Bayesian Models

Probability and Statistics for Data Science

Carlos Fernandez-Granda

NYU Courant Institute of Mathematical Sciences

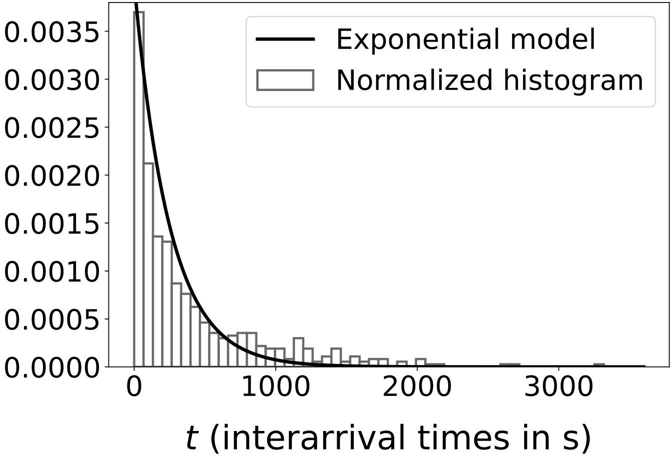
NYU Data Science

These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at https://www.ps4ds.net

Goal

Describe the framework of Bayesian modeling

Parametric modeling



Bayesian parametric modeling

Key idea: Interpret parameters as random variables

Modeling decisions

Parameters: $\tilde{\theta}$

Data: $\tilde{x}$

We need to make 2 decisions:

1. Prior distribution of parameters: $f_{\tilde{\theta}}$
2. Conditional distribution or likelihood of the data given the parameters $p_{\tilde{x} \mid \tilde{\theta}}$ or $f_{\tilde{x} \mid \tilde{\theta}}$

Goal: Compute posterior distribution of parameters given data

Coin flip

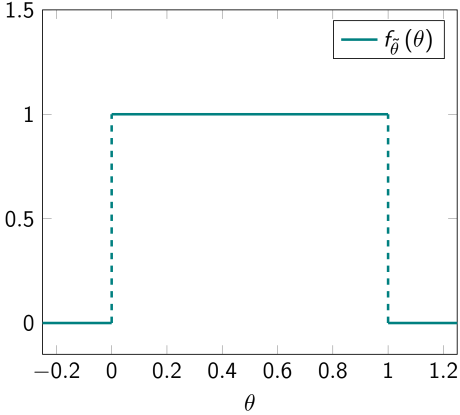
Bet on a coin flip, what is the probability of heads?

We are not sure whether coin flip is fair or not...

How can we encode uncertainty?

Through prior distribution of the probability $\tilde{\theta}$ that coin is heads

Uniform prior



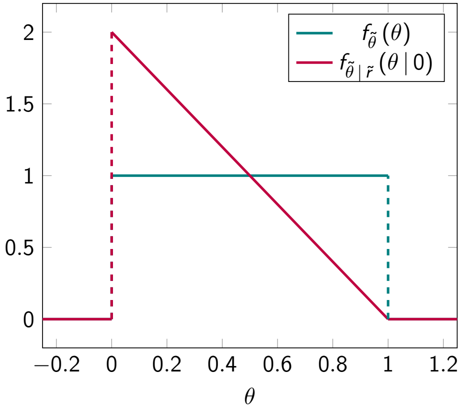
Posterior

Data: Coin lands on tails

Posterior pdf of $\tilde{\theta}$ given this information?

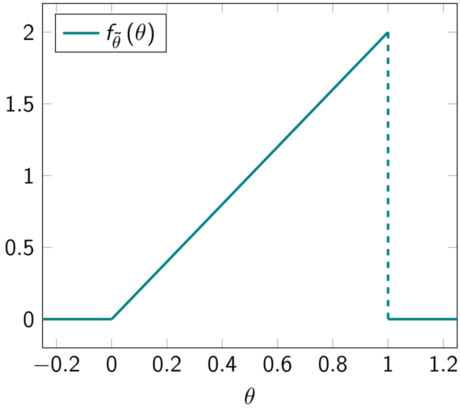
$$
\begin{aligned}
f_{\tilde{\theta}\,|\,\tilde{r}}(\theta\,|\,0) &= \frac{f_{\tilde{\theta}}(\theta)\, p_{\tilde{r}\,|\,\tilde{\theta}}(0\,|\,\theta)}{p_{\tilde{r}}(0)} \\
&= \frac{1-\theta}{\int_{u=-\infty}^{\infty} f_{\tilde{\theta}}(u)\, p_{\tilde{r}\,|\,\tilde{\theta}}(0\,|\,u)\,\mathrm{d}u} \\
&= \frac{1-\theta}{\int_{u=0}^{1}(1-u)\,\mathrm{d}u} \\
&= 2\,(1-\theta)
\end{aligned}
$$

Posterior



$f_{\tilde{\theta}}(\theta)$
$f_{\tilde{\theta}\,|\,\tilde{r}}(\theta\,|\,0)$
2
1.5
1
0.5
0
−0.2
0
0.2
0.4
0.6
0.8
1
1.2
θ

Triangular prior



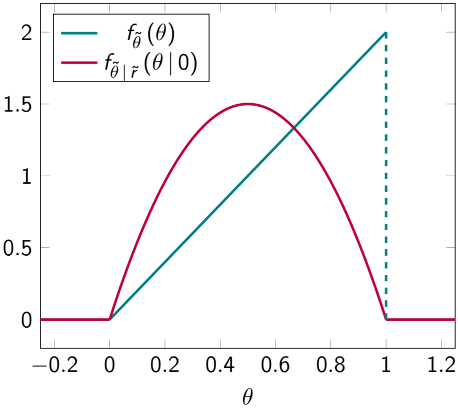
Posterior

Data: Coin lands on tails

Posterior pdf of $\tilde{\theta}$ given this information?

$$
\begin{aligned}
f_{\tilde{\theta}\,|\,\tilde{r}}(\theta\,|\,0) &= \frac{f_{\tilde{\theta}}(\theta)\, p_{\tilde{r}\,|\,\tilde{\theta}}(0\,|\,\theta)}{p_{\tilde{r}}(0)} \\
&= \frac{2\theta(1-\theta)}{\int_{u=-\infty}^{\infty} f_{\tilde{\theta}}(u)\, p_{\tilde{r}\,|\,\tilde{\theta}}(0\,|\,u)\, \mathrm{d}u} \\
&= \frac{2\theta(1-\theta)}{\int_{u=0}^{1} 2u(1-u)\, \mathrm{d}u} \\
&= 6\theta(1-\theta)
\end{aligned}
$$

Posterior



$f_{\tilde{\theta}}(\theta)$
$f_{\tilde{\theta}\,|\,\tilde{r}}(\theta\,|\,0)$
2
1.5
1
0.5
0
−0.2
0
0.2
0.4
0.6
0.8
1
1.2
θ

Beta distribution

Unimodal prior for parameters that represent probabilities

The pdf of a beta random variable $\tilde{t}$ with parameters a and b is

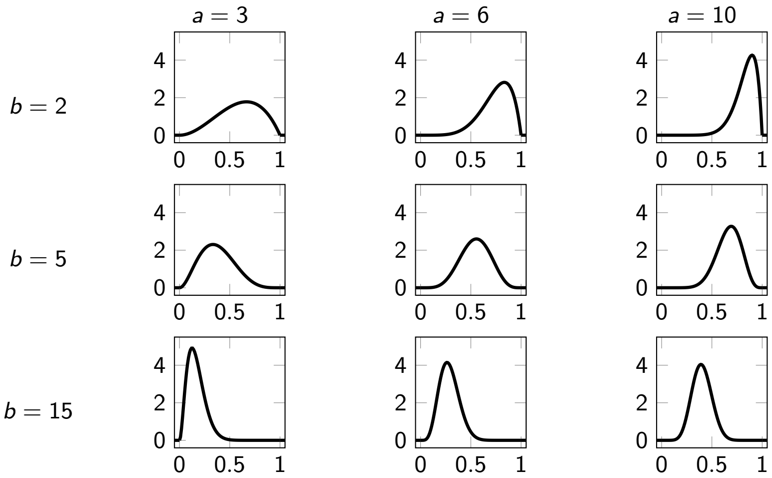
$$f_{\tilde{t}}(t) := \frac{t^{a-1}(1-t)^{b-1}}{\beta(a,b)} \qquad \text{if} \quad 0 \leq t \leq 1$$

and zero otherwise, where

$$\beta(a,b) := \int_u u^{a-1}(1-u)^{b-1}\,\mathrm{d}u$$

is a beta function or Euler integral of the first kind

Beta distribution



Conditional independence

What if we have more data?

Common assumption: Data are conditionally independent given parameters

Same effect as iid assumption: likelihood factorizes

$$p_{\tilde{x} \mid \tilde{\theta}}(x \mid \theta) = \prod_{i=1}^{n} p_{\tilde{x}[i] \mid \tilde{\theta}}(x[i] \mid \theta)$$

$$f_{\tilde{x} \mid \tilde{\theta}}(x \mid \theta) = \prod_{i=1}^{n} f_{\tilde{x}[i] \mid \tilde{\theta}}(x[i] \mid \theta)$$

Estimating parameter of Bernoulli

Data: Sequence of n binary outcomes (0 or 1)

Prior: Parameter $\tilde{\theta}$ modeled as beta with parameters a and b

Likelihood: Given $\tilde{\theta} = \theta$, each data point is Bernoulli with parameter θ

Distribution of number of 1s if data are conditionally independent given $\tilde{\theta}$?

Binomial with parameters n and θ

Posterior distribution

$$
\begin{aligned}
f_{\tilde{\theta} \mid \tilde{x}}(\theta \mid x) &= \frac{f_{\tilde{\theta}}(\theta)\, p_{\tilde{x} \mid \tilde{\theta}}(x \mid \theta)}{p_{\tilde{x}}(x)} \\
&= \frac{f_{\tilde{\theta}}(\theta)\, p_{\tilde{x} \mid \tilde{\theta}}(x \mid \theta)}{\int_u f_{\tilde{\theta}}(u)\, p_{\tilde{x} \mid \tilde{\theta}}(x \mid u)\, \mathrm{d}u} \\
&= \frac{\theta^{a-1}(1-\theta)^{b-1} \binom{n}{x} \theta^{x} (1-\theta)^{n-x}}{\int_u u^{a-1}(1-u)^{b-1} \binom{n}{x} u^{x} (1-u)^{n-x}\, \mathrm{d}u} \\
&= \frac{\theta^{x+a-1}(1-\theta)^{n-x+b-1}}{\int_u u^{x+a-1}(1-u)^{n-x+b-1}\, \mathrm{d}u}
\end{aligned}
$$

Beta random variable with parameters $x + a$ and $n - x + b$

Estimating parameter of Bernoulli

Data: Sequence of n binary outcomes (0 or 1)

Prior: Parameter $\tilde{\theta}$ modeled as beta with parameters a and b

Likelihood: Given $\tilde{\theta} = \theta$, each data point is Bernoulli with parameter θ, so number of 1s is binomial with parameters n and θ

Posterior distribution is beta with parameters $x + a$ and $n - x + b$

The beta distribution is a conjugate prior for the binomial likelihood

Real poll (Pennsylvania)

Data: 281 people intend to vote for Trump, 300 for Biden

Fraction of people that intend to vote for Trump: $\tilde{\theta}$

If n people are chosen independently at random with replacement from the population, likelihood of x voting for Trump?

Conditioned on $\tilde{\theta} = \theta$, binomial with parameters n and θ

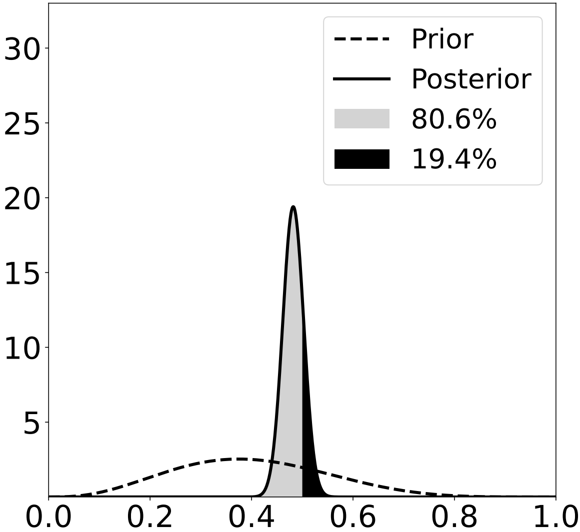
Posterior distribution of $\tilde{\theta}$ depends on prior!

If prior is beta with parameters a and b, posterior is beta with parameters $a + 281$ and $b + 300$

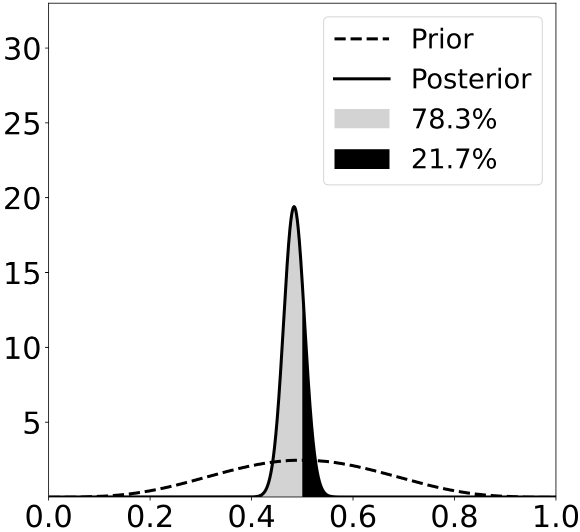
Real poll (Pennsylvania)

Probability that Trump wins in Pennsylvania? $P(\tilde{\theta} > 0.5 \,|\, \tilde{x} = x)$

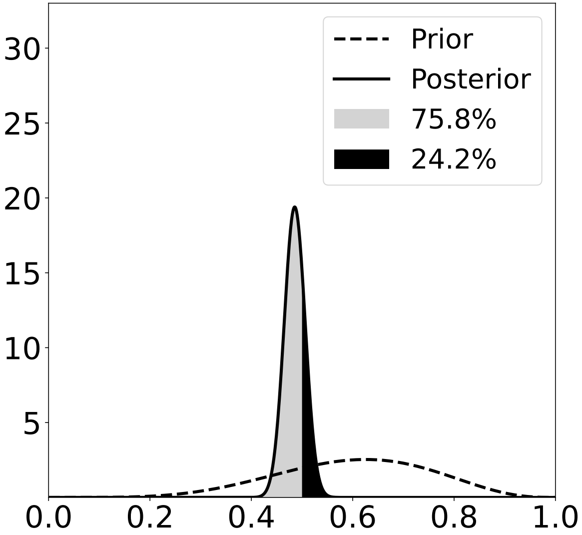
Poll in Pennsylvania



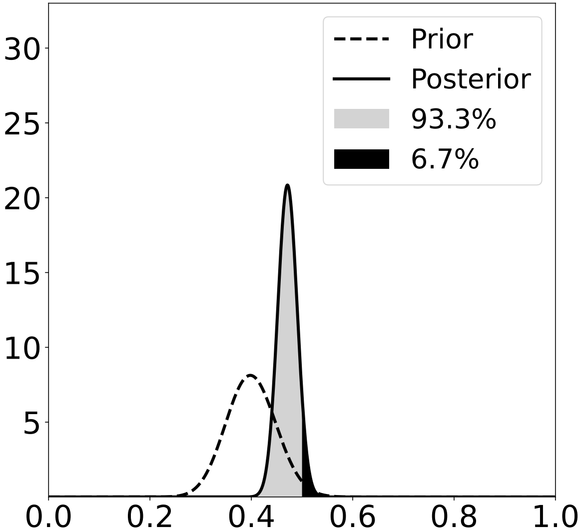
Poll in Pennsylvania



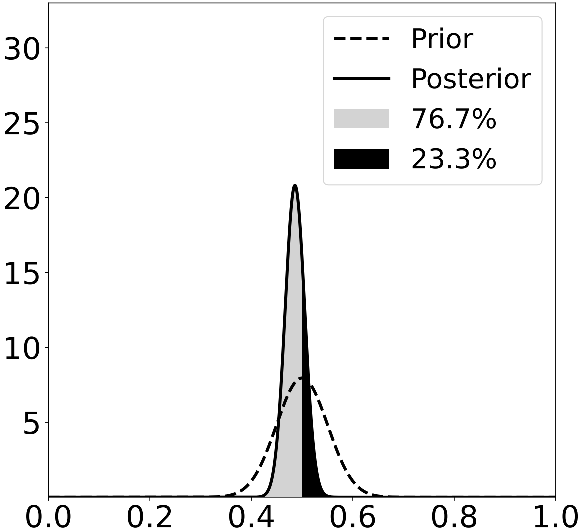
Poll in Pennsylvania



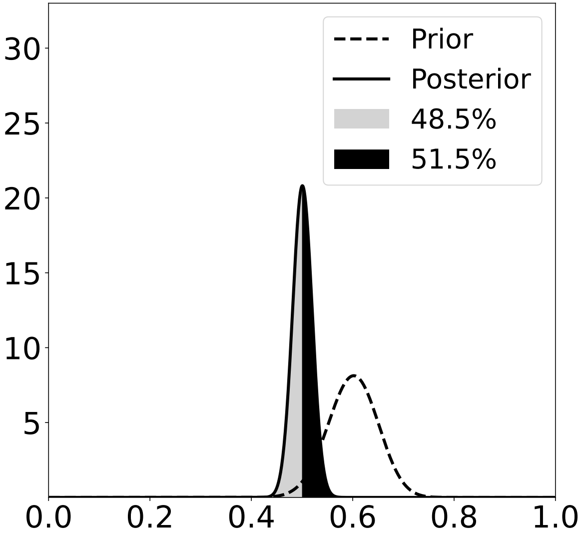
Poll in Pennsylvania



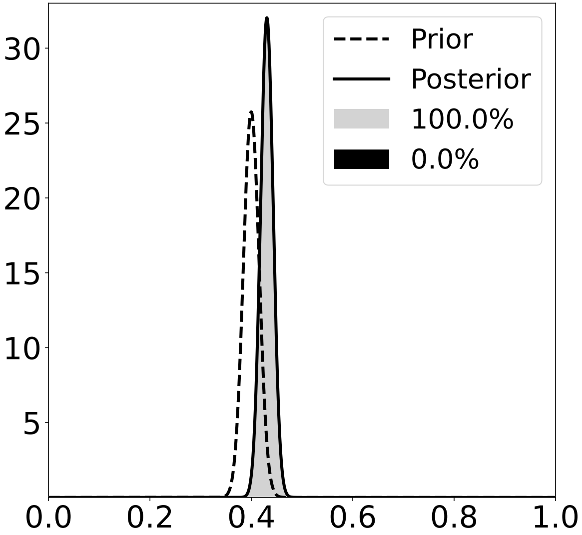
Poll in Pennsylvania



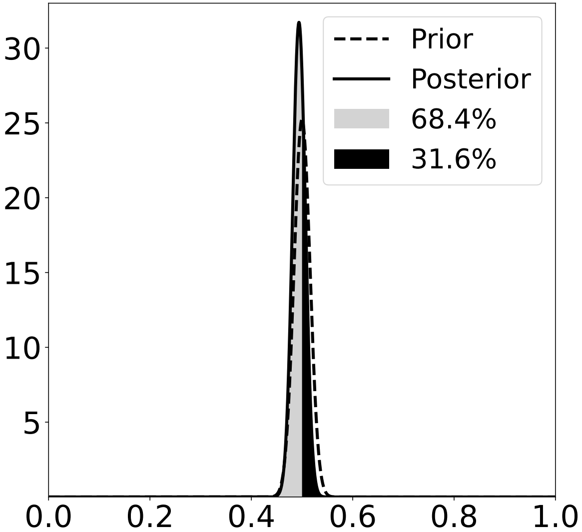
Poll in Pennsylvania



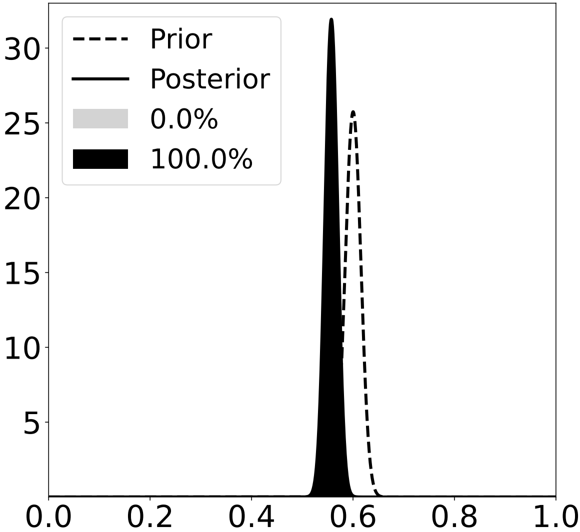
Poll in Pennsylvania



Poll in Pennsylvania



Poll in Pennsylvania



What have we learned?

Bayesian framework for parametric modeling

Conjugate priors