WARNING: REFERENCES TO OTHER PARTS OF THE DOCUMENT WON'T WORK HERE

1 Model

The general dynamic linear model (DLM) is a linear, gaussian, state space model. A state space model has two components — a sequence of real valued random vectors $\{y_t\}$ denoting an observation for each period and another sequence of real valued random vectors $\{\theta_t\}$ denoting a latent state for each period. The observations range from $t = 1, \cdots, T$, i.e. the length of the full time series, and the states range from $t = 0, \cdots, T$. The states form a Markov chain so that $p(\theta_{t+1}|\theta_{0:T}) = p(\theta_{t+1}|\theta_t)$ where $p(x|z)$ denotes the conditional density of x given z. Furthermore, the observations are conditionally independent given the states and in particular $p(y_{1:T}|\theta_{0:T}) = p(y_1|\theta_1) \times \cdots \times p(y_T|\theta_T)$. The state space model is then completed by specifying the observation and system equations: for $t = 1, 2, \cdots, T$

$$y_t = f_t(\theta_t, v_t) \tag{1}$$
$$\theta_t = g_t(\theta_{t-1}, w_t) \tag{2}$$

where $v_{1:T}$ and $w_{1:T}$ are independent and are each iid draws from some distribution. Equation (1) is known as the observation equation since it describes how the observations depend on the current latent state and (2) is known the system equation since it describes how the latent states, or the underlying system, evolve over time. The random vector v_t is called the observation error and w_t is called the system error or the system disturbance. The functions f_t and g_t and the distributions of $v_{1:T}$ and $w_{1:T}$ may depend on some unknown parameter vector ϕ that we wish to estimate.

The dynamic linear model adds a couple of constraints to the state space model. First, it requires that both f_t and g_t be linear functions. Second, it requires that $(v_{1:T}, w_{1:T})$ is normally distributed, usually with a mean of zero. We can then rewrite the DLM as

$$\begin{aligned} y_t|\theta_{0:T} &\overset{ind}{\sim} N(F_t\theta_t, V_t) \\ \theta_t|\theta_{0:t-1} &\sim N(G_t\theta_{t-1}, W_t) \end{aligned}$$

for $t = 1, 2, \cdots, T$ where F_t and G_t are matrices, and V_t and W_t are symmetric and positive definite covariance matrices. If θ_t is $p \times 1$ and y_t is $k \times 1$, then F_t is $k \times p$ and G is $p \times p$ while V_t is $k \times k$ and W_t is $p \times p$. The observation errors ("errors"), $v_t = y_t - F_t\theta_t$ for $t = 1, 2, \cdots, T$, and the system disturbances ("disturbances"), $w_t = \theta_t - G_t\theta_{t-1}$ for $t = 1, 2, \cdots, T$ are independent. Let ϕ denote the unknown parameter vector. Then possibly $F_{1:T}$, $G_{1:T}$, $V_{1:T}$, and $W_{1:T}$ are all functions of ϕ. We'll focus our attention on a simpler version of the DLM. Typically additional model structure is used to learn about $V_{1:T}$ and $W_{1:T}$ if time dependence is enforced – e.g. a stochastic volatility prior which would require a statespace model describing the $V_{1:T}$'s and $W_{1:T}$'s as data. Because of this additional complexity, we focus on the time-constant variances model, though many of our results may be useful in more complicated time-varying variance models. Thus we enforce $V_t = V$ and $W_t = W$ for $t = 1, 2, \cdots, T$. We will also suppose that F_t and G_t are known matrices for $t = 1, 2, \cdots, T$, though this constraint is immaterial since relaxing it will simple add one or more Gibbs steps to the algorithms we explore so long as no parameter that enters any F_t or G_t also enters V or W.

Thus $\phi = (V, W)$ is our unkown parameter vector and we can write the model as:

$$y_t|\theta_{0:T} \overset{ind}{\sim} N(F_t\theta_t, V) \tag{3}$$
$$\theta_t|\theta_{0:t-1} \sim N(G_t\theta_{t-1}, W) \tag{4}$$

To complete the model specification in a Bayesian context, we need priors on θ_0, $V_{1:T}$, and $W_{1:T}$. We'll use the standard approach for now and assume that they're mutually independent a priori and that $\theta_0 \sim N(m_0, C_0)$, $V \sim IW(\Lambda_V, \lambda_V)$ and $W \sim IW(\Lambda_W, \lambda_W)$ where m_0, C_0, Λ_V, λ_V, Λ_W, and λ_W are known hyperparameters and $IW(\Lambda, \lambda)$ denotes the inverse Wishart distribution with degrees of freedom λ and positive definite scale matrix Λ. This allows us to write the full joint distribution of $(V, W, \theta_{0:T}, y_{1:T})$ as

$$\begin{aligned}
p(V, W, \theta_{0:T}, y_{1:T}) \propto & \exp\left[-\frac{1}{2}(\theta_0 - m_0)' C_0^{-1}(\theta_0 - m_0)\right] \\
& \times |V|^{-(\lambda_V + k + T + 2)/2} \exp\left[-\frac{1}{2}\operatorname{tr}\left(\Lambda_V V^{-1}\right)\right] \exp\left[-\frac{1}{2}\sum_{t=1}^{T}(y_t - F_t\theta_t)' V^{-1}(y_t - F_t\theta_t)\right] \\
& \times |W|^{-(\lambda_W + p + T + 2)/2} \exp\left[-\frac{1}{2}\operatorname{tr}\left(\Lambda_W W^{-1}\right)\right] \exp\left[-\frac{1}{2}\sum_{t=1}^{T}(\theta_t - G_t\theta_{t-1})' W^{-1}(\theta_t - G_t\theta_{t-1})\right]
\end{aligned} \tag{5}$$

where $p = dim(\theta_t)$, $k = dim(y_t)$, and tr(.) is the matrix trace operator.