STAT 536 - OCT 26

Bayesian Inference

The Basics

There are three core parts of any Bayesian analysis:

1. *Sampling Model:* This is the generative model or distribution for the data. This is often denoted as $p(y|\theta)$ or $\pi(y|\theta)$.

2. *Prior Distribution:* This expresses prior beliefs about a parameter using a statistical distribution to account for uncertainty. The distributed is written as $p(\theta)$ of $\pi(\theta)$

3. *The Posterior Distribution:* Is the distribution for the parameters given the data and is completely identified by the combination of the prior and the sampling model. Formally, the posterior distribution can be computed as
$$\pi(\theta|y) = \frac{\pi(y|\theta)\pi(\theta)}{\int \pi(y|\theta)\pi(\theta)d\theta}.$$
The posterior distribution, $p(\theta|y)$ or $\pi(\theta|y)$, is a distribution that capture the uncertainty in the parameters.

The Bayesian framework also can be used for constructing a predictive distribution.

- *Predictive Distribution:* The predictive distribution for a new prediction, $\pi(y^*|y)$ or more specifically in a time series setting $\pi(y_{t+1}|y_t)$. This distribution does not contain any unknown parameters as they are integrated out:
$$\pi(y_{t+1}|y_t) = \int \pi(y_{t+1}|\theta)\pi(\theta|y)d\theta$$

Bayesian Integration

- In some cases, the integrals above can be analytically solved. One example is the set of Kalman filtering equations that you have already worked through.

- In other cases, computational algorithms such as Markov Chain Monte Carlo (MCMC) or Sequential Monte Carlo (SMC) are used to approximate those quantities.

State Space Models and Dynamic Linear Models

Overview

A state space model is typically expressed using two levels, the observation equation and state equation:

1. *Observation Equation:* The observation equation is focused on the data generating mechanism for the data you actually observe.

2. *State or Evolution Equation:* This captures the dynamics of a unobservable (latent) variable.

Consider the example from the second lab:

$$\begin{aligned} y_t &= \theta_t + \epsilon_t \qquad \text{(Observation Equation)} \\ \theta_t &= \theta_{t-1} + \nu + w_t \qquad \text{(Evolution Equation)}, \end{aligned}$$

where $\epsilon_t \sim N(0, \sigma^2)$ and $w_t \sim N(0, \sigma^2_w)$. Note that ν, σ^2, and σ^2_w are all assumed known, but they can easily be estimated in an MCMC procedure.

Parameter Estimation

Parameter estimates follow a few simple steps: initial (prior) values, prediction (propogation), and estimation (filter).

1. *Prior step:* Assume we have $\pi(\theta_t|y_{1:t})$. That is we have a distributional belief about the location of the latent variable at time t, given data observed up to time t.

2. *Propogation step:* Using the evolution equation we can propogate or predict the location of the latent variable at the next time point to get $\pi(\theta_{t+1}|y_{1:t})$. Mathematically this follows as

$$\pi(\theta_{t+1}|y_{1:t}) = \int \pi(\theta_{t+1}|\theta_{t+1})\pi(\theta_t|y_{1:t})d\theta.$$

   Then using the updated distributional belief about θ_{t+1} predictions can also be made for next observation:

$$p(y_{t+1}|y_{1:t}) = \int \pi(y_{t+1}|\theta_{t+1})\pi(\theta_{t+1}|y_{1:t})d\theta_{t+1}.$$

   The uncertainty in θ_{t+1} is accounted for with the integration.

3. *Estimation or filtering step:* After observing data at time $t+1$, the belief about the location of θ can be updated, which results in $\pi(\theta_{t+1}|y_{1:t+1})$.

Specifically the filtering density is estimated as:

$$\pi(\theta_{t+1}|y_{1:t+1}) = \frac{\pi(y_{t+1}|\theta_{t+1})\pi(\theta_{t+1}|y_{1:t})}{\int \pi(y_{t+1}|\theta_{t+1})\pi(\theta_{t+1}|y_{1:t})d\theta_{t+1}} = \frac{\pi(y_{t+1}|\theta_{t+1})\pi(\theta_{t+1}|y_{1:t})}{\pi(y_{t+1}|y_{1:t}}$$

- Note the filtering equation results in the prior distribution for the next time point.

- Hence for any data point at t we can use this three step iterative procedure to make predictions and update our beliefs.

About State Space Models

- θ_t is called the state process. This comes from engineering applications where a machine is assumed to be in certain "state".

A state space model has two assumptions:

1. (θ_t) is a Markov chain, which means $\pi(\theta_t|\theta_{1:t-1}) = \pi(\theta_t|\theta_{t-1})$, in other words it only depends on the immediately preceeding value and

2. Conditional on (θ_t), the $Y_t's$ are independent and Y_t only depends on θ_t.

- Hidden markov models refer to situations where the state variables are unobserved discrete-valued random variables

Dynamic Linear Models

A dynamic linear model is a special case of a state-space model, where:

- The prior on the state vector for time 0, follows a normal distribution:

$$\theta_0 \sim N_p(\mu_0, C_0)$$

- and the following equations are specified for the observation and evolution equations:

$$\begin{aligned} Y_t &= F_t\theta_t + v_t, \quad v_t \sim N(0, V_t) \\ \theta_t &= G_t\theta_{t-1} + w_t, \quad w_t \sim N(0, W_t) \end{aligned}$$

An example of a simple dynamic linear model would be the random walk plus noise model,

$$\begin{aligned} Y_t &= \theta_t + v_t, \quad v_t \sim N(0, V) \\ \theta_t &= \theta_{t-1} + w_t, \quad w_t \sim N(0, W), \end{aligned}$$

This model works well with no clear trend or season variation, and is often called the local level model.

The DLMs can be regarded as a generalization of a linear regression model, which allows time varying coefficients. Hence, rather than

$$Y_t = \theta_1 + \theta_2 x_t + \epsilon_t \quad \epsilon \sim N(0, \sigma^2)$$

we have

$$Y_t = \theta_{t,1} + \theta_{t,2} x_t + \epsilon_t \quad \epsilon \sim N(0, \sigma_t^2)$$

where the dynamics of $\widetilde{\theta_t}$ are controlled by the evolution equation.

- We will see that all of the AR - MA and ARMA type models are special cases of dynamic linear models.

Project Overview

My preference is that you use the project as an opportunity to use the state-space model framework that we have been developing.

For writing the project summaries, we will heavily rely on ideas in Schimel's *Writing Science*.

1. Opening (O): It frames an interesting question. What do you need to understand about the situation to follow the story? What is the larger problem you are addressing?

2. Challenge (C): It presents your research plan. What specific question do you propose to answer?

3. Action (A): It discusses the results. What happens to address the challenge? In a paper, this describes the work you did.

4. Resolution (R): It leaves the reader with an important conclusion about how our understanding of the world has changed as a result of the work. This is your conclusion – what did you learn from your work?

For Friday November 2, as part of HW 6, turn in an updated project proposal. The proposal should address the following points:

1. What is your opening? This should identify the larger problem to which you are contributing, give readers a sense of the direction your paper is going and make it clear why it is important. It should engage the widest audience practical.

2. What is your specific question or hypothesis?

3. Include a set of data visualization settings that related the question or hypotheses addressed above.