

Bayesian perspectives on statistical modeling

BE BOUNDLESS



What's Next

In the next 5 to 10 minutes, I will...

- > **What's Bayesian?**
- > **Essential difference between Frequentist and Bayesianist**
- > **Advantage of Bayesianist over Frequentist**
- > **Model example of covid death rate in Washington State**

What's Bayesian Statistics

Let's say we have a bag contains 4 marbles in blue and white. we take 3 marbles out of bags. They are blue, white, and blue respectively. What is the proportion of blue balls?



Suppose we have y as observed data vector (Blue, white, blue), Θ is the parameter of model, then we have

- > A actual proportion of blue marbles, Θ , is usually called a **parameter** value.
- > The previous belief about Θ before we see the data is usually called the **prior probability**. $P(\Theta)$
- > The relative possibility that a value Θ can produce the data is usually called a **likelihood**. $P(y | \Theta)$
- > The new, updated plausibility of any specific Θ is usually called the **posterior** $P(\Theta | y) \propto P(y | \Theta) P(\Theta)$



	θ	$P(\theta)$	$P(y \theta)$	$P(\theta y)$
0	0	0.2	0	0
1	0.25	0.2	3/64	3/20
2	0.5	0.2	8/64	8/20
3	0.75	0.2	9/64	9/20
4	1	0.2	0	0

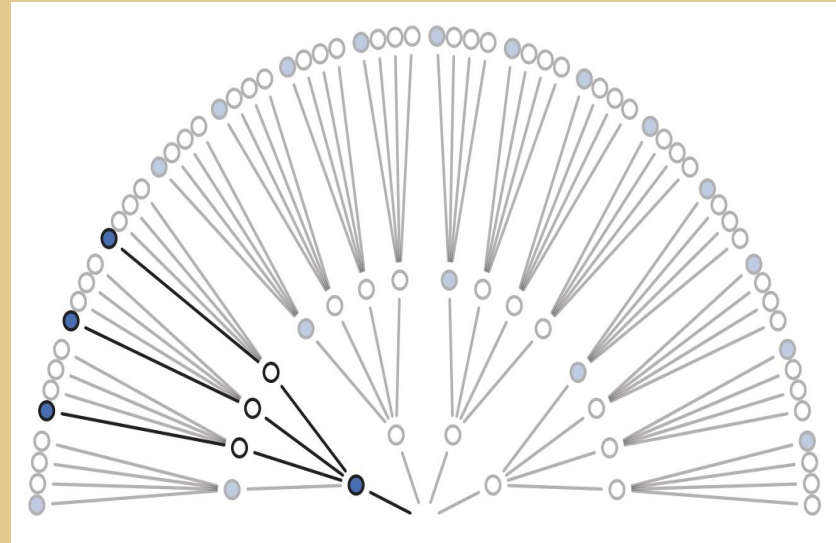
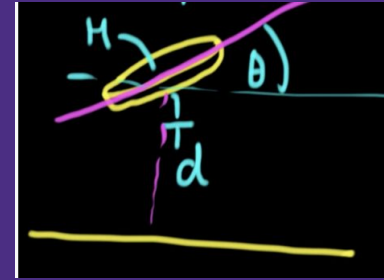


Image adapted from Statistical Rethinking, Richard McElreath



Frequentist/Bayesian Divide

What's probability?



- Fundamentally related to the frequencies of repeated events.
 - Frequentists
- Fundamentally related to our certainty or uncertainty of events.
 - Bayesian.

Covid death rate in Washington State

data:

- 1, We collect the data for the first 40 days in 2022 from USAFACTS.
- 2, Correcting data to reflect that deaths lag behind cases

	State	County Name	State	2022_case	2022_death	CFR	wa_cfr
0	WA	Adams County	WA	1011	0	0.0	0.0029403503844398400
1	WA	Asotin County	WA	777	6	0.007722007722007720	0.0029403503844398400
2	WA	Benton County	WA	12216	32	0.0026195153896529100	0.0029403503844398400
3	WA	Chelan County	WA	4622	10	0.0021635655560363500	0.0029403503844398400
4	WA	Clallam County	WA	3662	29	0.007919169852539600	0.0029403503844398400
5	WA	Clark County	WA	24268	111	0.0045739245096423300	0.0029403503844398400
6	WA	Columbia County	WA	71	4	0.056338028169014100	0.0029403503844398400
7	WA	Cowlitz County	WA	4452	32	0.0071877807726864300	0.0029403503844398400



The model

Likelihood

Prior distribution

$$\begin{aligned} \text{death}_i &\sim \text{Binomial}(\text{cases}_i, \theta_i) \\ \theta_i &\sim \text{Beta}(a, b) \end{aligned}$$

how do we choose a and b?

Choose a and b so that

mean of prior = mean of data, variance of prior = variance of data

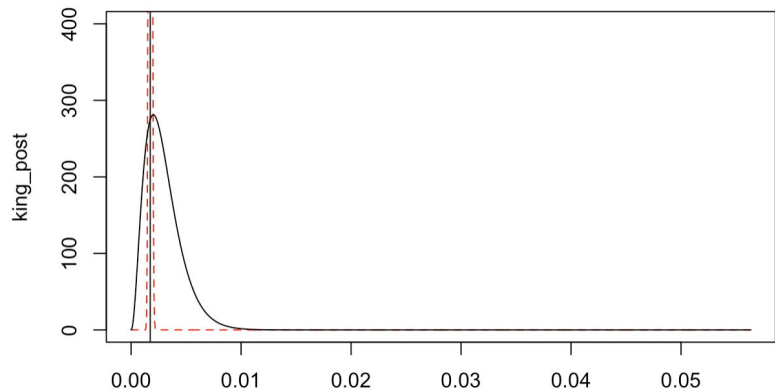
mean of beta dist: $a / (a+b)$, variance of beta: $ab / (a+b)^2(a+b+1)$

we get $a = 3.177$, $b = 1077$

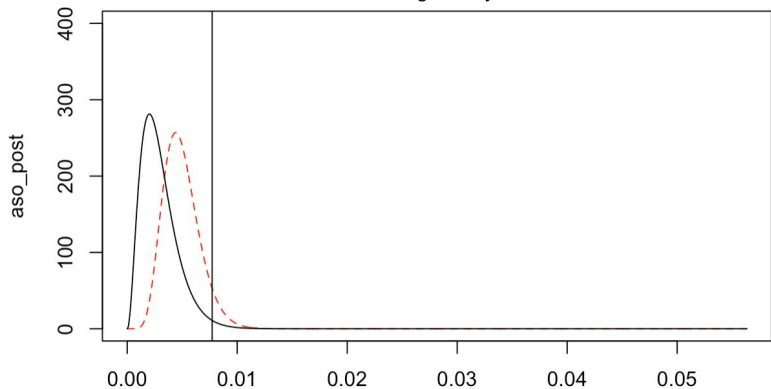
Posterior mean:

$$E[\theta_i | \text{data}] = \frac{a + \text{death}_i}{a + b + \text{cases}_i}$$

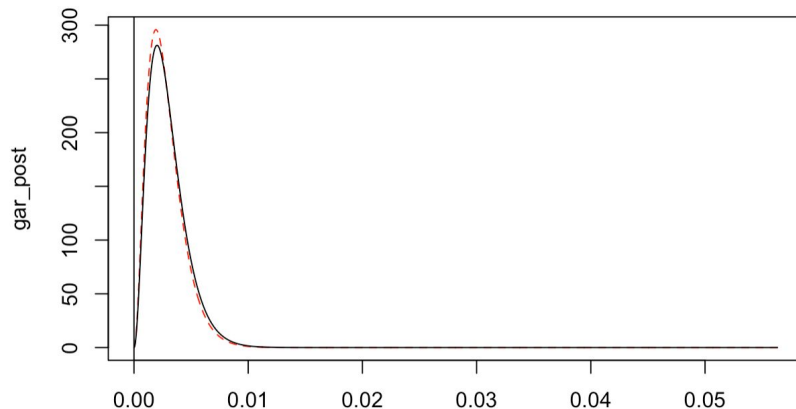
Advantage of Bayesian



King county

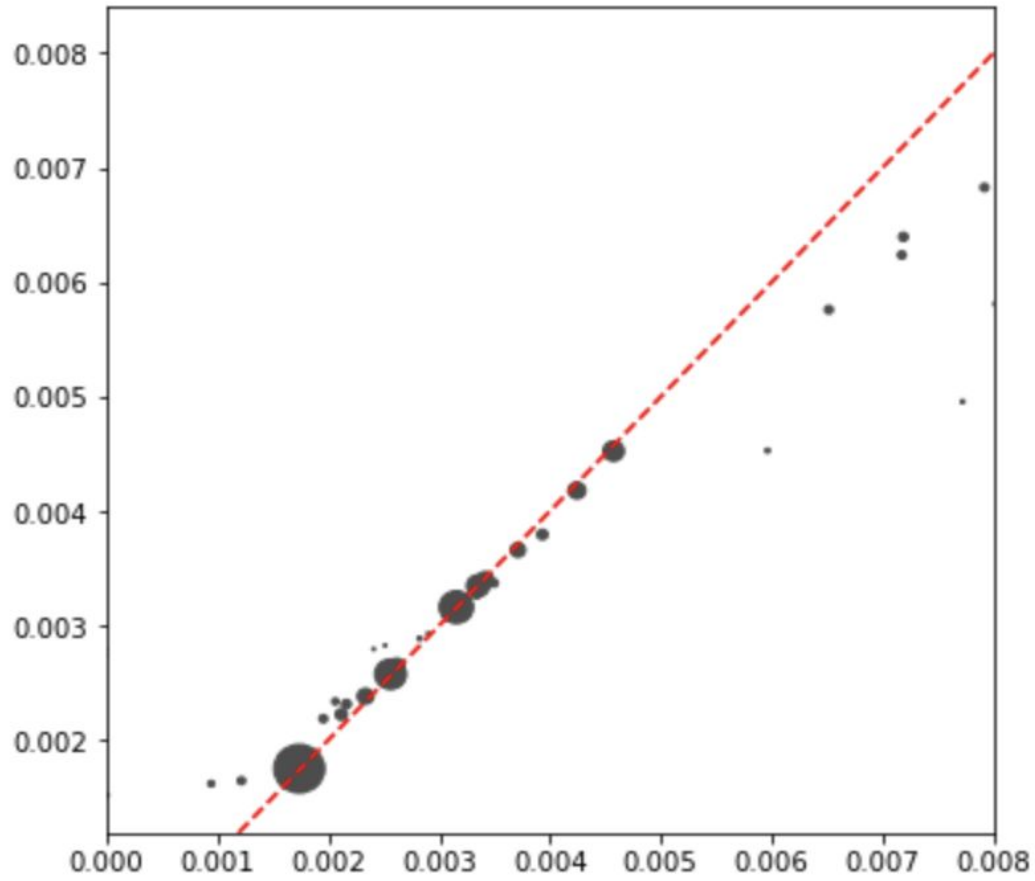


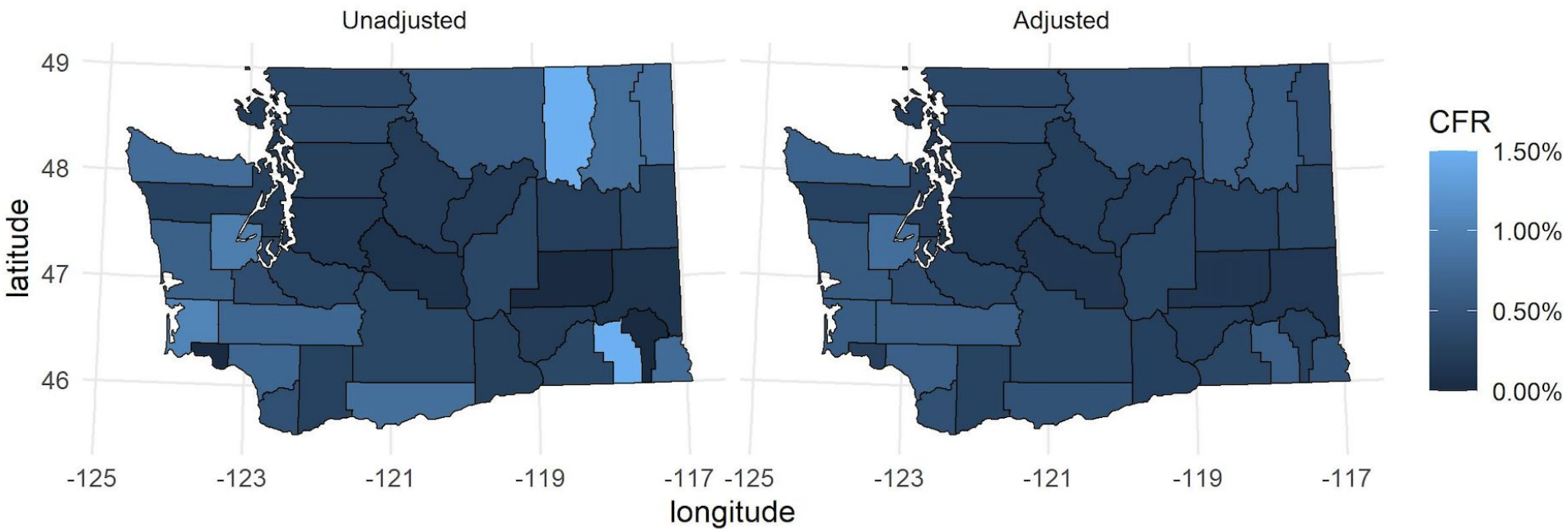
Asotin county



Garfield county

	case	death
King	144380	251
Garfield	56	0
Asotin	777	6





Note: For any county where the unadjusted CFR was above 1.5% (0.015), we replaced it with 1.5% so that the map scales could be compared.

