\sim} \mathcal{N}\left(0, \mathbf{\Sigma}\right).$$

  where $\mathbf{\Sigma}$ is $d \times d$ covariance matrix.

In this context, the set of unknown parameters is given by

$$\theta = \{\mathbf{P}, m(k),\ \mathbf{A}(k), \mathbf{\Sigma}\}_{k=1,\dots,K}.$$

3.2 Inference

In this context, the inference task is twofold:

1. Obtaining maximum likelihood estimate of θ;
2. Retracing the hidden sequence $Z_0, \dots Z_n$ given the observations $Y_{0:n}$.

It is well known that these two tasks are indeed complementary. A common way to solve these problems is the Expectation Maximization (EM) algorithm (Dempster et al., 1977). The algorithm is shortly depicted here.

3.2.1 Comment about notations

In the following, we use the notation p for a generic probability distribution. The law to which it refers is explicit through arguments. For instance $p(y_{0:n}|z_{0:n})$ is the p.d.f. of a Gaussian vector, the random vector $Y_{0:n}|Z_{0:n}$, and $p(z_t|z_{t-1})$ is the law of the discrete random variable $Z_t|Z_{t-1}$ evaluated at z_t and z_{t-1}. In this context, $p(z_t|z_{t-1}) = \mathbb{P}(Z_t = z_t|\{Z_{t-1} = z_{t-1}\}$.

3.2.2 Likelihood

A straightforward way to compute the likelihood in this model can be obtained, just using the Markov properties of this model:

$$\begin{aligned}
L(\theta|y_{0:n}) &:= p(y_{0:n}|\theta) \\
&= \sum_{z_{0:n}} p(y_{0:n}, z_{0:n}) \\
&= \sum_{z_{0:n}} p(z_{0:n})p(y_{0:n}|z_{0:n}) \\
&= \sum_{z_{0:n}} p(z_0|\theta)p(y_0|z_0,\theta) \prod_{t=1}^{n} p(z_t|z_{t-1},\theta)p(y_t|y_{t-1}, z_t, \theta). \qquad (3.1)
\end{aligned}$$

For a known θ, every term in (3.1) can be computed. However, in a general setting, there exists K^{n+1} possible sequences $z_{0:n}$, which makes this direct computation hardly feasible for any common values of n.

3.2.3 Complete log-likelihood

A workaround to find the maximum likelihood estimate is the EM algorithm. In this context, we focus on a different function, the *complete log-likelihood*, i.e. the likelihood of the *completed* observations (what we wish we could observe), $(y_{0:n}, x_{0:n})$, we have that:

$$\begin{aligned}
\ell(\theta|y_{0:n}, z_{0:n}) :=& \log p(y_{0:n}, z_{0:n}|\theta)\\
=& \log p(y_{0:n}, z_{0:n})\\
=& \log p(z_{0:n}) + \log p(y_{0:n}|z_{0:n})\\
=& \log p(z_0|\theta) + \log p(y_0|z_0, \theta)\\
&+ \sum_{t=1}^{n} \log p(z_t|z_{t-1}, \theta) + \sum_{t=1}^{n} \log p(y_t|y_{t-1}, z_t, \theta). \qquad (3.2)
\end{aligned}$$

For a given set of parameters, say $\theta^{(0)}$, let's consider the following function of θ

$$\begin{aligned}
Q(\theta|\theta_0) :=& \mathbb{E}[\ell(\theta|Y_{0:n}, Z_{0:n})|Y_{0:n} = y_{0:n}, \theta^{(0)}]\\
=& \sum_{z_{0:n}} \ell(\theta|y_{0:n}, z_{0:n}) p(z_{0:n}|y_{0:n}, \theta^{(0)}) \mathrm{d}z_{0:n}\\
=& \sum_{k=1}^{K} (\log p(z_0 = k) + \log(p(y_0|z_0 = k))) + \sum_{k=1}^{K} (\log p(z_0 = k) + \log(p(y_0|z_0 = k))) + \sum_{k=1}^{K} \sum_{k'=1}^{K} \sum_{t=1}^{n} \log p(z_t|z_{t-1},
\end{aligned}$$

Chapter 4

Applications

Some *significant* applications are demonstrated in this chapter.

4.1 Example one

4.2 Example two

Chapter 5

Final Words

We have finished a nice book.

Bibliography

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Xie, Y. (2020). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.17.