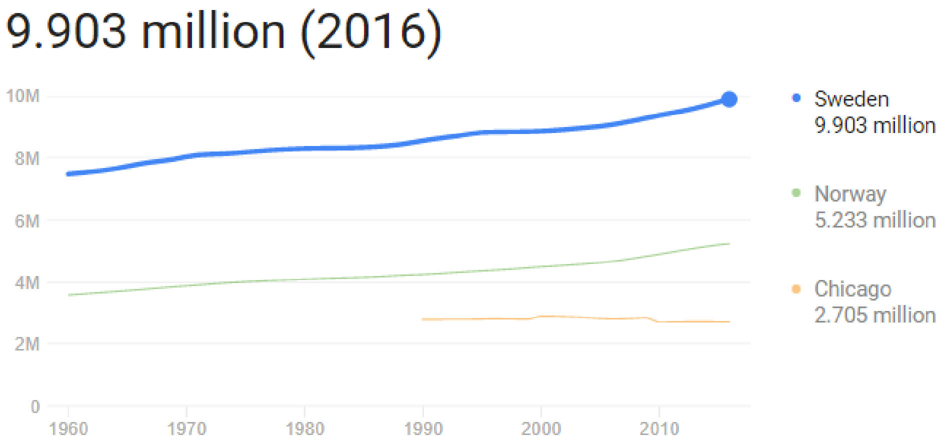
Sweden's Schools

Prediction of Swedish School Fires

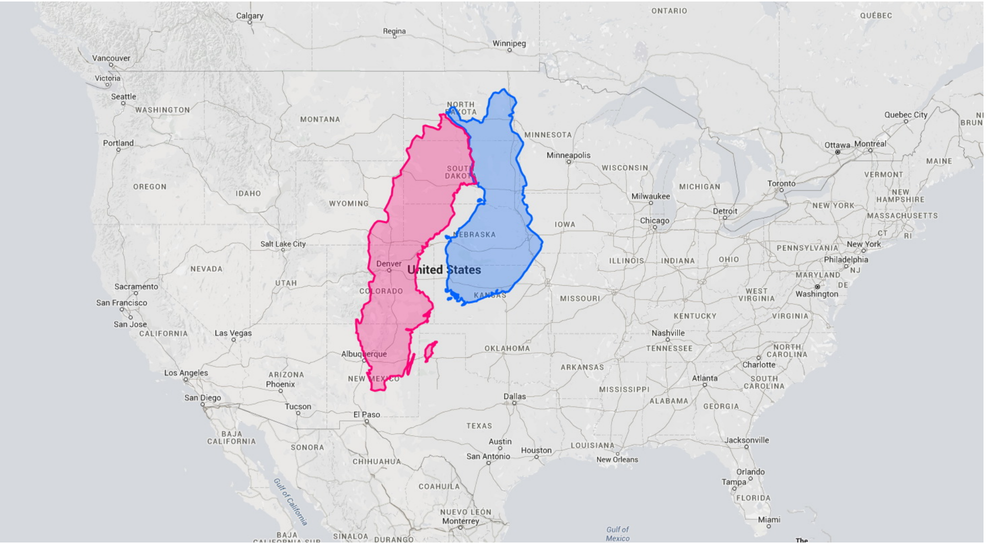
University of Michigan

Desmond Cole
Andrei Kopelevich
Mark Kurzeja
Teerth Patel

Introduction



Introduction



Introduction

Sweden has a serious problem. A problem with school fires. Every single day, an average of one to two schools are set ablaze.

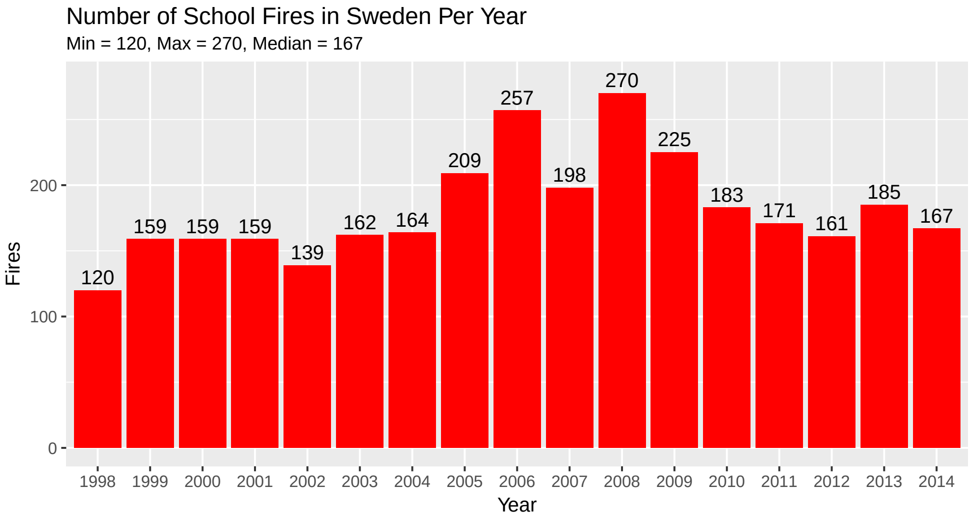
Introduction

Sweden has a serious problem. A problem with school fires. Every single day, an average of one to two schools are set ablaze.

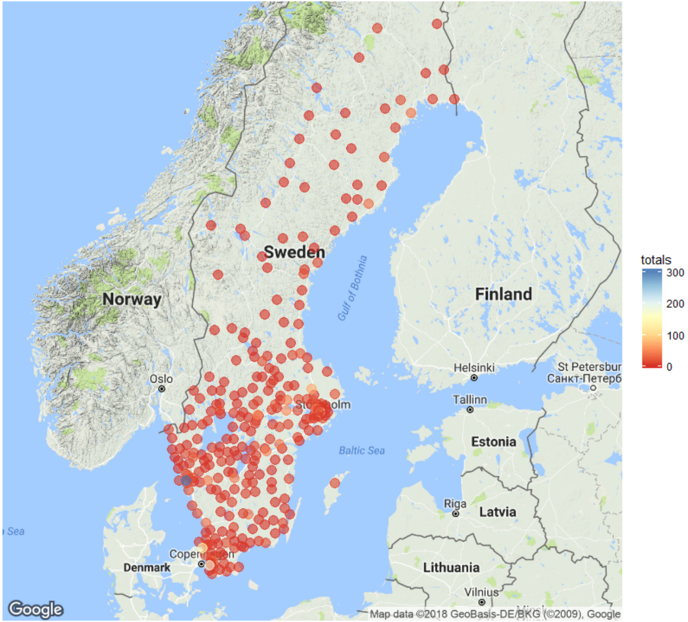
Introduction

Sweden has a serious problem. A problem with school fires. Every single day, an average of one to two schools are set ablaze.

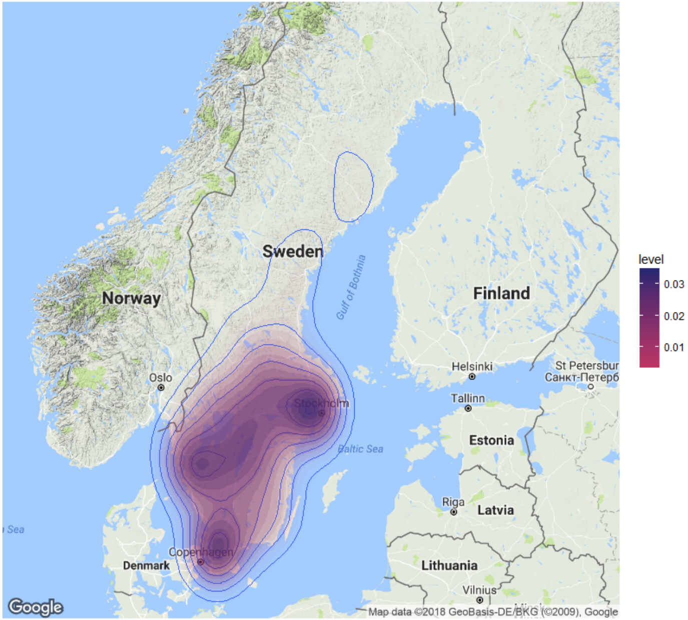
By the numbers



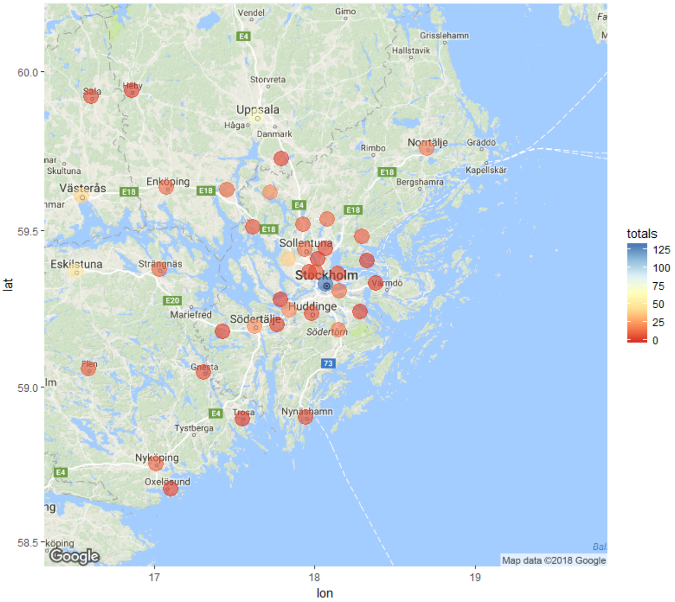
Geographic Heatmap



Geographic Heatmap



Geographic Heatmap



Introduction

Background:

- Sweden has a surprisingly large number of school fires for a small country ($<$ 10M inhabitants)
- "Almost every day between one and two school fires occur in Sweden. In most cases arson is the cause of the fire."
- The associated costs can be up to a billion SEK (around 120 million USD) per year.

Aim:

- To find out which properties and indicators of Swedish towns (municipalities, to be exact) might be related to a high frequency of school fires
- Explore data to determine risk factors
  - Geographic
  - Demographic
  - Cultural

Introduction

Background:

- Sweden has a surprisingly large number of school fires for a small country ($<$ 10M inhabitants)
- "Almost every day between one and two school fires occur in Sweden. In most cases arson is the cause of the fire."
- The associated costs can be up to a billion SEK (around 120 million USD) per year.

Aim:

- To find out which properties and indicators of Swedish towns (municipalities, to be exact) might be related to a high frequency of school fires
- Explore data to determine risk factors
  - Geographic
  - Demographic
  - Cultural

Introduction

Background:

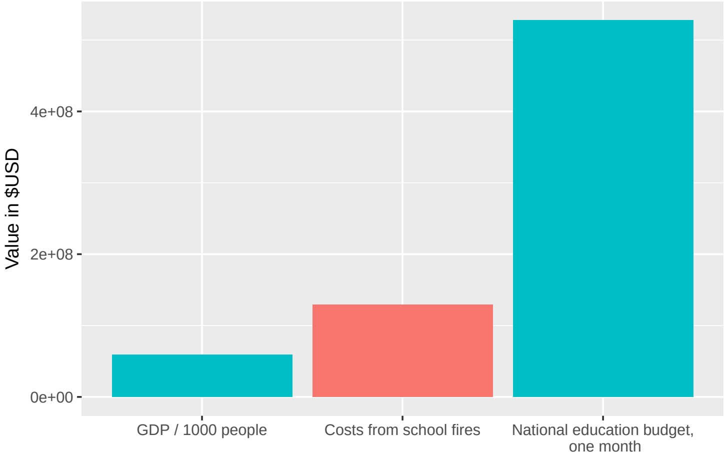
- Sweden has a surprisingly large number of school fires for a small country ($<$ 10M inhabitants)
- "Almost every day between one and two school fires occur in Sweden. In most cases arson is the cause of the fire."
- The associated costs can be up to a billion SEK (around 120 million USD) per year.

Aim:

- To find out which properties and indicators of Swedish towns (municipalities, to be exact) might be related to a high frequency of school fires
- Explore data to determine risk factors
  - Geographic
  - Demographic
  - Cultural

The Costs...

- The per-capita GDP of Sweden is $\approx \$51.6K$ USD
- Thus school fires cost approximately 2,160 people their entire productivity for an entire year...



The Data Set

The data set, indexed on a per-municipality basis, is quite rich. We have information on things as diverse as:

- Income information
- Unemployment information
- Crime Rates
- Foreign and domestic settlement rates
- Satisfaction and psychological measures
- Per Capita Transportation Information
  - Motorcycles
  - Autos
  - Snowmobiles
  - Tractors

The Data Set

Goal:

- Combine Demographic and Social Measures to identify high-risk areas and trends.
- Predictive Analysis
- Potential Cost Analysis

Potential Pitfalls:

- Repeated Measurements
- Noisy Data
- Correlation of Neighbors

The Data Set

Goal:

- Combine Demographic and Social Measures to identify high-risk areas and trends.
- Predictive Analysis
- Potential Cost Analysis

Potential Pitfalls:

- Repeated Measurements
- Noisy Data
- Correlation of Neighbors

Discussion of Methods

Why Bayesian:

- Multiple observations on a single municipality ⇒ Errors exhibit correlation by municipality
- Different years have different mean fire rates which need to be controlled
- Each year, each municipality has a different population size
- Potentially a correlation between years

Contemplated Methods:

- Linear Varying Effects Varying Intercepts Model
- Linear Varying Effects Varying Intercepts Model with Correlation
- Hierarchical Model with Time Series Components
- Inspiration from http://tharte.github.io/mbt/ and http://www.ling.uni-potsdam.de/ vasishth/statistics/BayesLMMs.html

Discussion of Methods

Why Bayesian:

- Multiple observations on a single municipality $\Rightarrow$ Errors exhibit correlation by municipality
- Different years have different mean fire rates which need to be controlled
- Each year, each municipality has a different population size
- Potentially a correlation between years

Contemplated Methods:

- Linear Varying Effects Varying Intercepts Model
- Linear Varying Effects Varying Intercepts Model with Correlation
- Hierarchical Model with Time Series Components
- Inspiration from http://tharte.github.io/mbt/ and http://www.ling.uni-potsdam.de/ vasishth/statistics/BayesLMMs.html

Discussion of Methods

Why Bayesian:

- Multiple observations on a single municipality $\Rightarrow$ Errors exhibit correlation by municipality
- Different years have different mean fire rates which need to be controlled
- Each year, each municipality has a different population size
- Potentially a correlation between years

Contemplated Methods:

- Linear Varying Effects Varying Intercepts Model
- Linear Varying Effects Varying Intercepts Model with Correlation
- Hierarchical Model with Time Series Components
- Inspiration from http://tharte.github.io/mbt/ and http://www.ling.uni-potsdam.de/ vasishth/statistics/BayesLMMs.html

Distribution of Work

This project will require quite a lot of work to get off the ground. Included below are the current allocation of tasks for the project.

Andrei	Overall documentation and model validation
Desmond	Pre-model exploration and processing
Mark	Model Specification and Implementation in Base Stan
Teerth	Data Visualization & Post-model analysis

Example Model Structure

First some notation for the example linear model

- Let M be the number of municipalities, indexed by m,
- Let Y be the number of years, indexed by y,
- Let V be the number of explanatory variables, indexed by v_i,
- Let β denote a parameter and let subscripts denote indexing: $\beta_{(m)}$ is the multiplier for a given municipality regardless of year or other variables, and $\beta_{(v_i,y)}$ would be the multiplier of the i^{th} variable during the y^{th} year
- $\mathbb{I}\left[\Phi = \phi\right]$ is an indicator function that is one when $\Phi = \phi$ and zero otherwise

Example Model Structure - Basic Linear Model

Key detail of the model: All years and municipalities have the same mean - it is a linear function of the (scaled) predictors

$$\beta_0, \beta_{v_i} \sim P_\bullet(\bullet) \quad \text{[Choice of Priors]}$$

$$\lambda = \beta_0 + \sum_{i=1}^{V} \beta_{v_i} v_i$$

$$Fires_{m,y} \sim Pois(\exp[\lambda])$$

Some choices for $P_\bullet(\bullet)$:

Weakly Informative Prior: Let data drive the analysis with a weak prior

Strong Normal Prior: Impose a strong normal prior around zero so that most coefficients will remain close to zero except for a select few

Spike-and-Slab: Impose a mixture model for the coefficients in the form of a spike-and-slab prior to force most distributions to zero and let others vary

Example Model Structure - Basic Linear Model

Key detail of the model: All years and municipalities have the same mean - it is a linear function of the (scaled) predictors

$$\beta_0, \beta_{v_i} \sim P_\bullet(\bullet) \quad \text{[Choice of Priors]}$$

$$\lambda = \beta_0 + \sum_{i=1}^{V} \beta_{v_i} v_i$$

$$Fires_{m,y} \sim Pois(\exp[\lambda])$$

Some issues with the Basic Linear model:

Correlated Errors
: We have multiple observations for each municipality, and thus the independence of the errors is violated

Years clearly have different effects
: The mean response for each year is clearly different from other years - see the histogram of effects at the beginning of the presentation

Municipality Differences
: Different Locations may respond differently to different predictors

Example Model Structure - Basic Linear Model

Key detail of the model: All years and municipalities have the same mean - it is a linear function of the (scaled) predictors

$$\beta_0, \beta_{v_i} \sim P_\bullet(\bullet) \quad \text{[Choice of Priors]}$$

$$\lambda = \beta_0 + \sum_{i=1}^{V} \beta_{v_i} v_i$$

$$Fires_{m,y} \sim Pois(\exp[\lambda])$$

Whats with the exp function wrapped around the mean variable?

- λ takes its support on the positive reals, but we do not impose a positive definite solution on our linear equation form
- We use the exponential to map $\mathbb{R}$ to $\mathbb{R}^+$ in a one-to-one fashion so that we do not need to impose domain restrictions in the model
- Truthfully, Stan just yelled at us for domain restrictions on a definite linear function... a lot ☹

Example Model Structure - Basic Linear Model

Key detail of the model: All years and municipalities have the same mean - it is a linear function of the (scaled) predictors

$$\beta_0, \beta_{v_i} \sim P_\bullet(\bullet) \quad \text{[Choice of Priors]}$$

$$\lambda = \beta_0 + \sum_{i=1}^{V} \beta_{v_i} v_i$$

$$Fires_{m,y} \sim Pois(\exp[\lambda])$$

Whats with the exp function wrapped around the mean variable?

- λ takes its support on the positive reals, but we do not impose a positive definite solution on our linear equation form
- We use the exponential to map $\mathbb{R}$ to $\mathbb{R}^+$ in a one-to-one fashion so that we do not need to impose domain restrictions in the model
- Truthfully, Stan just yelled at us for domain restrictions on a definite linear function... a lot ☹

Choices of Models

We have a lot of directions to build off of this fixed effects model. We then take inspiration from Alice on this one:

"Would you tell me, please, which way I ought to go from here?"
"That depends a good deal on where you want to get to."
"I don't much care where –"
"Then it doesn't matter which way you go."
– Lewis Carroll, Alice in Wonderland

We choose to go down the mixed-effects modeling path and we will proceed with adding to our model to account for issues that arise in our data.

Choices of Models

We have a lot of directions to build off of this fixed effects model. We then take inspiration from Alice on this one:

> *"Would you tell me, please, which way I ought to go from here?"*
> *"That depends a good deal on where you want to get to."*
> *"I don't much care where –"*
> *"Then it doesn't matter which way you go."*
> *– Lewis Carroll, Alice in Wonderland*

We choose to go down the mixed-effects modeling path and we will proceed with adding to our model to account for issues that arise in our data.

Choices of Models

We have a lot of directions to build off of this fixed effects model. We then take inspiration from Alice on this one:

> *"Would you tell me, please, which way I ought to go from here?"*
> *"That depends a good deal on where you want to get to."*
> *"I don't much care where –"*
> *"Then it doesn't matter which way you go."*
> *– Lewis Carroll, Alice in Wonderland*

We choose to go down the mixed-effects modeling path and we will proceed with adding to our model to account for issues that arise in our data.

Example Model Structure - Mixed Effects Model

Key detail of the model: Each year and municipality has a different mean that is incorporated - all variables are still included and each have the same overall effect regardless of municipality or year

$$\beta_0, \beta_{v_i}, \beta_{(m)}, \beta_{(y)} \sim P_\bullet(\bullet) \quad \text{[Choice of Priors]}$$

$$\lambda_{m,y} = \beta_0 + \beta_{(m)} + \beta_{(y)} + \sum_{i=1}^{V} \beta_{v_i} v_i$$

$$Fires_{m,y} \sim Pois(\exp[\lambda_{m,y}])$$

Some choices for $P_\bullet(\bullet)$ (as before):

- Weakly Informative Prior
- Strong Normal Prior
- Spike-and-Slab

Example Model Structure - Mixed Effects Model

Key detail of the model: Each year and municipality has a different mean that is incorporated - all variables are still included and each have the same overall effect regardless of municipality or year

$$\beta_0, \beta_{v_i}, \beta_{(m)}, \beta_{(y)} \sim P_\bullet(\bullet) \quad \text{[Choice of Priors]}$$

$$\lambda_{m,y} = \beta_0 + \beta_{(m)} + \beta_{(y)} + \sum_{i=1}^{V} \beta_{v_i} v_i$$

$$Fires_{m,y} \sim Pois(\exp[\lambda_{m,y}])$$

Some issues with the Mixed Effects Model

Yearly Trends The effect of different variables could vary by year which could be a source of error

Noise With so many predictors, economic noise could really begin to be an issue if we do not introduce some regularizations

Example Model Structure - Mixed Effects Model 2

Key detail of the model: Each year and municipality has a different mean that is incorporated - all variables are still included and each has a different effect based on the year

$$\beta_0, \beta_{(v_i,y)}, \beta_{(m)}, \beta_{(y)} \sim P_\bullet(\bullet) \quad \text{[Choice of Independent Priors]}$$

$$\lambda_{m,y} = \beta_0 + \beta_{(m)} + \beta_{(y)} + \sum_{i=1}^{V} \beta_{(v_i,y)} v_i$$

$$Fires_{m,y} \sim Pois(\exp[\lambda_{m,y}])$$

Some choices for $P_\bullet(\bullet)$ (as before):

- Weakly Informative Prior
- Strong Normal Prior
- Spike-and-Slab

Example Model Structure - Mixed Effects Model 2

Key detail of the model: Each year and municipality has a different mean that is incorporated - all variables are still included and each has a different effect based on the year

$$\beta_0, \beta_{(v_i,y)}, \beta_{(m)}, \beta_{(y)} \sim P_\bullet(\bullet) \quad \text{[Choice of Independent Priors]}$$

$$\lambda_{m,y} = \beta_0 + \beta_{(m)} + \beta_{(y)} + \sum_{i=1}^{V} \beta_{(v_i,y)} v_i$$

$$Fires_{m,y} \sim Pois(\exp[\lambda_{m,y}])$$

Some issues with the Mixed Effects Model 2:

Noise With so many predictors, economic noise could really begin to be an issue if we do not introduce some regularizations

Correlations We still have not dealt with geographic correlations

Example Model Structure - Mixed Effects Model 3

Key detail of the model: Each year and municipality has a different mean that is incorporated (and regularized by a hierarchical prior) - all variables are still included and each has a different effect based on the year.

$$\begin{aligned}
\beta_0 &\sim P_\bullet(\bullet) \quad \text{[Choice of Independent Priors]} \\
\beta_{(v_i,y)} &\sim N(\mu_{variable_i}, \sigma_{variable_i}) \quad \text{[Hierarchical Prior]} \\
\beta_{(y)} &\sim N(\mu_{years}, \sigma_{years}) \quad \text{[Hierarchical Prior]} \\
\beta_{(m)} &\sim N(\mu_{muni}, \sigma_{muni}) \quad \text{[Hierarchical Prior]} \\
\lambda_{m,y} &= \beta_0 + \beta_{(m)} + \beta_{(y)} + \sum_{i=1}^{V} \beta_{(v_i,y)} v_i \\
Fires_{m,y} &\sim Pois(\exp[\lambda_{m,y}])
\end{aligned}$$

Shrinkage from the hierarchical prior prevents unconstrained departures from the group average effect making analysis more tractable. This model captures a great amount of variability.

Example Model Structure - Mixed Effects Model 3

$$\beta_0 \sim P_\bullet(\bullet) \quad \text{[Choice of Independent Priors]}$$
$$\beta_{(v_i,y)} \sim N(\mu_{variable_i}, \sigma_{variable_i}) \quad \text{[Hierarchical Prior]}$$
$$\beta_{(y)} \sim N(\mu_{years}, \sigma_{years}) \quad \text{[Hierarchical Prior]}$$
$$\beta_{(m)} \sim N(\mu_{muni}, \sigma_{muni}) \quad \text{[Hierarchical Prior]}$$
$$\lambda_{m,y} = \beta_0 + \beta_{(m)} + \beta_{(y)} + \sum_{i=1}^{V} \beta_{(v_i,y)} v_i$$
$$Fires_{m,y} \sim Pois(\exp[\lambda_{m,y}])$$

Some issues with the Mixed Effects Model 3

Lots of slope variables: We still have north of 1000 slope variables and some predictions may not be significantly different from zero... Perhaps take inspiration from the Hoff problem in Homework 4...

Correlations: We still have not dealt with geographic correlations

Example Model Structure - Mixed Effects Model 3

$$\beta_0 \sim P_\bullet(\bullet) \quad \text{[Choice of Independent Priors]}$$

$$\beta_{(v_i,y)} \sim N(\mu_{variable_i}, \sigma_{variable_i}) \quad \text{[Hierarchical Prior]}$$

$$\beta_{(y)} \sim N(\mu_{years}, \sigma_{years}) \quad \text{[Hierarchical Prior]}$$

$$\beta_{(m)} \sim N(\mu_{muni}, \sigma_{muni}) \quad \text{[Hierarchical Prior]}$$

$$\lambda_{m,y} = \beta_0 + \beta_{(m)} + \beta_{(y)} + \sum_{i=1}^{V} \beta_{(v_i,y)} v_i$$

$$Fires_{m,y} \sim Pois(\exp[\lambda_{m,y}])$$

Some issues with the Mixed Effects Model 3

Lots of slope variables We still have north of 1000 slope variables and some predictions may not be significantly different from zero... Perhaps take inspiration from the Hoff problem in Homework 4...

Correlations We still have not dealt with geographic correlations

Example Model Structure - Mixed Effects Model 4

Key detail of the model: Each year and municipality has a different mean that is incorporated (and regularized by a hierarchical prior) - all variables can now "disappear" based on their overall effect in the model and year

$$\begin{aligned}
\beta_0 &\sim P_{\bullet}(\bullet) \quad \text{[Choice of Independent Priors]} \\
\beta_{(v_i,y)} &\sim N(\mu_{variable_i}, \sigma_{variable_i}) \quad \text{[Hierarchical Prior]} \\
p_{v_i} &\sim Beta\left({}^1\!/_2, {}^1\!/_2\right) \\
\zeta_{v_i} &\sim Bin(n = 1, p = p_{v_i}) \quad \text{i.e.} \quad \zeta_{v_i} \in [0, 1] \\
\beta_{(y)} &\sim N(\mu_{years}, \sigma_{years}) \quad \text{[Hierarchical Prior]} \\
\beta_{(m)} &\sim N(\mu_{muni}, \sigma_{muni}) \quad \text{[Hierarchical Prior]} \\
\lambda_{m,y} &= \beta_0 + \beta_{(m)} + \beta_{(y)} + \sum_{i=1}^{V} \beta_{(v_i,y)} \zeta_{v_i} v_i \\
Fires_{m,y} &\sim Pois(\exp[\lambda_{m,y}])
\end{aligned}$$

Example Model Structure - Mixed Effects Model 4

$$
\begin{aligned}
\beta_0 &\sim P_\bullet(\bullet) \quad \text{[Choice of Independent Priors]} \\
\beta_{(v_i,y)} &\sim N(\mu_{variable_i}, \sigma_{variable_i}) \quad \text{[Hierarchical Prior]} \\
p_{v_i} &\sim Beta\left(1/2, 1/2\right) \\
\zeta_{v_i} &\sim Bin(n = 1, p = p_{v_i}) \quad \text{i.e.} \quad \zeta_{v_i} \in [0, 1] \\
\beta_{(y)} &\sim N(\mu_{years}, \sigma_{years}) \quad \text{[Hierarchical Prior]} \\
\beta_{(m)} &\sim N(\mu_{muni}, \sigma_{muni}) \quad \text{[Hierarchical Prior]} \\
\lambda_{m,y} &= \beta_0 + \beta_{(m)} + \beta_{(y)} + \sum_{i=1}^{V} \beta_{(v_i,y)} \zeta_{v_i} v_i \\
Fires_{m,y} &\sim Pois(\exp[\lambda_{m,y}])
\end{aligned}
$$

The p_{v_i} variable can be interpreted to be the “probability” that a given variable is non-zero regardless of the year. This allows us to perform implicit feature selection and improve the fit. Bad variables will have low probability and good variables should have high probabilities.

Example Model Structure - Mixed Effects Model 4

$$
\begin{aligned}
\beta_0 &\sim P_\bullet(\bullet) \quad \text{[Choice of Independent Priors]} \\
\beta_{(v_i,y)} &\sim N(\mu_{variable_i}, \sigma_{variable_i}) \quad \text{[Hierarchical Prior]} \\
p_{v_i} &\sim Beta\left(1/2, 1/2\right) \\
\zeta_{v_i} &\sim Bin(n = 1, p = p_{v_i}) \quad \text{i.e.} \quad \zeta_{v_i} \in [0, 1] \\
\beta_{(y)} &\sim N(\mu_{years}, \sigma_{years}) \quad \text{[Hierarchical Prior]} \\
\beta_{(m)} &\sim N(\mu_{muni}, \sigma_{muni}) \quad \text{[Hierarchical Prior]} \\
\lambda_{m,y} &= \beta_0 + \beta_{(m)} + \beta_{(y)} + \sum_{i=1}^{V} \beta_{(v_i,y)} \zeta_{v_i} v_i \\
Fires_{m,y} &\sim Pois(\exp\left[\lambda_{m,y}\right])
\end{aligned}
$$

Some issues with the Mixed Effects Model 4

Correlations We still have not dealt with geographic nor temporal correlations

Graphical Model

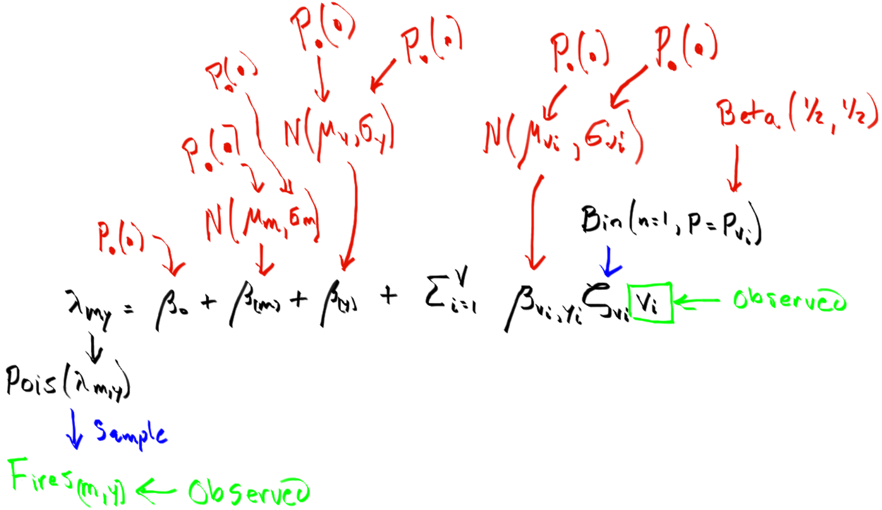


Figure: Let $P_\bullet(\bullet)$ be a (currently unspecified) modeling decision for a prior distribution. Red Denotes Priors, Green is Observed Data, and Black denotes structural forms

The King Once Said

Some issues with the Mixed Effects Model 4

Correlations We still have not dealt with geographic nor temporal correlations

"Begin at the beginning," the King said, very gravely, "and go on till you come to the end: then stop."
– Lewis Carroll, Alice in Wonderland

It is here that we have decided to end our journey down the modeling "rabbit-hole" and "come up for air". If our model is not able to sufficiently capture the complexity of the interactions, we will choose to incorporate temporal correlation and maybe spacial correlation. Until then, we chose to not complicate the model further.

The King Once Said

Some issues with the Mixed Effects Model 4

Correlations We still have not dealt with geographic nor temporal correlations

"Begin at the beginning," the King said, very gravely, "and go on till you come to the end: then stop."
– Lewis Carroll, Alice in Wonderland

It is here that we have decided to end our journey down the modeling "rabbit-hole" and "come up for air". If our model is not able to sufficiently capture the complexity of the interactions, we will choose to incorporate temporal correlation and maybe spacial correlation. Until then, we chose to not complicate the model further.

The King Once Said

Some issues with the Mixed Effects Model 4

Correlations We still have not dealt with geographic nor temporal correlations

"Begin at the beginning," the King said, very gravely, "and go on till you come to the end: then stop."
– Lewis Carroll, Alice in Wonderland

It is here that we have decided to end our journey down the modeling "rabbit-hole" and "come up for air". If our model is not able to sufficiently capture the complexity of the interactions, we will choose to incorporate temporal correlation and maybe spacial correlation. Until then, we chose to not complicate the model further.

Preliminary Findings - Model

To get an initial sense of predictor-outcome relationships, we ran a preliminary model using the rstanarm package. This preliminary model uses total recent fires (from 2010-2014) as the outcome, does not include year/municipality-specific parameters, and uses 2-knot spline functions S of each included predictor.

$$Fires_m \sim Poisson(\lambda)$$

$$Fires_m = \beta_0 + \sum_{i=1}^{V} \beta_i S(v_i)$$

$$\beta_1, ..., \beta_V \sim N(0, 1)$$

Preliminary Findings - Parameters