Lecture 8: Estimation Methods Bayesian Estimation & Empirical Bayes

Big Data and Machine Learning for Applied Economics Econ 4676

Ignacio Sarmiento-Barbieri

Universidad de los Andes

September 3, 2020

Announcement & Recap

- **Next Thursday September 10, Jacob will be teaching a complementary class**
- Maximum Likelihood Estimation
- Conditional Maximum Likelihood Estimation
- Bayesian Estimation

Agenda

1. Motivation & Extended Recap.
2. Empirical Bayes
   - Robbins' Formula
   - Sabermetrics
3. Further Readings

Motivation & Extended Recap.

Bayesian Estimation

- The Bayesian approach to stats is fundamentally different from the classical approach we have been taking
- In the classical approach, the parameter θ is thought to be an unknown, but fixed quantity, e.g., $X_i \sim f(\theta)$
- In the Bayesian approach θ is considered to be a quantity whose variation can be described by a probability distribution (*prior distribution*)
- Then a sample is taken from a population indexed by θ and the prior is updated with this sample
- The resulting updated prior is the *posterior distribution*

Recap: Bayes Theorem

For this updating we use *Bayes Theorem*

$$\pi(\theta|X) = \frac{f(X|\theta)p(\theta)}{m(X)} \tag{1}$$

with $m(X)$ is the marginal distribution of X, i.e.

$$m(X) = \int f(X|\theta)p(\theta)d\theta \tag{2}$$

Recap: Bayesian Linear Regression

Consider

$$y_i = \beta x_i + u_i \;\; u_i \sim_{iid} N(0, \sigma^2 I) \tag{3}$$

with prior

$$p(\beta) = \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2\tau^2}(\beta - \beta_0)^2} \tag{4}$$

The Posterior distribution then $\beta \sim N(m, V)$

Recap: Bayesian Linear Regression

$$m = \left(\frac{\frac{\sum x_i^2}{\sigma^2}}{\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2}} \right) \frac{\sum x_i y_i}{\sum x_i^2} + \left(\frac{\frac{1}{\tau^2}}{\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2}} \right) \beta_0 \tag{5}$$

$$m = \omega \hat{\beta}_{MLE} + (1-\omega)\beta_0 \tag{6}$$

Remarks

- If prior belief is strong $\tau \downarrow 0 \rightarrow \omega \downarrow 0 \implies m = \beta_0$
- If prior belief is weak $\tau \uparrow \infty \rightarrow \omega \uparrow 1 \implies m = \beta_{MLE}$

Bayesian Estimation

Conjugate Priors:

Definition Let $\mathcal{F}$ denote the class of densities $f(x|\theta)$. A class $\mathcal{C}$ of prior distributions is a conjugate family for $\mathcal{F}$ if the posterior distribution is in the class $\mathcal{C}$ for all $f \in \mathcal{F}$, all priors in $\mathcal{C}$, and all $x \in X$

For example:

- the normal distribution is a conjugate for the normal family
- the beta distribution for the binomial family
- the gamma distribution for the poisson family

Good and bad news:

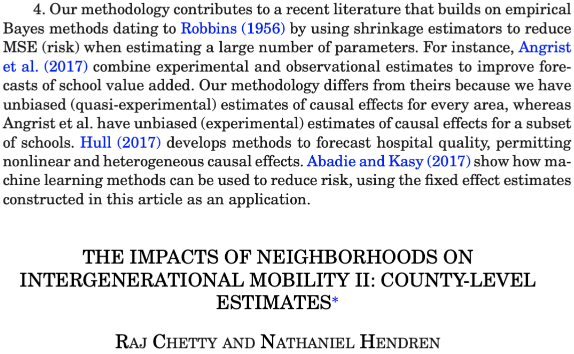
- Nice because gives us a nice close form for the posterior. However, whether a conjugate family is a reasonable choice is left to you!
- Downside, if we choose another families, then these results are no longer available. Then we have to use sampling-based methods (MCMC, Gibbs Sampler, etc)

Empirical Bayes

- The constraints of slow mechanical computation molded classical statistics into a mathematically ingenious theory of sharply delimited scope.
- After WW2, computers allowed a more expansive and useful statistical methodology.
- However, Some revolutions start slowly. The journals of the 1950s continued to emphasize classical themes
- Change came gradually, but by the 1990s a new statistical technology, computer enabled, was firmly in place.
- Empirical Bayes methodology, has been a particularly slow developer despite an early start in the 1940s.
- The roadblock here was not so much the computational demands of the theory as a lack of appropriate data sets.

Empirical Bayes

- In Economics this revolution is starting to catch up, fueled by Big Data



4. Our methodology contributes to a recent literature that builds on empirical Bayes methods dating to Robbins (1956) by using shrinkage estimators to reduce MSE (risk) when estimating a large number of parameters. For instance, Angrist et al. (2017) combine experimental and observational estimates to improve forecasts of school value added. Our methodology differs from theirs because we have unbiased (quasi-experimental) estimates of causal effects for every area, whereas Angrist et al. have unbiased (experimental) estimates of causal effects for a subset of schools. Hull (2017) develops methods to forecast hospital quality, permitting nonlinear and heterogeneous causal effects. Abadie and Kasy (2017) show how machine learning methods can be used to reduce risk, using the fixed effect estimates constructed in this article as an application.

THE IMPACTS OF NEIGHBORHOODS ON INTERGENERATIONAL MOBILITY II: COUNTY-LEVEL ESTIMATES*

RAJ CHETTY AND NATHANIEL HENDREN

Chetty, R., & Hendren, N. QJE (2018).

Empirical Bayes

Consider the following standard Bayesian model:

$$X|\theta \sim N(\theta, 1) \tag{7}$$

$$\theta|\tau^2 \sim N(0, \tau^2) \tag{8}$$

- Standard approach the experimenter would specify a prior value for τ^2
- Note that the marginal distribution of X is $N(0, \tau^2 + 1)$
- Empirical Bayes uses this “shortcut”. Uses the data to obtain the “unknown parameters”

Robbins' Formula

Example: an insurance company is concerned about the claims each policy holder will make in the next year.

Table 1: Claims data for a European automobile insurance company

Claims	0	1	2	3	4	5	6	7
Counts	7840	1317	239	42	14	4	4	1

Robbins' Formula

- It seems that we can use Bayes formula to get next years expected number of accidents
- We suppose that x_k, the number of claims to be made in a single year by policy holder k,
- This follows a Poisson distribution with parameter θ_k
- Recall that the mean and variance are θ_k

$$Pr(x_k = x) = p_{\theta_k}(x) = \frac{e^{-\theta_k}\theta_k^x}{x!} \;\; for \;\; x = 0, 1, 2, 3, ... \tag{9}$$

Robbins' Formula

Suppose now, that we know the prior density $g(\theta)$. Then using Bayes rule we would have

$$
\begin{align}
E(\theta|x) &= \int_0^\infty \theta\pi(\theta)d\theta \tag{10} \\
&= \frac{\int_0^\infty \theta p_{\theta_k}(x)g(\theta)d\theta}{\int_0^\infty p_{\theta_k}(x)g(\theta)d\theta} \tag{11}
\end{align}
$$

is the expected value of θ of a customer observed to make x claims in a sinble year. This would answer the insurance company's questions of what numbers of claims X to expect the next year from the same customer

Robbins' Formula

What happens if we don't know the prior? Note the following:

$$E(\theta|x) = \frac{\int_0^\infty \theta[e^{-\theta_k}\theta_k^x/x!]g(\theta)d\theta}{\int_0^\infty [e^{-\theta_k}\theta_k^x/x!]g(\theta)d\theta} \tag{12}$$

$$E(\theta|x) = \frac{(x+1)\int_0^\infty [e^{-\theta_k}\theta_k^{x+1}/(x+1)!]g(\theta)d\theta}{\int_0^\infty [e^{-\theta_k}\theta_k^x/x!]g(\theta)d\theta} \tag{13}$$

$$E(\theta|x) = \frac{(x+1)f(x+1)}{f(x)} \tag{14}$$

Robbins' Formula

The obvious estimate of the marginal density $f(x)/$is the proportion of total counts in category x,

$$\hat{f}(x) = \frac{y_x}{N} \tag{15}$$

where $N = \sum_x y_x$

Table 2: Claims data for a European automobile insurance company

Claims	0	1	2	3	4	5	6	7
Counts	7840	1317	239	42	14	4	4	1
Mean	.168	.363	.527	1.33	1.43	46	1.75	.

Sabermetrics: Batting Averages

Sabermetrics: Batting Averages

- One of the most commonly used statistics in baseball is the batting average

$$\text{Batting Average} = \frac{\text{number of hits (H)}}{\text{number of at-bats (AB)}} \tag{16}$$

Today we are going to explore two additional problems and use EB:

1. You want to recruit two players: One has achieved 4 hits in 10 chances, the other 300 hits in 1000 chances.
2. Based on first few performances, can we predict what is going to be the season-long batting averages

Sabermetrics: Recruiting

- So you want to recruit two players: One has achieved 4 hits in 10 chances, the other 300 hits in 1000 chances.
- We know by history that most batting averages are between .210 and .360
- How can we incorporate this info using Bayesian statistics?

We can model

$$\text{Batting Average} \sim Binomial(n, \theta) \tag{17}$$

- where n is the times at bat and θ is the proportion of successes

Sabermetrics: Recruiting

And the prior? We can use a conjugate prior for simplicity.

$$p(\theta) \sim Beta(\alpha_0, \beta_0) \tag{18}$$

The posterior is:

$$\pi(\theta) \sim Beta(\alpha_0 + hits, \beta_0 + N - hits) \tag{19}$$

Sabermetrics: Recruiting

Here I’m using a ”clean” version of `Batting` data from the `Lahman` package

```r
require("dplyr")
require("tidyr")
require("ggplot2")
```

```r
career<-readRDS("baseball.rds")
head(career)
```

```
## # A tibble: 6 x 4
##   name                H    AB average
##   <chr>           <int> <int>   <dbl>
## 1 Hank Aaron       3771 12364  0.305
## 2 Tommie Aaron      216   944  0.229
## 3 Andy Abad           2    21  0.0952
## 4 John Abadie        11    49  0.224
## 5 Ed Abbaticchio    772  3044  0.254
## 6 Fred Abbott       107   513  0.209
```

Sabermetrics: Recruiting

- Who are the best?

```
tail(career %>% arrange(average))
```

```
## # A tibble: 6 x 4
##   name             H    AB average
##   <chr>        <int> <int>   <dbl>
## 1 Roe Skidmore     1     1       1
## 2 Charlie Snow     1     1       1
## 3 Matt Tupman      1     1       1
## 4 Allie Watt       1     1       1
## 5 Al Wright        1     1       1
## 6 George Yantz     1     1       1
```

- And the worst?

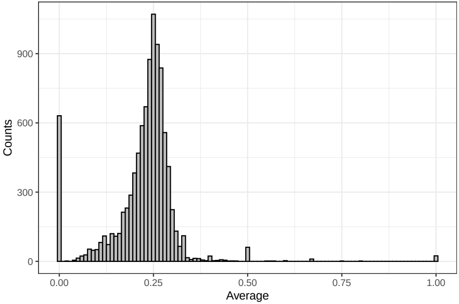
```
head(career %>% arrange(average))
```

```
## # A tibble: 6 x 4
##   name                 H    AB average
##   <chr>            <int> <int>   <dbl>
## 1 Frank Abercrombie     0     4       0
## 2 Horace Allen          0     7       0
## 3 Pete Allen            0     4       0
## 4 Walter Alston         0     1       0
## 5 Bill Andrus           0     9       0
## 6 Wyman Andrus          0     4       0
```

Sabermetrics: Recruiting

- Empirical Bayes in action
- Estimate a prior from all your data

$$X \sim Beta(\alpha_0, \beta_0) \tag{20}$$



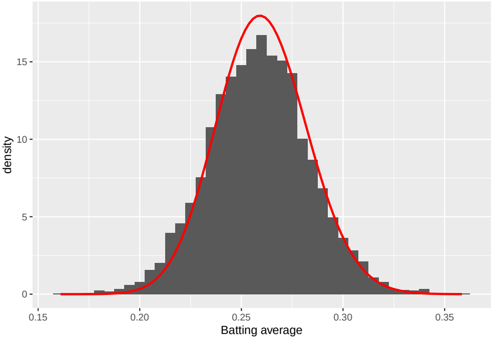
Sabermetrics: Recruiting

```
# Here, we have to filter for the players we actually
# have a decent estimate of the average
career_filtered <- career %>%
    filter(AB >= 500)

require("stats4")
require("VGAM")
```

```
# log-likelihood function
ll <- function(alpha, beta) {
x <- career_filtered$H
total <- career_filtered$AB
-sum(VGAM::dbetabinom.ab(x, total, alpha, beta, log = TRUE))
}

# maximum likelihood estimation
m <- mle(ll, start = list(alpha = 1, beta = 10),
method = "L-BFGS-B", lower = c(0.0001, .1))
ab <- coef(m)
```

Sabermetrics: Recruiting

```
alpha0 <- ab[1]
101.7319
beta0 <- ab[2]
289.046
```



Sabermetrics: Recruiting

Now we can update the estimated average based on the posterior mean

$$E(\theta|X) = \frac{\alpha + hits}{\alpha + \beta + N} \tag{21}$$

In R

```
career_eb <- career %>%
    mutate(eb_estimate = (H + alpha0) / (AB + alpha0 + beta0))
```

Sabermetrics: Recruiting

- Now we can ask again: who are the best batters by this improved estimate?

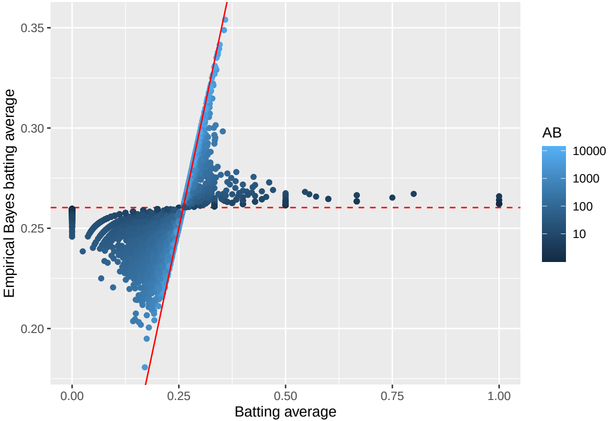
```
## # A tibble: 5 x 5
##   name                     H    AB average eb_estimate
##   <chr>                <int> <int>   <dbl>       <dbl>
## 1 Rogers Hornsby        2930  8173   0.358       0.354
## 2 Shoeless Joe Jackson  1772  4981   0.356       0.349
## 3 Ed Delahanty          2597  7510   0.346       0.342
## 4 Billy Hamilton        2164  6283   0.344       0.339
## 5 Willie Keeler         2932  8591   0.341       0.338
```

- Who are the *worst* batters?

```
## # A tibble: 5 x 5
##   name                H    AB average eb_estimate
##   <chr>           <int> <int>   <dbl>       <dbl>
## 1 Bill Bergen       516  3028   0.170       0.181
## 2 Ray Oyler         221  1265   0.175       0.195
## 3 Henry Easterday   203  1129   0.180       0.201
## 4 John Vukovich      90   559   0.161       0.202
## 5 George Baker       74   474   0.156       0.203
```

Sabermetrics: Recruiting

We can see how EB changed all of the batting average estimates:



Sabermetrics: Predicting Batting Averages

- Now supposed you want to know the end of season final batting average of players, after observing them their 45 first times at bat.

Player	Observed	Final
1	0.395	0.346
2	0.355	0.279
3	0.313	0.276
4	0.291	0.266
5	0.247	0.271
6	0.224	0.266
7	0.175	0.318

Sabermetrics: Predicting Batting Averages

- Recall that we can think each time at bat can be thought as a binomial trial, with θ the probability of success equal to the player's true batting average.
- With 45 trials, we can "reasonably" use a Normal Approximation.

$$X_i \sim N(\theta_i, \sigma^2) \tag{22}$$

where

- θ_i is the true batting average for player i
- σ^2 is the known variance that equals $(0.0659)^2$

We are going to use also a normal prior

$$\theta_i \sim N(\mu, \tau^2) \tag{23}$$

Sabermetrics: Predicting Batting Averages

With this model the posterior mean for θ_i is $E(\theta_i|X_i)$

$$E(\theta_i|X_i) = \frac{\sigma^2}{\sigma^2+\tau^2}\mu + \frac{\tau^2}{\sigma^2+\tau^2}X_i \tag{24}$$

Note that the marginal of X_i

$$m(X_i) \sim N(\mu, \sigma^2+\tau^2) \;\; i = 1, \dots, n \tag{25}$$

with these we can construct estimates of $E(\theta_i|X_i)$, note that

$$E(\bar{X}) = \mu \tag{26}$$

$$E\left[\frac{(n-3)\sigma^2}{\sum(X_i-\bar{X})^2}\right] = \frac{\sigma^2}{\sigma^2+\tau^2} \tag{27}$$

Sabermetrics: Predicting Batting Averages

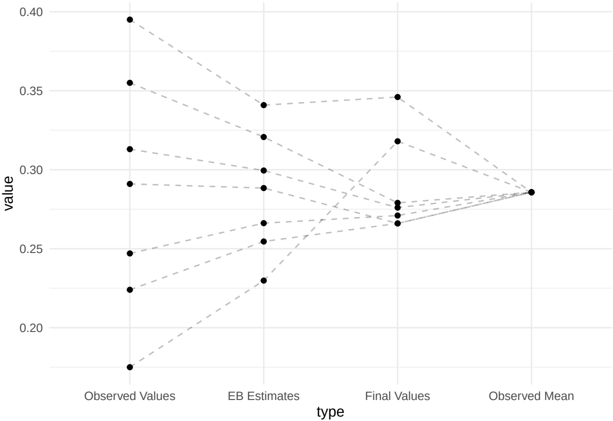
The empirical Bayes estimator of θ_i is then

$$\delta(X_i) = \left[\frac{(n-3)\sigma^2}{\sum((X_i-\bar{X})^2}\right]\bar{X} + \left[1 - \frac{(n-3)\sigma^2}{\sum((X_i-\bar{X})^2}\right]X_i \tag{28}$$

Player	Observed	Final	Empirical Bayes
1	0.395	0.346	0.341
2	0.355	0.279	0.321
3	0.313	0.276	0.299
4	0.291	0.266	0.288
5	0.247	0.271	0.266
6	0.224	0.266	0.255
7	0.175	0.318	0.230

- RMSE Observed 6.861903
- RMSE EB 3.918203

Sabermetrics: Predicting Batting Averages



Review & Next Steps

- Recap Bayesian
- Empirical Bayes Examples
- **Next Class:** Spatial Econometrics
- Questions? Questions about software?

Further Readings

- Casella, G., & Berger, R. L. (2002). Statistical inference (Vol. 2, pp. 337-472). Pacific Grove, CA: Duxbury. Chapter 7
- Casella, G. (1985). An introduction to empirical Bayes data analysis. The American Statistician, 39(2), 83-87.
- Chetty, R., & Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility II: County-level estimates. The Quarterly Journal of Economics, 133(3), 1163-1228.
- Efron, B., & Hastie, T. (2016). Computer age statistical inference (Vol. 5). Cambridge University Press. Chapter 6
- Robinson, D. (2017). Introduction to Empirical Bayes: Examples from Baseball Statistics. 2017.
- Gu, J., & Koenker, R. (2017). Empirical Bayesball remixed: Empirical Bayes methods for longitudinal data. Journal of Applied Econometrics, 32(3), 575-599.