Bayesian Updating: Continuous Priors

18.05 Spring 2014

Jeremy Orloff and Jonathan Bloom

Beta distribution

$Beta(a, b)$ has density

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

http://ocw.mit.edu/ans7870/18/18.05/s14/applets/beta-jmo.html

Observation:

The coefficient is a normalizing factor, so if we have a pdf

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

then

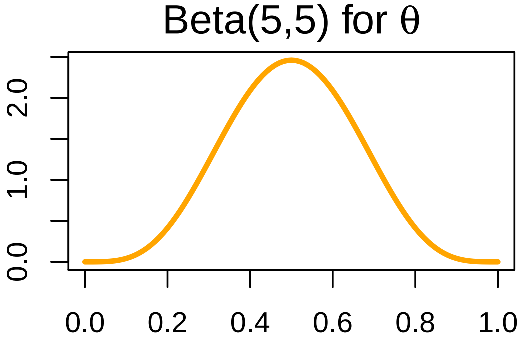
$$\theta \sim \text{beta}(a, b)$$

and

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

Board question preamble: beta priors

Suppose you have a coin with unknown probability of heads θ. You don't know that it's fair, but your prior belief is that it's probably not too unfair. You capture this intuition in with a beta(5,5) prior on θ.



In order to sharpen this distribution you take data and update the prior.

Question on next slide.

Board question: beta priors

- $Beta(a, b)$: $f(\theta) = \frac{(a+b-1)!}{(a-1)!(b-1)!}\theta^{a-1}(1-\theta)^{b-1}$
- Coin has prior $f(\theta) \sim \text{beta}(5, 5)$

1. Suppose you flip 10 times and get 6 heads. Find the posterior distribution on θ. Identify the type of the posterior distribution.

2. Suppose you recorded the order of the flips and got H H H T T H H H T T. Find the posterior based on this data.

3. Using your answer to (2) give an integral for the posterior predictive probability of heads on the next toss.

4. Use what you know about pdf's to evaluate the integral without computing it directly

Solution

1. Prior pdf is $f(\theta) = \frac{9!}{4!\,4!}\theta^4(1-\theta)^4 = c_1\theta^4(1-\theta)^4$.

hypoth.	prior	likelihood	un. post.	posterior
$\theta \pm d\theta$	$c_1\theta^4(1-\theta)^4\,d\theta$	$\binom{10}{6}\theta^6(1-\theta)^4$	$c_3\theta^{10}(1-\theta)^8\,d\theta$	beta$(11,9)$

We know the normalized posterior is a beta distribution because it has the form of a beta distribution ($c\theta^{a-}(1-\theta)^{b-1}$ on [0,1]) so by our earlier observation it must be a beta distribution.

2. The answer is the same. The only change is that the likelihood has a coefficient of 1 instead of a binomial coefficent.

3. The posterior on θ is beta$(11,9)$ which has density

$$f(\theta\,|,\text{data}) = \frac{19!}{10!\,8!}\theta^{10}(1-\theta)^8.$$

Solution to (3) continued on next slide

Solution continued

The law of total probability says that the posterior predictive probability of heads is

$$P(\text{heads} \mid \text{data}) = \int_0^1 f(\text{heads} \mid \theta) \cdot f(\theta \mid \text{data})\, d\theta$$
$$= \int_0^1 \theta \cdot \frac{19!}{10!\, 8!} \theta^{10} (1-\theta)^8 \, d\theta = \int_0^1 \frac{19!}{10!\, 8!} \theta^{11} (1-\theta)^8 \, d\theta$$

4. We compute the integral in (3) by relating it to the pdf of beta$(12, 9)$: $\frac{20!}{11!\, 8!} \theta^{11} (1-\theta)^7$. Since all pdf's integrate to 1 we have

$$\int_0^1 \frac{20!}{11!\, 8!} \theta^{11} (1-\theta)^7 = 1 \quad \Rightarrow \quad \int_0^1 \theta^{11} (1-\theta)^7 = \frac{11!\, 8!}{20!}.$$

Thus

$$\int_0^1 \frac{19!}{10!\, 8!} \theta^{11} (1-\theta)^8 \, d\theta = \frac{19!}{10!\, 8!} \cdot \frac{11!\, 8!}{20!}. = \boxed{\frac{11}{20}}.$$

Predictive probabilities

Continuous hypotheses θ, discrete data $x_1, x_2, \ldots$
(Assume trials are independent.)

Prior predictive probability

$$p(x_1) = \int p(x_1 \,|\, \theta) f(\theta)\, d\theta$$

Posterior predictive probability

$$p(x_2 \,|\, x_1) = \int p(x_2 \,|\, \theta) f(\theta \,|\, x_1)\, d\theta$$

Analogous to discrete hypotheses: $\mathcal{H}_1, \mathcal{H}_2, \ldots$.

$$p(x_1) = \sum_{i=1}^{n} p(x_1 \,|\, \mathcal{H}_i) P(\mathcal{H}_i) \qquad p(x_2 \,|\, x_1) = \sum_{i=1}^{n} p(x_2 \,|\, \mathcal{H}_i) p(\mathcal{H}_i \,|\, x_1).$$

Concept Question

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 $f(\theta|x)$?

1. Beta(2,5)

2. Beta(3,6)

3. Beta(6,8)

4. Beta(8,13)

We saw in the previous board question that 2 heads and 5 tails will update a beta(a, b) prior to a beta$(a + 2, b + 5)$ posterior.

<u>**answer:**</u> (4) beta$(8, 13)$.

Continuous priors, continuous data

Bayesian update tables with and without infinitesimals

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

hypoth.	prior	likeli.	unnormalized posterior	posterior
$\theta \pm \frac{d\theta}{2}$	$f(\theta)\,d\theta$	$f(x\,\vert\,\theta)\,dx$	$f(x\,\vert\,\theta)f(\theta)\,d\theta\,dx$	$f(\theta\,\vert\,x)\,d\theta = \dfrac{f(x\,\vert\,\theta)f(\theta)\,d\theta\,dx}{f(x)\,dx}$
total	1		$f(x)\,dx$	1

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

Normal prior, normal data

$\mathsf{N}(\mu, \sigma^2)$ has density

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-(y-\mu)^2/2\sigma^2}.$$

Observation:

The coefficient is a normalizing factor, so if we have a pdf

$$f(y) = c\mathrm{e}^{-(y-\mu)^2/2\sigma^2}$$

then

$$y \sim \mathsf{N}(\mu, \sigma^2)$$

and

$$c = \frac{1}{\sigma\sqrt{2\pi}}$$

Board question: normal prior, normal data

- $\mathrm{N}(\mu, \sigma^2)$ has pdf: $\quad f(y) = \dfrac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-(y-\mu)^2/2\sigma^2}$.
- Suppose our data follows a $\mathrm{N}(\theta, 4)$ distribution with unknown mean θ and variance 4. That is

$$f(x \mid \theta) = \text{pdf of } \mathrm{N}(\theta, 4)$$

- Suppose our prior on θ is $\mathrm{N}(3, 1)$.

Suppose we obtain data $x_1 = 5$.

1. Use the data to find the posterior pdf for θ.

Write out your tables clearly. Use (and understand) infinitesimals.

Solution

We have:

Prior: $\theta \sim \mathrm{N}(3,1)$: $\quad f(\theta) = c_1 \mathrm{e}^{-(\theta-3)^2/2}$

Likelihood $x \sim \mathrm{N}(\theta, 4)$: $\quad f(x \,|\, \theta) = c_2 \mathrm{e}^{-(x-\theta)^2/8}$

For $x = 5$ the likelihood is $c_2 \mathrm{e}^{-(5-\theta)^2/8}$

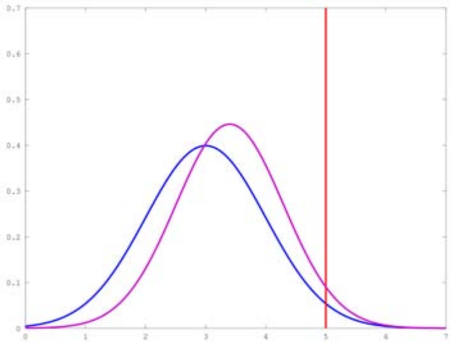
hypoth.	prior	likelihood	un. post.
$\theta \pm \frac{d\theta}{2}$	$c_1 \mathrm{e}^{-(\theta-3)^2/2}\, d\theta$	$c_2 \mathrm{e}^{-(5-\theta)^2/8}\, dx$	$c_3 \mathrm{e}^{-(\theta-3)^2/2} \mathrm{e}^{-(5-\theta)^2/8}\, d\theta\, dx$

A bit of algebraic manipulation of the unnormalized posterior gives

$$
\begin{aligned}
c_3 \mathrm{e}^{-(\theta-3)^2/2} \mathrm{e}^{-(5-\theta)^2/8}\, d\theta\, dx &= c_3 \mathrm{e}^{-\frac{5}{8}[\theta^2 - \frac{34}{5}\theta + 61]} = c_3 \mathrm{e}^{-\frac{5}{8}[(\theta-17/5)^2 + 61 - (17/5)^2]} \\
&= c_3 \mathrm{e}^{-\frac{5}{8}(61-(17/5)^2)} \mathrm{e}^{-\frac{5}{8}(\theta-17/5)^2} \\
&= c_4 \mathrm{e}^{-\frac{5}{8}(\theta-17/5)^2} = c_4 \mathrm{e}^{-\frac{(\theta-17/5)^2}{2\cdot\frac{4}{5}}}
\end{aligned}
$$

The last expression shows the posterior is $\mathrm{N}\left(\frac{17}{5}, \frac{4}{5}\right)$.

Solution graphs



prior = blue; posterior = purple; data = red

Data:	$x_1 = 5$	
Prior:	$\mu_{\text{prior}} = 3$;	$\sigma_{\text{prior}} = 1$
Posterior is normal	$\mu_{\text{posterior}} = 3.4$;	$\sigma_{\text{posterior}} = 0.894$

Board question: Romeo and Juliet

Romeo is always late. How late follows a uniform distribution uniform$(0, \theta)$ with unknown parameter θ in hours.

Juliet knows that $\theta \leq 1$ hour and she assumes a flat prior for θ on $[0, 1]$.

On their first date Romeo is 15 minutes late.

(a) find and graph the prior and posterior pdf's for θ

(b) find and graph the prior predictive and posterior predictive pdf's of how late Romeo will be on the second data (if he gets one!).

See next slides for solution

Solution

Parameter of interest: $\theta =$ upper bound on R's lateness.
Data: $x_1 = .25$.
Goals: (a) Posterior pdf for θ
(b) Predictive pdf's –requires pdf's for θ
In the update table we split the hypotheses into the two different cases $\theta < .25$ and $\theta \geq .25$:

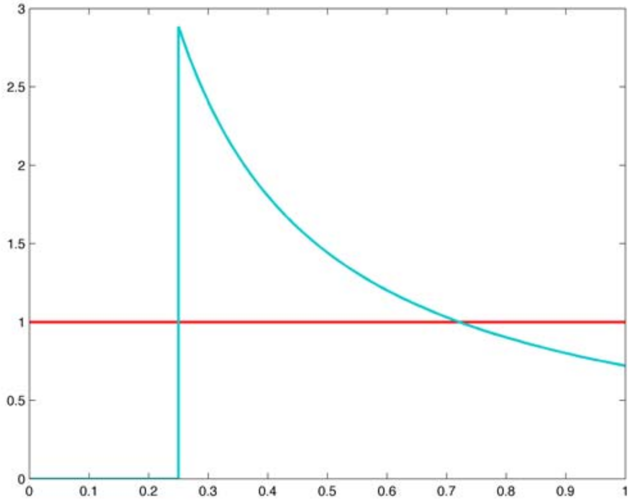
hyp.	prior $f(\theta)$	likelihood $f(x_1\|\theta)$	unnormalized posterior	posterior $f(\theta\|x_1)$
$\theta < .25$	$d\theta$	0	0	0
$\theta \geq .25$	$d\theta$	$\frac{1}{\theta}$	$\frac{d\theta}{\theta}$	$\frac{c}{\theta}\,d\theta$
Tot.	1		T	1

The normalizing constant c must make the total posterior probability 1, so

$$c\int_{.25}^{1}\frac{d\theta}{\theta} = 1 \;\Rightarrow\; c = \frac{1}{\ln(4)}.$$

Continued on next slide.

Solution continued



Prior and posterior pdf's for θ.

Solution continued

(b) Prior prediction: The likelihood function is a function of θ for fixed x_2

$$f(x_2|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } \theta \geq x_2 \\ 0 & \text{if } \theta < x_2 \end{cases}$$

Therefore the prior predictive pdf of x_2 is

$$f(x_2) = \int f(x_2|\theta) f(\theta)\, d\theta = \int_{x_2}^{1} \frac{1}{\theta}\, d\theta = -\ln(x_2).$$

continued on next slide

Solution continued

Posterior prediction:
The likelihood function is the same as before:

$$f(x_2|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } \theta \geq x_2 \\ 0 & \text{if } \theta < x_2. \end{cases}$$

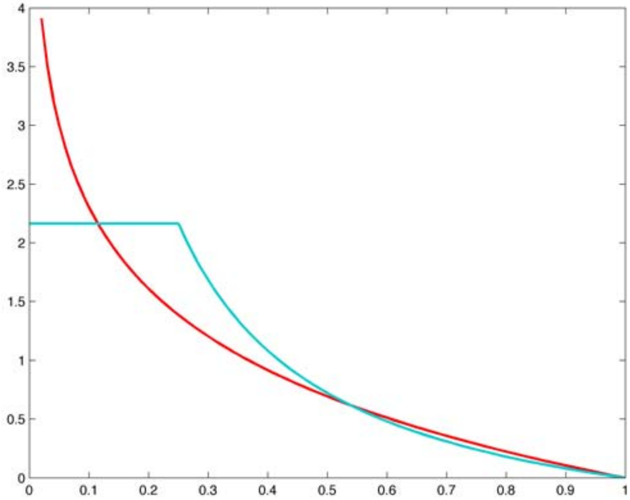
The posterior predictive pdf $f(x_2|x_1) = \int f(x_2|\theta)f(\theta|x_1)\,d\theta$. The integrand is 0 unless $\theta > x_2$ and $\theta > .25$. We compute it for the two cases:

$$\text{If } x_2 < .25: \qquad f(x_2|x_1) = \int_{.25}^{1} \frac{c}{\theta^2}\,d\theta = 3c = 3/\ln(4).$$

$$\text{If } x_2 \geq .25: \qquad f(x_2|x_1) = \int_{x_2}^{1} \frac{c}{\theta^2}\,d\theta = (\frac{1}{x_2} - 1)/\ln(4)$$

Plots of the predictive pdf's are on the next slide.

Solution continued



Prior (red) and posterior (blue) predictive pdf's for x_2

From discrete to continuous Bayesian updating

Bent coin with unknown probability of heads θ.

Data x_1: heads on one toss.

Start with a flat prior and update:

hyp.	prior	likelihood	unnormalized posterior	posterior
θ	$d\theta$	θ	$\theta\, d\theta$	$2\theta\, d\theta$
Total	1		$\int_0^1 \theta\, d\theta = 1/2$	1

Posterior pdf: $f(\theta \,|\, x_1) = 2\theta$.

Approximate continuous by discrete

- approximate the continuous range of hypotheses by a finite number of hypotheses.
- create the discrete updating table for the finite number of hypotheses.
- consider how the table changes as the number of hypotheses goes to infinity.

Chop $[0, 1]$ into 4 intervals

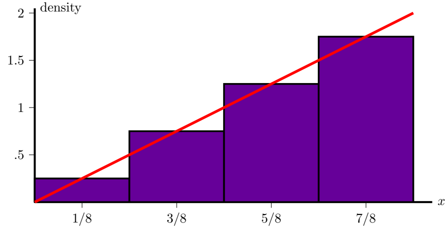
hypothesis	prior	likelihood	un. posterior	posterior
$\theta = 1/8$	$1/4$	$1/8$	$(1/4) \times (1/8)$	$1/16$
$\theta = 3/8$	$1/4$	$3/8$	$(1/4) \times (3/8)$	$3/16$
$\theta = 5/8$	$1/4$	$5/8$	$(1/4) \times (5/8)$	$5/16$
$\theta = 7/8$	$1/4$	$7/8$	$(1/4) \times (7/8)$	$7/16$
Total	1	–	$\sum_{i=1}^{n} \theta_i \, \Delta\theta$	1

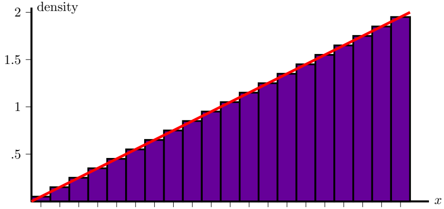
Chop [0, 1] into 12 intervals

hypothesis	prior	likelihood	un. posterior	posterior
$\theta = 1/24$	$1/12$	$1/24$	$(1/12) \times (1/24)$	$1/144$
$\theta = 3/24$	$1/12$	$3/24$	$(1/12) \times (3/24)$	$3/144$
$\theta = 5/24$	$1/12$	$5/24$	$(1/12) \times (5/24)$	$5/144$
$\theta = 7/24$	$1/12$	$7/24$	$(1/12) \times (7/24)$	$7/144$
$\theta = 9/24$	$1/12$	$9/24$	$(1/12) \times (9/24)$	$9/144$
$\theta = 11/24$	$1/12$	$11/24$	$(1/12) \times (11/24)$	$11/144$
$\theta = 13/24$	$1/12$	$13/24$	$(1/12) \times (13/24)$	$13/144$
$\theta = 15/24$	$1/12$	$15/24$	$(1/12) \times (15/24)$	$15/144$
$\theta = 17/24$	$1/12$	$17/24$	$(1/12) \times (17/24)$	$17/144$
$\theta = 19/24$	$1/12$	$19/24$	$(1/12) \times (19/24)$	$19/144$
$\theta = 21/24$	$1/12$	$21/24$	$(1/12) \times (21/24)$	$21/144$
$\theta = 23/24$	$1/12$	$23/24$	$(1/12) \times (23/24)$	$23/144$
Total	1	–	$\sum_{i=1}^{n} \theta_i \, \Delta\theta$	1

Density historgram

Density historgram for posterior pmf with 4 and 20 slices.





The original posterior pdf is shown in red.