A Simple Example Of Maximum Likelihood Estimation

Pier-André Bouchard St-Amant [*]

I present here a simple version of a maximum likelihood estimation of a an AR(1) process :

$$y_t = \rho y_{t-1} + u_t \qquad\qquad u_t \sim N(0, \sigma^2).$$

1 Finding ρ and σ^2

To estimate it, I use a dataset of six observations on y_t and these values are :

$$y = \left[\underbrace{-2.13}_{=y_1}, -1.77, -1.06, 1.78, 1.71, \underbrace{1.46}_{=y_6} \right]'$$

Based on these numbers and the information we have about u_t, we seek to estimate ρ and σ. The white noise u_t can be written as $u_t = y_t - \rho y_{t-1}$. Since u_t is distributed normally, I know that the likelihood (probability) of observing one given value of u_t is given by :

$$P(u_t = U) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}U^2}.$$

Hence, the probability of observing :

$$\begin{aligned} u_2 &= y_2 - \rho y_1 \\ &= -1.77 + \rho 2.13 \end{aligned}$$

is given by :

$$P(u_2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(-1.77+\rho 2.13)^2}$$

* pabsta@econ.queensu.ca.

Similarly, the probability of observing $u_3, u_4, \ldots, u_8$ is given by :

$$P(u_3) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(-1.06+\rho 1.77)^2} \qquad P(u_4) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(1.78+\rho 1.06)^2}$$

$$P(u_5) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(1.71-\rho 1.78)^2} \qquad P(u_2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(1.46-\rho 1.71)^2}$$

I also know that observing the $u_2, u_3, \ldots, u_6$ jointly is the product of their single probabilities since a white noise process is uncorrelated across time.

Hence, the probability of observing all these error altogether is :

$$\begin{aligned} P(u_2, \ldots, u_5) &= P(u_2)P(u_3)P(u_4)P(u_5) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^5 e^{-\frac{1}{2\sigma^2}\left((-1.77+\rho 2.13)^2+(-1.06+\rho 1.77)^2+(1.78+\rho 1.06)^2+(1.71-\rho 1.78)^2+(1.46-\rho 1.71)^2\right)} \end{aligned}$$

There is still some information I can use since I have not used the value y_1 twice. I did not observe y_0, so I need to use another strategy. Under the assumption of stationarity, I however know that y has a stationary distribution. Hence, I can infer on the probability of observing y_1 alone. The distribution of y is $N(0, \frac{\sigma^2}{1-\rho})$ (see the slides for the derivation). Hence, the probability of observing $y_1 = 2.13$ is given by :

$$P(y_1) = \frac{\sqrt{1-\rho}}{\sqrt{2\pi\sigma^2}} e^{-\frac{1-\rho}{2\sigma^2}2.13^2}.$$

Therefore, the joint probability of observing the full sample is :

$$P(y_1, \ldots, u_5) = \sqrt{1-\rho}\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^6 e^{-\frac{1}{2\sigma^2}\left((1-\rho)2.13^2+(1.77-\rho 2.13)^2+(1.06-\rho 1.77)^2+(1.78+\rho 1.06)^2+(1.71-\rho 1.78)^2+(1.46-\rho 1.7}$$

This joint probability $P(y_1, u_2, u_3, u_4, u_5, u_6)$ is called the *likelihood function*. It gives the joint probability of observing the sample in hand. Notice that it is a function of y, ρ and σ^2, so it is usually noted $L(y, \rho, \sigma^2)$.

$$\begin{aligned} L(y, \rho, \sigma^2) &= P(y_1, \ldots, u_5) \\ &= \sqrt{1-\rho}\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^6 e^{-\frac{1}{2\sigma^2}\left((1-\rho)2.13^2+(1.77-\rho 2.13)^2+(1.06-\rho 1.77)^2+(1.78+\rho 1.06)^2+(1.71-\rho 1.78)^2+(1.46-\rho 1.71)^2\right)} \end{aligned}$$

Since this sample has been generated by the process, I seek the values of the parameters ρ and σ^2 that *maximizes* the probability of observing this sample. These values are called the *maximum likelihood estimators* of ρ and σ^2. I thus seek to solve the program :

$$\max_{\rho, \sigma^2} L(y, \rho, \sigma^2).$$

Before I do so, I will clean up the Likelihood function to shorten the argument of the exponential function a bit. One can indeed check that :

$$\big((1-\rho)2.13^2 + (1.77-\rho 2.13)^2 + (1.06-\rho 1.77)^2 + (1.78+\rho 1.06)^2 + (1.71-\rho 1.78)^2 + (1.46-\rho 1.71)^2\big)$$

is equal to :

$$17.017 - 23.137\rho + 14.886\rho^2$$

So what I maximize is :

$$\max_{\rho,\sigma^2} \sqrt{1-\rho}\frac{1}{8\pi^3(\sigma^2)^3} e^{-\frac{1}{2\sigma^2}\left(17.017-23.137\rho+14.886\rho^2\right)}$$

If I plot this in the rho and sigma space, I get the following picture :

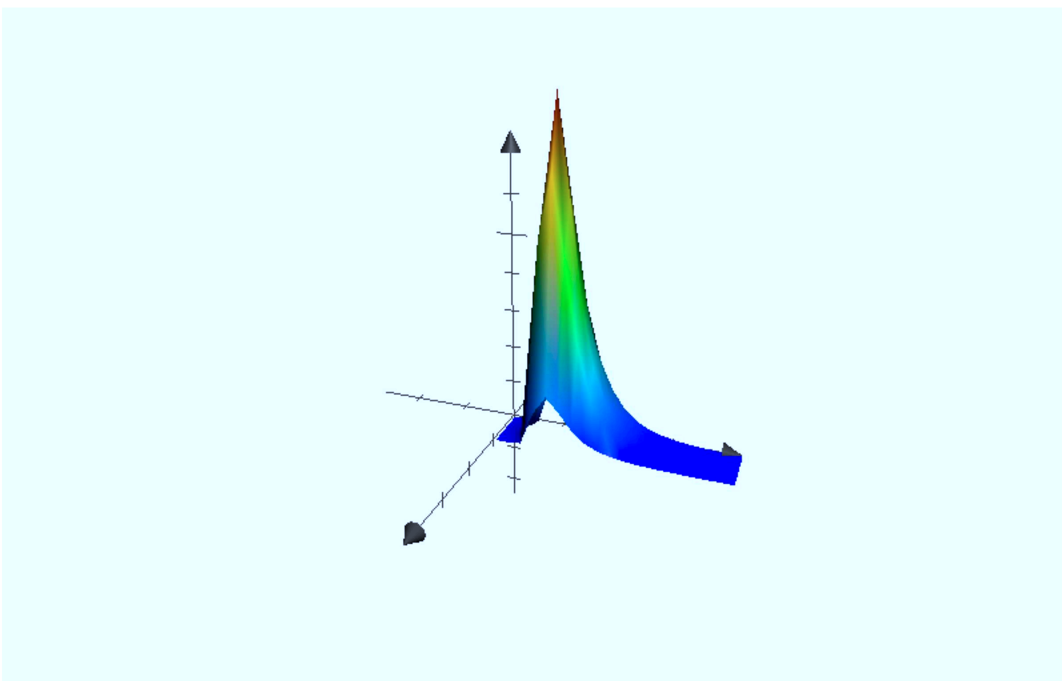

FIGURE 1 – **The Likelihood Function In the** (ρ, σ) **Space**

One can clearly see that there is a maximal value for both ρ and σ. The job is thus to take the first order conditions and find out these values :

$$
\begin{aligned}
\frac{\partial L(y,\rho,\sigma^2)}{\partial \rho} &= -\frac{1}{2}\frac{1}{\sqrt{1-\rho}}\frac{1}{8\pi^3(\sigma^2)^3}e^{-\frac{1}{2\sigma^2}\left(17.017-23.137\rho+14.886\rho^2\right)} \\
&\quad -\frac{1}{2\sigma^2}(29.732\rho-23.137)\sqrt{1-\rho}\frac{1}{8\pi^3(\sigma^2)^3}e^{-\frac{1}{2\sigma^2}\left(17.017-23.137\rho+14.886\rho^2\right)} \\
&= 0 \\
\frac{\partial L(y,\rho,\sigma^2)}{\partial \sigma^2} &= -3\sqrt{1-\rho}\frac{1}{8\pi^3(\sigma^2)^4}e^{-\frac{1}{2\sigma^2}\left(17.017-23.137\rho+14.886\rho^2\right)} \\
&\quad +\frac{1}{2(\sigma^2)^2}\left(17.017-23.137\rho+14.886\rho^2\right)\sqrt{1-\rho}\frac{1}{8\pi^3(\sigma^2)^3}e^{-\frac{1}{2\sigma^2}\left(17.017-23.137\rho+14.886\rho^2\right)} \\
&= 0
\end{aligned}
$$

Let's simplify the first one, by removing the exponential part :

$$
\begin{aligned}
0 &= -\frac{1}{2}\frac{1}{\sqrt{1-\rho}}\frac{1}{8\pi^3(\sigma^2)^3} - \frac{1}{2\sigma^2}(29.732\rho-23.137)\sqrt{1-\rho}\frac{1}{8\pi^3(\sigma^2)^3} \\
&= \frac{1}{\sqrt{1-\rho}} + \frac{1}{\sigma^2}(29.732\rho-23.137)\sqrt{1-\rho} \\
&= \sigma^2 + (29.732\rho-23.137)(1-\rho) \\
&= \sigma^2 - 23.137 + 52.869\rho - 29.732\rho^2 \\
\Rightarrow \sigma^2 &= 23.137 - 52.869\rho + 29.732\rho^2
\end{aligned}
$$

Now, simplify the second one by also removing the exponential part :

$$
\begin{aligned}
0 &= -3\sqrt{1-\rho}\frac{1}{8\pi^3(\sigma^2)^4} + \frac{1}{2(\sigma^2)^2}\left(17.017-23.137\rho+14.886\rho^2\right)\sqrt{1-\rho}\frac{1}{8\pi^3(\sigma^2)^3} \\
&= -3 + \frac{1}{2(\sigma^2)}\left(17.017-23.137\rho+14.886\rho^2\right) \\
&= -6\sigma^2 + 17.017 - 23.137\rho + 14.886\rho^2 \\
\Rightarrow \sigma^2 &= 2.8362 - 3.8562\rho + 2.4810\rho^2
\end{aligned}
$$

Equating those two equations yields :

$$
\begin{aligned}
23.137 - 52.869\rho + 29.732\rho^2 &= 2.8362 - 3.8562\rho + 2.4810\rho^2 \\
\Rightarrow 0 &= 20.301 - 49.013\rho + 27.251\rho^2
\end{aligned}
$$

This is a polynomial of degree two which has two roots. One can check that these roots are 1.15179 and 0.64679. The root greater than one is forbidden by $\sqrt{1-\rho}$ that I multiplied out in the simplification. Hence, the root I retain is $\hat{\rho} = 0.64679$. This imply in turn that the estimated variance is given by $\hat{\sigma}^2 = 23.137 - 52.869\hat{\rho} + 29.732\hat{\rho}^2 = 1.3799$.

2 A Few Words About The Log-Likelihood

As explained in the slides, the logarithm of the likelihood function is used instead of the likelihood itself. This is so because the likelihood yields small numerical values that are untractable for computers.

The log likelihood of the function is given by :

$$\begin{aligned}\ln(L(y,\rho,\sigma^2)) &= \ln\left(\sqrt{1-\rho}\frac{1}{8\pi^3(\sigma^2)^3}e^{-\frac{1}{2\sigma^2}\left(17.017-23.137\rho+14.886\rho^2\right)}\right),\\ &= \frac{1}{2}\ln(1-\rho)-\ln(8\pi^3)-3\ln(\sigma^2)-\frac{1}{2\sigma^2}\left(17.017-23.137\rho+14.886\rho^2\right).\end{aligned}$$

Notice that the first order conditions of this problem yields :

$$\begin{aligned}\frac{\partial\ln L(y,\rho,\sigma^2)}{\partial\rho} &= -\frac{1}{2}\frac{1}{1-\rho}-\frac{1}{2\sigma^2}\left(29.732\rho-23.137\right)=0\\ \frac{\partial\ln L(y,\rho,\sigma^2)}{\partial\sigma^2} &= -\frac{3}{\sigma^2}+\frac{1}{2(\sigma^2)^2}\left(17.017-23.137\rho+14.886\rho^2\right)=0\end{aligned}$$

After a few manipulations, one can check that these FOCs are the same as the previous one. Hence, the solution is also identical.

This is so because the logarithm is a strictly increasing function, so one can see this as a simple change of scale. The maximum of one function will be reached for some values irregardless of the scale used.

This can be seen formally by noticing that for any strictly increasing (and differentiable) function f :

$$\frac{\partial f(L)}{\partial x}=\frac{\partial L}{\partial x}f'(L).$$

By setting it equal to zero, I get :

$$0=\frac{\partial L}{\partial x}f'(L)$$

and since f is strictly increasing, $f'(L)$ is guaranteed to be positive. Hence, I can multiply by $(f'(L)^{-1}$ on both sides to get :

$$0=\frac{\partial L}{\partial x}.$$

That is, the original first order condition.