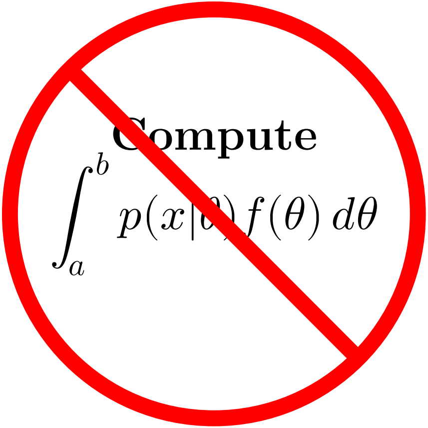
Conjugate Priors: Beta and Normal
18.05 Spring 2022



Announcements/Agenda

Announcements

- Studio comments

Agenda

- Lots of concept questions today: let's be efficient.
- Beta distribution
- Beta-binomial conjugate priors
- Normal-normal conjugate priors

What is Bayesian updating good for

- Deep Bayesian Learning
- Speech recognition: training models
- Medical diagnosis, visualization
- Experimental design
- Lots more!

Beta distribution

Beta(a, b) has density

$$f(\theta) = \frac{(a+b-1)!}{(a-1)!(b-1)!}\theta^{a-1}(1-\theta)^{b-1}$$

We will use a and b positive integers but real a, $b > 0$ are allowed.

- Beta(a, b) has nice computational properties.
- By choosing a and b, the Beta distribution allows us to choose a prior on the range [0,1] that has its mode at any value and has small or large variance around the mode.

https://mathlets.org/mathlets/beta-distribution/

Observation

The factorials out front give a normalizing factor. That is,

$$f(\theta) = \frac{(a+b-1)!}{(a-1)!(b-1)!}\theta^{a-1}(1-\theta)^{b-1} = c\theta^{a-1}(1-\theta)^{b-1}$$

Where there is no choice in the value of c, i.e., it must make

$$\int_0^1 f(\theta)\, d\theta = 1.$$

So if we have a pdf on $[0,1]$ of the form $f(\theta) = c\theta^{a-1}(1-\theta)^{b-1}$, then

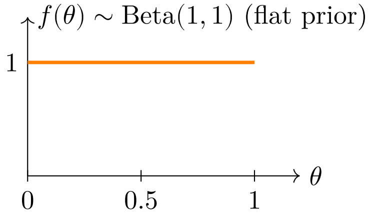
$$\theta \sim \text{Beta}(a,b), \text{ and } c = \frac{(a+b-1)!}{(a-1)!(b-1)!}.$$

Aside on flat priors

If you don't have any prior information of a location parameter like θ, you can use a flat prior also called an uninformative prior.

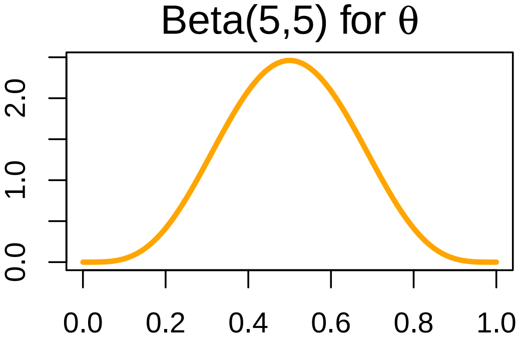
$$f(\theta) = 1$$

Note: This is the pdf of a Beta(1,1) distribution.



Board question preamble: Beta priors

Suppose you are testing a new medical treatment with unknown probability of success θ. You don't know θ, but your prior belief is that it's probably not too far from 0.5. You capture this intuition with a Beta(5,5) prior on θ.



To sharpen this distribution you take data and update the prior.

Question on next slide.

Board question: Beta priors

- Beta(a,b): $f(\theta) = \dfrac{(a+b-1)!}{(a-1)!(b-1)!}\theta^{a-1}(1-\theta)^{b-1}$
- Treatment has prior $f(\theta) \sim$ Beta$(5,5)$

(a) Suppose you test it on 25 patients and have 20 successes.
– Find the posterior distribution on θ.
– Identify the type of the posterior distribution.

(b) Suppose you recorded the order of the results and got

$$S\,S\,S\,S\,F\,S\,S\,S\,S\,S\,F\,F\,S\,S\,S\,F\,S\,F\,S\,S\,S\,S\,S\,S\,S$$

(20 S and 5 F). Find the posterior based on this data.

(c) Using your answer to (b) give an integral for the posterior predictive probability of success with the next patient.

Conjugate priors

We had

- Prior $f(\theta)\,d\theta =$ Beta(5,5): Beta distribution
- Likelihood $p(x|\theta)$: binomial distribution $x =$ 20 success, 5 failure
- Posterior $f(\theta|x)\,d\theta =$ Beta(25,10): Beta distribution

The Beta distribution is called a **conjugate prior** for the binomial likelihood.

That is, the Beta prior becomes a Beta posterior and repeated updating is easy!
Only the parameters have been changed to reflect the data.

Concept question: More Beta

Suppose your prior $f(\theta)$ in the bent coin example is Beta$(6,8)$. You flip the coin 7 times, getting 2 heads and 5 tails. What is the posterior pdf $f(\theta|x)$?

(a) Beta(2,5)

(b) Beta(11,10)

(c) Beta(6,8)

(d) Beta(8,13)

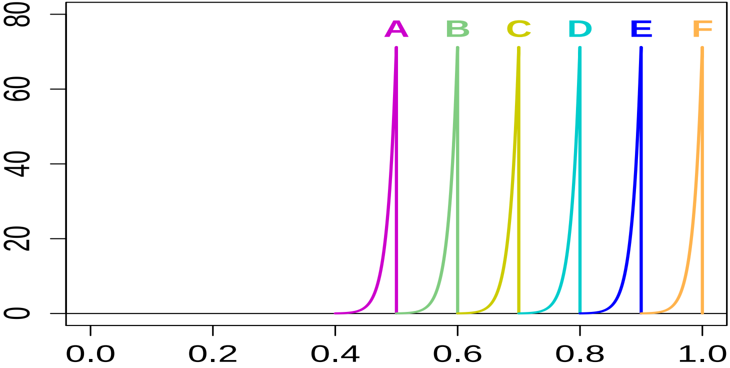
Concept question: strong priors

Say we have a bent coin with unknown probability of heads θ.

We are convinced that $\theta \leq 0.7$.
Our prior is uniform on $[0, 0.7]$ and 0 from 0.7 to 1.

We flip the coin 65 times and get 60 heads.

Which of the graphs below is the posterior pdf for θ?



Updating with normal prior and normal likelihood

A normal prior is conjugate to a normal likelihood with known σ.

- Data: $x_1, x_2, \ldots, x_n$
- **Normal likelihood.** $x_1, x_2, \ldots, x_n \sim \mathsf{N}(\theta, \sigma^2)$

  Assume θ is our unknown parameter of interest, σ is known.
- **Normal prior.** $\theta \sim \mathsf{N}(\mu_{\text{prior}}, \sigma^2_{\text{prior}})$.
- **Normal Posterior.** $\theta \sim \mathsf{N}(\mu_{\text{post}}, \sigma^2_{\text{post}})$.

- We have simple updating formulas that allow us to avoid complicated algebra or integrals (see next slide).

$$a = \frac{1}{\sigma^2_{\text{prior}}}, \quad b = \frac{n}{\sigma^2}, \quad \mu_{\text{post}} = \frac{a\mu_{\text{prior}} + b\bar{x}}{a+b}, \quad \sigma^2_{\text{post}} = \frac{1}{a+b}.$$

Updating with normal prior and normal likelihood

Formulas:

$$a = \frac{1}{\sigma^2_{\text{prior}}}, \quad b = \frac{n}{\sigma^2}, \quad \mu_{\text{post}} = \frac{a\mu_{\text{prior}} + b\bar{x}}{a+b}, \quad \sigma^2_{\text{post}} = \frac{1}{a+b}.$$

Notes:

- Posterior mean μ_{post} is a weighted average of the data mean and the prior mean.
- Bigger n puts more weight on the data.
- Variance always decreases, i.e. $\sigma^2_{\text{prior}} < \sigma^2_{\text{post}}$.

The update table with continuous data looks the same as always.

hypoth.	prior	likelihood	posterior
θ	$f(\theta) \sim \text{N}(\mu_{\text{prior}}, \sigma^2_{\text{prior}})$	$\phi(x\vert\theta) \sim \text{N}(\theta, \sigma^2)$	$f(\theta\vert x) \sim \text{N}(\mu_{\text{post}}, \sigma^2_{\text{post}})$
	$= c_1 \exp\left(\frac{-(\theta-\mu_{\text{prior}})^2}{2\sigma^2_{\text{prior}}}\right)$	$= c_2 \exp\left(\frac{-(x-\theta)^2}{2\sigma^2}\right)$	$= c_3 \exp\left(\frac{-(\theta-\mu_{\text{post}})^2}{2\sigma^2_{\text{post}}}\right)$

Board question: Normal-normal updating formulas

For data $x_1, \ldots, x_n$ with data mean $\bar{x} = \frac{x_1 + \ldots + x_n}{n}$

$$a = \frac{1}{\sigma^2_{\text{prior}}}, \quad b = \frac{n}{\sigma^2}, \quad \mu_{\text{post}} = \frac{a\mu_{\text{prior}} + b\bar{x}}{a+b}, \quad \sigma^2_{\text{post}} = \frac{1}{a+b}.$$

Suppose we have one data point $x = 2$ drawn from $\mathrm{N}(\theta, 3^2)$

Suppose θ is our parameter of interest with prior $\theta \sim \mathrm{N}(4, 2^2)$.

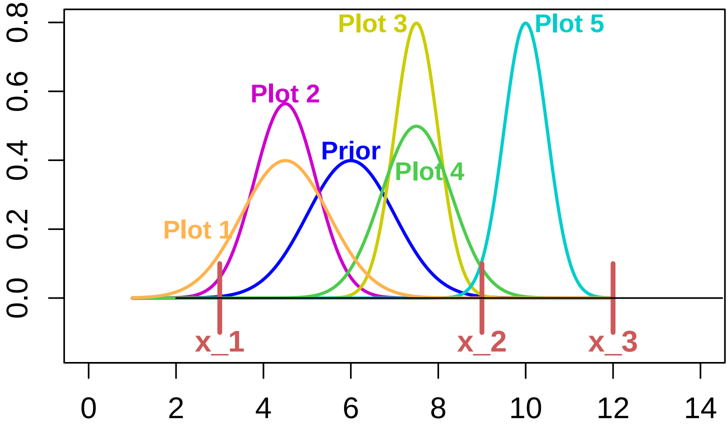
(a) Identify μ_{prior}, σ_{prior}, σ, n, and $\bar{x}$.

(b) Make a Bayesian update table, but leave the posterior as an unsimplified product.

(c) Use the updating formulas to find the posterior.

Concept question: (a) Normal priors, normal likelihood

(a)

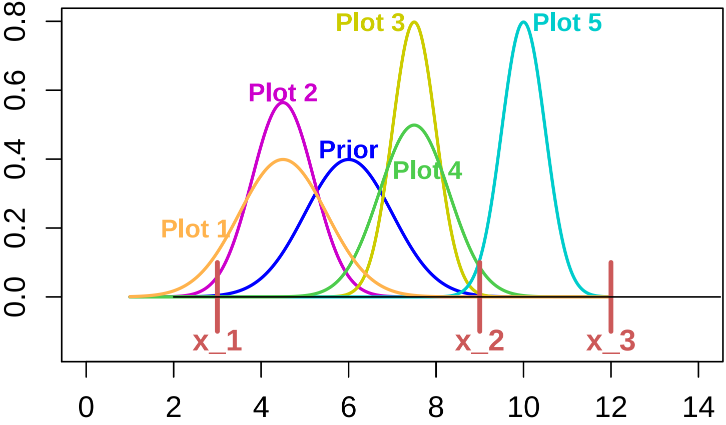


Blue graph = prior, Red lines = data in order: 3, 9, 12

Which plot is the posterior to just the first data value?

Concept question: (b) Normal priors, normal likelihood

(b)



Blue graph = prior, Red lines = data in order: 3, 9, 12
Which graph is posterior to all 3 data values?

Board question: normal/normal

For data $x_1, \dots, x_n$ with data mean $\bar{x} = \frac{x_1 + \dots + x_n}{n}$

$$a = \frac{1}{\sigma^2_{\text{prior}}}, \quad b = \frac{n}{\sigma^2}, \quad \mu_{\text{post}} = \frac{a\mu_{\text{prior}} + b\bar{x}}{a+b}, \quad \sigma^2_{\text{post}} = \frac{1}{a+b}.$$

Question. On a basketball team the average career free throw percentage over all players follows a $\mathsf{N}(75, 6^2)$ distribution. In a given year individual players free throw percentage is $\mathsf{N}(\theta, 4^2)$ where θ is their career average.

This season Sophie Lie made 85 percent of her free throws. What is the posterior expected value of her career percentage θ?

Table discussion: likelihood principle

Suppose the prior has been set. Let x_1 and x_2 be two sets of data. Which of the following are true?

(a) If the likelihoods $\phi(x_1|\theta)$ and $\phi(x_2|\theta)$ are the same then they result in the same posterior.

(b) If the likelihoods $\phi(x_1|\theta)$ and $\phi(x_2|\theta)$ are proportional (as functions of θ) then they result in the same posterior.

(c) If two likelihood functions are proportional then they are equal.

Conjugate priors

A prior is conjugate to a likelihood if the posterior is the same type of distribution as the prior.
Updating becomes algebra instead of calculus.

	hypothesis	data	prior	likelihood	posterior
Bernoulli/Beta	$\theta \in [0,1]$	x	$\text{beta}(a,b)$	$\text{Bernoulli}(\theta)$	$\text{beta}(a+1,b)$ or $\text{beta}(a,b+1)$
	θ	$x=1$	$c_1\theta^{a-1}(1-\theta)^{b-1}$	θ	$c_3\theta^{a}(1-\theta)^{b-1}$
	θ	$x=0$	$c_1\theta^{a-1}(1-\theta)^{b-1}$	$1-\theta$	$c_3\theta^{a-1}(1-\theta)^{b}$
Binomial/Beta	$\theta \in [0,1]$	x	$\text{beta}(a,b)$	$\text{binomial}(N,\theta)$	$\text{beta}(a+x,b+N-x)$
(fixed N)	θ	x	$c_1\theta^{a-1}(1-\theta)^{b-1}$	$c_2\theta^{x}(1-\theta)^{N-x}$	$c_3\theta^{a+x-1}(1-\theta)^{b+N-x-1}$
Geometric/Beta	$\theta \in [0,1]$	x	$\text{beta}(a,b)$	$\text{geometric}(\theta)$	$\text{beta}(a+x,b+1)$
	θ	x	$c_1\theta^{a-1}(1-\theta)^{b-1}$	$\theta^{x}(1-\theta)$	$c_3\theta^{a+x-1}(1-\theta)^{b}$
Normal/Normal	$\theta \in (-\infty,\infty)$	x	$\text{N}(\mu_{\text{prior}},\sigma^2_{\text{prior}})$	$\text{N}(\theta,\sigma^2)$	$\text{N}(\mu_{\text{post}},\sigma^2_{\text{post}})$
(fixed σ^2)	θ	x	$c_1\exp\left(\frac{-(\theta-\mu_{\text{prior}})^2}{2\sigma^2_{\text{prior}}}\right)$	$c_2\exp\left(\frac{-(x-\theta)^2}{2\sigma^2}\right)$	$c_3\exp\left(\frac{(\theta-\mu_{\text{post}})^2}{2\sigma^2_{\text{post}}}\right)$

There are many other likelihood/conjugate prior pairs.

Finger question: conjugate priors

Which of the following likelihood/prior pairs are conjugate?

	hypothesis	data	prior	likelihood
(a) Exponential/Normal	$\theta \in [0, \infty)$	x	$\mathrm{N}(\mu_{\text{prior}}, \sigma^2_{\text{prior}})$	$\exp(\theta)$
	θ	x	$c_1 \exp\left(-\frac{(\theta-\mu_{\text{prior}})^2}{2\sigma^2_{\text{prior}}}\right)$	$\theta \mathrm{e}^{-\theta x}$
(b) Exponential/Gamma	$\theta \in [0, \infty)$	x	$\mathrm{Gamma}(a, b)$	$\exp(\theta)$
	θ	x	$c_1 \theta^{a-1} \mathrm{e}^{-b\theta}$	$\theta \mathrm{e}^{-\theta x}$
(c) Binomial/Normal	$\theta \in [0, 1]$	x	$\mathrm{N}(\mu_{\text{prior}}, \sigma^2_{\text{prior}})$	$\mathrm{binomial}(N, \theta)$
(fixed N)	θ	x	$c_1 \exp\left(-\frac{(\theta-\mu_{\text{prior}})^2}{2\sigma^2_{\text{prior}}}\right)$	$c_2\, \theta^x (1-\theta)^{N-x}$

1. none 2. a 3. b 4. c
5. a,b 6. a,c 7. b,c 8. a,b,c