Multiple regression, PyMC3, and a tiny bit of causal inference

ISTA 410 / INFO 510: Bayesian Modeling and Inference

U. of Arizona School of Information

September 9, 2020

Outline

- Return to the Waffle House example
- Specifying models in PyMC3
- DAGs and a little causal inference

Next assignment will be posted later tonight (due 9/18), feedback from HW1 by Friday

Recap

Last week

Things we covered last week:

- Normal model, known variance
- Normal model, unknown variance
- A little bit of regression

Things we didn't cover:

- Conjugate prior for normal model, unknown variance (see section 3.3 for details)

Basic specification of a linear regression

Basic specification of a linear regression model (a little more broad than last time):

$$
\begin{aligned}
y_i|\beta, \sigma, \mathsf{x} &\sim \mathrm{Normal}(\alpha + \beta \cdot \mathsf{x}, \sigma) \\
\alpha &\sim \mathrm{Normal}(0, 0.2) \\
\beta_j &\sim \mathrm{Normal}(0, 0.5) \\
\sigma &\sim \mathrm{HalfCauchy}(10)
\end{aligned}
$$

Notes

- Typical constant-variance assumption is not strictly required
- Priors assume standardized data – more on next slide
- Half Cauchy prior on σ is qualitatively similar to the uninformative σ^{-2} prior but is proper. Exponential could be another reasonable prior.

Priors

We set the following priors for α, β:

- $\alpha \sim \mathrm{Normal}(0, 0.2)$
  With standardized data α really has to be 0! In simulations it should be very close to 0. We're including it as an exercise to show that.
- $\beta_j \sim \mathrm{Normal}(0, 0.5)$
  We shouldn't see impossibly strong relationships. Flat priors would allow arbitrarily large slopes – impossible with standardized data. This prior is pretty vague besides that.

Priors and regularization

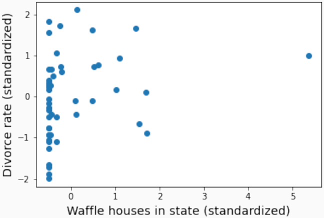
Remember from last time:

- Flat prior on $\beta \leftrightarrow$ ordinary least squares
- Normal prior on $\beta \leftrightarrow$ Ridge regression
- $|\beta| \sim \text{Exponential}(\lambda) \leftrightarrow$ lasso regression (this prior is also called a Laplace distribution)

Waffle house, divorce, regression

Waffle House and divorce

Last time we saw an example of Waffle House as a predictor for divorce rate.



Other predictors

Fortunately the Waffle House dataset contains a few other variables that we can use as predictors:

- Marriage rate
- Median age at (first) marriage

We'll explore the effects of those later.

There are a few more potential predictors in the data set – feel free to experiment.

Other predictors

Multiple regression models will allow us to:

- Remove spurious associations
- Uncover masked associations
- Cause spurious associations and hide real associations

Because these models can do both good and bad things – they can help or hurt our understanding – we will need to think carefully about what variables to include.

Specifying a model in PyMC3

PyMC3

PyMC3 is a Bayesian modeling and computation framework

- MC is for "Markov chain" (or maybe "Monte Carlo") but we'll hold the MCMC for a few weeks from now
- `conda install -c conda-forge pymc3`
  or `pip install pymc3`
- Uses Theano as a backend for tensor calculations, but you usually won't have to interact directly with these
- There is a PyMC4 that uses TensorFlow, but it's still in alpha

PyMC3: specifying a model

On paper, we've specified models by specifying the distribution of the data conditional on parameters, and then prior distributions for those parameters, as in:

$$y_i \sim \mathrm{Normal}(\theta, \sigma)$$
$$\theta \sim \mathrm{Normal}(\mu_0 = 192, \tau_0 = 17)$$

(the known-variance normal model from last week)

Specifying the same model in PyMC3 looks very similar!

PyMC3: specifying a model

The same model in PyMC3 looks like:

```
import pymc3 as pm

with pm.Model() as normal_model:
    y = pm.Normal('y', mu=theta, sigma = 14,
                  observed = bball['Combined'])
    theta = pm.Normal('theta', mu=192, sigma=17)
    # do stuff with the model here
```

PyMC3 uses Python context management extensively, so everything happens behind a `with` statement

Specifying a linear model

A linear model for the divorce-Waffle House model:

$$\begin{aligned}
D &\sim \mathrm{Normal}(\mu, \sigma) \\
\mu &= \alpha + \beta_W W \\
\alpha &\sim \mathrm{Normal}(0, 0.2) \\
\beta_W &\sim \mathrm{Normal}(0, 0.5) \\
\sigma &\sim \mathrm{HalfCauchy}(10)
\end{aligned}$$

Specifying a linear model

```
with pm.Model() as linear_model:
    alpha = pm.Normal('alpha', mu=0, sigma=0.2)
    betaW = pm.Normal('betaW', mu=0, sigma=0.5)
    s = pm.HalfCauchy('s', beta = 10)
    mu = alpha + betaW * waffles['WaffleHouses']
    d = pm.Normal('d', mu = mu, sigma = s,
                  observed = waffles['Divorce'])
    # inference stuff goes here
```

Normal approximation

What is our “inference stuff”?

For now, quadratic (normal) approximation:

- Find the MAP estimate
- Approximate the log posterior as a quadratic around that point

Log posterior quadratic $\Rightarrow$ posterior Gaussian, so this is just a normal approximation

In the future: sampling from the posterior using HMC

Directed acyclic graphs and causal inference

It's not the waffles

Probably we all guessed this, but:

- Waffle Houses don't cause divorce
- Some other property influences the presence of WH and also the rate of divorce

It’s not the waffles

Probably we all guessed this, but:

- Waffle Houses don’t cause divorce
- Some other property influences the presence of WH and also the rate of divorce

- The US South

Anything idiosyncratic to the South will be associated with Waffle House – let’s look at a couple of others

Other predictors

We have two other predictors in the data set that have more plausible causal relationships with divorce rate

- Marriage rate
    - You have to get married to get divorced (+)
    - Society values marriage, so opposes divorce (-)
- Median age at (first) marriage
    - Younger people make worse decisions
    - People change

DAGs

What's a DAG?

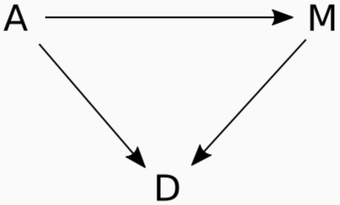
- Directed (edges are arrows)
- Acyclic (No directed loops)
- Graph (nodes and edges)

Use as a heuristic model for causal relationships

- Not a mechanical model – does not include explanation of how or why the causal relationship exists

Our first DAG

Here is a DAG representing a possible model for the relationships between age at marriage, marriage rate, and divorce rate:



In this model, total causal effect of A on D:

1. $A \rightarrow D$ – direct causal effect
2. $A \rightarrow M \rightarrow D$ – indirect effect

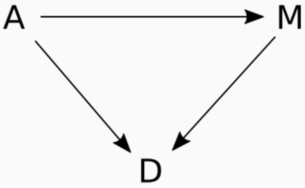
Multiple regression and control

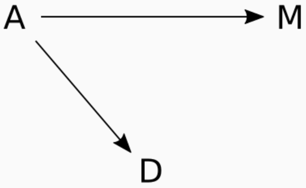
Multiple regression provides statistical "control." This means *conditioning on the information in one variable*, not *setting the value of one variable*

- Multiple regression answers: once we know all other predictors, how is each predictor associated with the outcome?
- To interpret the effect of statistical control, we need a clear model of what the causal relationships might be – this is what the DAG offers us

Two competing DAGs

Here are two DAGs:

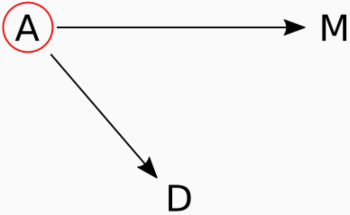




Simple regression can't distinguish between these, but multiple regression can, so let's go to the computer

Conditioning on the

Here, adding variables to the regression helped us determine the effect:



Conditioning on A eliminated the spurious association of M with D

In some cases, conditioning on a variable can *introduce* a spurious association – depends on the structure of the DAG

Summary

Summary:

- Specifying models in PyMC3 looks like specifying models on paper
- Multiple regression allows statistical “control”
- DAGs help us think about causality

Next time

- More on DAGs
- The “backdoor criterion”
- Different sources of confounding