Continuous Data with Continuous Priors
Class 14, 18.05, Spring 2014

Jeremy Orloff and Jonathan Bloom

1 Learning Goals

1. Be able to construct a Bayesian update table for continuous hypotheses and continuous data.

2. Be able to recognize the pdf of a normal distribution and determine its mean and variance.

2 Introduction

We are now ready to do Bayesian updating when both the hypotheses and the data take continuous values. The pattern is the same as what we've done before, so let's first review the previous two cases.

3 Previous cases

1. Discrete hypotheses, discrete data

Notation

- Hypotheses $\mathcal{H}$
- Data x
- Prior $P(\mathcal{H})$
- Likelihood $p(x \mid \mathcal{H})$
- Posterior $P(\mathcal{H} \mid x)$.

Example 1. Suppose we have data x and three possible explanations (hypotheses) for the data that we'll call A, B, C. Suppose also that the data can take two possible values, -1 and 1.

In order to use the data to help estimate the probabilities of the different hypotheses we need a prior pmf and a likelihood table. For this example we are only concerned with the formal process of of Bayesian updating so we can just make these up. We put them in the following tables.

hypothesis $\mathcal{H}$	prior $P(\mathcal{H})$
A	.1
B	.3
C	.6

Prior probabilities

hypothesis	likelihood $p(x \mid \mathcal{H})$	
$\mathcal{H}$	$x = -1$	$x = 1$
A	.2	.8
B	.5	.5
C	.7	.3

Likelihoods

Naturally, each entry in the likelihood table is a likelihood $p(x \mid \mathcal{H})$. For instance the .2 in row A and column $x = -1$ is the likelihood $p(x = -1 \mid A)$.

Question: Suppose we run one trial and obtain the data $x_1 = 1$. Use this to find the posterior probabilities for the hypotheses.

answer: The data picks out one column from the likelihood table which we then use in our Bayesian update table.

hypothesis	prior	likelihood	unnormalized posterior	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$p(x = 1 \mid \mathcal{H})$	$p(x \mid \mathcal{H})P(\mathcal{H})$	$P(\mathcal{H} \mid x) = \frac{p(x \mid \mathcal{H})P(\mathcal{H})}{p(x)}$
A	.1	.8	.08	.195
B	.3	.5	.15	.366
C	.6	.3	.18	.439
total	1		$p(x) = .41$	1

To summarize: the prior probabilities (of hypotheses) and likelihoods (of data given hypothesis) were given; the unnormalized posterior is the product of the prior and likelihood; the total probability $p(x)$ is the sum of the probabilities in the unnormalized posterior column; and we divide by $p(x)$ to normalize the unnormalized posterior.

2. Continuous hypotheses, discrete data

Now suppose that we have data x that can take a discrete set of values and a continuous parameter θ that determines the distribution the data is drawn from.

Notation

- Hypotheses θ
- Data x
- Prior $f(\theta)\, d\theta$
- Likelihood $p(x \mid \theta)$
- Posterior $f(\theta \mid x)\, d\theta$.

Note: Here we multiplied by $d\theta$ to express the prior and posterior as probabilities. As densities, we have the prior pdf $f(\theta)$ and the posterior pdf $f(\theta \mid x)$.

Example 2. Assume that $x \sim \text{Binomial}(5, \theta)$. So θ is in the range $[0, 1]$ and the data x can take six possible values, 0, 1, ..., 5.

Since there is a continuous range of values we use a pdf to describe the prior on θ. Let's suppose the prior is $f(\theta) = 2\theta$. We can still make a likelihood table, though it only has one row representing an arbitrary hypothesis θ.

hypothesis	likelihood $p(x \mid \mathcal{H})$					
	$x = 0$	$x = 1$	$x = 2$	$x = 3$	$x = 4$	$x = 5$
θ	$\binom{5}{0}(1-\theta)^5$	$\binom{5}{1}\theta(1-\theta)^4$	$\binom{5}{2}\theta^2(1-\theta)^3$	$\binom{5}{3}\theta^3(1-\theta)^2$	$\binom{5}{4}\theta^4(1-\theta)$	$\binom{5}{5}\theta^5$

Likelihoods

Question: Suppose we run one trial and obtain the data $x_1 = 2$. Use this to find the posterior pdf for the parameter (hypotheses) θ.

answer: As before, the data picks out one column from the likelihood table which we can use in our Bayesian update table. Since we want to work with probabilities we write $f(\theta)d\,\theta$ and $f(\theta \mid x_1)\,d\theta$ for the pdf's.

hypothesis	prior	likelihood	unnormalized posterior	posterior
$\theta \pm \frac{d\theta}{2}$	$f(\theta)\,d\theta$	$p(x = 2 \mid \theta)$	$p(x \mid \theta)f(\theta)\,d\theta$	$f(\theta \mid x)\,d\theta = \dfrac{p(x \mid \theta)f(\theta)\,d\theta}{p(x)}$
$\theta \pm \frac{d\theta}{2}$	$2\theta\,d\theta$	$\binom{5}{2}\theta^2(1-\theta)^3$	$2\binom{5}{2}\theta^3(1-\theta)^3\,d\theta$	$f(\theta \mid x)\,d\theta = \dfrac{3!\,3!}{7!}\theta^3(1-\theta)^3\,d\theta$
total	1		$p(x) = \int_0^1 2\binom{5}{2}\theta^2(1-\theta)^3\,d\theta = 2\binom{5}{2}\frac{3!\,3!}{7!}$	1

To summarize: the prior probabilities (of hypotheses) and likelihoods (of data given hypothesis) were given; the unnormalized posterior is the product of the prior and likelihood; the total probability $p(x)$ is the integral of the probabilities in the unnormalized posterior column; and we divide by $p(x)$ to normalize the unnormalized posterior.

4 Continuous hypotheses and continuous data

When both data and hypotheses are continuous, the only change to the previous example is that the likelihood function uses a pdf $f(x \mid \theta)$ instead of a pmf $p(x \mid \theta)$. The general shape of the Bayesian update table is the same.

Notation

- Hypotheses θ
- Data x
- Prior $f(\theta)d\theta$

- Likelihood $f(x\,|\,\theta)\,dx$
- Posterior $f(\theta\,|\,x)\,d\theta$.

Simplifying the notation. In the previous cases we included $d\theta$ so that we were working with probabilities instead of densities. When both data and hypotheses are continuous we will need both $d\theta$ and dx. This makes things conceptually simpler, but notationally cumbersome. To simplify the notation we will allow ourselves to drop $d\theta$ and dx in our tables.

For comparison, we first show the general table in simplified notation followed immediately afterward by the table showing the infinitesimals.

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

Bayesian update table without $d\theta$ and dx

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

Bayesian update table with $d\theta$ and dx

To summarize: the prior probabilities (of hypotheses) and likelihoods (of data given hypothesis) were given; the unormalized posterior is the product of the prior and likelihood; the total probability $f(x)\,dx$ is the integral of the probabilities in the unnormalized posterior column; we divide by $f(x)\,dx$ to normalize the unnormalized posterior.

5 Normal hypothesis, normal data

A standard example of continuous hypotheses and continuous data assumes that both the data and prior follow normal distributions. The following example assumes that the variance of the data is known.

Example 3. Suppose our data x is drawn from a normal distribution with unknown mean θ and standard deviation 1.

$$x \sim \mathrm{N}(\theta, 1)$$

Suppose further that our prior distribution on θ is $\theta \sim \mathrm{N}(2, 1)$.

Let x represent an arbitrary data value.

(a) Make a Bayesian table with prior, likelihood, and unnormalized posterior.

(b) Show that the posterior distribution for θ is normal as well.

(c) Find the mean and variance of the posterior distribution.

answer: A good compromise on the notation is to include $d\theta$ but not dx. The reason for this is that the total probability is computed by integrating over θ and the $d\theta$ reminds of us that.

Our prior pdf is

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

The likelihood function is

$$f(x \,|\, \theta) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-(x-\theta)^2/2}.$$

We know we are going to multiply the prior and the likelihood, so we carry out that algebra first. In the very last step we combine all the constant factors into one constant we call c_1.

$$\begin{aligned}
\text{prior} \cdot \text{likelihood} &= \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-(\theta-2)^2/2} \cdot \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-(x-\theta)^2/2} \\
&= \frac{1}{2\pi} \mathrm{e}^{-(2\theta^2-(4+2x)\theta+4+x^2)/2} \\
&= \frac{1}{2\pi} \mathrm{e}^{-(2(\theta-(2+x)\theta)+4+x^2)/2} \\
&= \frac{1}{2\pi} \mathrm{e}^{-(2(\theta-(1+x/2))^2-(1+x/2)^2+4+x^2)/2} \quad \text{(complete the square)} \\
&= \frac{1}{2\pi} \mathrm{e}^{-(-(1+x/2)^2+4+x^2)/2} \; \mathrm{e}^{-(2(\theta-(1+x/2))^2)/2} \\
&= c_1 \mathrm{e}^{-(\theta-(1+x/2))^2}
\end{aligned}$$

Remember that the data x will be a known value when we do the updating so we can consider it a constant, so $\mathrm{e}^{-(-(1+x/2)^2+4+x^2)}$ is also just a constant.

hypothesis	prior	likelihood	unnormalized posterior	posterior
$\theta \pm \frac{d\theta}{2}$	$f(\theta)\, d\theta$	$f(x \,\|\, \theta)$	$f(x \,\|\, \theta) f(\theta)\, d\theta$	$f(\theta \,\|\, x)\, d\theta = \dfrac{f(x \,\|\, \theta) f(\theta)\, d\theta}{f(x)}$
$\theta \pm \frac{d\theta}{2}$	$\frac{1}{\sqrt{2\pi}} \mathrm{e}^{-(\theta-2)^2/2}\, d\theta$	$\frac{1}{\sqrt{2\pi}} \mathrm{e}^{-(x-\theta)^2/2}$	$c_1 \mathrm{e}^{-(\theta-(1+x/2))^2}$	$c_2 \mathrm{e}^{-(\theta-(1+x/2))^2}$
total	1		$f(x) = \int f(x \,\|\, \theta) f(\theta)\, d\theta$	1

We can see by the form of the posterior pdf that it must be a normal distribution. Therefore we don't need to bother computing the total probability; it is just used for normalization and we already know the normalization constant $\frac{1}{\sigma\sqrt{2\pi}}$ for a normal distribution.

Indeed, looking at the posterior we see that it must come from a normal distribution with mean $\mu_{\text{post}} = 1 + x/2$ and variance $\sigma^2_{\text{post}} = 1/2$.

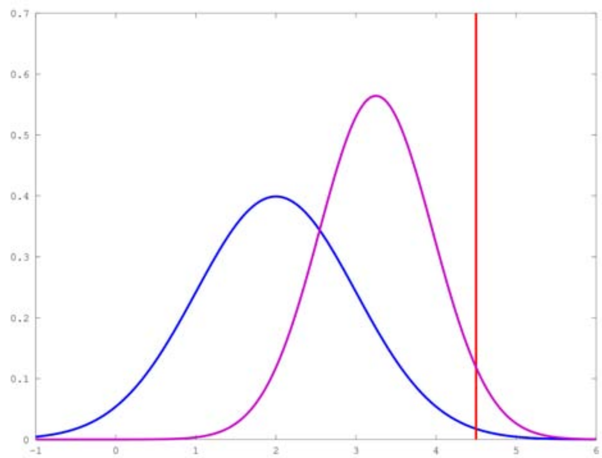
To see the posterior mean and variance are what we claim we write

$$c_2 e^{-(\theta-(1+x/2))^2} = c_2 e^{-\frac{(\theta-(1+x/2))^2}{2(1/2)}}$$

Matching this with the form $c_2 e^{-(\theta-\mu)^2/2\sigma^2}$ we see that the posterior

$$f(\theta \mid x) \sim \text{N}(1 + x/2, 1/2).$$

Here is the graph of the prior and the posterior pdf's when the data $x = 4.5$. Note how the data 'pulled' the prior towards the data.



prior = blue; posterior = purple; data = red

18.05 Introduction to Probability and Statistics
Spring 2014