What's Going On?

Graham Gibson

August 17, 2017

Abstract

What are we doing?????????

Contents

1 Models for Disease Process

1.1 GP

Insert model here. Are there multiple variations on this theme?

- GP + noise model for $X_t \sim X_{t-1}, \ldots, X_{t-L}$. What is variance a function of?
- Hierarchical structure for adjacent spatial units like clusters of districts in Thailand?

1.2 DP

Insert models here

- For $X_t \sim X_{t-1}, \ldots, X_{t-L}$
- Possible methods paper, see below
- Could also end up being used for seasonal disease predictions

1.3 RBF MDN

Equivalent to mixture of normals for (X, Y), conditioned on X

1.4 Tying together horizons

- iterate 1-step-ahead predictions
- Copulas
- all remaining horizons in season jointly estimated at once
- linkages (like copulas, but multivariate marginals) – reserve for a MHC student to explore??

for now, pick something and go with it – maybe a paper here later?

2 Models for Reporting Process

2.1 Krzysztof's model for dengue in Thailand

Can be used for either individual level data or case counts.

Insert description here. Basic idea:

$X_{ij} \sim Poisson(\lambda_i p_{ij})$ $p_{ij} = CDF(j - i|\alpha, \beta)$, where CDF comes from K's survival model.

2.2 Flu – Errors survival model

Reporting delay errors occur according to some model that allows for positive or negative values, and they persist for some number of weeks.

Possibly integrate with seasonal fluctuations in reporting.

3 Paper Concepts

In some kind of order from firstish to lastish.

3.1 Flexible Stick Breaking for Dirichlet Process and Probit Stick Breaking Process Mixture of Regression Models

In DP and PSBP (is that the right acronym?) models, number of mixture components assigned non-negligible weight is sensitive to interactions between a variety of factors, including the functional form used in the stick breaking process, priors on parameters used in the stick breaking process, and priors on mixture component regression coefficients. Most commonly (??), in Dirichlet Process mixture of regressions (and possibly PBSP too??) a logistic-linear specification in x is used for stick breaking. We propose a novel, more flexible functional form for stick breaking using a BNN. Basically, this allows the model to estimate upper and lower bounds on the weight that should be assigned to a single mixture component. This lets the model distribute weight across multiple mixture components more easily. (BUT WE STILL NEED TO EVALUATE RELATIVE TO SOME OTHER OPTIONS/MAKE SURE THIS STORY PANS OUT).

- Use BNN for breaking process probabilities: more flexible than standard linear form.
- Compare to 2-3 common alternatives chosen from the following candidates:
  - stick breaking prob = logistic(a + bx)
  - paper with empirical bayes priors for DP
  - probit stick breaking process?
  - basic idea: 1-2 "default" implementations and 1-2 "fancy/new" implementations?
- Simulation Study
- Apply to infectious disease prediction h steps ahead (not seasonal targets?). Also compare this to a linear model a la SARIMA?
- Apply to something else DP mixtures of regressions have been used for previously?

3.2 Hierarchical Non-Parametric Density Estimation

We would like to borrow information across adjacent spatial units using hierarchical model structures. We would like to do (conditional) density estimation in a flexible way, using non-parametric or flexible semi-parametric methods. How can we do that?

- Hierarchical GP is the most obvious answer
- Is there a way to do this with a DP flavored method?

- simulation study
- Potential application: Groups of adjacent/similar districts in Thailand
- Can we also estimate clustering/hierarchy based on similarity of time series??

3.3 Integrating Reporting Delays with Prediction in Thailand

Predictions in Thailand would benefit from integrating a reporting delay model with a disease process model. In a Bayesian setup, this could use any of the non-parametric methods outlined above and combine with K's reporting delay model. Note that K's model can work with case count data as briefly outlined above.

- Province-level prediction in Thailand: 76 provinces
- at the level of case count data
- this probably comes after a more basic application of the disease process model that doesn't deal with reporting delays...

3.4 Integrating Reporting Delays with Prediction for Flu

Evan will be thinking about this, reporting delay stuff different from