Bayesian Hierarchical Modeling in Stan

Dina Sinclair

21 June 2018

Outline

Bayesian?? Hierarchical Modeling?? Stan??

Frequentist Approach

Frequentists base their conclusions on *significance testing*. This involves:

- A hypothesis to test
- A significance level α
- a p-value
- statistical significance when $p < \alpha$

Bayesian Approach

Bayesians like to use *Bayes' theorem*, which looks at the probability event θ will happen given event y happened already:

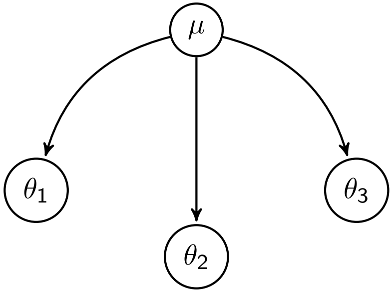
$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)}$$

This allows us to condition on y, our prior knowledge.

"To the non-Bayesian, a Bayesian is someone who pollutes clean data with a subjective prior distribution. But, to the Bayesian, a classical statistician is someone who arbitrarily partitions all the available information into something called "the data" which can be analyzed and something called "prior information" which is off limits" - Andrew Gelman

Hierarchical Models

Bayesian models can also impose *structure* on the data.



Computational Problems Begin

So what happens if you're calculating your Bayesian probabilities and you get stuck trying to find $p(y)$, where

$$p(y) = \int_\theta p(y|\theta)p(\theta)d\theta$$

The problem: Calculating this integral can get messy! [1]

[1] Dustin Stansbury, *The Clever Machine*

Monte Carlo Methods for Integration

Expected value of some function f(x) of a random variable X distributed through p(x):

$$E[x] = \int_{p(x)} p(x) f(x) dx \approx \int_{\theta} p(y|\theta) p(\theta) d\theta$$

Which has the same structure as our first integral.

Monte Carlo Methods for Integration

Since we can approximate expected values using

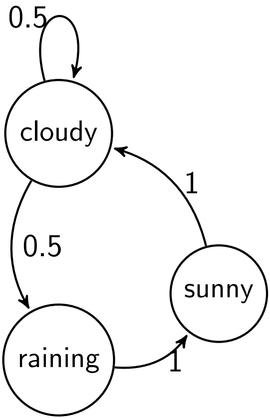
$$E_{p(x)}\left[f(x)\right] \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i)$$

we can do the same here!

$$E_{\theta}\left[p(y|\theta_i)\right] \approx \frac{1}{N}\sum_{i=1}^{N} p(y|\theta_i)$$

Markov Chain Sampling

So how do we sample from a distribution quickly and effectively? Markov chains!



So coding this sounds... fun, right?

So coding this sounds... fun, right?

That depends on the *language* you code in.

Java

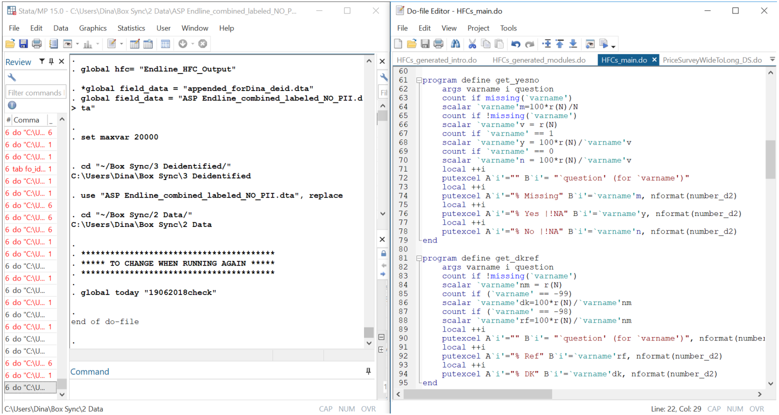
```
public class HelloWorld
{
    public static void main (String[] args)
    {
        System.out.println("Hello, world!");
    }
}
```

Java

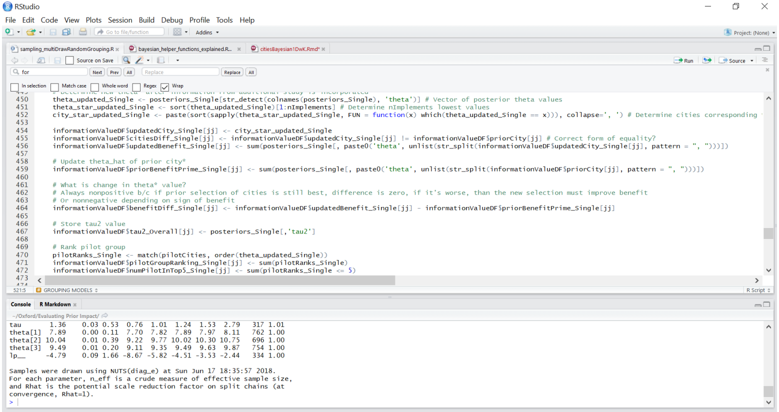
```
public class HelloWorld
{
    public static void main (String[] args)
    {
        System.out.println("Hello, world!");
    }
}
```

Python

```
print "Hello, world!"
```


Stata/MP 15.0 - C:\Users\Dina\Box Sync\2 Data\ASP Endline_combined_labeled_NO_P...
File Edit Data Graphics Statistics User Window Help
Review
Filter commands
Comma
. global hfc= "Endline_HFC_Output"
. *global field_data = "appended_forDina_deid.dta"
. global field_data = "ASP Endline_combined_labeled_NO_PII.d
> ta"
. set maxvar 20000
. cd "~/Box Sync/3 Deidentified/"
C:\Users\Dina\Box Sync\3 Deidentified
. use "ASP Endline_combined_labeled_NO_PII.dta", replace
. cd "~/Box Sync/2 Data/"
C:\Users\Dina\Box Sync\2 Data
. **
. ***** TO CHANGE WHEN RUNNING AGAIN *****
. **
. global today "19062018check"
end of do-file
Command
C:\Users\Dina\Box Sync\2 Data
CAP NUM OVR
Do-file Editor - HFCs_main.do
File Edit View Project Tools
HFCs_generated_intro.do HFCs_generated_modules.do HFCs_main.do PriceSurveyWideToLong_DS.do
program define get_yesno
args varname i question
count if missing(`varname')
scalar `varname'm=100*r(N)/N
count if !missing(`varname')
scalar `varname'v = r(N)
count if `varname' == 1
scalar `varname'y = 100*r(N)/`varname'v
count if `varname' == 0
scalar `varname'n = 100*r(N)/`varname'v
local ++i
putexcel A`i'="" B`i'= "`question' (for `varname')"
local ++i
putexcel A`i'="% Missing" B`i'=`varname'm, nformat(number_d2)
local ++i
putexcel A`i'="% Yes |!NA" B`i'=`varname'y, nformat(number_d2)
local ++i
putexcel A`i'="% No |!NA" B`i'=`varname'n, nformat(number_d2)
end
program define get_dkref
args varname i question
count if !missing(`varname')
scalar `varname'nm = r(N)
count if (`varname' == -99)
scalar `varname'dk=100*r(N)/`varname'nm
count if (`varname' == -98)
scalar `varname'rf=100*r(N)/`varname'nm
local ++i
putexcel A`i'="" B`i'= "`question' (for `varname')", nformat(numbe
local ++i
putexcel A`i'="% Ref" B`i'=`varname'rf, nformat(number_d2)
local ++i
putexcel A`i'="% DK" B`i'=`varname'dk, nformat(number_d2)
end
Line: 22, Col: 29 CAP NUM OVR

Stan


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
sampling_multiDrawRandomGrouping.R
bayesian_helper_functions_explained.R...
citiesBayesian1DwK.Rmd*
Source on Save
Run
Source
for
Next Prev All
Replace
Replace All
In selection Match case Whole word Regex Wrap
theta_updated_Single <- posteriors_Single[str_detect(colnames(posteriors_Single), 'theta')] # Vector of posterior theta values
theta_star_updated_Single <- sort(theta_updated_Single)[1:nImplements] # Determine nImplements lowest values
city_star_updated_Single <- paste(sort(sapply(theta_star_updated_Single, FUN = function(x) which(theta_updated_Single == x))), collapse=', ') # Determine cities corresponding
informationValueDF$updatedCity_Single[jj] <- city_star_updated_Single
informationValueDF$citiesDiff_Single[jj] <- informationValueDF$updatedCity_Single[jj] != informationValueDF$priorCity[jj] # Correct form of equality?
informationValueDF$updatedBenefit_Single[jj] <- sum(posteriors_Single[, paste0('theta', unlist(str_split(informationValueDF$updatedCity_Single[jj], pattern = ", ")))])
Update theta_hat of prior city*
informationValueDF$priorBenefitPrime_Single[jj] <- sum(posteriors_Single[, paste0('theta', unlist(str_split(informationValueDF$priorCity[jj], pattern = ", ")))])
What is change in theta* value?
Always nonpositive b/c if prior selection of cities is still best, difference is zero, if it's worse, than the new selection must improve benefit
Or nonnegative depending on sign of benefit
informationValueDF$benefitDiff_Single[jj] <- informationValueDF$updatedBenefit_Single[jj] - informationValueDF$priorBenefitPrime_Single[jj]
Store tau2 value
informationValueDF$tau2_Overall[jj] <- posteriors_Single[,'tau2']
# Rank pilot group
pilotRanks_Single <- match(pilotCities, order(theta_updated_Single))
informationValueDF$pilotGroupRanking_Single[jj] <- sum(pilotRanks_Single)
informationValueDF$numPilotInTop5_Single[jj] <- sum(pilotRanks_Single <= 5)
521:5
GROUPING MODELS
R Script
Console
R Markdown
~/Oxford/Evaluating Prior Impact/
tau 1.36 0.03 0.53 0.76 1.01 1.24 1.53 2.79 317 1.01
theta[1] 7.89 0.00 0.11 7.70 7.82 7.89 7.97 8.11 762 1.00
theta[2] 10.04 0.01 0.39 9.22 9.77 10.02 10.30 10.75 696 1.00
theta[3] 9.49 0.01 0.20 9.11 9.35 9.49 9.63 9.87 754 1.00
lp__ -4.79 0.09 1.66 -8.67 -5.82 -4.51 -3.53 -2.44 334 1.00
Samples were drawn using NUTS(diag_e) at Sun Jun 17 18:35:57 2018.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
>

Stan

```
data {
  int<lower=0> J; // number of individuals per study
  int<lower=0> I; // number of studies
  real Y[I, J]; // data points from study i individual j
  real<lower=0> sigmaSq[I, J]; // var of effect estimates
}
parameters {
  real mu;
  real<lower=0> tau;
  real theta[I];
}
transformed parameters {
}
model {
  for (i in 1:I){
    target += normal_lpdf(theta[i] | mu, tau^2);
    for (j in 1:J){
      target += normal_lpdf( Y[i, j] | theta[i], sigmaSq[i, j]);
    }
  }
}
```

Why choose Stan?

- Free, open source
- Faster to write
- Faster to run
- Easier to learn
- Easier to understand/reproduce other people's code

We want to implement a program in a subset of cities. Where will be the most effective to implement it?

We want to implement a program in a subset of cities. Where will be the most effective to implement it?

Initial priors

We want to implement a program in a subset of cities. Where will be the most effective to implement it?

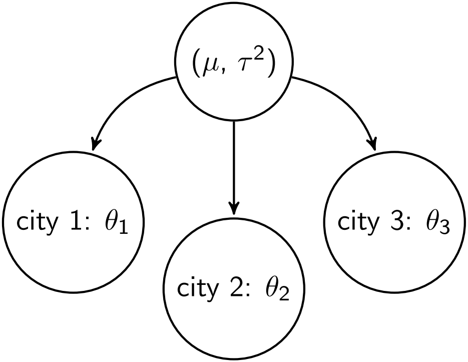
Initial priors $\rightarrow$ Pilot programs

We want to implement a program in a subset of cities. Where will be the most effective to implement it?

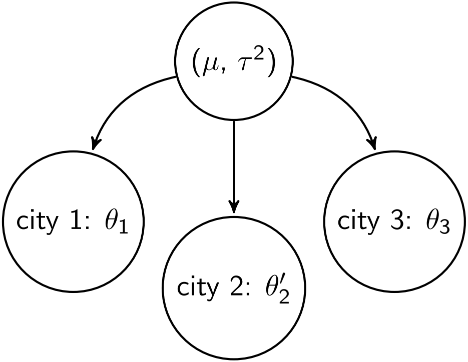
Initial priors $\rightarrow$ Pilot programs $\rightarrow$ Updated priors

We want to implement a program in a subset of cities. Where will be the most effective to implement it?

Initial priors $\rightarrow$ Pilot programs $\rightarrow$ Updated priors $\rightarrow$ Action


(μ, τ^2)
city 1: θ_1
city 2: θ_2
city 3: θ_3

City Model



The city example is simple, but it gets at some key ideas.

- Hierarchical modeling can help us improve experimental design.
- With Stan, hierarchical modeling is a realistic path.
- Statistics and other fields use these tools regularly - why not economics?

For those that are curious, Stan has a great website and growing number of tutorials at mc-stan.org

If all goes well, my cities example will be online by the time I've left GPI as well!