

# Continuous Data with Continuous Priors

## Class 14, 18.05

### 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  $p(x | \mathcal{H})$
- Posterior  $P(\mathcal{H} | x)$ .

**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. Assume the prior and likelihoods are given in the following table. (For this example we are only concerned with the formal process of Bayesian updating. So we just made up the prior and likelihoods.)

hypothesis $\mathcal{H}$	prior $P(\mathcal{H})$
A	0.1
B	0.3
C	0.6

Prior probabilities

hypothesis $\mathcal{H}$	likelihood $p(x   \mathcal{H})$	
	$x = -1$	$x = 1$
A	0.2	0.8
B	0.5	0.5
C	0.7	0.3

Likelihoods

Naturally, each entry in the likelihood table is a likelihood  $p(x | \mathcal{H})$ . For instance the 0.2 row  $A$  and column  $x = -1$  is the likelihood  $p(x = -1 | A)$ .

**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 $\mathcal{H}$	prior $P(\mathcal{H})$	likelihood $p(x = 1   \mathcal{H})$	Bayes	
			numerator $p(x   \mathcal{H})P(\mathcal{H})$	posterior $P(\mathcal{H}   x) = \frac{p(x   \mathcal{H})P(\mathcal{H})}{p(x)}$
$A$	0.1	0.8	0.08	0.195
$B$	0.3	0.5	0.15	0.366
$C$	0.6	0.3	0.18	0.439
total	1		$p(x) = 0.41$	1

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

## 2. Continuous hypotheses, discrete data

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

### Notation

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

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  $f(\theta | x)$ .

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

Since there is a continuous range of values we use a pdf to describe the prior on  $\theta$ . 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  $\theta$ .

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

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)  $\theta$ .

**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  $f(\theta | x_1) d\theta$  for the pdf's.

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

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

## 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  $\phi(x | \theta)$  instead of a pmf  $p(x | \theta)$ . The general shape of the Bayesian update table is the same.

### Notation

- **Hypotheses  $\theta$ .** When the hypothesis is a real number, it may be more realistic to hypothesize that “the parameter is in a small interval of size  $d\theta$  around  $\theta$ .”
- **Data  $x$ .** When the data is a real number, it may be more realistic to say “the data is in a small interval of size  $dx$  around  $x$ .”

- **Prior**  $f(\theta)d\theta$ . This is our initial belief about the probability that the parameter is in a small interval of size  $d\theta$  around  $\theta$ .
- **Likelihood**  $\phi(x|\theta)dx$ . This is the (calculated) probability that the data is in a small interval of size  $dx$  around  $x$ , ASSUMING the hypothesis  $\theta$ .
- **Posterior**  $\phi(\theta|x)d\theta$ . This is the (calculated) probability that the parameter is in a small interval of size  $d\theta$  around  $\theta$ , ASSUMING the data  $x$ .

**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 **sometimes allow ourselves to drop  $dx$  in our tables**. This is fine because the data  $x$  is fixed in each calculation. We keep the  $d\theta$  because the hypothesis  $\theta$  is allowed to vary.

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

Bayes				
hypothesis	prior	likelihood	numerator	posterior
$\theta$	$f(\theta) d\theta$	$\phi(x \theta)$	$\phi(x \theta)f(\theta) d\theta$	$f(\theta x) d\theta = \frac{\phi(x \theta)f(\theta) d\theta}{\phi(x)}$
total (int. over $\theta$ )	1		$\phi(x) = \int \phi(x \theta)f(\theta) d\theta$ = prior prob. density for data $x$	1

Bayesian update table without  $dx$ 

Bayes				
hypothesis	prior	likelihood	numerator	posterior
$\theta$	$f(\theta) d\theta$	$\phi(x \theta) dx$	$\phi(x \theta)f(\theta) d\theta dx$	$f(\theta x) d\theta = \frac{\phi(x \theta)f(\theta) d\theta dx}{\phi(x) dx}$ $= \frac{\phi(x \theta)f(\theta) d\theta}{\phi(x)}$
total	1		$\phi(x) dx = (\int \phi(x \theta)f(\theta) d\theta) dx$	1

Bayesian update table with  $d\theta$  and  $dx$ 

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

## 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 we have data  $x = 5$  which was drawn from a normal distribution with unknown mean  $\theta$  and standard deviation 1.

$$x \sim N(\theta, 1)$$

Suppose further that our prior distribution for  $\theta$  is  $\theta \sim N(2, 1)$ .

Let  $x$  represent an arbitrary data value.

- Make a Bayesian table with prior, likelihood, and Bayes numerator.
- Show that the posterior distribution for  $\theta$  is normal as well.
- Find the mean and variance of the posterior distribution.

**answer:** As we did with the tables above, 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  $\theta$  and the  $d\theta$  reminds of us that.

Our prior pdf is

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

The likelihood function is

$$\phi(x = 5 | \theta) = \frac{1}{\sqrt{2\pi}} e^{-(5-\theta)^2/2}.$$

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 give the complicated constant factor the name  $c_1$ .

$$\begin{aligned} \text{prior} \cdot \text{likelihood} &= \frac{1}{\sqrt{2\pi}} e^{-(\theta-2)^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-(5-\theta)^2/2} \\ &= \frac{1}{2\pi} e^{-(2\theta^2 - 14\theta + 29)/2} \\ &= \frac{1}{2\pi} e^{-(\theta^2 - 7\theta + 29/2)} \quad (\text{complete the square}) \\ &= \frac{1}{2\pi} e^{-((\theta-7/2)^2 + 9/4)} \\ &= \frac{e^{-9/4}}{2\pi} e^{-(\theta-7/2)^2} \\ &= c_1 e^{-(\theta-7/2)^2} \end{aligned}$$

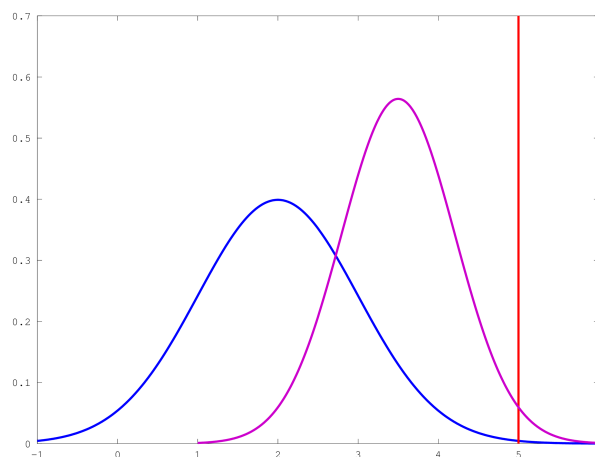
In the last step we named the complicated constant factor  $c_1$ .

hypothesis	prior	likelihood	Bayes numerator	posterior $f(\theta   x = 5) d\theta$
$\theta$	$f(\theta) d\theta$	$\phi(x = 5   \theta)$	$\phi(x = 5   \theta)f(\theta) d\theta$	$\frac{\phi(x = 5   \theta)f(\theta) d\theta}{\phi(x = 5)}$
$\theta$	$\frac{1}{\sqrt{2\pi}}e^{-(\theta-2)^2/2} d\theta$	$\frac{1}{\sqrt{2\pi}}e^{-(5-\theta)^2/2}$	$c_1e^{-(\theta-7/2)^2}$	$c_2e^{-(\theta-7/2)^2}$
total	1	$\phi(x = 5) = \int \phi(x = 5   \theta)f(\theta) d\theta$		1

We can see by the form of the posterior pdf that it is a normal distribution. Because the exponential for a normal distribution is  $e^{-(\theta-\mu)^2/2\sigma^2}$  we have mean  $\mu = 7/2$  and  $2\sigma^2 = 1$ , so variance  $\sigma^2 = 1/2$ .

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.

Here is the graph of the prior and the posterior pdf's for this example. Note how the data 'pulls' the prior (the wider bell on the left) towards the data: the posterior is the narrower bell on the right. After collecting data, we have a new opinion about the mean, and we are surer of this new opinion.



prior = blue; posterior = purple; data = red

Now we'll repeat the previous example for general  $x$ . When reading this if you mentally substitute 5 for  $x$  you will understand the algebra.

**Example 4.** Suppose our data  $x$  is drawn from a normal distribution with unknown mean  $\theta$  and standard deviation 1.

$$x \sim N(\theta, 1)$$

**answer:** As before, we show the algebra used to simplify the Bayes numerator: The prior pdf and likelihood function are

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

The Bayes numerator is the product of the prior and the likelihood:

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

Just as in the previous example, in the last step we replaced all the constants, including the exponentials that just involve  $x$ , by the simple constant  $c_1$ .

Now the Bayesian update table becomes

hypothesis	prior	likelihood	Bayes numerator	posterior $f(\theta   x) d\theta$
$\theta$	$f(\theta) d\theta$	$\phi(x   \theta)$	$\phi(x   \theta) f(\theta) d\theta$	$\frac{\phi(x   \theta) f(\theta) d\theta}{\phi(x)}$
$\theta$	$\frac{1}{\sqrt{2\pi}} e^{-(\theta-2)^2/2} d\theta$	$\frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2}$	$c_1 e^{-(\theta - (1+x/2))^2}$	$c_2 e^{-(\theta - (1+x/2))^2}$
total	1		$\phi(x) = \int \phi(x   \theta) f(\theta) d\theta$	1

As in the previous example we can see by the form of the posterior that it must be a normal distribution with mean  $1 + x/2$  and variance  $1/2$ . (Compare this with the case  $x = 5$  in the previous example.)

## 6 Predictive probabilities

Since the data  $x$  is continuous it has prior and posterior predictive pdfs. The [prior predictive pdf](#) is the total probability density computed at the bottom of the Bayes numerator column:

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

where the integral is computed over the entire range of  $\theta$ .

The [posterior predictive pdf](#) has the same form as the prior predictive pdf, except it uses the posterior probabilities for  $\theta$ :

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

As usual, we usually assume  $x_1$  and  $x_2$  are [conditionally independent](#). That is,

$$\phi(x_2|\theta, x_1) = \phi(x_2|\theta).$$

In this case the formula for the posterior predictive pdf is a little simpler:

$$\phi(x_2|x_1) = \int \phi(x_2|\theta)f(\theta|x_1) d\theta.$$