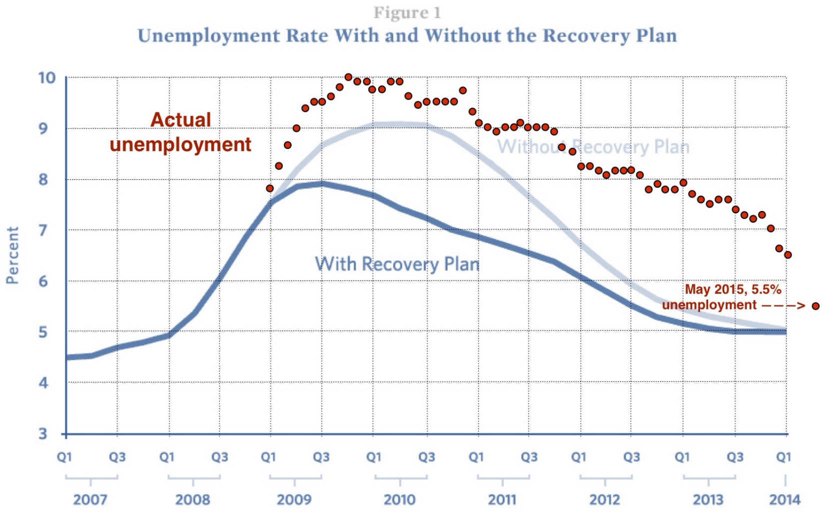
Likelihoods and Bayes
(See Statistics)

Trevor Gallen

Fall 2015

Economics vs. Reality or, Economics is a Science



Example

- Take a two-state Markov chain, in which you observe the following data:

$$X_t = \{2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 2, 2, 2, 2, 2, 1\}$$

- Say we wanted to estimate the markov process:

$$\pi = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix}$$

- How do we do it? (What is your estimate?)

Maximum Likelihood

- We can write down the likelihood of seeing each type of transition:

$$\mathcal{L}_{1\to 1} = p_{11}$$
$$\mathcal{L}_{1\to 2} = p_{12}$$
$$\mathcal{L}_{2\to 1} = p_{21}$$
$$\mathcal{L}_{2\to 2} = p_{22}$$

Or:

$$\mathcal{L} = \prod_{i=1}^{T} (\mathcal{L}_{1\to 1})^{x_{11}} (\mathcal{L}_{1\to 2})^{x_{12}} (\mathcal{L}_{2\to 1})^{x_{21}} (\mathcal{L}_{2\to 2})^{x_{22}}$$

Maximum Likelihood

- The likelihood:

$$\mathcal{L} = \prod_{i=1}^{T} (p_{11})^{x_{11}} (1 - p_{11})^{x_{12}} (1 - p_{22})^{x_{21}} (p_{22})^{x_{22}}$$

- Taking logs:

$$\log \mathcal{L} = \sum_{i=1}^{T} x_{11} \log(p_{11}) + x_{12} \log(1 - p_{11}) + x_{21} \log(1 - p_{22}) + x_{22} \log(p_{22})$$

- Is this all we can do?

Maximum Likelihood

- The likelihood:

$$\mathcal{L} = \prod_{i=1}^{T} (p_{11})^{x_{11}} (1 - p_{11})^{x_{12}} (1 - p_{22})^{x_{21}} (p_{22})^{x_{22}}$$

- Taking logs:

$$\log \mathcal{L} = \sum_{i=1}^{T} x_{11} \log(p_{11}) + x_{12} \log(1 - p_{11}) + x_{21} \log(1 - p_{22}) + x_{22} \log(p_{22})$$

- Is this all we can do? (Hint: no.)

Maximum Likelihood

- The likelihood:

$$\mathcal{L} = \prod_{i=1}^{T} (p_{11})^{x_{11}} (1 - p_{11})^{x_{12}} (1 - p_{22})^{x_{21}} (p_{22})^{x_{22}}$$

- Taking logs:

$$\log \mathcal{L} = \sum_{i=1}^{T} x_{11} \log(p_{11}) + x_{12} \log(1 - p_{11}) + x_{21} \log(1 - p_{22}) + x_{22} \log(p_{22})$$

- Is this all we can do? (Hint: no.)

- The first observation gives us data!

Better Maximum Likelihood

- The likelihood:

$$\log \mathcal{L} = \sum_{i=1}^{T} x_{11} \log(p_{11}) + x_{12} \log(1 - p_{11}) + x_{21} \log(1 - p_{22}) + x_{22} \log(p_{22})$$

- Denote a dummy for the first observation as $x^{[1]}$ or $x^{[2]}$, depending on the value:

$$\log \mathcal{L} = \sum_{i=1}^{T} x_{11} \log(p_{11}) + x_{12} \log(1 - p_{11}) + x_{21} \log(1 - p_{22}) + x_{22} \log(p_{22}) + x^{[1]} \left(\frac{1 - p_{22}}{2 - p_{11} - p_{22}} \right) + x^{[2]} \left(\frac{1 - p_{11}}{2 - p_{11} - p_{22}} \right)$$

- Why?

Results

- We could do things closed form (how?)
- Or numerically (see Markov.m)
- What else can we do?

Bayes Rule

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

More Fun with Markovs

- Now imagine we didn't observe the states x_t directly, but observed some noise process y_t
- If state $x_t = 1$, then $y_t \sim \mathcal{N}(\mu_1, \sigma_1^2)$
- If state $x_t = 2$, then $y_t \mathcal{N}(\mu_2, \sigma_2^2)$
- To skip annoying notation for the first step, let's say we knew the first state but from then on out we knew nothing else.

Example

$$x_1 = 1$$

$$\pi = \begin{bmatrix} 0.9 & 0.1 \\ 0.95 & 0.05 \end{bmatrix}$$

$$y_t | x_t = 1 \sim \mathcal{N}(0, 10)$$

$$y_t | x_t = 2 \sim \mathcal{N}(1, 4)$$

$$P(x_t | y_t) = \frac{P(x_t)P(y_t|x_t)}{P(y_t)}$$

See Markov.2

Kalman Filter: Motivation

- Life is full of scenarios in which we see a *signal* about some true underlying process but never observe the truth
  - Missiles
  - Polls
  - Recessions
  - Economic variables
- We typically have some belief of what the underlying object is, where it's going to go, and get some signal related to the object
- How do we put all our information together?

Kalman Filter: Preview to Lemma

Let X explain Y:

$$Y = X\beta + \epsilon$$

Then:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

And:

$$\hat{Y} = X\beta$$

So:

$$\begin{aligned} Var(Y - \hat{Y}|X) &= Var(Y - X\beta|X) \\ &= Var(Y|X) + \beta^2 Var(X|X) - 2\beta Cov(Y, X|X)) \\ &= Var(Y|X) + 2\beta^2 Var(Y)^{-1} \end{aligned}$$

Kalman Filter: Lemma*

If:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} S_{XX'} & S_{XY'} \\ S_{YX'} & S_{YY'} \end{bmatrix}\right)$$

Then:

$$Y|X \sim \mathcal{N}\left(S_{XY'}S_{XX'}^{-1}X, S_{YY'|X}\right)$$

Where, letting $A = S_{XY'}S_{XX}^{-1}$

$$S_{YY'|X} = S_{YY'} - S_{XY'}S_{XX}^{-1}S_{YX'} = S_{YY'} - AS_{XX'}^{-1}A'$$

In other words, our expectation of X given Y comes from a regression, and our conditional variance is our unconditional variance minus the regression coefficient squared times the variance of our signal.

* This and the next two slides are inspired by Harald Uhlig's notation & wonderful slides.

Kalman System

- You have an observation equation:

$$Y_t = H_t \xi_t + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(0, \Sigma_t)$$

- And a state equation:

$$\xi_{t+1} = F_{t+1} \xi_t + \eta_{t+1} \qquad \eta_{t+1} \sim \mathcal{N}(0, \Phi_{t+1})$$

- We assume that ϵ_t and η_t are independent.
- Y_t is a noisy observation of ξ_t, which moves around with noise.

Updating our beliefs

- We start with some beliefs from last period about where ξ would be this period (called $\xi_{t|t-1}$).
- We summarize these as:

$$\xi_{t|t-1} \sim \mathcal{N}\left(\hat{\xi}_{t|t-1}, \Omega_{t|t-1}\right)$$

- We want to look at information today and say what we think ξ is, calling this $\xi_{t|t}$.

First Step: best guess of what the signal will be

- We start with beliefs:

$$\xi_t \sim \mathcal{N}\left(\hat{\xi}_{t|t-1}, \Omega_{t|t-1}\right)$$

- And we know, as a law:

$$Y_t = H_t \xi_t + \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \Sigma_t)$$

- Then we have our best guess of what Y will be, along with its variance:

$$\hat{Y}_t = H_t \xi_{t|t-1}$$

$$S_{YY|t} = H_t \Omega_{t|t-1} H_t + \Sigma_t$$

Second Step: use surprise info to update prior beliefs

- We have our *unexpected* information:

$$\hat{\epsilon}_t = Y_t - \hat{Y}_t$$

- Then our best fit is, like a regression fit:

$$\hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + S_{\xi Y'|t} S_{YY'|t}^{-1} \hat{\epsilon}_t$$

- Where our "signal" is:

$$S_{\xi Y'|t} = \Omega_{t|t-1} H_t'$$

- And our beliefs are updated:

$$\Omega_{t|t} = \Omega_{t|t-1} - S_{\xi Y'|t} S_{YY'|t}^{-1} S_{Y\xi'|t}$$

Third Step: use current best beliefs to find tomorrow's best beliefs

- We have our beliefs for today, $\hat{\xi}_{t|t}$ and $\Omega_{t|t}$
- We want:

$$\xi_{t+1} \sim \mathcal{N}(\hat{\xi}_{t+1|t}, \Omega_{t+1|t})$$

- Update using the law of motion:

$$\hat{\xi}_{t+1|t} = F_{t+1}\hat{\xi}_{t|t}$$

$$\Omega_{t+1|t} = F_{t+1}\Omega_{t|t}F'_{t+1} + \Phi_{t+1}$$

Summarizing the Kalman Filter

$$Y_t \sim \mathcal{N}\left(H_t \xi_t, \Sigma_t\right), \qquad \xi_t \sim \mathcal{N}\left(F_{t+1}\xi_t, \Phi_{t+1}\right)$$

- Given $\xi_t \sim \mathcal{N}(\hat{\xi}_{t|t-1}, \Omega_{t|t-1})$,
  1. Forecast Y_t given what you know:

     $$\hat{Y}_t = H_t \hat{\xi}_{t|t-1} \qquad S_{YY'|t} H_t \Omega_{t|t-1} H_t' + \Sigma_t$$

  2. Update ξ_t given surprise:

     $$\hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + S_{\xi Y'|t} S_{YY'|t}^{-1} (Y_t - \hat{Y}_t)$$

     $$\hat{\Omega}_{t|t} = \hat{\Omega}_{t|t-1} + S_{\xi Y'|t} S_{YY'|t}^{-1} S_{\xi Y'|t}$$

     Where: $S_{\xi Y'|t} = \Omega_{t|t-1} H_t'$
  3. Forecast and set up for tomorrow

     $$\hat{\xi}_{t+1|t} = F_{t+1} \hat{\xi}_{t|t} \qquad \Omega_{t+t|t} = F_{t+1} \Omega_{t|t} F_{t+1}' + \Phi_{t+1}$$

Coding it

See Kalman.m

Uses

- Estimating underlying data, like polls, recessions
- Alternatively, think of your *regression coefficients* as your unknown ξ and your data as Y_t
- Then for:

$$Y_t \sim \mathcal{N}\left(H_t \xi_t, \Sigma_t\right), \quad \xi_t \sim \mathcal{N}\left(F_{t+1} \xi_t, \Phi_{t+1}\right)$$

  - Y_t is your dependent variable
  - H_t is your independent variable
  - Σ_t is your noise term
  - F_{t+1} is just 1, if your coefficients are constant
  - Φ_{t+1} is zero, if your coefficients are constant.
- Now you can run Kalman filter point-by-point on your data to uncover your belief *distribution* over your coefficients.