Bayesian Categorical Data Analysis

Andy Grogan-Kaylor

14 Jul 2020 15:05:48

Introduction

```
. clear all
```

The Importance of Thinking About Prior Information

Thinking Through Bayesian Ideas

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    =      5,370
                                                 Acceptance rate  =      .2129
                                                 Efficiency:  min =     .02443
                                                              avg =     .03574
Log marginal likelihood = -3183.4711                          max =     .04685

------------------------------------------------------------------------------
                                                             Equal-tailed
     liberal |      Mean   Std. Dev.     MCSE     Median  [95% Cred. Interval]
-------------+----------------------------------------------------------------
        race |
       black |   .548588   .0831709   .003843   .5476248   .3904489   .7067804
       other |  .2686216   .1423001   .009104   .2736241  -.0167725   .5361192
             |
       class |
working class| -.1566057   .1268648   .006531  -.1554297  -.4091349   .0968553
middle class |  -.016449   .1335839   .008162   -.013235  -.2721034   .2440431
 upper class |  .2101377   .2057211   .010693   .2076646  -.2120719    .611688
             |
       _cons | -.9900013   .1243759   .006097  -.9904929  -1.238569  -.7409336
------------------------------------------------------------------------------
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    =      5,370
                                                Acceptance rate  =      .2792
                                                Efficiency:  min =      .0218
                                                             avg =     .03738
Log marginal likelihood = -3165.5355                         max =     .05414
```

liberal	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
race						
black	.5434675	.0861621	.003901	.5428555	.3787866	.7144542
other	.2799266	.1351239	.005807	.2873817	.0103018	.5362551
class						
working class	-.1525076	.1330829	.007231	-.1559297	-.4001128	.1047808
middle class	-.0050202	.1346973	.007531	-.0052376	-.25755	.2621892
upper class	.1991016	.2045538	.013855	.1896267	-.2107889	.5841896
_cons	-.9959125	.1288296	.007016	-.9929873	-1.244611	-.7511895

```
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	9.845097	1.015819	7.351274	12.5474

```
. 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)
```

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
```

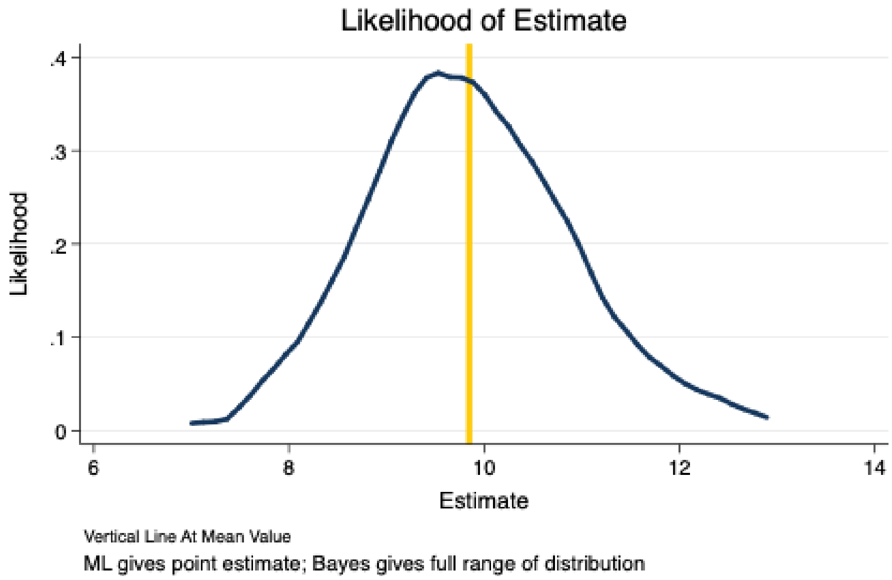

Figure 1: MCMC vs. ML

```
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    =      5,359
                                                 Acceptance rate  =      .2154
                                                 Efficiency:  min =     .01525
                                                              avg =     .04114
Log marginal likelihood = -3156.3553                          max =     .06464

------------------------------------------------------------------------------
             |                                                Equal-tailed
     liberal |      Mean   Std. Dev.     MCSE     Median  [95% Cred. Interval]
-------------+----------------------------------------------------------------
        race |
       black |  .5371258    .088161   .005346    .533988   .3686128   .7075168
       other |  .0581145   .1310966   .010615   .0584126  -.2096021   .2948507
             |
       class |
working class |  .1179909   .1478062   .007693   .1155219  -.1624791   .4134657
 middle class |  .2649346   .1482299   .005997   .2634409   -.006191   .5795305
  upper class |  .1883275   .2130126   .008378    .191786  -.2337438   .6073955
             |
       _cons | -1.233312   .1444018   .007068  -1.230555  -1.523936  -.9601146
------------------------------------------------------------------------------
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)
```

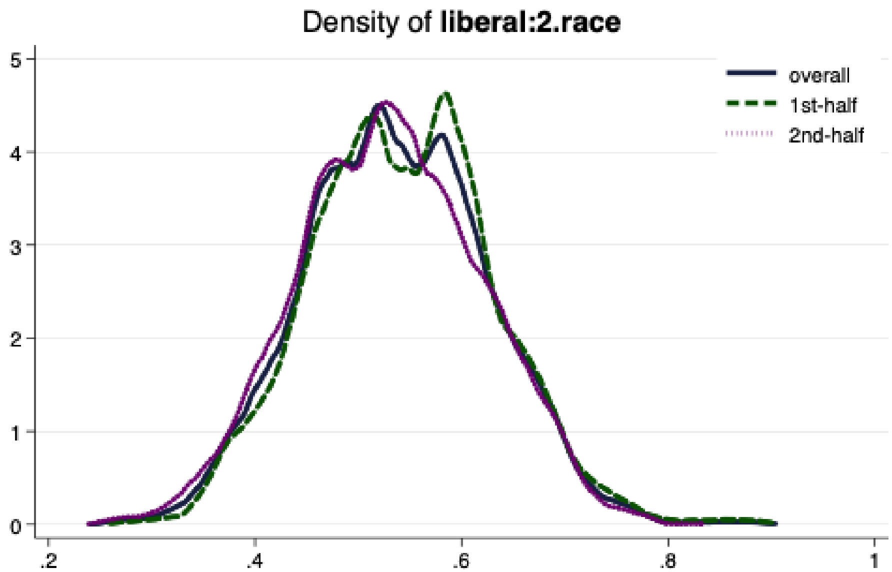

Figure 2: Density Plot of Parameter