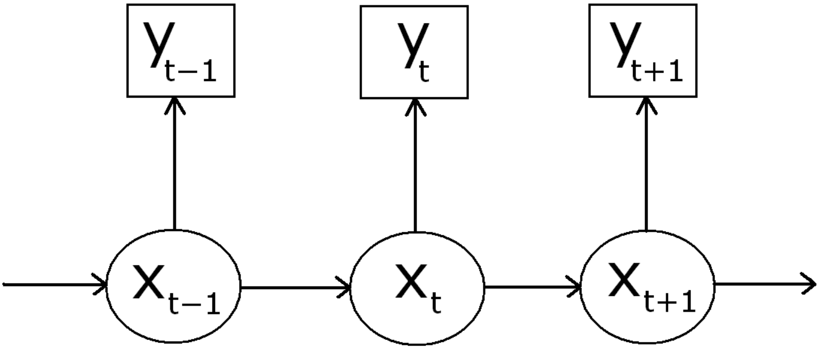
Time Series and Dynamic Models

Section 4
The AR(1) + noise model
Our first DLM...

Carlos M. Carvalho
The University of Texas at Austin

AR(1) Process Observed with Noise

In general:



AR(1) Process Observed with Noise

The Model:

$$\begin{aligned} y_t &= \theta_t + \nu_t \\ \theta_t &= \alpha + \beta\theta_{t-1} + \epsilon_t \end{aligned}$$

- $\nu_t \sim N(0, \tau^2)$ and $\epsilon_t \sim N(0, \omega^2)$... independent.
- $p(\theta_0 | D_0) = N(m_0, C_0)$ – This is the posterior for θ_0 at time 0
- $D_t = \{D_{t-1}, y_t\}$ is the information set up to t.
- Assume, for now, knowledge of α, β, ω^2 and τ^2
- Is this model weakly stationary?

AR(1) Process Observed with Noise

1. Prior for θ_1:

$$\begin{aligned} p(\theta_1|D_0) &= \int p(\theta_1|\theta_0, D_0)p(\theta_0|D_0)d\theta_0 \\ &= N(a_1, R_1) \end{aligned}$$

where

$$a_1 = \alpha + \beta m_0 \quad \text{and} \quad R_1 = \beta^2 C_0 + \omega^2$$

2. Predictive for y_1:

$$\begin{aligned} p(y_1|D_0) &= \int p(y_1|\theta_1, D_0)p(\theta_1|D_0)d\theta_1 \\ &= N(f_1, Q_1) \end{aligned}$$

where

$$f_1 = a_1 \quad \text{and} \quad Q_1 = R_1 + \tau^2$$

AR(1) Process Observed with Noise

3. Posterior for θ_1:

$$p(\theta_1|D_1) \quad \propto \quad p(y_1|\theta_1, D_0)p(\theta_1|D_0)$$

$$\left[\begin{array}{c} \theta_1 \\ y_1 \end{array} |D_0\right] \sim N\left(\left[\begin{array}{c} a_1 \\ f_1 \end{array}\right], \left(\begin{array}{cc} R_1 & R_1 \\ R_1 & Q_1 \end{array}\right)\right)$$

(Why?) Therefore,

$$[\theta_1|D_1] \sim N(m_1, C_1)$$

where

$$m_1 = a_1 + A_1 e_1$$
$$C_1 = R_1 - R_1 A_1$$

with

$$A_1 = R_1/Q_1 \quad \text{and} \quad e_1 = (y_1 - f_1)$$

AR(1) Process Observed with Noise

- What about $(\theta_0|D_1)$? (what does that mean)?

$$
\begin{aligned}
p(\theta_0|D_1) &= \int p(\theta_0|\theta_1, D_1)p(\theta_1|D_1)d\theta_1 \\
&= \int p(\theta_0|\theta_1, D_0)p(\theta_1|D_1)d\theta_1
\end{aligned}
$$

where

$$
p(\theta_0|\theta_1, D_0) \propto p(\theta_1|\theta_0, D_0)p(\theta_0|D_0)
$$

...okay, in this model we should be able to solve the above integral... an easier way is to work with the multivariate normal representation:

$$
\left[\begin{array}{c} \theta_0 \\ y_1 \end{array} |D_0\right] \sim N\left(\left[\begin{array}{c} m_0 \\ f_1 \end{array}\right], \left(\begin{array}{cc} C_0 & \beta C_0 \\ \beta C_0 & Q_1 \end{array}\right)\right)
$$

(Why?)

AR(1) Process Observed with Noise

Therefore,

$$[\theta_0 | D_1] \sim N(h_0, H_0)$$

where

$$\begin{aligned} h_0 &= m_0 + \beta C_0 Q_1^{-1} e_1 \\ H_0 &= C_0 - \beta^2 C_0^2 Q_1^{-1} \end{aligned}$$

Filter-Forward Recursions

We can generalize the above discussion by the following:

- Posterior at $t-1$: $[\theta_{t-1}|D_{t-1}] \sim N(m_{t-1}, C_{t-1})$
- Prior at $t-1$: $[\theta_t|D_{t-1}] \sim N(a_t, R_t)$
  with
  $$a_t = \alpha + \beta m_{t-1} \quad \text{and} \quad R_t = \beta^2 C_t + \omega^2$$
- Predictive at $t-1$: $[y_t|D_{t-1}] \sim N(f_t, Q_t)$
  with
  $$f_t = a_t \quad \text{and} \quad Q_t = R_t + \tau^2$$
- Posterior at t: $[\theta_t|D_t] \sim N(m_t, C_t)$
  $$m_t = a_t + A_t e_t \quad \text{and} \quad C_t = R_t - R_t A_t$$
  with
  $$A_t = R_t / Q_t \quad \text{and} \quad e_t = y_t - f_t$$

But...

- What about α, β, ω^2 and τ^2?
- Could we handle

$$[\alpha, \beta, \omega^2, \tau^2 | \theta_{1:T}, D_T]$$

- If so, a Gibbs sampler iterates through (drawing)
    1. $[\alpha, \beta, \omega^2, \tau^2 | \theta_{1:T}, D_T]$
    2. $[\theta_{1:T} | \alpha, \beta, \omega^2, \tau^2, D_T]$

Our First FFBS

FFBS stands for *Filter Forward Backward Sampling*
This is what the Gibbs Sample for models like the AR(1) plus noise is called due to its form... we'll see below.

Our Goal: Obtain samples from the joint posterior of $(\alpha, \beta, \tau^2, \omega^2, \theta_{1:T})$

How: Build a Gibbs sampler that iterates through (drawing)

1. $p(\alpha, \beta, \omega^2, \tau^2 | \theta_{1:T}, D_T)$
2. $p(\theta_{1:T} | \alpha, \beta, \omega^2, \tau^2, D_T)$

- Step 1 should be easy, right?
- Step 2 requires some work...

$p(\theta_{1:T}|\alpha, \beta, \omega^2, \tau^2, D_T)$

For notation simplicity, let us drop the fixed parameters from the conditioning set, i.e, $p(\theta_{1:T}|\alpha, \beta, \omega^2, \tau^2, D_T) = p(\theta_{1:T}|D_T)$

$$
\begin{aligned}
p(\theta_{1:T}|D_T) &= p(\theta_1, \theta_2, \ldots, \theta_T|\alpha, \beta, \omega^2, \tau^2, D_T) \\
&= p(\theta_1|\theta_2, \ldots, \theta_T, D_T)p(\theta_2|\theta_3, \ldots, \theta_T, D_T) \ldots p(\theta_T|D_T) \\
&= \prod_{t=1}^{T-1} p(\theta_t|\theta_{t+1:T}, D_T)p(\theta_T|D_T)
\end{aligned}
$$

- So, we know $p(\theta_T|D_T)$, right? $p(\theta_T|D_T) = N(m_T, C_T)$
- How about $p(\theta_t|\theta_{t+1:T}, D_T)$?

$p(\theta_{1:T}|\alpha, \beta, \omega^2, \tau^2, D_T)$

Given the conditional independence structure of the model we can write

$$p(\theta_t|\theta_{t+1:T}, D_T) = p(\theta_t|\theta_{t+1}, D_t)$$

(Why?)

Okay, now this should be straightforward as

$$p(\theta_t|\theta_{t+1}, D_t) \propto p(\theta_{t+1}|\theta_t, D_t)p(\theta_t|D_t)$$

$$\left[\begin{array}{c} \theta_t \\ \theta_{t+1} \end{array} |D_t\right] \sim N\left(\left[\begin{array}{c} m_t \\ a_{t+1} \end{array}\right], \left(\begin{array}{cc} C_t & \beta C_t \\ \beta C_t & R_{t+1} \end{array}\right)\right)$$

$p(\theta_{1:T}|\alpha, \beta, \omega^2, \tau^2, D_T)$

Hence,

$$p(\theta_t|\theta_{t+1}, D_t) = N(h_t, B_t)$$

where

$$\begin{aligned} h_t &= m_t + \beta C_t / R_{t+1}(\theta_{t+1} - a_{t+1}) \\ B_t &= C_t - \beta^2 C_t^2 / R_{t+1} \end{aligned}$$

Now what?

- Okay, now what??
- It looks like we can, conditional on the fixed parameters defining the system, filter forward and get to $p(\theta_T|D_T)$
- We then draw $\theta_T^{(1)}$ from $p(\theta_T|D_T)$...
- Now, we should be able to sample $\theta_{T-1}^{(1)}$ from $p(\theta_{T-1}|\theta_T^{(1)}D_{T-1})$... keep going until we get to...
- $\{\theta_1^{(1)}, \theta_2^{(1)}, \ldots, \theta_{T-1}^{(1)}, \theta_T^{(1)}\}$ a draw from the joint distribution $p(\theta_{1:T}|\alpha, \beta, \omega^2, \tau^2, D_T)$

Homework

- Estimate an AR(1) plus noise model to the "daily temperature in Austin data" available in the class website. Only use the last 1,000 observations.
- Report histograms of your posterior distribution for the fixed parameters
- Plot the time series of the observed temperature along with the posterior mean and posterior 95% range for each latent state.
- Predict the temperature for the next 20 days... plot the predictions and prediction intervals.