AR(1) process

Assume that we want to model time series data $\{y_1, \ldots, y_n\}$ as a random realization from a simple Gaussian AR(1) process:

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

A key practical issue concerns the (unobservable) initial value y_0. Should it be fixed? If so, at what value? If not, how to handle it probabilistically?

Let us assume that

$$y_0 \sim N(b, B)$$

When $|\phi| < 1$, b and B can be built based on the equilibrium (unconditional) distribution of y_t for any t:

Picking b and B via stationarity

For any t, the (unconditional) mean and variance of y_t are

$$\begin{aligned} b &= E(y_t) = \alpha + \phi E(y_{t-1}) + E(\epsilon_t) = \alpha + \phi b \\ B &= V(y_t) = \phi^2 V(y_{t-1}) + V(\epsilon_t) = \phi^2 B + \sigma^2 \end{aligned}$$

so

$$b = \frac{\alpha}{1-\phi} \quad \text{and} \quad B = \frac{\sigma^2}{1-\phi^2}$$

Note: In fact, conditionally on stationarity, we do not even need to consider y_0. Instead, the above result can be directly applied to y_1. In this case the joint likelihood for $\theta = (\alpha, \phi, \sigma^2)$ is

$$L(\theta) = \left[\prod_{t=2}^{n} p_N(y_t|\alpha + \phi y_{t-1}, \sigma^2)\right] p_N\left(y_1; \frac{\alpha}{1-\phi}, \frac{\sigma^2}{1-\phi^2}\right).$$

Don't care about stationarity? Don't worry!

The Gaussian distribution of y_1, say $N(d, D)$, can be easily derived from the (linear and Gaussian) AR(1) model and the joint Gaussian distribution of y_0, $N(b, B)$, by using two standard probability results:

$$E(Y) = E[E(Y|X)] \quad \text{and} \quad V(Y) = E[V(Y|X)] + V[E(Y|X)],$$

so

$$\begin{aligned} d = E(y_1) &= E[E(y_1|y_0)] = E(\alpha + \phi y_0) = \alpha + \phi b \\ D = V(y_1) &= V[E(y_1|y_0)] + E[V(y_1|y_0)] \\ &= V(\alpha + \phi y_0) + E(\sigma^2) = \phi^2 B + \sigma^2. \end{aligned}$$

The joint likelihood for $\theta = (\alpha, \phi, \sigma^2)$ is

$$L(\theta) = \left[\prod_{t=2}^{n} p_N(y_t|\alpha + \phi y_{t-1}, \sigma^2)\right] p_N(y_1; \alpha + \phi b, \phi^2 B + \sigma^2).$$

which is well defined for $|\phi| < 1$ as well as for $|\phi| \geq 1$.

In practice

$$p_N\left(y_1; \frac{\alpha}{1-\phi}, \frac{\sigma^2}{1-\phi^2}\right) \qquad |\phi| < 1$$

or

$$p_N(y_1; \alpha + \phi b, \phi^2 B + \sigma^2) \qquad \phi \in \Re,$$

are ignored and (approximate) inference proceeds "as if" y_1 is the initial values and there are only $n-1$ observations, $y_2, \ldots, y_n$.

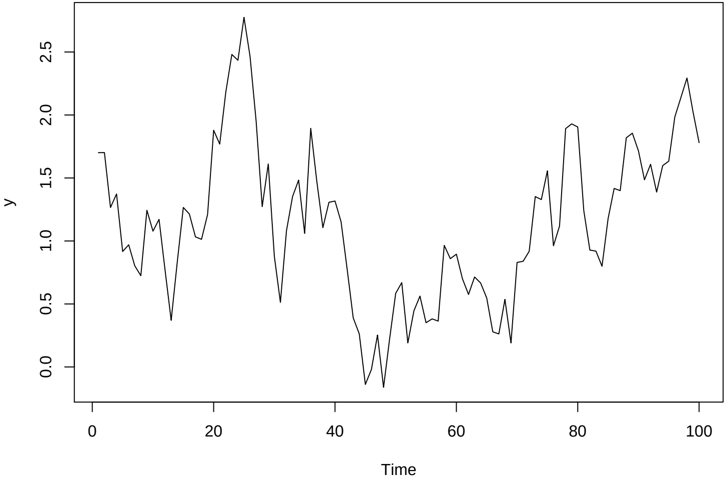
In these cases the (approximate likelihood) becomes

$$L(\theta) = \prod_{t=3}^{n} p_N(y_t | \alpha + \phi y_{t-1}, \sigma^2),$$

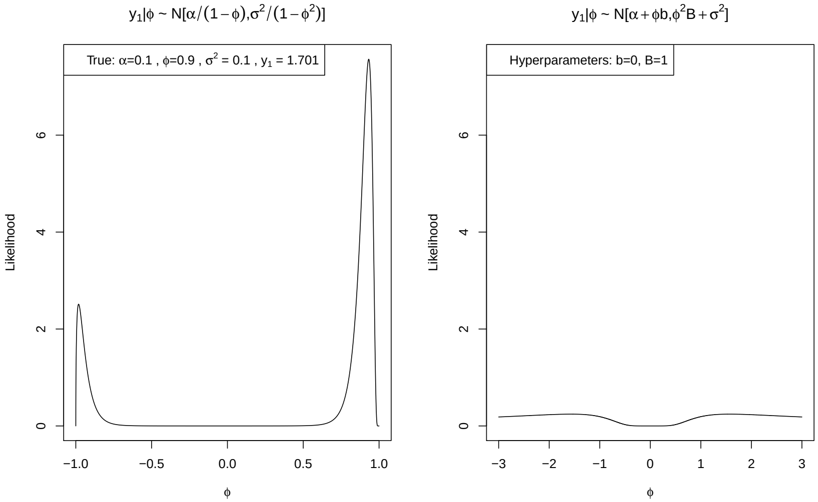
which is usually straightforward to handle regardless from both frequentist or Bayesian viewpoints.

Example: $n = 100$

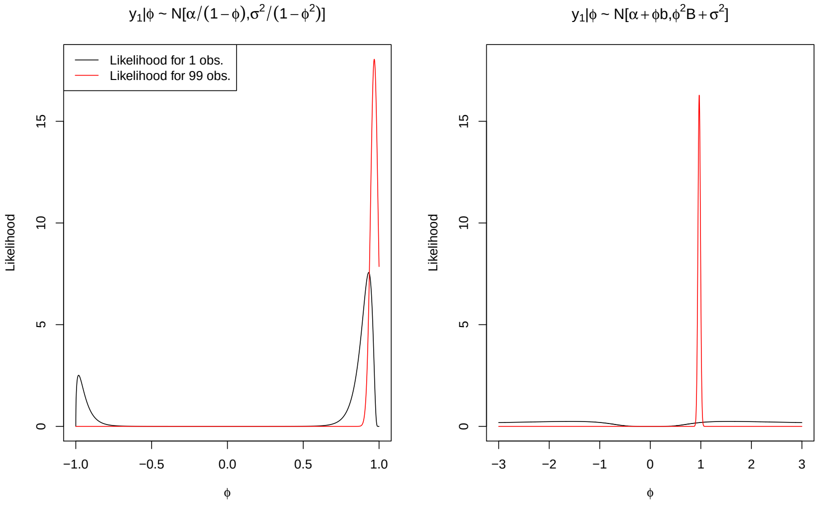
$\alpha = 0.1$, $\phi = 0.9$, $\sigma^2 = 1$



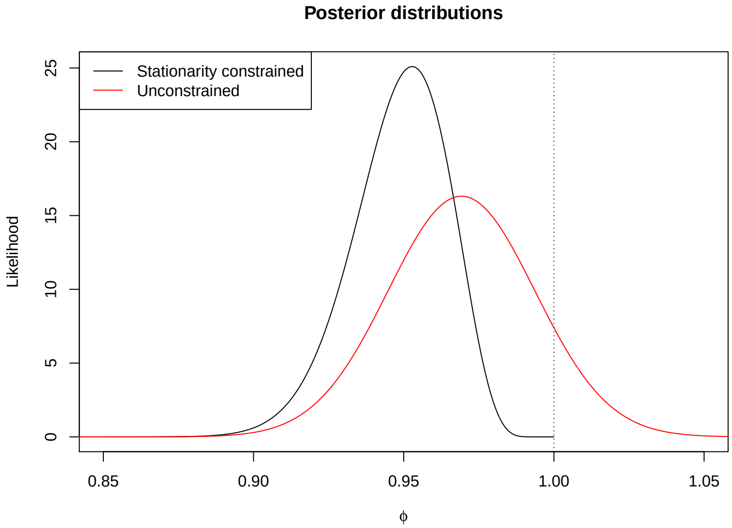
Likelihood of ϕ based on y_1



Likelihoods of ϕ based on y_1 and $y_2, \ldots, y_n$



Posterior of ϕ (based on noninformative prior)[1]



[1] Basically the likelihood of ϕ based on $y_1, \ldots, y_n$.

AR(2) process

Assume that we want to model time series data $\{y_1, \ldots, y_n\}$ as a random realization from a simple Gaussian AR(2) process:

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \qquad \epsilon_t \sim N(0, \sigma^2).$$

A key practical issue concerns the (unobservable) initial values y_0 and y_{-1}. Should they be fixed? If so, at what values? If not, how to handle them probabilistically?

Let us assume that

$$(y_0, y_{-1})' \sim N(b, B),$$

where

$$b = (b_1, b_2)' \quad \text{and} \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{12} & B_{22} \end{pmatrix}$$

Picking b and B via stationarity

When (ϕ_1, ϕ_2) are within the boundaries of stationarity (remember the fancy triangular region?), b and B can be built based on the equilibrium (unconditional) distribution of any two contiguous observations, say y_t and y_{t-1}, from an AR(2) process.

Stationary mean:

$$\begin{aligned}\mu = E(y_t) &= \alpha + \phi_1 E(y_{t-1}) + \phi_2 E(y_{t-2}) + E(\epsilon_t) \\ &= \alpha + \phi_1 \mu + \phi_2 \mu + 0,\end{aligned}$$

implying that the stationary mean is

$$\mu = \frac{\alpha}{1 - \phi_1 - \phi_2}$$

Covariance and correlation functions

Replacing α by $\mu(1-\phi_1-\phi_2)$ in the AR(2) leads to

$$\begin{aligned} y_t &= \mu(1-\phi_1-\phi_2)+\phi_1 y_{t-1}+\phi_2 y_{t-2}+\epsilon_t \\ y_t-\mu &= \phi_1(y_{t-1}-\mu)+\phi_2(y_{t-2}-\mu)+\epsilon_t, \end{aligned}$$

so the autocovariance functions are

$$\begin{aligned} \gamma_j &= E[(y_t-\mu)(y_{t-j}-\mu)] = \phi_1 E[(y_{t-1}-\mu)(y_{t-j}-\mu)] \\ &+ \phi_2 E[(y_{t-2}-\mu)(y_{t-j}-\mu)] + E[\epsilon_t(y_{t-j}-\mu)] \\ &= \begin{cases} \phi_1\gamma_{j-1}+\phi_2\gamma_{j-2} & j=1,2,\ldots, \\ \phi_1\gamma_1+\phi_2\gamma_2+\sigma^2 & j=0 \end{cases} \end{aligned}$$

In terms of autocorrelations:

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \phi_1\rho_{j-1}+\phi_2\rho_{j-2} \qquad j=1,2,\ldots,$$

with $\rho_0 = 1$.

Since

$$\begin{aligned}\rho_1 &= \phi_1\rho_0 + \phi_2\rho_{-1} = \phi_1 + \phi_2\rho_1 \\ \rho_2 &= \phi_1\rho_1 + \phi_2 \\ \gamma_0 &= \phi_1\rho_1\gamma_0 + \phi_2\rho_2\gamma_0 + \sigma^2,\end{aligned}$$

it follows that

$$\rho_1 = \frac{\phi_1}{1-\phi_2} \quad \text{and} \quad \rho_2 = \frac{\phi_1(1+\phi_1-\phi_2)}{1-\phi_2}$$

and

$$\begin{aligned}\gamma_0 &= \frac{(1-\phi_2)}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]}\sigma^2 \\ \gamma_1 &= \rho_1\gamma_0 = \frac{\phi_1}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]}\sigma^2\end{aligned}$$

Now we are ready to go back to $(y_0, y_{-1})' \sim N(b, B)$:

$$\begin{aligned} b_1 = b_2 = \mu &= \frac{\alpha}{1 - \phi_1 - \phi_2} \\ B_{11} = B_{22} = \gamma_0 &= \frac{(1-\phi_2)}{(1+\phi_2)[(1-\phi_2)^2 - \phi_1^2]}\sigma^2 \\ B_{12} = \gamma_1 &= \frac{\phi_1}{(1+\phi_2)[(1-\phi_2)^2 - \phi_1^2]}\sigma^2 \end{aligned}$$

Note: In fact, conditionally on stationarity, we do not even need to consider y_0 and y_{-1}. Instead, the above result can be directly applied to the pair $\tilde{y} = (y_1, y_2)$. In this case the joint likelihood for $\theta = (\alpha, \phi_1, \phi_2, \sigma^2)$ is

$$L(\theta) = \left[\prod_{t=3}^{n} p_N(y_t|\alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2}, \sigma^2)\right] p_N(\tilde{y}|b, B).$$

Don't care about stationarity? Don't worry!

The joint Gaussian distribution of $\tilde{y} = (y_1, y_2)'$, say $N(d, D)$, can be easily derived from the (linear and Gaussian) AR(2) model and the joint Gaussian distribution of $(y_0, y_{-1})'$, $N(b, B)$, by using two standard probability results:

$$\begin{aligned} E(Y) &= E[E(Y|X)] \\ V(Y) &= E[V(Y|X)] + V[E(Y|X)] \end{aligned}$$

so

$$\begin{aligned} d_1 = E(y_1) &= E[E(y_1|y_0, y_{-1})] \\ &= E(\alpha + \phi_1 y_0 + \phi_2 y_{-1}) \\ &= \alpha + \phi_1 b_1 + \phi_2 b_2 \\ d_2 = E(y_2) &= E[E(y_2|y_1, y_0)] \\ &= \alpha + \phi_1 E(y_1) + \phi_2 b_1 \\ &= \alpha(1 + \phi_1) + b_1(\phi_1^2 + \phi_2) + b_2 \phi_1 \phi_2 \end{aligned}$$

$$
\begin{aligned}
D_{11} = V(y_1) &= V[(E(y_1|y_0, y_{-1})] + E[V(y_1|y_0, y_{-1})] \\
&= V(\alpha + \phi_1 y_0 + \phi_2 y_{-1}) + E(\sigma^2) \\
&= \phi_1^2 V(y_0) + \phi_2^2 V(y_{-1}) + 2\phi_1\phi_2 Cov(y_0, y_{-1}) + \sigma^2 \\
&= \phi_1^2 B_{11} + \phi_2^2 B_{22} + 2\phi_1\phi_2 B_{12} + \sigma^2 \\
D_{22} = V(y_2) &= V[(E(y_2|y_1, y_0)] + E[V(y_2|y_1, y_0)] \\
&= \sigma^2 + \phi_1^2 V(y_1) + \phi_2 B_{11} + 2\phi_1\phi_2 Cov(y_1, y_0) \\
&= \sigma^2 + \phi_1^2 V(y_1) + \phi_2 B_{11} + 2\phi_1\phi_2(\phi_1 B_{11} + \phi_2 B_{12}) \\
D_{12} = Cov(y_1, y_2) &= Cov(\alpha + \phi_1 y_0 + \phi_2 y_{-1}, \alpha + \phi_1 y_1 + \phi_2 y_0) \\
&= \phi_1\phi_2 V(y_0) + \phi_1^2 Cov(y_0, y_1) \\
&+ \phi_1\phi_2 Cov(y_1, y_{-1}) + \phi_2^2 Cov(y_0, y_{-1}) \\
&= \phi_1\phi_2 B_{11} + \phi_1^2(\phi_1 B_{11} + \phi_2 B_{12}) \\
&+ \phi_1\phi_2(\phi_1 B_{12} + \phi_2 B_{22}) + \phi_2^2 B_{12}
\end{aligned}
$$

since

$$
\begin{aligned}
Cov(y_1, y_0) &= Cov(\alpha + \phi_1 y_0 + \phi_2 y_{-1}, y_0) = \phi_1 B_{11} + \phi_2 B_{12} \\
Cov(y_1, y_{-1}) &= \phi_1 B_{12} + \phi_2 B_{22}
\end{aligned}
$$

Therefore, the joint likelihood for $\theta = (\alpha, \phi_1, \phi_2, \sigma^2)$ is

$$L(\theta) = \left[\prod_{t=3}^{n} p_N(y_t|\alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2}, \sigma^2)\right] p_N(\tilde{y}|d, D).$$

Note 1: The above likelihood function is well defined for pairs (ϕ_1, ϕ_2) inside the stationary region as well as outside the stationary region.

Note 2: In practice $p_N(\tilde{y}|b, B)$ or $p_N(\tilde{y}|d, D)$ are ignored and (approximate) inference proceeds "as if" y_1 and y_2 are the initial values and there are only $n-2$ observations. In these cases the (approximate likelihood) becomes

$$L(\theta) = \prod_{t=3}^{n} p_N(y_t|\alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2}, \sigma^2),$$

which is usually much easier to handle regardless from a frequentist or a Bayesian viewpoint.