Bayesian Categorical Data Analysis

Andy Grogan-Kaylor

16 Jul 2020 15:42:27

Introduction

```
. clear all
```

The Importance of Thinking About Prior Information

Thinking Through Bayesian Ideas

More About Priors From SAS Corporation

"In addition to data, analysts often have at their disposal useful auxiliary information about inputs into their model—for example, knowledge that high prices typically decrease demand or that sunny weather increases outdoor mall foot traffic. If used and incorporated correctly into the analysis, the auxiliary information can significantly improve the quality of the analysis. But this information is often ignored. Bayesian analysis provides a principled means of incorporating this information into the model through the prior distribution, but it does not provide a road map for translating auxiliary information into a useful prior."

–SAS Corporation

Formal Derivation of Bayes Theorem

Following inspiration from Kruschke (2011).

Probability	A	Not A
B	P_1	P_2
Not B	P_3	P_4

Filling in the probabilities.

Probability	A	Not A
B	$P(A, B)$	$P(\text{not}A, B)$
Not B	$P(A, \text{not}B)$	$P(\text{not}A, \text{not}B)$

From the definition of conditional probability:

$P(A|B) = P(A,B)/P(B)$

$P(B|A) = P(A,B)/P(A)$

Then:

$P(A|B)P(B) = P(A,B)$

$P(B|A)P(A) = P(A,B)$

Then:

$P(A|B)P(B) = P(B|A)P(A)$

Then:

$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

Applying the Derivation to Data Analysis

Probability	Data	Not Data
Hypothesis	$P(D,H)$	$P(\text{not}D,H)$
Not Hypothesis	$P(D,\text{not}H)$	$P(\text{not}D,\text{not}H)$

From the definition of conditional probability:

$P(D|H) = P(D,H)/P(H)$

$P(H|D) = P(D,H)/P(D)$

Then:

$P(D|H)P(H) = P(D,H)$

$P(H|D)P(D) = P(D,H)$

Then:

$P(D|H)P(H) = P(H|D)P(D)$

Then:

$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$

posterior $\sim$ likelihood $\times$ prior

Accepting the Null Hypothesis

We Are Directly Estimating The Probability of the Hypothesis Given The Data

- Could be large e.g. .8.
- Could be small e.g. .1.
- Could be effectively 0. (*Essentially, we can accept a null hypothesis*)

We Are *Not* Rejecting a Null Hypothesis

We are *not* imagining a hypothetical *null hypothesis* (*that may not even be substantively meaningful*), and asking the question of whether the data we observe are extreme enough that we wish to reject this null hypothesis.

- H_0: $\bar{x} = 0$ or $\beta = 0$
- Posit H_A: $\bar{x} \neq 0$ or $\beta \neq 0$
- Observe data and calculate a test statistic (e.g. t). If test statistic > critical value, e.g. $t > 1.96$ then reject H_0.
- We can never *accept* H_0, only *reject* H_A.

Accepting Null Hypotheses

What is the effect on science and publication of having a statistical practice where we can never affirm $\bar{x} = 0$ or $\beta = 0$, but only reject $\bar{x} = 0$ or $\beta = 0$?

- Only affirm difference not similarity
- Publication bias

See https://agrogan1.github.io/Bayes/accepting-H0/accepting-H0.html

Bayesian statistics allow us to accept the null hypothesis H_0.

Bayesian Categorical Data Analysis in Stata

```
. clear all

. use "../logistic-regression/GSSsmall.dta", clear
```

Frequentist Logistic Regression

```
. logit liberal i.race i.class
Iteration 0:   log likelihood = -31538.733
Iteration 1:   log likelihood = -31370.507
Iteration 2:   log likelihood = -31369.841
Iteration 3:   log likelihood = -31369.841
Logistic regression                             Number of obs     =     53,625
                                                LR chi2(5)        =     337.78
                                                Prob > chi2       =     0.0000
Log likelihood = -31369.841                     Pseudo R2         =     0.0054

       liberal |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
               |
          race |
         black |   .4443531   .0272062    16.33   0.000       .39103    .4976762
         other |   .3190896   .0413275     7.72   0.000     .2380891    .4000901
               |
         class |
 working class |  -.1397848    .041515    -3.37   0.001    -.2211527   -.0584169
  middle class |  -.0117948   .0416509    -0.28   0.777     -.093429    .0698394
   upper class |   .1512565   .0648962     2.33   0.020     .0240624    .2784507
               |
         _cons |  -.9900441   .0397384   -24.91   0.000     -1.06793   -.9121582
```

Bayesian Logistic Regression

Takes a few minutes since using MCMC (5-10 minutes).

```
. sample 10 // Random Sample To Speed This Example: DON´T DO THIS IN PRACTICE!!!
(58,332 observations deleted)
```

How do we interpret the result for some of the **social class** categories where the credibility interval includes 0?

```
. bayes: logit liberal i.race i.class

Burn-in ...
Simulation ...
Model summary
------------------------------------------------------------------------------
Likelihood:
  liberal ~ logit(xb_liberal)
Prior:
  {liberal:i.race i.class _cons} ~ normal(0,10000)                        (1)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_liberal.
Bayesian logistic regression                    MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling        Burn-in          =      2,500
                                                MCMC sample size =     10,000
                                                Number of obs    =      5,393
                                                Acceptance rate  =      .2257
                                                Efficiency:  min =     .01383
                                                             avg =     .03232
Log marginal likelihood = -3180.1718                         max =      .0583

------------------------------------------------------------------------------
             |                                                Equal-tailed
     liberal |      Mean   Std. Dev.     MCSE     Median  [95% Cred. Interval]
-------------+----------------------------------------------------------------
        race |
       black |  .3947493   .0822819   .005895   .3933429   .2289023   .5567789
       other |  .5018242   .1295803   .011017    .503707    .255992   .7522975
             |
       class |
working class | -.4241844  .1243631   .007727   -.422026  -.6713128    -.20072
 middle class | -.2619162  .1232291   .006378  -.2664475  -.5197306  -.0254913
  upper class | -.1508993  .1983133   .010032  -.1534356  -.5326012   .2457839
             |
       _cons | -.7522187   .1149148   .004759  -.7508257  -.9723163  -.5232482
------------------------------------------------------------------------------
Note: Default priors are used for model parameters.
```

Blocking May Improve Estimation

```
. * bayes, block({liberal:i.race}): logit liberal i.race i.class // blocking may improve
> estimation
```

Bayesian Logistic Regression With Priors

Priors:

- Encode prior information: strong theory; strong clinical or practice wisdom; strong previous empirical results.
- May be helpful in quantitatively encoding the results of prior literature.

- May be especially helpful when your sample is small.

```
. bayes, normalprior(5): logit liberal i.race i.class

Burn-in ...
Simulation ...
Model summary
------------------------------------------------------------------------------
Likelihood:
  liberal ~ logit(xb_liberal)
Prior:
  {liberal:i.race i.class _cons} ~ normal(0,25)                            (1)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_liberal.
Bayesian logistic regression                     MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =      5,393
                                                 Acceptance rate  =      .2018
                                                 Efficiency:  min =     .01531
                                                              avg =      .0254
Log marginal likelihood = -3162.2981                          max =     .04229

------------------------------------------------------------------------------
             |                                                Equal-tailed
     liberal |      Mean   Std. Dev.     MCSE     Median  [95% Cred. Interval]
-------------+----------------------------------------------------------------
        race |
       black |  .4063884   .0830629   .004039   .4098444   .2476307   .5649653
       other |  .5087264   .1206866   .007204   .5028487   .2772664   .7407117
             |
       class |
working class | -.4232222   .1221491   .008601   -.424899  -.6637786  -.1720494
middle class | -.2570835   .1197257   .009677  -.2614378  -.4845897  -.0119759
 upper class |  -.135455    .200802   .012247  -.1340006  -.5505343   .2664472
             |
       _cons | -.7615401   .1153949   .008224  -.7556258   -.998298  -.5372218
------------------------------------------------------------------------------
Note: Default priors are used for model parameters.
```

MCMC vs. ML

```
. clear all

. set obs 100
number of observations (_N) was 0, now 100

. generate myestimate = rnormal() + 10 // simulated values of estimate

. summarize myestimate
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
  myestimate |        100     9.99037    .9061558   7.483831   11.92453

. local mymean = r(mean)

. kdensity myestimate ,  ///
> title("Likelihood of Estimate") ///
> xtitle("Estimate") ytitle("Likelihood") ///
> note("Vertical Line At Mean Value") ///
> caption("ML gives point estimate; Bayes gives full range of distribution") ///
> xline(`mymean´, lwidth(1) lcolor(gold)) scheme(michigan)

. graph export MCMC-ML.png, width(500) replace
(file MCMC-ML.png written in PNG format)
```

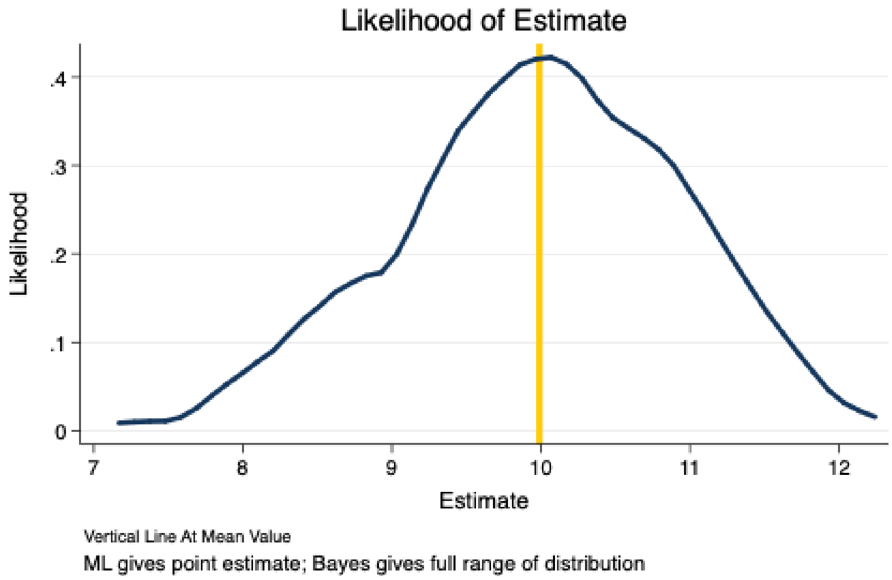

Figure 1: MCMC vs. ML

Full Distribution of Parameters

```
. clear all

. use "../logistic-regression/GSSsmall.dta", clear

. sample 10 // Random Sample for These Slides: DON´T DO THIS IN PRACTICE!!!
(58,332 observations deleted)

. bayes, normalprior(5): logit liberal i.race i.class

Burn-in ...
Simulation ...
Model summary
------------------------------------------------------------------------------
Likelihood:
  liberal ~ logit(xb_liberal)
Prior:
  {liberal:i.race i.class _cons} ~ normal(0,25)                            (1)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_liberal.
Bayesian logistic regression                     MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =      5,345
                                                 Acceptance rate  =      .2082
                                                 Efficiency:  min =     .02443
                                                              avg =     .03407
Log marginal likelihood = -3105.9749                          max =     .06349

------------------------------------------------------------------------------
             |                                                Equal-tailed
     liberal |      Mean   Std. Dev.     MCSE     Median  [95% Cred. Interval]
-------------+----------------------------------------------------------------
        race |
       black |  .5422929   .0827044   .004872   .5434182   .3823331   .7078339
       other |  .4157509   .1281866   .005087   .4171814    .163697   .6570723
             |
       class |
```

```
working class | -.2171419   .1361014   .008708  -.2179448  -.4807177   .0627315
 middle class | -.1583411   .1396594   .008239  -.1582649  -.4454747   .1111331
  upper class | -.0034842   .2229609   .012502  -.0032252  -.4430678   .4161878
              |
        _cons | -.9453812   .1325399   .008048  -.9430257  -1.212341  -.6728245
-------------------------------------------------------------------------------
Note: Default priors are used for model parameters.

. bayesgraph kdensity {liberal:2.race}, scheme(michigan)

. graph export mybayesgraph.png, width(500) replace
(file mybayesgraph.png written in PNG format)
```

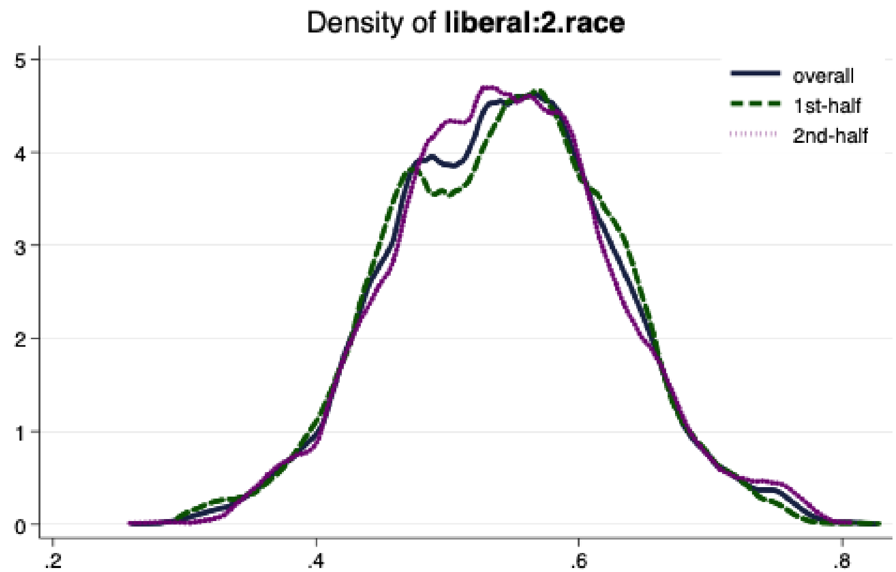

Figure 2: Density Plot of Parameter