State Space Models, Kalman Filter, and FFBS

Rob McCulloch

1. State Space Models
2. Forward Filtering
3. Forward Filtering for the Linear Model
4. Backward Sampling
5. Backward Sampling for the Linear Model

1. State Space Models

State Space Models

We consider a class of models of this type:

First,
it has the
basic
Hierarchical
structure.

$$X \mid \theta$$

$$\theta$$

$$X = (X_1, X_2, \ldots X_T)$$

$$\theta = (\theta_1, \theta_2, \ldots \theta_T)$$

Markov on the States:

We call the θ's the *states* and the prior/model on them is Markov, that is, we specify:

$$\theta_t \mid \theta_{t-1}$$

and let's suppose we specify a prior on θ_0, so that the full joint distribution of θ_i, i=0,1,2.... is defined.

For every θ_t we see an X_t:

We then have the observation equations:

$$p(X \mid \theta) = \prod p(X_t \mid \theta_t)$$

That is, conditional on all the θ's, the X's are independent, and X_t only depends on θ_t.

Usually we are thinking about time series data.
The X's are our series of observations.
We do not observe the θ's.

Example:

$$Y_t = \alpha + \beta_t x_t + v_t$$

$$\beta_t = \beta_{t-1} + w_t$$

time varying regression coefficient !!

Graphical representation:

The general picture:

$$
\begin{array}{cccccccccc}
\theta_0 \rightarrow & \theta_1 \rightarrow & \theta_2 \rightarrow & \cdots & \theta_{t-1} \rightarrow & \theta_t \rightarrow & \theta_{t+1} \rightarrow & \cdots & \rightarrow & \theta_T \\
 & \downarrow & \downarrow & & \downarrow & \downarrow & \downarrow & & & \downarrow \\
 & X_1 & X_2 & & X_{t-1} & X_t & X_{t+1} & & & X_T
\end{array}
$$

Each X is a "peek" at the corresponding θ.

If you margin out the θ's get a model in which future X's depend on past X's.

The General Linear Form:

<u>Example</u>

the linear/normal form of the model:

Observation equation →	$X_t = F_t'\theta_t + \alpha_t + \nu_t, \quad \nu_t \sim N(0, V_t),$
state equation →	$\theta_t = G_t\theta_{t-1} + \gamma_t + \omega_t, \quad \omega_t \sim N(0, W_t)$

All the ν and ω are independent.
For now, think of the (F,α,V) and (G,γ,W) as known.

Vector Autoregression on all the Coefficients:

Example:

Have time series regression,

$$y_t = x_t\beta + \varepsilon_t$$

worried that the coefficients may not be constant:

$$y_t = x_t\beta_t + \varepsilon_t$$
$$\beta_t = A\beta_{t-1} + \omega_t$$

2. Forward Filtering

FFBS

Our goal is to have a method for drawing from:

$$\theta \mid X$$

In general, we could use Gibbs sampling and draw:

$$\theta_i \mid \theta_{-i}, X$$

But, if the q's are highly dependent (and they should be!)
then convergence will be slow.
We'd like to be able to draw from the joint.

We will "filter forward and then backward sample" : FFBS.

Forward Filtering:

<u>Forward Filtering</u>

$$\begin{array}{cccccccccc}
\theta_0 \rightarrow & \theta_1 \rightarrow & \theta_2 \rightarrow & \cdots & \theta_{t-1} \rightarrow & \theta_t \rightarrow & \theta_{t+1} \rightarrow & \cdots & \rightarrow & \theta_T \\
 & \downarrow & \downarrow & & \downarrow & \downarrow & \downarrow & & & \downarrow \\
 & X_1 & X_2 & & X_{t-1} & X_t & X_{t+1} & & & X_T
\end{array}$$

Our prior on θ_0, and the state equation, gives us a prior on θ_1.

Given X_1 we can then compute the posterior on θ_1.

Inference for first state:

Put another way, we have the joint distribution of

$$p(\theta_0, \theta_1, X_1) = p(\theta_0)p(\theta_1 | \theta_0)p(X_1 | \theta_1)$$

from which we compute the marginal:

$$p(\theta_1, X_1)$$

from which we compute the conditional:

$$p(\theta_1 | X_1)$$

Inference for state t:

$$\begin{array}{ccccccccc} \theta_0 \rightarrow & \theta_1 \rightarrow & \theta_2 \rightarrow & \cdots & \theta_{t-1} \rightarrow & \theta_t \rightarrow & \theta_{t+1} \rightarrow & \cdots \quad \rightarrow & \theta_T \\ & \downarrow & \downarrow & & \downarrow & \downarrow & \downarrow & & \downarrow \\ & X_1 & X_2 & & X_{t-1} & X_t & X_{t+1} & & X_T \end{array}$$

Now let $D_t = (X_1, X_2, \ldots X_t)$

And let us suppose that we have "computed"

$$\theta_{t-1} \,|\, D_{t-1}$$

Then we can treat this as prior info and compute:

$$\theta_t \,|\, D_t$$

Inference for state t:

$$p(\theta_{t-1}, \theta_t, X_t \mid D_{t-1}) = p(\theta_{t-1} \mid D_{t-1}) p(\theta_t \mid \theta_{t-1}) p(X_t \mid \theta_t)$$

from which we compute the marginal:

$$p(\theta_t, X_t \mid D_{t-1}) = p(\theta_t \mid D_{t-1}) p(X_t \mid \theta_t)$$

(from state equation) *(from observation equation)*

from which we compute the conditional:

$$p(\theta_t \mid X_t, D_{t-1}) = p(\theta_t \mid D_t)$$

Inference for state t:

By iterating the process forward, we obtain

$$\theta_t \,|\, D_t \quad t = 1, 2, \ldots T$$

assuming that the model has a form which enables us to make the calculations.

3. Forward Filtering for the Linear Model

Forward Filtering for the Linear Model

Recall:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N(), \Rightarrow X \mid Y \sim N(\mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(Y - \mu_Y), \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX})$$

$$\equiv N(\mu_X + A(Y - \mu_Y), \Sigma_{XX} - A\Sigma_{YY}A')$$

Note:

Under the linear model all the θ's and X's are multivariate normal !

$\theta_t | D_t$: (m_t, C_t):

Notation:

$$\theta_t \,|\, D_t \sim N(m_t, C_t)$$

Now assume we know m_{t-1}, C_{t-1}

How do we update? We need:

$$p(\theta_t, X_t \,|\, D_{t-1})$$

Because everything is normal, we just have to compute first and second moments.

marginal of $\theta_t | D_{t-1}$: (a_t, R_t):

marginal of θ_t:

$$X_t = F_t' \theta_t + \alpha_t + \nu_t, \quad \nu_t \sim N(0, V_t),$$
$$\theta_t = G_t \theta_{t-1} + \gamma_t + \omega_t, \quad \omega_t \sim N(0, W_t)$$

$$E(\theta_t | D_{t-1}) = G_t E(\theta_{t-1} | D_{t-1}) + \gamma_t = G_t m_{t-1} + \gamma_t$$
$$a_t \equiv G_t m_{t-1} + \gamma_t$$

$$Var(\theta_t | D_{t-1}) = G_t Var(\theta_{t-1} | D_{t-1}) G_t' + W_t = G_t C_{t-1} G_t' + W_t$$
$$R_t \equiv G_t C_{t-1} G_t' + W_t$$

$$\theta_t | D_{t-1} \sim N(a_t, R_t)$$

marginal of $X_t | D_{t-1}$: (f_t, Q_t):

marginal for X_t:

$$X_t = F_t' \theta_t + \alpha_t + \nu_t, \quad \nu_t \sim N(0, V_t),$$
$$\theta_t = G_t \theta_{t-1} + \gamma_t + \omega_t, \quad \omega_t \sim N(0, W_t)$$

$$X_t \,|\, D_{t-1} \sim N(f_t, Q_t), \quad f_t \equiv F_t' a_t + \alpha_t \quad Q_t \equiv F_t' R_t F_t + V_t$$

$cov(X_t, \theta_t)|D_{t-1}$:

Finally, we need the covariance:

$$X_t = F_t'\theta_t + \alpha_t + \nu_t, \quad \nu_t \sim N(0, V_t),$$
$$\theta_t = G_t\theta_{t-1} + \gamma_t + \omega_t, \quad \omega_t \sim N(0, W_t)$$

Assume (wlog) all the means are 0:

$$\mathrm{Cov}(\theta_t, X_t \,|\, D_{t-1}) = E(\theta_t X_t')$$
$$= E(\theta_t \theta_t' F_t) = R_t F_t$$

(m, C) update and A_t (regression of θ_t on X_t given D_t):

For X on Y, $A = \Sigma_{XY}\Sigma_{YY}^{-1}$.

Now we can apply:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N(), \Rightarrow X \mid Y \sim N(\mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(Y-\mu_Y), \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX})$$

$$\equiv N(\mu_X + A(Y-\mu_Y), \Sigma_{XX} - A\Sigma_{YY}A')$$

$$A_t = R_t F_t Q_t^{-1}$$

$$\theta_t \mid D_t = \theta_t \mid D_{t-1}, X_t \sim N(a_t + A_t(X_t - f_t), R_t - A_t Q_t A_t')$$

$$m_t = a_t + A_t(X_t - f_t),\ C_t = R_t - A_t Q_t A'$$

Claim it is simple!!

Although the blizzard of matrices can look a little forbidding, the basic process is quite easy and easy to code up.

4. Backward Sampling

Backward Sampling

Want to draw from $\theta \mid X = \theta \mid D_T$

Have:

$$p(\theta_1, \theta_2, \ldots \theta_T \mid D_T) =$$
$$p(\theta_T \mid D_T) p(\theta_{T-1} \mid \theta_T, D_T) \cdots p(\theta_{t-1} \mid \theta_t, \theta_{t+1}, \cdots \theta_T, D_T) \cdots p(\theta_1 \mid \theta_2, \cdots \theta_T, D_T)$$

BS: Key idea:

$$\begin{array}{ccccccccc}
\theta_0 \rightarrow & \theta_1 \rightarrow & \theta_2 \rightarrow & \cdots & \theta_{t-1} \rightarrow & \theta_t \rightarrow & \theta_{t+1} \rightarrow & \cdots \quad \rightarrow & \theta_T \\
 & \downarrow & \downarrow & & \downarrow & \downarrow & \downarrow & & \downarrow \\
 & X_1 & X_2 & & X_{t-1} & X_t & X_{t+1} & & X_T
\end{array}$$

$$D_t = (X_1, X_2, \ldots X_t) \quad Y_t = (X_{t+1}, X_{t+2}, \ldots X_T)$$

Claim:

$$p(\theta_t \mid \theta_{t+1}, \cdots \theta_T, D_T) = p(\theta_t \mid \theta_{t+1}, D_t)$$

This is the key idea.

Obvious???

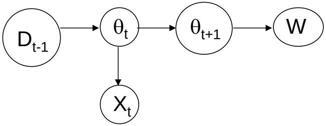
θ_{t+1} has all the data information from the future and D_t has all the data information from the present and past.

BS: reduce it to a few variables:

Write the whole model,

$$
\begin{array}{ccccccccc}
\theta_0 \rightarrow & \theta_1 \rightarrow & \theta_2 \rightarrow & \cdots & \theta_{t-1} \rightarrow & \theta_t \rightarrow & \theta_{t+1} \rightarrow & \cdots \quad \rightarrow & \theta_T \\
 & \downarrow & \downarrow & & \downarrow & \downarrow & \downarrow & & \downarrow \\
 & X_1 & X_2 & & X_{t-1} & X_t & X_{t+1} & & X_T
\end{array}
$$

as,



We get D_{t-1} in the left hand node simple by margining out θ_j j<t.
Just define W to include θ_j j>t+1 and X_j, j>t.

then, $$p(\theta_t \mid \theta_{t+1},\cdots\theta_T,D_T) = p(\theta_t \mid D_{t-1},X_t,\theta_{t+1},W)$$

D_{t-1} and W drop out!

then,

$$p(\theta_t \mid D_{t-1}, X_t, \theta_{t+1}, W) \propto p(\theta_t, D_{t-1}, X_t, \theta_{t+1}, W)$$

$$= p(D_{t-1}) p(\theta_t \mid D_{t-1}) p(X_t \mid \theta_t) p(\theta_{t+1} \mid \theta_t) p(W \mid \theta_{t+1})$$

$$\propto p(\theta_t \mid D_{t-1}, X_t, \theta_{t+1})$$

$$= p(\theta_t \mid D_t, \theta_{t+1})$$

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

So to do backward sampling we do the draw:

$$p(\theta_1, \theta_2, \ldots \theta_T \mid D_T) = p(\theta_T \mid D_T) p(\theta_{T-1} \mid \theta_T, D_{T-1}) \cdots p(\theta_t \mid \theta_{t+1}, D_t) \cdots p(\theta_1 \mid \theta_2, D_1)$$

From the forward filtering we have,

$p(\theta_t \mid D_t)$ t=1,2,...T.

We get

$p(\theta_t \mid \theta_{t+1}, D_t)$ from $p(\theta_t, \theta_{t+1} \mid D_t) = p(\theta_t \mid D_t) p(\theta_{t+1} \mid \theta_t)$

5. Backward Sampling for the Linear Model

Backward Sampling for the linear model:

$$X_t = F_t'\theta_t + \alpha_t + \nu_t, \quad \nu_t \sim N(0, V_t),$$
$$\theta_t = G_t\theta_{t-1} + \gamma_t + \omega_t, \quad \omega_t \sim N(0, W_t)$$

$$\theta_t \mid D_t \sim N(m_t, C_t), \quad \theta_{t+1} \mid D_t \sim N(a_{t+1}, R_{t+1})$$

$$Cov(\theta_t, \theta_{t+1} \mid D_t) = E(\theta_t \theta_{t+1}' \mid D_t) \quad \text{(assuming 0 means)}$$
$$= E(\theta_t (G_{t+1}\theta_t)' \mid D_t) = C_t G_{t+1}'$$

Where we have everything we need from the FF.

B_t: regression coefficients of θ_t on θ_{t+1} given D_t:

$$B_t = cov(\theta_t, \theta_{t+1})[var(\theta_{t+1})]^{-1} = C_t\, G'_{t+1}\, R^{-1}_{t+1}.$$

$$\theta_t \mid \theta_{t+1}, D_t \sim N(m_t + C_t G'_{t+1} R^{-1}_{t+1}(\theta_{t+1} - a_{t+1}), C_t - C_t G'_{t+1} R^{-1}_{t+1} G_{t+1} C_t)$$
$$\equiv N(m_t + B_t(\theta_{t+1} - a_{t+1}), C_t - B_t R_{t+1} B'_t)$$

Again, you can find books that make this very hard to understand but it is easy to understand and (more importantly) easy to code up.

A simple non-linear example

Any time we can easily compute FF,

$$p(\theta_t \mid X_t, D_{t-1}) = p(\theta_t \mid D_t)$$

and BS,

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

then the method is viable.

The basic case is the linear one we have discussed.

The other basic case is that of a discrete state.

Rather than writing out the general "formulas" for the discrete case let's do a simple example from Carter and Kohn.

Two state markov switching stochastic volatility:

Observe time series X_t.
The state θ_t is either 1 or 2.

$$X_t \mid \theta_t \sim \begin{cases} N(\mu, \sigma^2) & \theta_t = 1 \\ N(\mu, \kappa^2\sigma^2) & \theta_t = 2 \end{cases}$$

$$p(\theta_{t+1} = 2 \mid \theta_t = i) = p_{i2}$$

k>1 so there are two states, state 1 is the low variance state and state 2 is the high variance state.

Let

$f_1(x)$ be the $N(\mu, \sigma^2)$ density.

$f_2(x)$ be the $N(\mu, \kappa^2\sigma^2)$ density.

so,

$$X_t \mid \theta_t \sim \begin{cases} X_t \sim f_1 & \theta_t = 1 \\ X_t \sim f_2 & \theta_t = 2 \end{cases}$$

$$p(\theta_t \mid X_t, D_{t-1}) = p(\theta_t \mid D_t)$$

$$p(\theta_t = 2 \mid D_{t-1}) =$$
$$p(\theta_{t-1} = 1 \mid D_{t-1})p_{12} + p(\theta_{t-1} = 2 \mid D_{t-1})p_{22}$$

$$p(\theta_t = 2 \mid D_t) = \frac{p(\theta_t = 2 \mid D_{t-1})f_2(x_t)}{p(\theta_t = 2 \mid D_{t-1})f_2(x_t) + p(\theta_t = 1 \mid D_{t-1})f_1(x_t)}$$

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

From FF have $p(\theta_t = 2 \mid D_t)$

Thus, the joint is:

		θ_{t+1} = 1	θ_{t+1} = 2
θ_t	1	$p(\theta_t = 1 \mid D_t) p_{11}$	$p(\theta_t = 1 \mid D_t) p_{12}$
	2	$p(\theta_t = 2 \mid D_t) p_{21}$	$p(\theta_t = 2 \mid D_t) p_{22}$

so,

$$p(\theta_t = 2 \mid \theta_{t+1} = i, D_t) = \frac{p(\theta_t = 2 \mid D_t) p_{2i}}{p(\theta_t = 2 \mid D_t) p_{2i} + p(\theta_t = 1 \mid D_t) p_{1i}}$$

Gibbs sampling with state space models

Of course, we can think use the state space model as a component embedded with in a larger DAG model. The draw of the state is then just one of the conditionals.

Example

$$X_t \mid \theta_t \sim \begin{cases} N(\mu, \sigma^2) & \theta_t = 1 \\ N(\mu, \kappa^2\sigma^2) & \theta_t = 2 \end{cases}$$

$$p(\theta_{t+1} = 2 \mid \theta_t = i) = p_{i2}$$

$\mu \sim N(\bar{\mu}, \zeta^2)$ $\quad p_{i2} \sim \text{Beta}(a_i, b_i)$

$\sigma^2 \sim \frac{\nu_1 \lambda_1}{\chi^2_{\nu_1}}$ $\quad$ all indep

$\kappa^2 \sim \frac{\nu_2 \lambda_2}{\chi^2_{\nu_2}}$

Gibbs:

$$\theta \mid \mu, \sigma, \kappa, p, X$$
$$\mu \mid \theta, \sigma, \kappa, p, X$$
$$\sigma \mid \theta, \mu, \kappa, p, X$$
$$\kappa \mid \theta, \sigma, \mu, p, X$$
$$p \mid \theta, \mu, \sigma, \kappa, X$$

how could you handle κ>1?

Example

$$X_t = F_t'\theta_t + \alpha_t + \nu_t, \quad \nu_t \sim N(0, V_t),$$
$$\theta_t = G_t\theta_{t-1} + \gamma_t + \omega_t, \quad \omega_t \sim N(0, W_t)$$

Draw F,α,V and G,γ,W given the state.

Prediction

For each draw from the posterior,
just simulate the model out.