A note on gaussian AR(1) process

Artur Wegrzyn

July 27, 2021

Abstract

This note presents elementary properties of autoregressive AR(1) gaussian process. It is intended as a future reference for the author, hence not too much effort is made to ensure that propounding of all steps is sufficient to satisfy the tastes of more fastidious readers. Also, notation is at times quite sloppy and is assumed to be self-explanatory.

Contents

1 AR(1) process - overview

Let $\{Y_t\}_{t=-\infty}^{\infty}$ be a weakly stationary stochastic process that obeys the equation:

$$Y_t = \phi Y_{t-1} + c + \epsilon_t \tag{1}$$

with gaussian error term, i.e.:

$$\epsilon_t \sim N(0, \sigma^2) \tag{2}$$

and all the error terms being uncorrelated. Furthermore, suppose that $T+1$ consecutive realizations of the process $\{Y_t\}$ are observed for times $t = 0, 1, ..., T$ and that they are denoted $y_0, y_1, y_2, ..., y_T$.

2 Likelihood function

I start by looking at the conditional distribution of Y_t with respect to Y_{t-1}. To this end, suppose that the value of Y_{t-1} is known and equal to y_{t-1}. Then:

$$Y_t = c + \phi y_{t-1} + \epsilon_t$$

Unsurprisingly, then:

$$E(Y_t|Y_{t-1} = y_{t-1}) = c + \phi y_{t-1}$$

$$Var(Y_t|Y_{t-1} = y_{t-1}) = \sigma^2$$

As $\{\epsilon_t\}$ has been assumed to be gaussian, it follows that the conditional probability density function of Y_t with respect to $Y_{t-1} = y_{t-1}$ is given by:

$$f_{Y_t|Y_{t-1}}(y_t|y_{t-1}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(y_t - c - \phi y_{t-1})^2}{\sigma^2}\right) \tag{3}$$

This formula will be of great use in the next step which is rewriting of joint probability density function of $(Y_T, Y_{T-1}, ..., Y_1, Y_0)$.

Consequently, let's deal with the joint density. Using the simple fact presented in [1], I have:

$$\begin{aligned} f_{Y_T,Y_{T-1},...,,Y_1,Y_0}(y_T, y_{T-1}, ..., y_1, y_0) =& f_{Y_T|Y_{T-1},...,Y_1,Y_0}(y_T|y_{T-1}, ..., y_1, y_0)\cdot \\ & f_{Y_{T-1}|Y_{T-2},...,Y_1,Y_0}(y_{T-1}|y_{n-2}, ..., y_1, y_0) \cdot ... \cdot \\ & f_{Y_1|Y_0}(y_1|y_0) \cdot f_{Y_0}(y_0) \end{aligned}$$

Written more succintly:

$$f_{Y_T,Y_{T-1},...,,Y_1,Y_0}(y_T, y_{T-1}, ..., y_1, y_0) = f_{Y_0}(y_0)\cdot\prod_{t=1}^{T} f_{Y_t|Y_{t-1},...,y_0}(y_t|y_{t-1}, ..., y_0)$$

Since I aim for *conditional* likelihood, I devide both sided by $f_{Y_0}(y_0)$ and obtain:

$$f_{Y_T,Y_{T-1},...,,Y_1|Y_0}(y_T, y_{T-1}, ..., y_1|y_0) = \prod_{t=1}^{T} f_{Y_t|Y_{t-1},...,y_0}(y_t|y_{t-1}, ..., y_0)$$

Now, by equation (3) one can easily see that since the value of the process at time t is conditional only on its value at time $t-1$ the conditional densities in the product in the equation above reduce as follows:

$$f_{Y_t|Y_{t-1},Y_{t-2}...,Y_0}(y_t|y_{t-1}, y_{t-2}, ..., y_0) = f_{Y_t|Y_{t-1}}(y_t|y_{t-1}) \tag{4}$$

The intuitive explanation of the above is that by (3) the density of Y_t is conditional on Y_{t-1} only and not on $Y_{t-2}, Y_{t-3}, ...$. More formal approach to this topic is presented in one of the later sections.
For now, let's focus on obtaining an alaytical formula for the conditional likelihood function. Substituting (4) into (2) yields:

$$f_{Y_T,Y_{T-1},...,,Y_1|Y_0}(y_T, y_{T-1}, ..., y_1|y_0) = \prod_{t=1}^{T} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(y_t - c - \phi y_{t-1})^2}{\sigma^2}\right) \tag{5}$$

Since we consider the above as a function of the model parameters ϕ, c, σ conditional on the sample $y_0, y_1, ..., y_T$ I introduce the conditional lieklihood $L(\phi, c, \sigma|y_T, t_{T-1}, ..., y_1, y_0)$ as defined using the equation (5) :

$$L(\phi, c, \sigma|y_T, t_{T-1}, ..., y_1, y_0) = \prod_{t=1}^{T} \sigma^{-1}(2\pi)^{-1/2} \exp\left(-\frac{1}{2}\frac{(y_t - c - \phi y_{t-1})^2}{\sigma^2}\right) \tag{6}$$

Later on I will be conditionality in the expression for L for brevity so I will be writing $L(\phi, c, \sigma)$ instead of $L(\phi, c, \sigma|y_T, t_{T-1}, ..., y_1, y_0)$.

3 Derivation of condtional maximum likelihood estimators of the parameters

Let's pick up the expression (6) and take natural logarithm of both sides to obtain the log-likelihhod function $\mathcal{L}$:

$$\begin{aligned}\mathcal{L}(c,\phi,\sigma) = & \quad \sum_{t=1}^{T}\left[-\ln(\sigma)-\ln(2\pi)-\frac{1}{2}\frac{(y_t-c-\phi y_{t-1})^2}{\sigma^2}\right] \\ = & \quad -T\ln(\sigma)-\frac{T}{2}\ln(2\pi)-\frac{1}{2}\sum_{t=1}^{T}\frac{(y_t-c-\phi y_{t-1})^2}{\sigma^2}\end{aligned}$$

The conditional maximum likelihood estimators are obtained by putting the first derivatives of $\mathcal{L}$ with respect to $\sigma, /phi, c$ to zero. These derivatives are:

$$\frac{\partial \mathcal{L}}{\partial \sigma} = -\frac{T}{\sigma} + \sum_{t=1}^{T}\frac{(y_t-c-\phi y_{t-1})^2}{\sigma^3} \tag{7}$$

$$\frac{\partial \mathcal{L}}{\partial \phi} = \frac{1}{\sigma^2}\sum_{t=1}^{T} y_{t-1}(y_t-c-\phi y_{t-1}) \tag{8}$$

$$\frac{\partial \mathcal{L}}{\partial c} = \frac{1}{\sigma^2}\sum_{t=1}^{T}(y_t-c-\phi y_{t-1}) \tag{9}$$

After putting to zero and solving the following estimators $\hat{c}, \hat{\phi}, \hat{\sigma}^2$ are obtained. Firstly, estimator $\hat{\sigma}^2$ is obtained from the equation (7):

$$\hat{\sigma}^2 = \frac{1}{(T+1)-1}\sum_{t=1}^{T}(y_t-\hat{c}-\hat{\phi}y_{t-1})^2$$

Equations (8) and (9) make it possible to establish that the following relations hold:

$$\sum_{t=1}^{T} y_{t-1}(y_t-\hat{c}-\hat{\phi}y_{t-1}) = 0$$

$$\sum_{t=1}^{T}(y_t-\hat{c}-\hat{\phi}y_{t-1}) = 0$$

So that (conditional) maximum likelihood estimators for $\hat{c}, \hat{\phi}$ are given by:

$$\hat{\phi} = \frac{\frac{1}{T}\sum_{t=1}^{T} y_t y_{t-1} - (\frac{1}{T}\sum_{t=1}^{T} y_t)(\frac{1}{T}\sum_{t=1}^{T} y_{t-1})}{\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 - (\frac{1}{T}\sum_{t=1}^{T} y_t)^2}$$

$$\hat{c} = \frac{1}{T}\sum_{t=1}^{T}(y_t-\hat{\phi}y_{t-1})$$

4 Rewriting AR(1) as MA(∞)

Let's go back to square one and aim for another interesting feature of AR(1) process. Start with the equation (1):

$$Y_t = c + \phi y_{t-1} + \epsilon_t$$

Writing this equation for time $t-1$:

$$Y_{t-1} = c + \phi y_{t-2} + \epsilon_{t-1}$$

Substitution of the formula for Y_{t-1} into the formula for Y_t yields:

$$\begin{aligned} Y_t = & \; c + \phi\,(c + \phi y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ = & \; c + \phi c + \epsilon_t + \phi \epsilon_{t-1} + \phi^2 y_{t-2} \end{aligned}$$

Iterating one more time I obtain:

$$Y_t = c(1 + \phi + \phi^2) + (\epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2}) + \phi^3 Y_{t-3}$$

Generalizing:

$$Y_t = c\sum_{k=0}^{n}\phi^k + \sum_{k=0}^{n}\phi^k\epsilon_{t-k} + \phi^{n+1}Y_{t-n-1}$$

Since $|\phi| < 1$, it follows that $\lim_{n\to\infty}\phi^{n+1} = 0$. Therefore[1]:

$$Y_t = \frac{c}{1-\phi} + \sum_{k=0}^{\infty}\phi^k\epsilon_{t-k}$$

5 Unconditional distribution of Y_t

I continue the analysis in the vein of (4). Since all ϵ_t are normally, independently distributed with $\epsilon_t \sim N(0, \sigma^2)$, then the Y_t is also normally distributed with mean:

$$E(Y_t) = \frac{c}{1-\phi}$$

[1] I deliberately restrain from intruducing the polynomial notation here to keep focus of what is of greatest interest here.

and variance:

$$\begin{aligned} Var(Y_t) = &\quad Var(\textstyle\sum_{n=0}^{\infty} \phi^n \epsilon_{t-n}) \\ = &\quad \textstyle\sum_{n=0}^{\infty} \phi^{2n} \sigma^2 \\ = &\quad \frac{\sigma^2}{1-\phi^2} \end{aligned}$$

This allows me to write down the unconditional density of Y_t:

$$f_{Y_t}(y_t) = \frac{1}{\sqrt{2\pi}(\sigma/\sqrt{1-\phi^2})} \exp\left(-\frac{1}{2}\frac{(y_t - c/(1-\phi))^2}{\sigma^2/(1-\phi^2)}\right) \tag{10}$$

6 Joint density of Y_{t+2}, Y_{t+1}, Y_t

Building on top of the results of two previous sections I aim at deriving the joint density of $\{Y_t\}$ at three consecutive time instances, i.e. Y_{t+2}, Y_{t+1}, Y_t. First, joint density of Y_{t+1}, Y_t is by substitution of (10) and (3) into the formula:

$$f_{Y_{t+1},Y_t}(y_{t+1}, y_t) = f_{Y_{t+1}|Y_t}(y_{t+1}|y_t) f_{Y_t}(y_t)$$

which yields:

$$f_{Y_{t+1},Y_t}(y_{t+1}, y_t) = \frac{\sqrt{1-\phi^2}}{\sigma^2\sqrt{2\pi}} \exp\left(-\frac{(y_t - c/(1-\phi))^2 (y_{t+1} - c - \phi y_t)^2}{2\sigma^2(1-\phi^2)}\right)$$

attack on the joint density of Y_{t+2}, Y_{t+1}, Y_t proceeds analogously via:

$$f_{Y_{t+2},Y_{t+1},Y_t}(y_{t+2}, y_{t+1}, y_t) = f_{Y_{t+2}|Y_{t+1},Y_t}(y_{t+2}|y_{t+1}, y_t) f_{Y_{t+1},Y_t}(y_{t+1}, y_t)$$

But how is the condtional density $f_{Y_{t+2}|Y_{t+1},Y_t}$ obtained? To get it, let's reconsider the following relationships that express Y_{t+2} as a function of Y_{t+1} or a function of Y_t:

$$Y_{t+2} = c + \phi Y_{t+1} + \epsilon_{t+2}$$

$$Y_{t+2} = c + \phi c + \epsilon_{t+2} + \phi\epsilon_{t+1} + \phi^2 Y_t$$

First of these equations allows me to establish the following fact:

$$Y_{t+2}|Y_{t+1} = y_{t+1}, Y_t = y_t \sim N(c + \phi y_{t+1}, \sigma^2)$$

Note that dependency of Y_{t+2} on Y_t is, by definition, already contained inside Y_{t+1}. Hence, the conditional density of interest is:

$$f_{Y_{t+2}|Y_{t+1},Y_t}(y_{t+2}|y_{t+1},y_t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(y_{t+2}-c-\phi y_{t+1})^2}{\sigma^2}\right)$$

And the formula for the joint density of $f_{Y_{t+2},Y_{t+1},Y_t}(y_{t+2},y_{t+1},y_t)$ follows:

$$\begin{aligned} f_{Y_{t+2},Y_{t+1},Y_t}(y_{t+2},y_{t+1},y_t) &= \tfrac{\sqrt{1-\phi^2}}{\sigma^3(2\pi)^{3/2}} \cdot \\ &\exp\left(-\tfrac{(y_t-c/(1-\phi))^2(y_{t+1}-c-\phi y_t)^2(y_{t+2}-c-\phi y_{t+1})^2}{2\sigma^3(1-\phi^2)}\right) \end{aligned}$$

7 References

The exposition in this note follows closely the one in [2], chapter 5. This note should just be considered a subjective rephrasing of Hamilton's exposition. Another useful reference is [3].

References

[1] A. Wegrzyn, "A note on conditional probabilities of multivariate continuous variables with applications to time series analysis." `https://github.com/wegar-2/latex_files/blob/master/probability_calculus/conditional_probabilities_for_time_series.pdf`, 2021.

[2] J. D. Hamilton, *Time Series Analysis.* Princeton University Press, 1994.

[3] W. A. Fuller, *Introduction to Statistical Time Series, 2nd Edition.* Wiley, 1996.