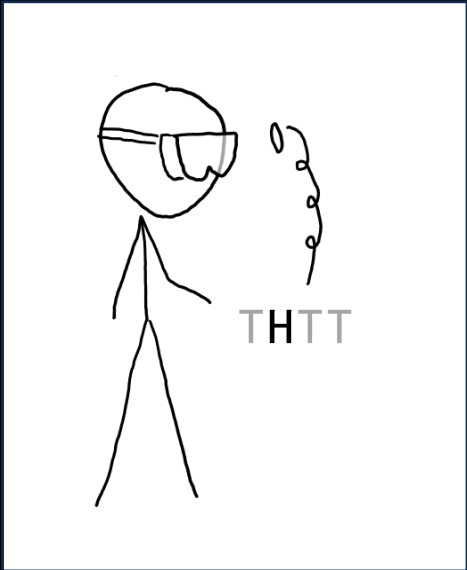


# Smoke testing our epistemology [theory of knowledge & evidence]



I tossed four coins and got one head.

What is it reasonable for me to infer about the probability of heads (call it  $\theta$ )?

- ~~"My estimate is  $\hat{\theta} = 25\%$ , thus the true probability of heads is 25%"~~ (if I tossed billions more coins that's the fraction of heads I'd see) **unjustified!**
- ~~"All we know for certain is that  $0 < \theta < 1$ "~~ **logical, but useless!**
- ???

What properties would we want, from *any* sane epistemology?

One sensible property: it should give us a way to distinguish between

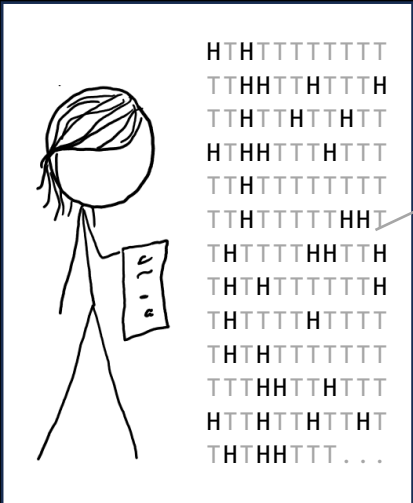
- uncertainty about the "laws of nature" (called **epistemic** uncertainty)
- & uncertainty about outcomes (called **aleatoric** uncertainty)



I tossed four coins  
and got one head.

Using a  $\text{Bin}(n, \theta)$  model I estimate the  
probability of heads is  $\hat{\theta} = 25\%$

very unsure about  $\theta$ :  
high *epistemic* uncertainty



I tossed four million coins  
and got one million heads.

Using a  $\text{Bin}(n, \theta)$  model, I estimate the  
probability of heads is  $\hat{\theta} = 25\%$

pretty confident about  $\theta$ :  
low *epistemic* uncertainty

even if we knew  $\theta$  exactly, there's still  
uncertainty about the next coin toss:  
this is *aleatoric* uncertainty



In Bayesianism we use a random variable to express our beliefs about  $\theta$ :

We use  $\Pr_{\Theta}$  to represent epistemic uncertainty,  $\Pr_X(x|\Theta)$  for aleatoric uncertainty.

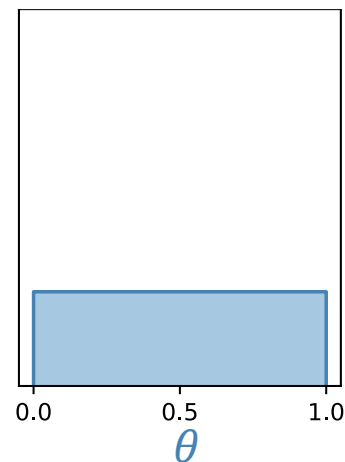


$$\Theta \sim U[0,1]$$



$$X \sim \text{Bin}(4, \Theta)$$

prior belief  
 $\Pr_{\Theta}(\theta)$

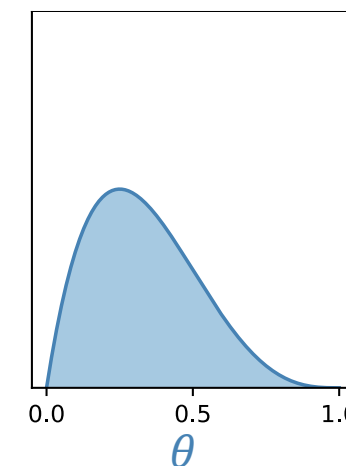


+

data

→

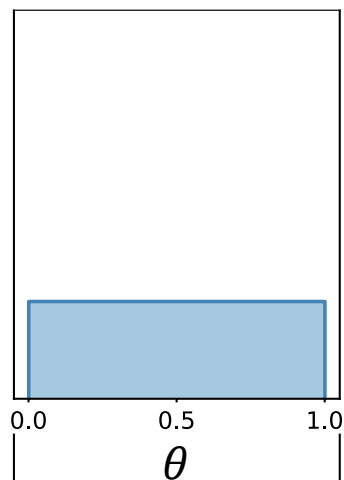
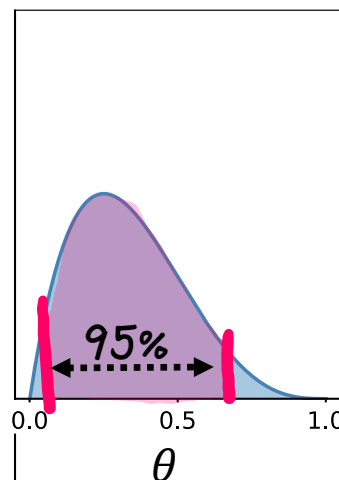
posterior belief  
 $\Pr_{\Theta}(\theta|X=1)$





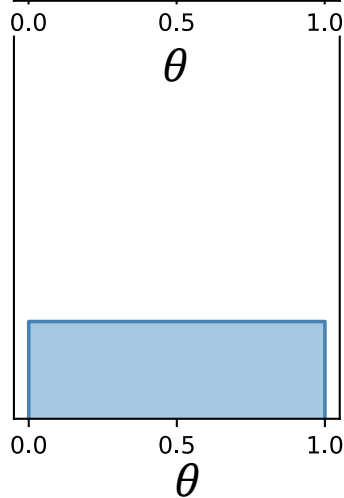
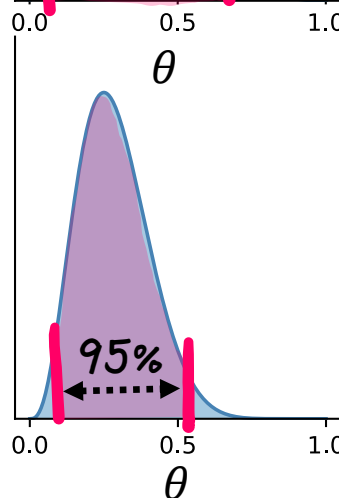
The posterior distribution depends on the amount of evidence we've seen.  
It quantifies our epistemic uncertainty.

prior belief  $\Pr_{\Theta}(\theta)$  + data  $x$   $\rightarrow$  posterior belief  $\Pr_{\Theta}(\theta|X = x)$



$\mathbb{P}(\Theta \in [0.053, 0.716] \mid \text{data}) = 95\%$   
"I'm 95% confident that  $\Theta \in [5.3\%, 71.6\%]$ "

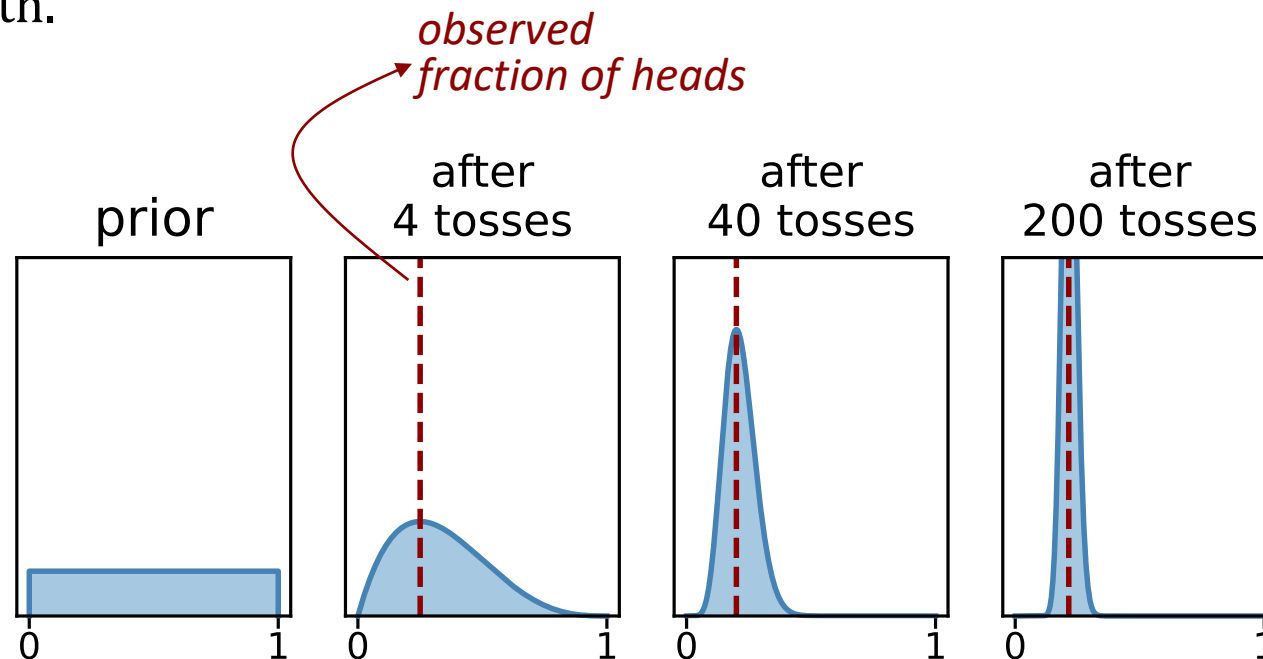
more data leads to  
lower epistemic  
uncertainty



$\mathbb{P}(\Theta \in [0.091, 0.538] \mid \text{data}) = 95\%$   
"I'm 95% confident that  $\Theta \in [9.1\%, 53.8\%]$ "



Typically, the more data we have, the closer our posterior gets to the truth.





We *must* have a prior belief about every unknown parameter. We *must* choose it before seeing the dataset in question.



$$\Theta \sim U[0,1]$$
$$\downarrow$$
$$X \sim \text{Bin}(4, \Theta)$$

But where does the prior come from?

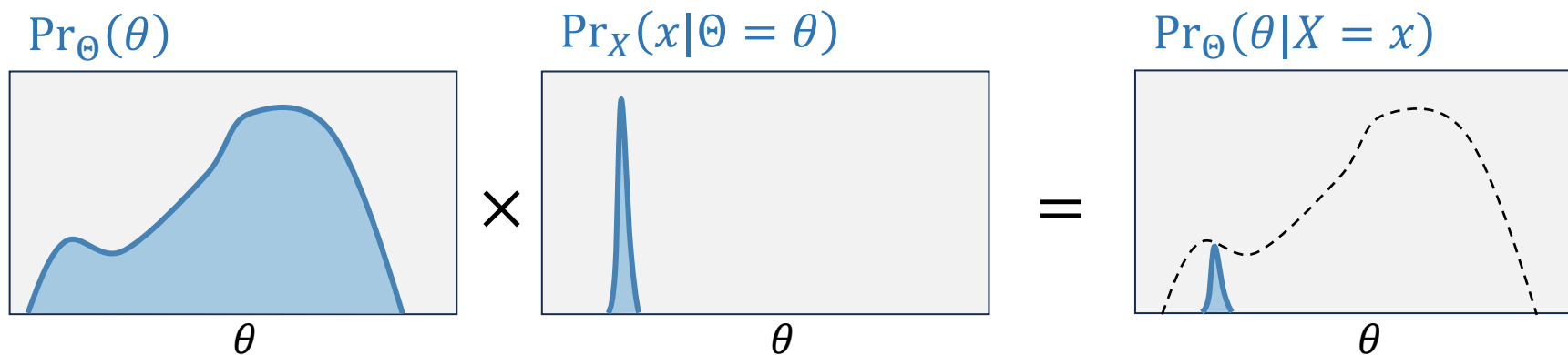
It comes from what you know already — it's how you can integrate your existing knowledge into your modelling.

If you don't have a prior to start off with, you have no business even thinking about the experiment!



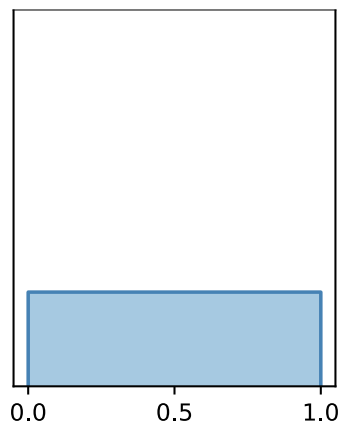
Often, with lots of data, the prior doesn't make much difference.

$$\Pr_{\Theta}(\theta|X = x) = \frac{\kappa \Pr_{\Theta}(\theta) \Pr_X(x|\Theta = \theta)}{\Pr_X(x)} \text{ constant (with respect to } \theta \text{)}$$

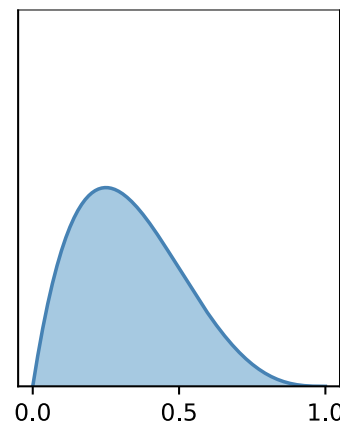
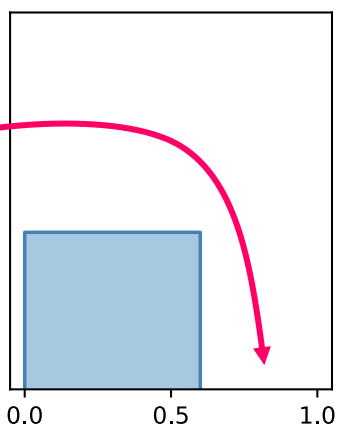
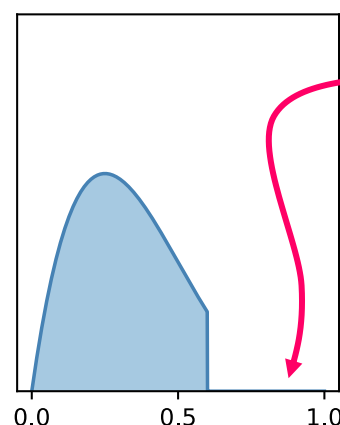


You are entitled to your own personal prior beliefs.  
They are entirely your choice.

$\Pr_{\Theta}(\theta)$



$\Pr_{\Theta}(\theta|X = x) = \kappa \Pr_{\Theta}(\theta) \Pr_X(x|\Theta = \theta)$

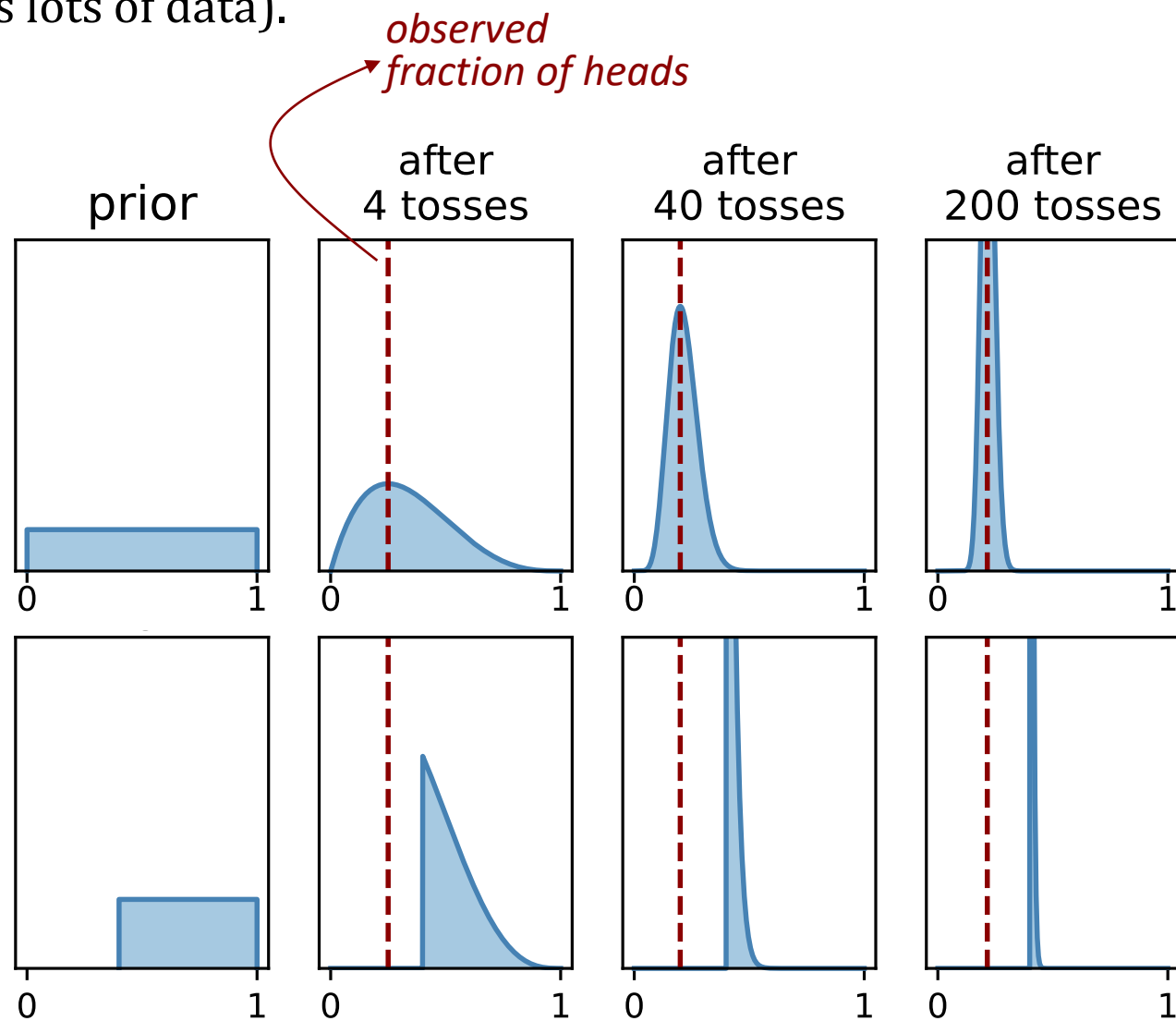


Preconception  
that  $\theta > 0.6$  is  
impossible

The preconception  
is unshakeable



If your prior is extreme, it will be reflected in your posterior (even if there's lots of data).



# Rationalist community

2 languages

Article Talk

Read Edit View history

From Wikipedia, the free encyclopedia

*This article is about the 21st-century movement. For the philosophical concept, see [Rationalism](#). For other uses, see [Rationalism \(disambiguation\)](#).*

The **rationalist community** is a 21st-century movement that formed around a group of internet blogs, primarily [LessWrong](#) and *Astral Codex Ten* (formerly known as *Slate Star Codex*). The movement initially gained prominence in the [San Francisco Bay Area](#). Its members seek to use insights from various academic fields (such as [epistemology](#), [probability theory](#), and [psychology](#)) and awareness of [cognitive biases](#) to achieve both epistemic rationality (coming to true beliefs about the world) and instrumental rationality (acting in a way to achieve one's objectives). Heavily overlapping communities include [effective altruism](#), [transhumanism](#), and [AI safety](#) (specifically mitigation of [AI extinction risk](#)).

The borders of the rationalist community are blurry and subject to debate among the community and adjacent groups.<sup>[1]</sup> Members who diverge from typical rationalist beliefs often self-describe as "rationalist-adjacent" or "post-rationalist" (also known as "ingroup" and "TPOT", an [acronym](#) for "this part of [Twitter](#)"<sup>[2]</sup>).

## Description [edit]

### Rationality [edit]

Rationalists define rationality to include epistemic rationality (coming to true beliefs about the world) and instrumental rationality (acting in a way to achieve one's objectives).<sup>[4]</sup>

The rationalists are concerned with applying science and probability to various topics, with [Bayesian inference](#).<sup>[5][6]</sup> According to [Ellen Huet](#), the rationalist community "aim[s] to keep even when the conclusions are scary".<sup>[3]</sup>



Bayesianism is an epistemology, i.e. a theory of knowledge and evidence, that passes all these “smoke tests”.

Surely (some say) it’s how any rational person should think!

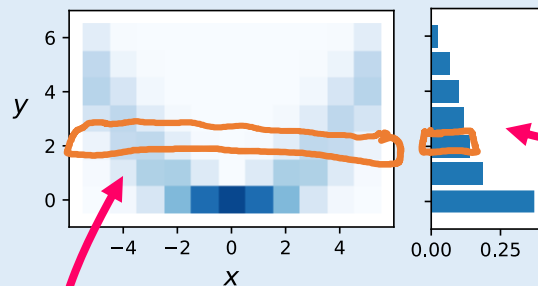
# How to apply the Bayesian update in practice

$$\Pr_X(x|Y=y) = \frac{\Pr_X(x) \Pr_Y(y|X=x)}{\Pr_Y(y)} \text{ when } \Pr_Y(y) > 0$$

First, what do these "conditional likelihoods" even mean?

# Conditional random variables

```
def rxy():
    x = np.random.randint(low=-5, high=6) # from -5 to +5 inclusive
    y = np.random.binomial(n=6, p=(x/6)**2)
    return (x,y)
```



The joint pmf of  $(X, Y)$

$$\Pr_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$$

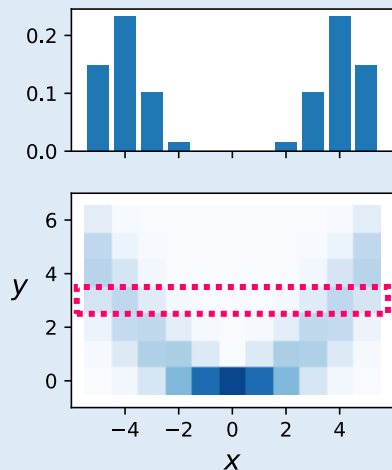
The marginal of  $Y$

$$\Pr_Y(y) = \mathbb{P}(Y = y)$$

$$= \sum_x \mathbb{P}(X = x, Y = y) \quad \text{by the Sum Rule}$$

$$= \sum_x \Pr_{X,Y}(x, y)$$

# Conditional random variables



```
def rxy():
    x = np.random.randint(low=-5, high=6) # from -5 to +5 inclusive
    y = np.random.binomial(n=6, p=(x/6)**2)
    return (x,y)
```

**QUESTION. What is  $X$  conditional on  $Y = 3$ ?**

$$\mathbb{P}(X = x | Y = 3) = \frac{\mathbb{P}(X = x, Y = 3)}{\mathbb{P}(Y = 3)} = \frac{\Pr_{X,Y}(x, 3)}{\Pr_Y(3)}$$

i.e. take the  $Y=3$  row,  
then rescale it to sum to 1

We can think of “ $X$  conditional on  $Y = 3$ ”  
as a random variable ...

It has a sample space, as usual:

$$\Omega = \{-5, -4, \dots, 4, 5\}$$

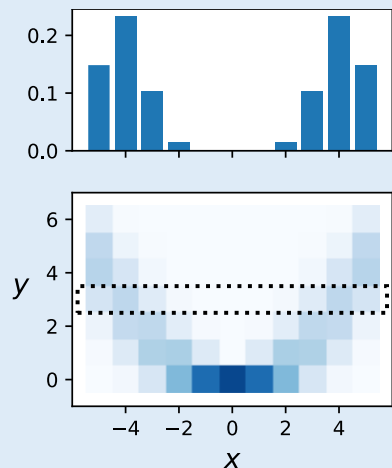
It has a valid probability mass function:

$$\text{pmf}_3(x) := \Pr_{X,Y}(x, 3) / \Pr_Y(3)$$

$$\text{pmf}_3(x) \geq 0 \text{ for all } x$$

$$\sum_x \text{pmf}_3(x) = 1$$

# Conditional random variables



```
def rxy():
    x = np.random.randint(low=-5, high=6) # from -5 to +5 inclusive
    y = np.random.binomial(n=6, p=(x/6)**2)
    return (x,y)
```

**We've specified a probability distribution.**

It's called a conditional random variable, and written  $(X|Y = y)$ .

We specified it by declaring its likelihood:

$$\Pr_{X|Y=y}(x) := \frac{\Pr_{X,Y}(x, y)}{\Pr_Y(y)}$$

This likelihood is also written  $\Pr_X(x|Y = y)$ .

# How to apply the Bayesian update in practice

$$\Pr_{X|Y=y}(x) = \kappa \frac{\Pr_X(x) \Pr_{Y|X=x}(y)}{\Pr_Y(y)} \quad \text{when } \Pr_Y(y) > 0$$

In practice, we just need to get the  $x$ -dependency right, and then the constant  $\kappa$  must be whatever is needed for  $\Pr_{X|Y=y}(x)$  to sum/integrate to 1, over  $x$ .

### Exercise.

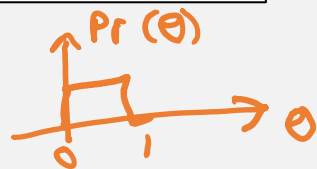
Consider the pair of random variables  $(\Theta, X)$  where

$$\Theta \sim U[0,1],$$

$$X \sim \text{Bin}(4, \Theta)$$

Find the distribution of  $(\Theta|X=1)$ .

$$\Pr_{\Theta}(\theta) = 1 \quad \text{for } 0 \leq \theta \leq 1$$



$$\Pr_X(x|\Theta = \theta) = \binom{4}{x} \theta^x (1-\theta)^{4-x}$$

$$\Pr_{\Theta}(\theta|X=1) = \kappa \Pr_{\Theta}(\theta) \Pr_X(1|\Theta = \theta)$$

$$= \kappa \times 1 \times \binom{4}{1} \theta^1 (1-\theta)^3 = \kappa' \theta (1-\theta)^3$$

annalgamare the non- $\theta$  term

Normalize:

$$\int_{\theta=0}^1 \kappa' \theta (1-\theta)^3 d\theta = 1 \Rightarrow \kappa' = \frac{1}{\int_{\theta=0}^1 \theta (1-\theta)^3 d\theta}.$$

### Exercise.

Consider the pair of random variables  $(\Theta, X)$  where

$$\Theta \sim U[0,1], \quad X \sim \text{Bin}(4, \Theta)$$

Find the distribution of  $(\Theta|X = 1)$ .

### Beta

| Probability density function |                                                                                                                                                                                                             |
|------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Notation                     | Beta( $\alpha, \beta$ )                                                                                                                                                                                     |
| PDF                          | $\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$ <p>where <math>B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}</math> and <math>\Gamma</math> is the Gamma function.</p> |

this is a standard pdf

$$\Pr_{\Theta}(\theta|X = 1) = \kappa \Pr_{\Theta}(\theta) \Pr_X(1|\Theta = \theta)$$

$$= \kappa' \theta (1-\theta)^3$$

$$= \underbrace{\kappa' B(\alpha, \beta)}_{\text{so this constant must be 1 (otherwise this pdf wouldn't integrate to 1 wr.t. } \theta)}$$

$$\frac{\theta^{\alpha-1} (1-\theta)^{\beta-1}}{B(\alpha, \beta)}$$

where  $\alpha = 2, \quad \beta = 4$

Thus  $(\Theta|X=1) \sim \text{Beta}(\alpha=2, \beta=4)$

### Exercise.

Consider the pair of random variables  $(\Theta, X)$  where

$$\Theta \sim U[0,1], \quad X \sim \text{Bin}(4, \Theta)$$

Find the distribution of  $(\Theta|X = 1)$ .

*In practice:*

$$\Pr_{\Theta}(\theta|X = 1) = \kappa \Pr_{\Theta}(\theta) \Pr_X(1|\Theta = \theta)$$

$$\propto \theta (1-\theta)^3,$$

This  $\theta$ -dependence matches that for  $\text{Beta}(2, 4)$ .

Thus the posterior is  $(\Theta|X=1) \sim \text{Beta}(2, 4)$ .

### Beta

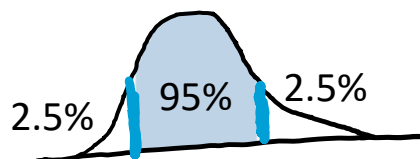
| Probability density function |                                                                                                                                                                                                             |
|------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Notation                     | Beta( $\alpha, \beta$ )                                                                                                                                                                                     |
| PDF                          | $\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$ <p>where <math>B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}</math> and <math>\Gamma</math> is the Gamma function.</p> |

# How should we report the posterior distribution?

Prior distribution for  $\Theta$



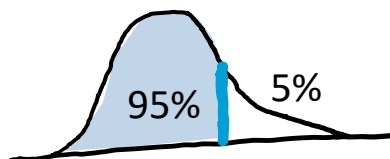
Posterior distribution for  $\Theta$



We could report a *95% confidence interval*  $[l_o, h_i]$  such that

$$\mathbb{P}(\Theta < l_o \mid \text{data}) = 2.5\%$$

$$\mathbb{P}(\Theta > h_i \mid \text{data}) = 2.5\%$$



or indeed any other 95% confidence interval e.g.

$$l_o = -\infty$$

$$\mathbb{P}(\Theta > h_i \mid \text{data}) = 5\%$$

(though for discrete  $\Theta$  we might not be able to hit the probabilities exactly, so it's generally better to find other ways to report the posterior)

prior belief

$$\Pr_{\Theta}(\theta)$$

+

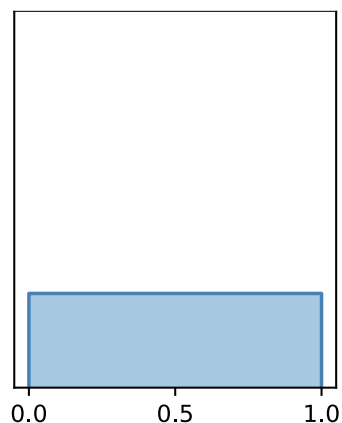
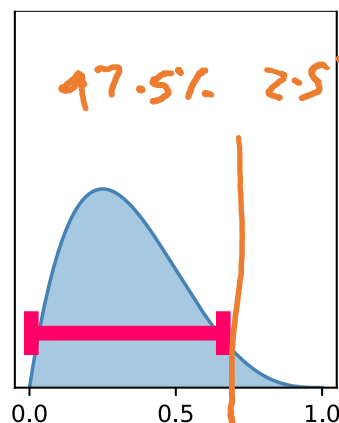
data

 $x$ 

→

posterior belief

$$\Pr_{\Theta}(\theta|X = x)$$

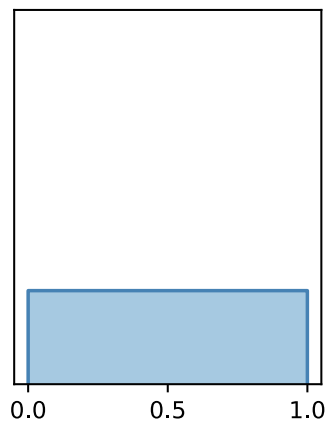
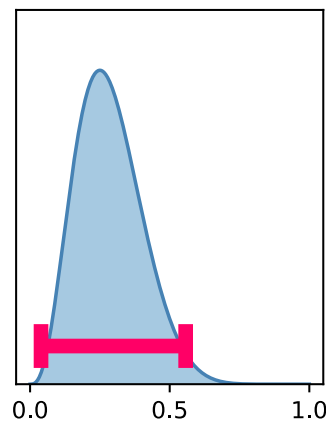


Posterior (after 1 head from 4 tosses) is  
 $(\Theta|X = 1) \sim \text{Beta}(2, 4)$

My 95% confidence interval is

```
D = scipy.stats.beta(a=2, b=4)
(lo, hi) = (D.ppf(.025), D.ppf(.975)) # (.053, .716)
```

*D.ppf(.975) returns "the value such that 97.5% of D's probability mass is below it"*



```
D = scipy.stats.beta(a=4, b=10)
(lo, hi) = (D.ppf(.025), D.ppf(.975)) # (.091, .538)
```

### Exercise 5.2.3 (Bayesian classification)

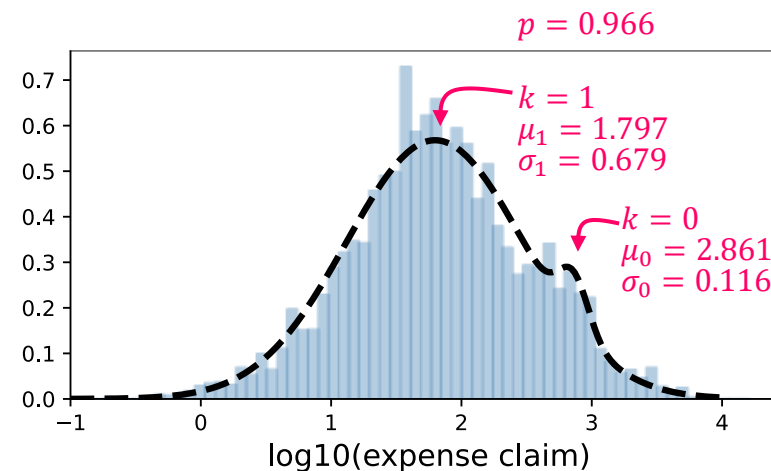
In a dataset of MP expense claims, let  $y_i$  be  $\log_{10}$  of the claim amount in record  $i$ . A histogram of the  $y_i$  suggests we use a Gaussian mixture model with two components,

$$K = \begin{cases} 1 & \text{with prob } p \\ 0 & \text{with prob } 1 - p \end{cases}$$

$$Y \sim \text{Normal}(\mu_K, \sigma_K^2)$$

$K$   
↓  
 $Y$

Find the probability that a claim amount £5000 belongs to the component  $k = 0$ .



In other words, find  $\mathbb{P}(K = 0 | Y = y)$ , where  $y = \log_{10} 5000 = 3.70$ . (The unknown parameter  $K$  is binary, so a confidence interval would be no use.)

$$\Pr_K(k) = \begin{cases} p & \text{if } k=1 \\ 1-p & \text{if } k=0 \end{cases} = p^k (1-p)^{1-k}$$

$$\Pr_{Y|K=k}(y) = \Phi_{\mu_k, \sigma_k}(y) \quad \text{where} \quad \Phi_{\mu, \sigma}(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

$$\Pr_{K|Y=y}(k) = \kappa \Pr_K(k) \Pr_{Y|K=k}(y)$$

$$= \kappa \begin{cases} p \Phi_{\mu_1, \sigma_1}(y) & \text{if } k=1 \\ (1-p) \Phi_{\mu_0, \sigma_0}(y) & \text{if } k=0. \end{cases}$$

Look at this as a function of  $k$ . Is it a well-known function? If not, solve for  $\kappa$ :

$$\Rightarrow \kappa = \dots$$

$$\Rightarrow \mathbb{P}(K=0 | Y=y) = \Pr_{K|Y=y}(0) = \frac{(1-p) \Phi_{\mu_0, \sigma_0}(y)}{(1-p) \Phi_{\mu_0, \sigma_0}(y) + p \Phi_{\mu_1, \sigma_1}(y)}$$