April 24, 2009

Switching Autoregressive Hidden Markov Model (SAR-HMM)

Tristan Fletcher

www.cs.ucl.ac.uk/staff/T.Fletcher/

1 Model Specification

Ephraim and Roberts' [1] Switching Autoregressive Hidden Markov Model (SAR-HMM) is similar to the standard HMM but instead of the observations of the time series v_t being generated by the hidden states s_t at any given time t, the latent states describe an autoregressive relationship between sequences of observations. The model switches between sets of autoregressive paramaters with probabilities determined by a state transition probability similar to that of a standard HMM:

$$v_t = \sum_{r=1}^{R} a_r(s_t) v_{t-r} + \eta_t \qquad \text{with} \qquad \eta_t \sim N(0, \sigma^2) \tag{1.1}$$

where $a_r(s_t)$ is the r^{th} autoregressor when in state $s \in \{1 \ldots k\}$ at time t and each η_t is an i.i.d. normally distributed innovation with mean 0 and variance σ^2.

2 Simulation

A time series $\mathbf{v}$ of T observations of v_t can be generated once the following parameters have been defined:

- number of states k
- number of autoregressors R
- a $R \times k$ matrix $\mathbf{A}$ of the autoregessor values $\mathbf{a}(s_t) = a_{1:R}(s_t)$ for each state
- the noise innovation variance σ^2
- the $k \times k$ state transition probability matrix $\mathbf{P}$ where $P_{ij} = p(s_t = i | s_{t-1} = j)$
- the probality of starting (i.e. when $t = 1$) in each of the k states, $a_i = p(s_1 = i)$ where $i \in \{1 \ldots k\}$
- the first R observations $v_{1:R}$

3 Inference

The definition in (1.1) allows us to to describe the probability of an observation at time t expressed as a function of the previous R observations and the current state s_t:

$$p(v_t|v_{t-R:t-1}, s_t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\left(v_t - \sum_{r=1}^{R} a_r(s_t)v_{t-r}\right)^2\right\} \tag{3.1}$$

Inferring $p(s_t|v_{1:T})$ is accomplished in standard HMM using the forward-backward procedure. However, because of the SAR-HMM's autoregressive relationship between the observations, creating forward dependencies R time steps into the future for any observation v_t, a different backward pass needs to be implemented.

3.1 Forward Pass

The forward pass calculates at time t the probability of being in a state s_t given all the observations *up to that point in time* $v_{1:t}$, i.e. $p(s_t|v_{1:t})$.

Starting with an initial probability of being in a state given the first observation:

$$p(s_1|v_1) \propto p(v_1|s_1)p(s_1)$$

values for $p(s_t|v_{1:t})$ for subsequent values of t can be found by iteration:

$$\begin{aligned}
p(s_t|v_{1:t}) &= p(s_t|v_t, v_{1:t-1}) \\
&= \frac{p(s_t, v_t|v_{1:t-1})}{p(v_t|v_{1:t-1})} \\
&\propto p(s_t, v_t|v_{1:t-1}) \\
&= \sum_{s_{t-1}} p(s_t, v_t|v_{1:t-1}, s_{t-1})p(s_{t-1}|v_{1:t-1}) \\
&= \sum_{s_{t-1}} p(v_t|v_{1:t-1}, s_{t-1}, s_t)p(s_t|s_{t-1}, v_{1:t-1})p(s_{t-1}|v_{1:t-1}) \\
s_t \perp v_{1:t-1}|s_{t-1} &\Rightarrow \sum_{s_{t-1}} p(v_t|v_{1:t-1}, s_{t-1}, s_t)p(s_t|s_{t-1})p(s_{t-1}|v_{1:t-1}) \\
v_t \perp s_{t-1}|v_{1:t-1}, s_t &\Rightarrow \sum_{s_{t-1}} p(v_t|v_{1:t-1}, s_t)p(s_t|s_{t-1})p(s_{t-1}|v_{1:t-1}) \\
p(v_t|v_{1:t-1}, s_t) \equiv p(v_t|v_{t-R:t-1}, s_t) & \\
&\Rightarrow \sum_{s_{t-1}} p(v_t|v_{t-R:t-1}, s_t)p(s_t|s_{t-1})p(s_{t-1}|v_{1:t-1})
\end{aligned} \tag{3.2}$$

where each $p(s_{t-1}|v_{1:t-1})$ is the result of (3.2) for the previous time step $t-1$ and $p(v_t|v_{t-R:t-1}, s_t)$ is calculated from (3.1).

In this implementation, the forward pass will be started at $t = R + 1$ with the previous posterior equalling the initial starting probability for each state, i.e. $p(s_R|v_{1:R}) = p(s_1)$.

3.2 Backward Pass

Once the forward pass has been completed for all time steps $t = 1 : T$, the backward pass calculates at time t the probability of being in a state s_t given all the observations *in the entire sequence of T time steps* $v_{1:T}$, i.e. $p(s_t|v_{1:T})$.

Commencing at the final time step ($t = T$) and working backwards to the start ($t = 1$), $p(s_t|v_{1:T})$ can be evaluated as follows:

$$
\begin{aligned}
p(s_t|v_{1:T}) &= \sum_{s_{t+1}} p(s_t|s_{t+1}, v_{1:T})p(s_{t+1}|v_{1:T}) \\
s_t \perp v_{t+1:T}|s_{t+1} &\Rightarrow p(s_t|s_{t+1}, v_{1:T}) = p(s_t|s_{t+1}, v_{1:t}) \\
\therefore\ p(s_t|v_{1:T}) &= \sum_{s_{t+1}} p(s_t|s_{t+1}, v_{1:t})p(s_{t+1}|v_{1:T}) \\
&= \frac{\sum_{s_{t+1}} p(s_{t+1}, s_t|v_{1:t})p(s_{t+1}|v_{1:T})}{p(s_{t+1}|v_{1:t})} \\
&\propto \sum_{s_{t+1}} p(s_{t+1}, s_t|v_{1:t})p(s_{t+1}|v_{1:T}) \\
&= \sum_{s_{t+1}} p(s_{t+1}|s_t, v_{1:t})p(s_t|v_{1:t})p(s_{t+1}|v_{1:T}) \\
s_{t+1} \perp v_{1:t}|s_t &\Rightarrow \sum_{s_{t+1}} p(s_{t+1}|s_t)p(s_t|v_{1:t})p(s_{t+1}|v_{1:T}) \qquad (3.3)
\end{aligned}
$$

where each $p(s_{t+1}|v_{1:T})$ is the result of (3.3) from a previous iteration (i.e. future time step $t+1$) and $p(s_t|v_{1:t})$ for each time step has been calculated by the forward pass in (3.2).

References

[1] Y. Ephraim and W. J. J. Roberts, “Revisiting autoregressive hidden markov modeling of speech signals,” *IEEE Signal Processing Letters*, vol. 12, pp. 166–169, 2005.