Time Series Count Models #1

Paul Johnson

Dec 3, 2001

The papers by Brandt et al require us to take 2 preliminary detours, one into probability and the other into the Kalman filter.

1 Conjugate prior and mixture distributions

See the separate handout on this.

2 Kalman filter

See the separate handout on this.

3 Know about moving averages?

3.1 flat moving average.

a k-period Moving Average: $MA(x) = \frac{x_t + x_{t-1} + \ldots + x_{t-k}}{k}$

3.2 Exponentially weighted moving average.

EWMA means you take the "old" value of the EWMA and calculate a new value by averaging in a new piece of input. Begin EWMA at time 0 with a value of 0, say.

Suppose some general formula like:

$$EWMA(t) = w * EWMA(t-1) + (1-w)x_t$$

The weight w is applied to past observations, and the remainder of the weight is put on current input. If w is a big number, the system heavily weights the past, otherwise it does not.

Why call it exponentially weighted? Write out a few steps. Lets start the process at time 0 with a value of 0 for the EWMA:

$EWMA(0) = 0$
$EWMA(1) = (1-w) * x_1 + (w) * EWMA(0) = (1-w)x_1$
$EWMA(2) = (1-w) * x_2 + w * EWMA(1)$

$= (1-w)*x_2 + (1-w)*w*x_1$
$EWMA(3) = (1-w)*x_3 + w*EWMA(2) = (1-w)x_3 + w(1-w)*x_2 + (1-w)w^2x_1$
Get it? The inputs have exponentially decreasing weight.

4 How Brandt et al graft a Poisson model onto a Kalman Filter

They say this is a variant of the extended Kalman filter (which is widely criticized, incidentally)

4.1 State equation.

There is an underlying expected number of counts, μ_t. It is assumed to have this TRANSITION EQUATION

$$\mu_t = e^{r_t}\mu_{t-1}\eta_t$$

This means the state depends on its own past values, multiplied by two coefficients. The variable η is a "shock" to the system, a random error input. The parameter r is the "growth rate". They assume that

1) η_t is Beta distributed, so $\eta \sim \beta(\omega a_{t-1}, (1-\omega)a_{t-1})$

This value a_{t-1} is an "estimated" quantity that is internal to the Kalman filter model.

Recall from Law and Kelton that the expected value of the Beta is the first coefficient divided by the sum of the coefficients, so $E(\eta_t) = \omega a_{t-1}/a_{t-1} = \omega$. Note this means that ω is a "hyperparameter" that discounts past values in making forecasts. That's so because $E(\mu_t) = e^r\mu_{t-1}*\omega$.

4.2 Measurement (observation) equation

The variable μ is the underlying expected number of events, and the observed is given by a Poisson that takes μ as input.

$$Pr(y_t|\mu_t) = \frac{\mu_t^{y_t}e^{-\mu_t}}{y_t!}$$

There is something I do not understand in Brandt, equation (2) page 828. Although the transition equation gives the formula for time dependence of μ, they also assert that the unobserved μ is "parameterized by a multiplicative equation:

$$\mu_t = \mu_t^* exp(X_t\delta)$$

"

Now I easily understand the exp part, but I do not understand the μ_t^* part. They say it is a "time varying component) and make an analogy to including an intercept in a model. You can see that, right? The μ^* is, they say, a "smoothed mean of previous observations".

What I don't understand is how to reconcile the two equations for μ_t.

4.3 Conjugate prior

The authors define the time varying component as a gamma variate:

$$\mu^*_{t-1} \sim Gamma(a_{t-1}, b_{t-1})$$

The variables a and b pop up here, and I know from other literature on the Kalman filter that they are Kalman-inspired. I'm trying to make this more clear than their notation, and failing, but suppose we say the Gamma formula is

$$f(\mu - star|a,b) = \frac{e^{-b\mu - star}(\mu - star)^{a-1}\beta^{\alpha}}{\Gamma(a)}$$

$\mu*$ (or $\mu - star$)!?! evolves.

Your previous mastery of the conjugate priori should come into play now. They originally said that $\mu - star$ depends on a and b. $\mu - star$ plays the part of the "prior" in this story, and in the end we are going to want a distribution for the number of events given a and b.

5 About that Kalman filter application.

Given a stream of data, refer to Y_{t-1} as a collection of observed y's and x's.

Oh, hell. The way this is written up it just seems like a tangle of math and I can't see any way to make it easy to see. The key thing is that you step through time, developing the new estimates for each step on the basis of the information in the previous step.

At time t, you need to create an estimate of the state, given Y_{t-1}. The state is gamma distributed after you assume Y_{t-1} is "resolved". They write

$$\mu_t|Y_{t-1} \sim \Gamma(a_{t|t-1}, b_{t|t-1})$$

So you can make new estimates of the a and b coefficients of the Gamma

$$a_{t|t-1} = \omega a_{t-1}$$

and

$$b_{t|t-1} = \omega b_{t-1} exp(-X_t\delta - r_t)$$

Feed these estimates in to calcuate the expected value of μ_t and the variance of μ_t.

6 And I ran out of gas here...