Conjugate Priors: Beta and Normal

18.05 Spring 2018

Review: Continuous priors, discrete data

'Bent' coin: unknown probability θ of heads.

Prior $f(\theta) = 2\theta$ on [0,1].

Data: heads on one toss.

Question: Find the posterior pdf to this data.

hypoth.	prior	likelihood	Bayes numerator	posterior
θ	$2\theta\,d\theta$	θ	$2\theta^2\,d\theta$	$3\theta^2\,d\theta$
Total	1		$T = \int_0^1 2\theta^2\,d\theta = 2/3$	1

Posterior pdf: $f(\theta|x) = 3\theta^2$.

Review: Continuous priors, continuous data

Bayesian update table

hypoth.	prior	likeli.	Bayes numerator	posterior
θ	$f(\theta)\,d\theta$	$\phi(x\,\vert\,\theta)$	$\phi(x\,\vert\,\theta)f(\theta)\,d\theta$	$f(\theta\,\vert\,x)\,d\theta = \dfrac{\phi(x\,\vert\,\theta)f(\theta)\,d\theta}{\phi(x)}$
total	1		$\phi(x)$	1

$$\phi(x) = \int \phi(x\,|\,\theta)f(\theta)\,d\theta$$

$$= \text{probability of data } x$$

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.** $x_1, x_2, \ldots, x_n \sim \mathrm{N}(\theta, \sigma^2)$

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

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

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

Board question: Normal-normal updating formulas

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

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

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

0. Identify μ_{prior}, σ_{prior}, σ, n, and $\bar{x}$.

1. Make a Bayesian update table, but leave the posterior as an unsimplified product.

2. Use the updating formulas to find the posterior.

3. By doing enough of the algebra, understand that the updating formulas come by using the updating table and doing a lot of algebra.

Solution

0. $\mu_{\text{prior}} = 4$, $\sigma_{\text{prior}} = 2$, $\sigma = 3$, $n = 1$, $\bar{x} = 2$.

1.

hypoth.	prior	likelihood	posterior
θ	$f(\theta) \sim \text{N}(4, 2^2)$	$f(x\|\theta) \sim \text{N}(\theta, 3^2)$	$f(\theta\|x) \sim \text{N}(\mu_{\text{post}}, \sigma^2_{\text{post}})$
θ	$c_1 \exp\left(\frac{-(\theta-4)^2}{8}\right)$	$c_2 \exp\left(\frac{-(2-\theta)^2}{18}\right)$	$c_3 \exp\left(\frac{-(\theta-4)^2}{8}\right) \exp\left(\frac{-(2-\theta)^2}{18}\right)$

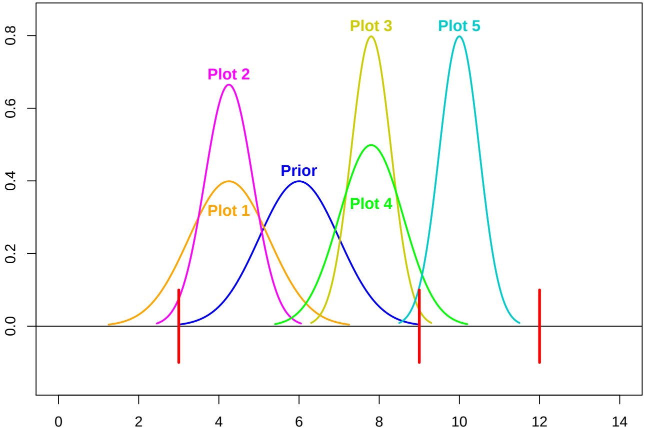
2. We have $a = 1/4, \quad b = 1/9, \quad a + b = 13/36.$ Therefore

$$\begin{aligned} \mu_{\text{post}} &= (1 + 2/9)/(13/36) = 44/13 = 3.3846 \\ \sigma^2_{\text{post}} &= 36/13 = 2.7692 \end{aligned}$$

The posterior pdf is $f(\theta|x = 2) \sim \text{N}(3.3846, 2.7692)$.

3. See the reading class15-prep-a.pdf example 2.

Concept question: normal priors, normal likelihood

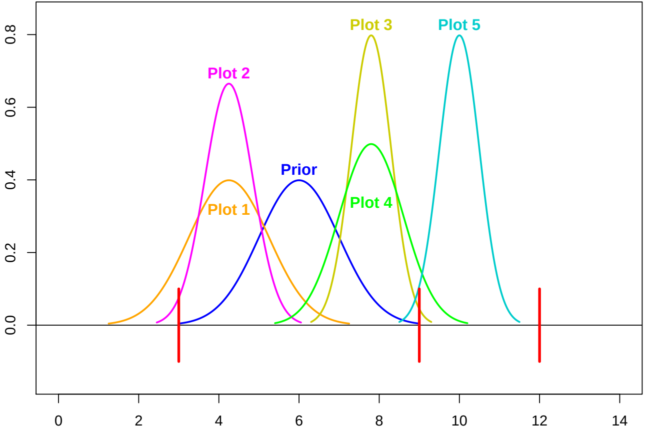


Blue graph = prior   Red lines = data in order: 3, 9, 12

(a) Which plot is the posterior to just the first data value?

(Solution in 2 slides)

Concept question: normal priors, normal likelihood



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

(b) Which graph is posterior to all 3 data values?

(Solution on next slide)

Solution to concept question

(a) Plot 2: The first data value is 3. Therefore the posterior must have its mean between 3 and the mean of the blue prior. The only possibilites for this are plots 1 and 2. We also know that the variance of the posterior is less than that of the posterior. Between the plots 1 and 2 graphs only plot 2 has smaller variance than the prior.

(b) Plot 3: The average of the 3 data values is 8. Therefore the posterior must have mean between the mean of the blue prior and 8. Therefore the only possibilities are the plots 3 and 4. Because the posterior is posterior to the magenta graph (plot 2) it must have smaller variance. This leaves only the Plot 3.

Board question: normal/normal

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

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

Question. On a basketball team the players are drawn from a pool in which the career average free throw percentage follows a $\mathrm{N}(75, 6^2)$ distribution. In a given year individual players free throw percentage is $\mathrm{N}(\theta, 4^2)$ 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?
<u>**answer:**</u> *Solution on next frame*

Solution

This is a normal/normal conjugate prior pair, so we use the update formulas.
Parameter of interest: $\theta =$ career average.
Data: $x = 85 =$ this year's percentage.
Prior: $\theta \sim \mathrm{N}(75, 36)$
Likelihood $x \sim \mathrm{N}(\theta, 16)$. So $f(x|\theta) = c_1 \mathrm{e}^{-(x-\theta)^2/2\cdot 16}$.
The updating weights are

$$a = 1/36, \quad b = 1/16, \quad a + b = 52/576 = 13/144.$$

Therefore

$$\mu_{\text{post}} = (75/36 + 85/16)/(52/576) = 81.9, \qquad \sigma^2_{\text{post}} = 36/13 = 11.1.$$

The posterior pdf is $f(\theta|x = 85) \sim \mathrm{N}(81.9, 11.1)$.

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.

Concept question: conjugate priors

Which are conjugate priors?

	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

Answer: 3. b

We have a conjugate prior if the posterior *as a function of* θ has the same form as the prior.

Exponential/Normal posterior:

$$f(\theta|x) = c_1 \theta \mathrm{e}^{-\frac{(\theta-\mu_{\text{prior}})^2}{2\sigma_{\text{prior}}^2} - \theta x}$$

The factor of θ before the exponential means this is not the pdf of a normal distribution. Therefore it is not a conjugate prior.

Exponential/Gamma posterior: Note, we have never learned about Gamma distributions, but it doesn't matter. We only have to check if the posterior has the same form:

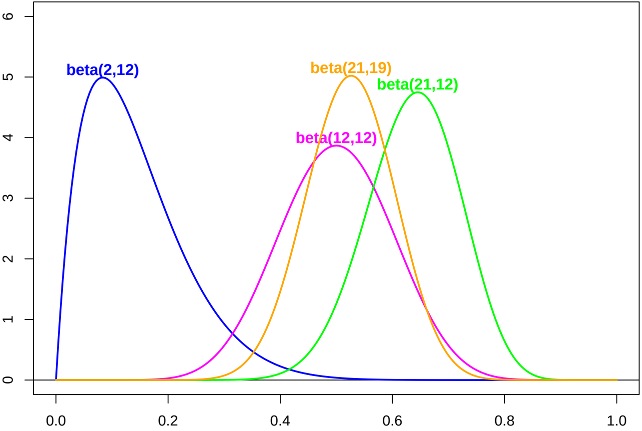
$$f(\theta|x) = c_1 \theta^a \mathrm{e}^{-(b+x)\theta}$$

The posterior has the form Gamma$(a+1, b+x)$. This is a conjugate prior.

Binomial/Normal: It is clear that the posterior does not have the form of a normal distribution.

Variance can increase

Normal-normal: variance **always** decreases with data.
Beta-binomial: variance **usually** decreases with data.



Variance of beta(2,12) (blue) is smaller than that of beta(12,12) (magenta), but beta(12,12) can be a posterior to beta(2,12)

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 x_1 and x_2 result in the same posterior then their likelihood functions are the same.

(c) 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.

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

answer: (4): a: true; b: false, the likelihoods are proportional.
c: true, scale factors don't matter d: false

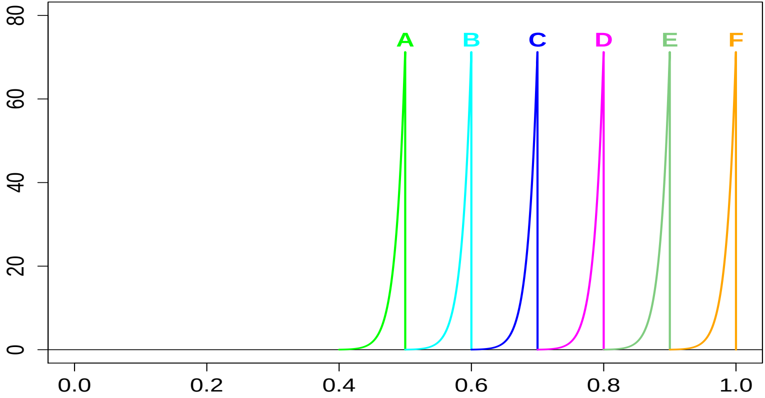
Concept question: strong priors

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?



Solution to concept question

answer: Graph C, the blue graph spiking near 0.7.

Sixty heads in 65 tosses indicates the true value of θ is close to 1. Our prior was 0 for $\theta > 0.7$. So no amount of data will make the posterior non-zero in that range. That is, we have foreclosed on the possibility of deciding that θ is close to 1. The Bayesian updating puts θ near the top of the allowed range.