

PLSC 504 – Fall 2023

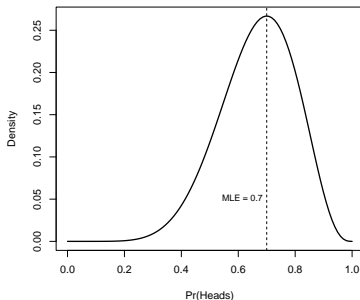
Bayesian Statistics

(A micro-intro)

November 15, 2023

Some characteristics:

- Probability = *Long-run relative frequency*
- $\Pr(X)$ is a *fixed* but *unknown* quantity
- Example: *Likelihood*
 - Suppose we flipped a coin 10 times, and got heads 7 of them.
 - Q: What is $\theta \equiv \Pr(\text{Heads})$?
 - A: The MLE is $\hat{\theta} = \frac{N_{\text{Heads}}}{N} = 7/10 = 0.7$



Components:

- Quantity of interest = θ
- Data = Y
- Sampling density = $\Pr(Y|\theta)$
- We want to know $\Pr(\theta|Y)$
- Likelihood $L(\theta|Y) \propto \Pr(Y|\theta)$

Begin by noting that:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \quad (1)$$

and

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}. \quad (2)$$

That means that:

$$\Pr(A \cap B) = \Pr(B|A) \Pr(A). \quad (3)$$

Substituting (3) into (1), we get

$$\Pr(A|B) = \frac{\Pr(B|A) \Pr(A)}{\Pr(B)}. \quad (4)$$

Bayes' Rule Applied

For our data example:

$$\begin{aligned}\Pr(\theta|Y) &= \frac{\Pr(\theta \cap Y)}{\Pr(Y)} \\ &= \frac{\Pr(Y|\theta)\Pr(\theta)}{\Pr(Y)}.\end{aligned}$$

- $\Pr(\theta|Y)$ is the *posterior density* of θ
- $\Pr(Y|\theta)$ is the *sampling density*
- $\Pr(\theta)$ is the *prior density* of θ
- $\Pr(Y)$ is the *marginal probability* of Y

Since Y is fixed in a single sample, we can write:

$$\Pr(\theta|Y) \propto \Pr(Y|\theta) \Pr(\theta).$$

Bayes and Subjective Probability

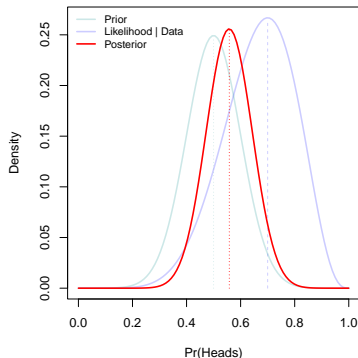
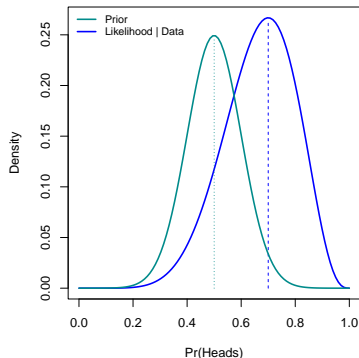
Bayesian probability:

- ...is **subjective**: Probability is a *belief about the world*...
 - $\Pr(\theta)$ is our prior / “pre-data” estimate of the distribution of θ
 - $\Pr(\theta|Y)$ is our posterior / “post-data” estimate
- ...**updates** our prior beliefs about θ based on the data, in a manner consistent with Bayes’ Theorem
- ...presents a probability **density** for $\hat{\theta}$, not just a point estimate

Bayesian Example

Suppose:

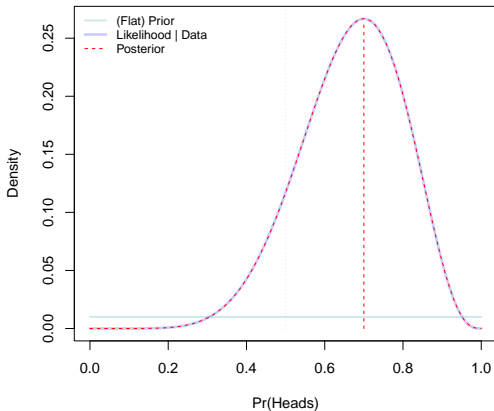
- I believe (with some uncertainty) that a particular coin is “fair” (i.e., that $\theta = 0.5$) *and*...
- ...we then flipped the coin 10 times, and got heads 7 of them.



Bayesian 🤝 Frequentist

Often Bayesian and frequentist results converge when priors are *uninformative* (these are sometimes called “flat priors”).

Example:



The process:

- Specify a probability model for the data + parameters.
- Posit one's prior beliefs.
- Calculate the posterior distribution using Bayes' Theorem.
- Summarize the posterior density.
- Conduct post-estimation model checking.

“MCMC” = Markov Chain Monte Carlo...

Goal: To characterize a (potentially complex, unknown) probability density in the parameter space.

Note that:

- The posterior density is a combination of the prior(s) and the data / likelihood...
- ...but it is also unknown until we incorporate information from the data.
- High-dimensional multivariate distributions are often mathematically + computationally intractable...

Intuition:



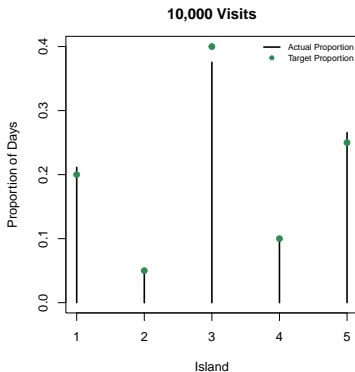
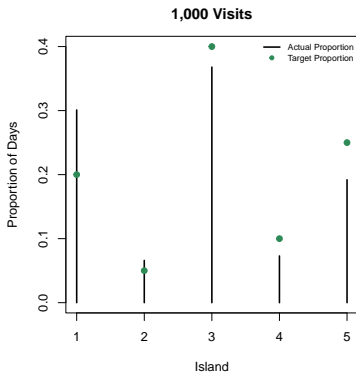
- A legislator campaigns along a chain of five east-west islands...
- Goal: Spend time in each in proportion to their (unknown) population, *but*
- The legislator learns the population of each island when she visits it.
- Algorithm: Each day...
 - ... flip a fair coin: *heads* = east, *tails* = west
 - If the island to the {*east*, *west*} has a population higher than the current island, go there
 - If the island to the {*east*, *west*} has a population lower than the current island:
 - Calculate $a = \frac{\text{Population of prospective island}}{\text{Population of current island}}$
 - Move with probability a ; stay with probability $1 - a$

MCMC: Intuition (continued)

Suppose the populations of the islands are:



After 1,000 or 10,000 days campaigning:



MCMC samples from the (multivariate) posterior distribution of $\hat{\theta}$...

- It's a *Markov chain*, because it's “memoryless,” but
- ...the stable / equilibrium distribution of the chain is the target distribution
- This means that the chain “focuses” (samples more frequently from) places in the parameter space where the target distribution has more density
- → the values of the chain at each iteration are (by construction) autocorrelated...
 - “Burn-in”: Initializing the chain in the parameter space
 - “Thinning”: Taking every p th observation from the chain for estimation/inference
- For details, check out [this](#), or [this](#), or [this](#), or [this interactive visualization](#), or [this long thing](#)...

Inference: Credible Intervals

Consider that:

Concept	Frequentist...	Bayesian...
Parameter (θ)	<i>Fixed</i>	Varies with prior
C.I. $[\theta_L, \theta_H]$	Varies with sample	<i>Fixed</i>

This means that...

- ...a Bayesian *credible interval* is an interval within which an unobserved parameter value falls with a particular probability.
- That is, a $k \times 100\%$ credible interval is the interval in which the (unobserved) parameter value falls with probability k .
- For more, read [this](#) or [this](#), or [this](#).

Example: Black State Representatives

Data from the [Correlates of State Policy Project](#) at MSU...

- Annual state-level data, 2013-2017 ($N = 50$, $T = 5$)
- Outcome: Count of Black representatives in the state legislature's lower house
- Predictor variables:
 - Percentage of the state population that is Black
 - Total state expenditures (logged)
 - Median state household income (logged)
 - State population (logged)
- Models: Linear regression + Poisson / negative binomial

Linear Regression via OLS

```
> OLSfit<-lm(BlackHouseMembers~PercentBlackPop+log(TotalExpenditures)+  
              log(MedianHHIncome)+log(Population),data=DF)  
> summary(OLSfit)
```

Call:

```
lm(formula = BlackHouseMembers ~ PercentBlackPop + log(TotalExpenditures) +  
    log(MedianHHIncome) + log(Population), data = DF)
```

Residuals:

Min	1Q	Median	3Q	Max
-15.215	-2.576	0.397	2.141	17.234

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-46.7852	18.5221	-2.53	0.01218 *
PercentBlackPop	0.9929	0.0335	29.59	< 2e-16 ***
log(TotalExpenditures)	-3.0865	1.3635	-2.26	0.02448 *
log(MedianHHIncome)	2.4575	1.7896	1.37	0.17096
log(Population)	4.7667	1.2504	3.81	0.00018 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.49 on 240 degrees of freedom

Multiple R-squared: 0.846, Adjusted R-squared: 0.843

F-statistic: 329 on 4 and 240 DF, p-value: <2e-16

Linear Regression via MCMC (using bayesm)

```
> DF$logTE<-log(DF$TotalExpenditures)
> DF$logHHInc<-log(DF$MedianHHIncome)
> DF$logPop<-log(DF$Population)
> DF$One<-1 # constant...

> # Model:
>
> Data<-list(y=DF$BlackHouseMembers,X=as.matrix(DF[,c(11,4,8,9,10)]))
> MCMC<-list(R=1e6,keep=10,nprint=0)

> BayesFit<-runireg(Data=Data,Mcmc=MCMC)
```

Starting IID Sampler for Univariate Regression Model
with 245 observations

Prior Parms:

betabar

[1] 0 0 0 0 0

A

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	0.01	0.00	0.00	0.00	0.00
[2,]	0.00	0.01	0.00	0.00	0.00
[3,]	0.00	0.00	0.01	0.00	0.00
[4,]	0.00	0.00	0.00	0.01	0.00
[5,]	0.00	0.00	0.00	0.00	0.01

nu = 3 ssq= 128.4295

MCMC parms:

R= 1e+06 keep= 10 nprint= 0

Bayesian Linear Regression Summary

```
> summary(BayesFit$betadraw)
```

Summary of Posterior Marginal Distributions

Moments

	mean	std dev	num se	rel eff	sam size
1	-39.89	17.624	0.06097	1.08	45000
2	0.99	0.034	0.00011	0.99	90000
3	-2.99	1.399	0.00486	1.09	45000
4	1.83	1.717	0.00594	1.08	45000
5	4.66	1.281	0.00446	1.09	45000

Quantiles

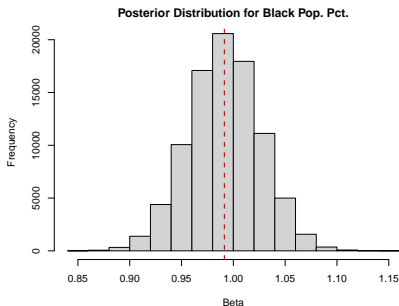
	2.5%	5%	50%	95%	97.5%
1	-74.39	-68.85	-39.88	-11.01	-5.44
2	0.92	0.93	0.99	1.05	1.06
3	-5.74	-5.29	-2.99	-0.68	-0.24
4	-1.54	-1.00	1.82	4.64	5.19
5	2.14	2.55	4.66	6.77	7.19

based on 90000 valid draws (burn-in=10000)

Plotting the Posterior (“By Hand”)

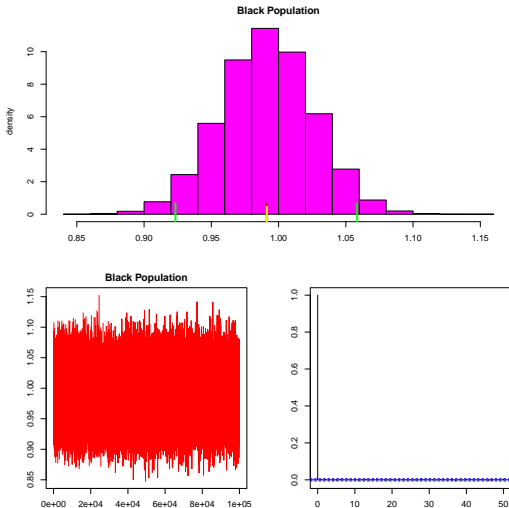
```
B<-10000 # Discard 10K burn-in draws...
ND<-1e5  # Total kept draws after thinning...

hist(BayesFit$betadraw[B:ND,1],xlab="Beta",
     main="Posterior Distribution for Black Pop. Pct.")
abline(v=median(BayesFit$betadraw[B:ND,1]),
       lty=2,lwd=2,col="red")
```

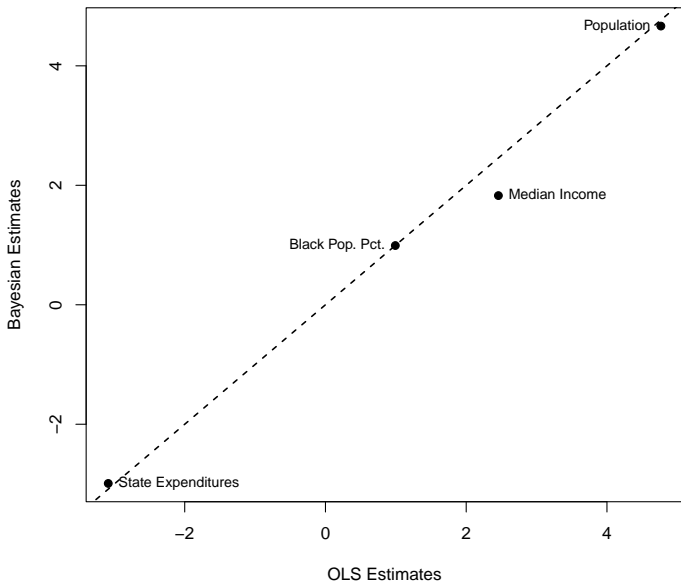


Plotting the Posterior (using `plot.bayesm`)

```
plot.bayesm.mat(BayesFit$betadraw[,1],names="Black Population")
```



Comparing Bayes & OLS $\hat{\beta}$ s



Negative Binomial (via MLE)

```
> NBfit<-glm.nb(BlackHouseMembers~PercentBlackPop+log(TotalExpenditures)+  
+               log(MedianHHIncome)+log(Population),data=DF,init.theta=4,maxit=1e4)
```

```
> summary(NBfit)
```

Call:

```
glm.nb(formula = BlackHouseMembers ~ PercentBlackPop + log(TotalExpenditures) +  
        log(MedianHHIncome) + log(Population), data = DF, maxit = 10000,  
        init.theta = 75.45975019, link = log)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-10.22986	1.42113	-7.20	6.1e-13 ***
PercentBlackPop	0.07961	0.00223	35.63	< 2e-16 ***
log(TotalExpenditures)	-0.24002	0.12054	-1.99	0.046 *
log(MedianHHIncome)	0.31709	0.13851	2.29	0.022 *
log(Population)	0.77393	0.11124	6.96	3.5e-12 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(75.46) family taken to be 1)

Null deviance: 2476.51 on 244 degrees of freedom
Residual deviance: 334.61 on 240 degrees of freedom
AIC: 1192

Number of Fisher Scoring iterations: 1

Theta: 75.5
Std. Err.: 63.7

2 x log-likelihood: -1179.7

→ Suggests Poisson...

```
> PoisFit<-glm(BlackHouseMembers~PercentBlackPop+log(TotalExpenditures)+  
+              log(MedianHHIncome)+log(Population),data=DF,family=poisson)
```

```
> summary(PoisFit)
```

Call:

```
glm(formula = BlackHouseMembers ~ PercentBlackPop + log(TotalExpenditures) +  
    log(MedianHHIncome) + log(Population), family = poisson,  
    data = DF)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-10.00217	1.25320	-7.98	1.4e-15	***
PercentBlackPop	0.07857	0.00199	39.51	< 2e-16	***
log(TotalExpenditures)	-0.21925	0.10728	-2.04	0.041	*
log(MedianHHIncome)	0.29760	0.12153	2.45	0.014	*
log(Population)	0.75093	0.09811	7.65	1.9e-14	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 2823.42 on 244 degrees of freedom
Residual deviance: 365.67 on 240 degrees of freedom
AIC: 1191

Number of Fisher Scoring iterations: 5

Bayesian Poisson (using bpr)

```
> BayesPois<-sample_bpr(BlackHouseMembers~PercentBlackPop+log(TotalExpenditures)+
+ log(MedianHHIncome)+log(Population),data=DF,
+ iter=1e5,burnin=5e2,thin=5,perc_burnin=0)
```

Running MH sampler with a gaussian prior distribution.
Chains initialized at the maximum likelihood estimates.

Sampling 1e+05 iterations

Sampling completed in 6.65438 secs

```
> summary(BayesPois)
```

Call:

```
sample_poisreg( formula = BlackHouseMembers ~ PercentBlackPop + log(TotalExpenditures) +
  log(MedianHHIncome) + log(Population),prior= gaussian,algorithm=Metropolis-
Hastings )
```

Coefficients:

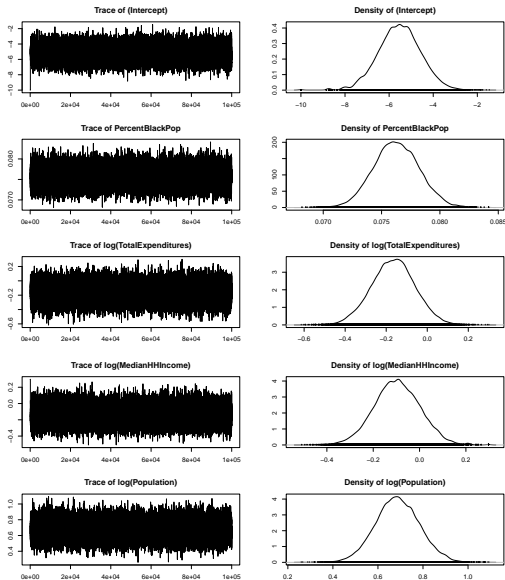
	Mean	Std. Error	Median	Lower CI	Upper CI
(Intercept)	-5.528200	0.9431559	-5.523252	-7.4372	-
3.744*					
PercentBlackPop	0.076275	0.0019659	0.076260	0.0725	0.0801*
log(TotalExpenditures)	-0.166010	0.1060195	-0.164954	-0.3731	0.0440
log(MedianHHIncome)	-0.096640	0.0984735	-0.097365	-0.2842	0.0989
log(Population)	0.687099	0.0969245	0.686325	0.4979	0.8784*

'*' if 95% credible interval does not include zero.

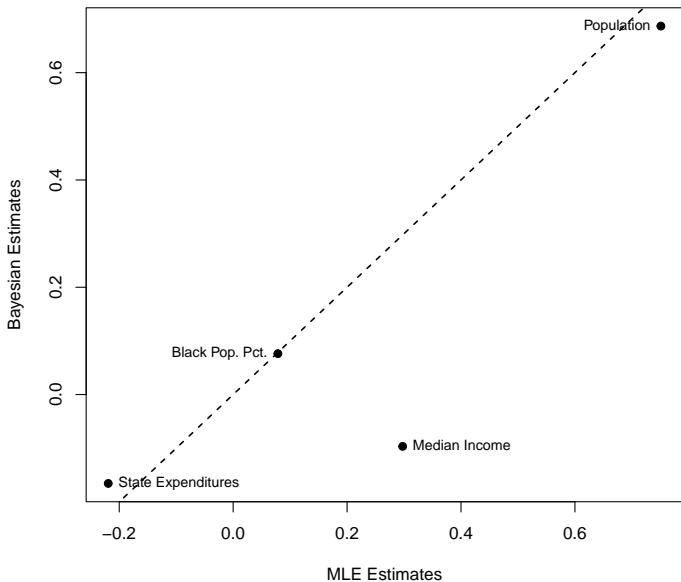
Algorithm:

Posterior estimates computed on 19900 iterations after discarding the first 500 iterations
as burn-in, with thinning frequency = 5.
Mean effective sample size is equal to 9728.
Acceptance rate is 0.3682.

Posteriors for $\beta...$



Comparing Bayes & MLE $\hat{\beta}$ s



The Good:

- Directly quantifies uncertainty
- Provides direct quantities of interest to researchers.
- Logically consistent and intuitive
- Allow the incorporation of prior information
- Allow the fitting of (very) complex models
- Flexibility

The (Potentially) Bad:

- Inherent subjectivity of choosing priors
- Computational complexity
- Difficulty in knowing when estimates have converged
- Lack of software¹

¹This is less and less a concern.

Bayesian Statistics in R

Note: The [Bayesian CRAN Task View](#) is usually a good place to start...

Packages:

- [arm](#) – from Gelman et al.'s book *Data Analysis Using Regression and Multilevel/Hierarchical Models*
- [bayesm](#) – lots of regression-like models (built for marketing, but quite general)
- [MCMCpack](#) – general + specific Bayesian models, fit via MCMC
- Other packages for causal inference, hierarchical / multilevel models, IRT/measurement, machine learning, time series, network models, spatial models, more...
- Learning tools: LearnBayes, BayesDA, BaM, rethinking, others...