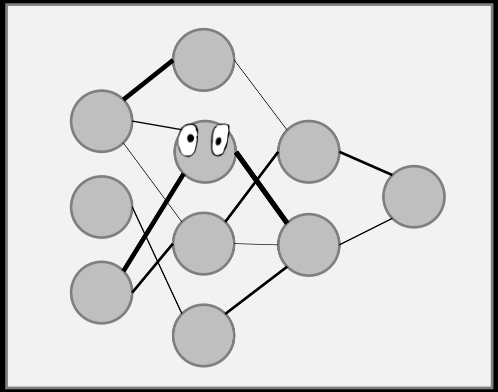
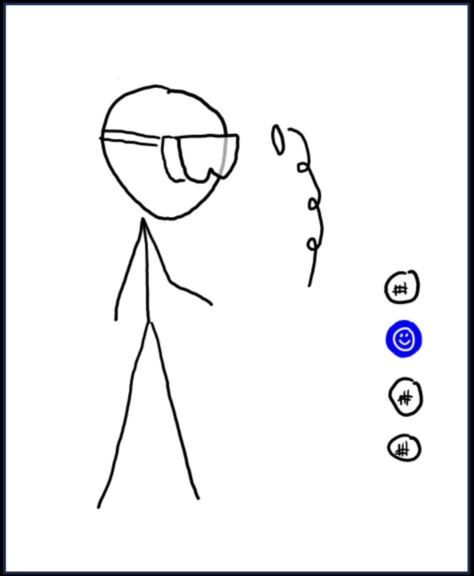
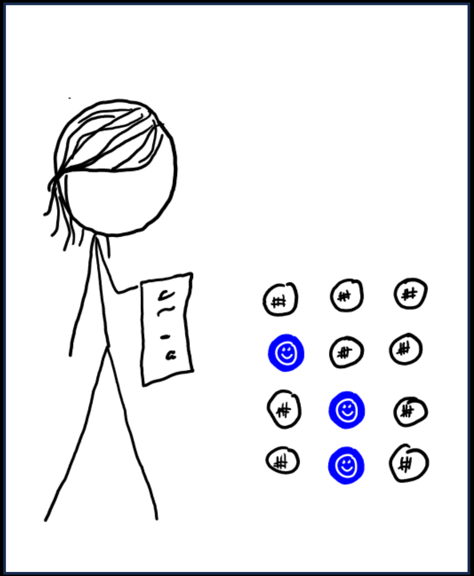


This is a 40mph speed limit, with probability 98%.

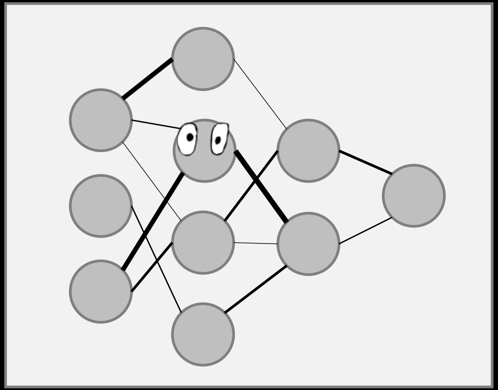


Using a Bin(n, p) model, I estimate the probability of heads is $\hat{p} = 25\%$



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

But surely, the more data we have, the more confident we should be!



This is a 40mph speed limit, with probability 98%.

Neural networks tell us *probabilities*, but they don't tell us their *confidence.*

No one has worked out how to extract confidences from neural networks. But, in Bayesian statistics, we do know how to …

Data from a population sample of 10,000 people:

	test +ve	test -ve	total
got COVID	376	24	400
not got COVID	996	98,604	99,600

Let's rewrite this data as a probability model:

Let $X = 1_{\text{have COVID}}$ and let $Y = 1_{\text{test+ve}}$

```
X ~ Bin(1, 0.004)
if X == 1:
    Y ~ Bin(1, 0.94)
else:
    Y ~ Bin(1, 0.01)
```

What are these probabilities?

- $\mathbb{P}(\text{have COVID} \mid \text{test +ve})$
- $\mathbb{P}(\text{have COVID} \mid \text{test} -\text{ve})$

$$\mathbb{P}(X = 1 \mid Y = 1)$$

$$= \frac{\mathbb{P}(X = 1)\, \mathbb{P}(Y = 1 | X = 1)}{\mathbb{P}(Y = 1)}$$

$$= \frac{0.004 \times 0.94}{0.004 \times 0.94 + 0.996 \times 0.01}$$

How does Bayes's rule apply to continuous random variables?

Let $X = 1_{\text{have COVID}}$
Let $Y = 1_{\text{test+ve}}$

What is the probability I have COVID, i.e. $X = 1$, if $Y = 1$?

By Bayes's rule,

$$\mathbb{P}(X = 1 \mid Y = 1) = \frac{\mathbb{P}(X = 1)\, \mathbb{P}(Y = 1 \mid X = 1)}{\mathbb{P}(Y = 1)}$$

Let $X = 1_{\text{have COVID}}$
Let $Y =$ amount of viral RNA in a PCR test **(CONTINUOUS)**

What is the probability I have COVID, for an amount $Y = y$?

$$\mathbb{P}(X = 1|Y = 2.1) = \frac{\mathbb{P}(X = 1)\, \underbrace{\mathbb{P}(Y = 2.1|X = 1)}_{0}}{\underbrace{\mathbb{P}(Y = 2.1)}_{=0}}$$

This version of Bayes's rule doesn't make sense for continuous random variables!

Reverend Thomas Bayes, 1701–1761

Bayes's rule for random variables

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

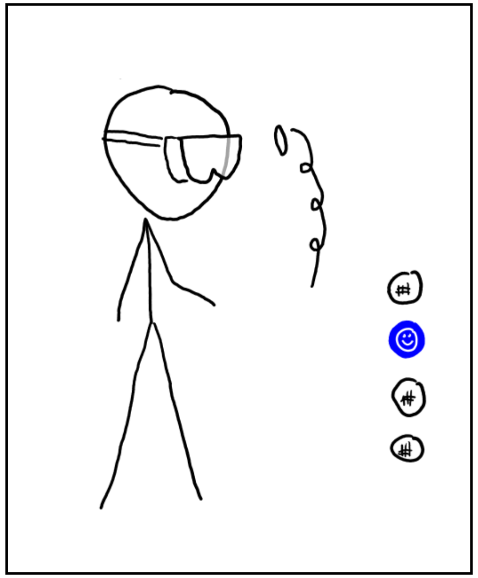
This version of Bayes's rule works for both continuous and discrete random variables.

See section 5.1 for why it works.

Bayesianism

Whenever there's an unknown parameter, you should express your uncertainty about it by treating it as a random variable.

SECTION 8. BAYESIANISM



I got $x = 1$ head out of $n = 4$ coin tosses.

I propose the probability model $X \sim \text{Bin}(n, \Theta)$.

I don't know Θ, so I'll treat it as a random variable.
I shall assume $\Theta \sim U[0,1]$ i.e. $\Pr_\Theta(\theta) = 1$.

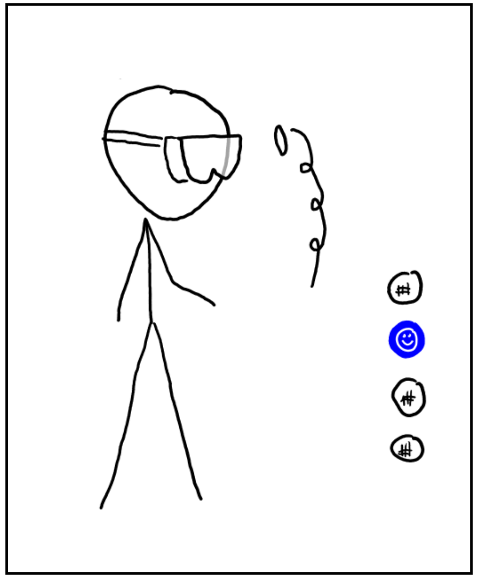
You must have a prior belief about every unknown parameter.

What has the data told me about Θ?
What is my posterior belief $\Pr_\Theta(\theta | X = x)$?

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

The only logical way to update your beliefs is by using Bayes's rule.

SECTION 8. BAYESIANISM

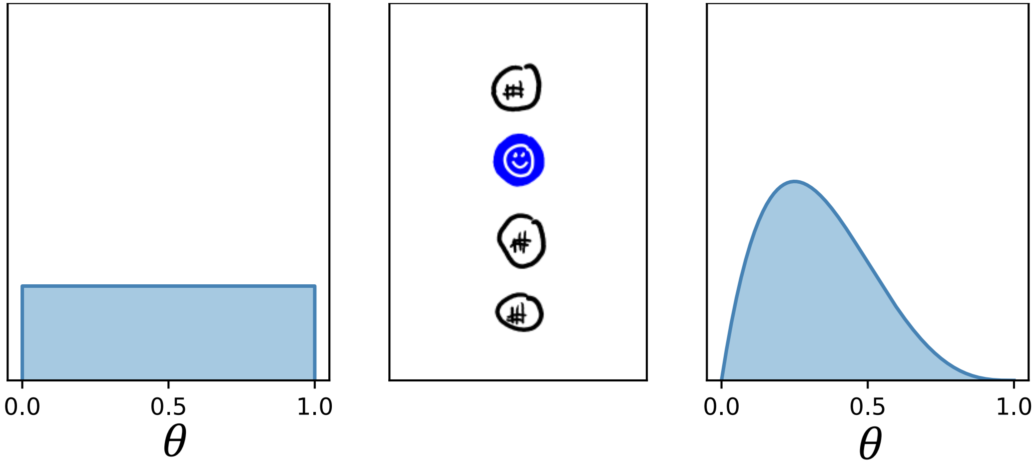


I got $x = 1$ head out of $n = 4$ coin tosses.

I propose the probability model $X \sim \mathrm{Bin}(n, \Theta)$.

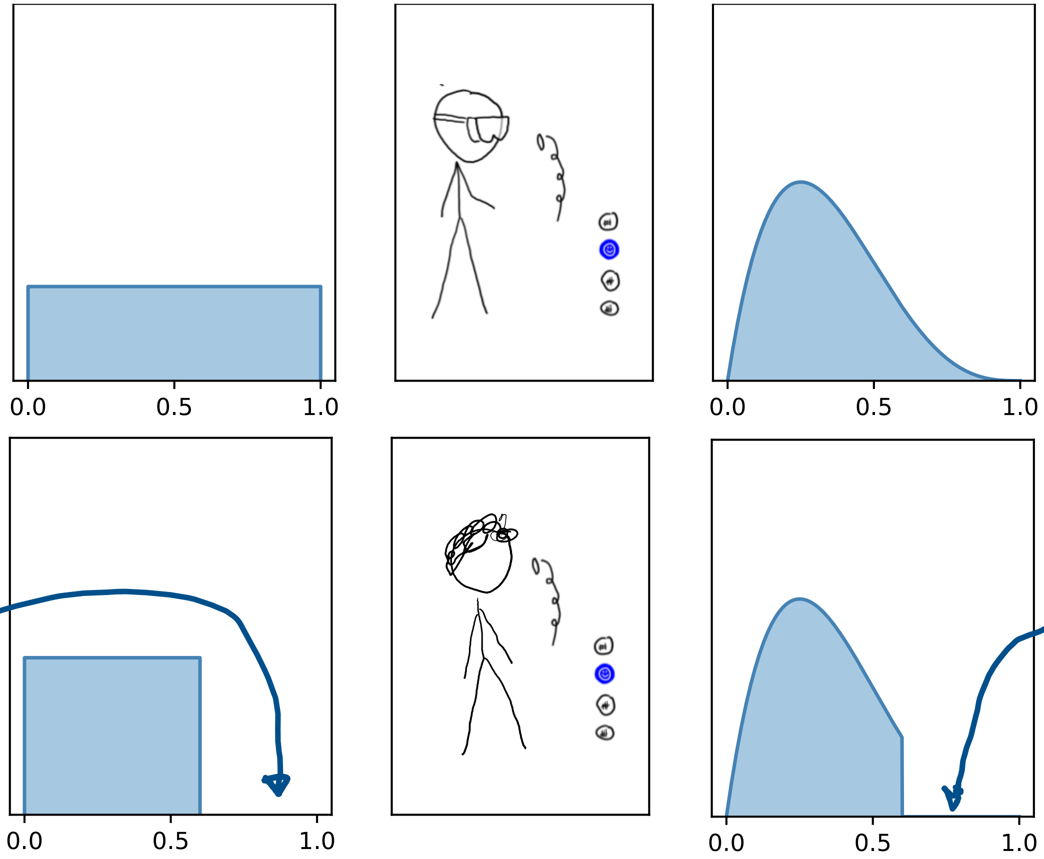
I don't know Θ, so I'll treat it as a random variable.
I shall assume $\Theta \sim U[0,1]$ i.e. $\Pr_\Theta(\theta) = 1$.

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



You are entitled to your own personal prior beliefs.
You have to invent them before you see the data.
They are entirely your choice.

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

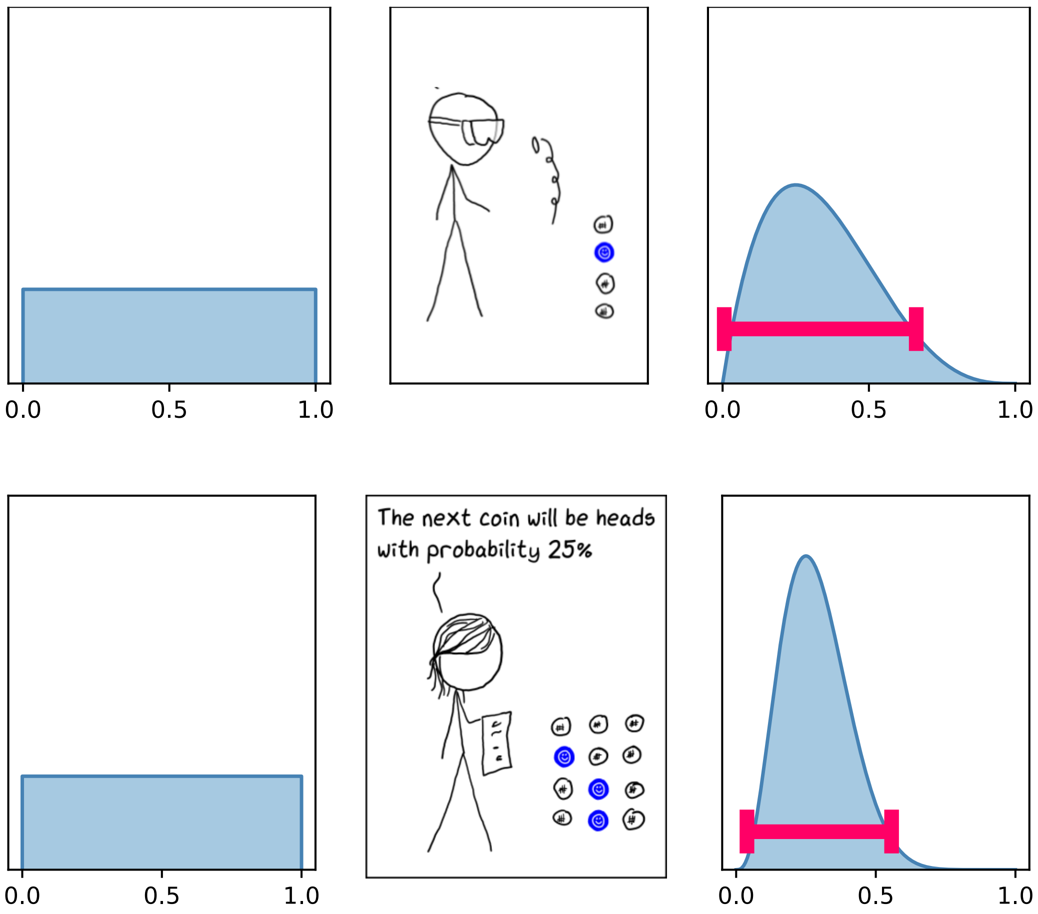


Preconception that $\theta > 0.6$ is impossible

The preconception is unshakeable

The data you see will affect your posterior belief about the parameter.

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



we can measure confidence by the width of posterior distribution

0. Write out the likelihood of the dataset $\Pr_X(x|\Theta = \theta)$

$X \sim \mathrm{Bin}(n, \theta)$ $\quad\quad$ $\Pr_X(x|\Theta = \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$

1. Invent a *prior distribution* and write out its likelihood, $\Pr_\Theta(\theta)$

$\Theta \sim U[0,1]$ $\quad\quad$ $\Pr_\Theta(\theta) = 1$

2. Apply the *Bayes update* to find the *posterior distribution*, i.e. the distribution of $(\Theta|X = x)$

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

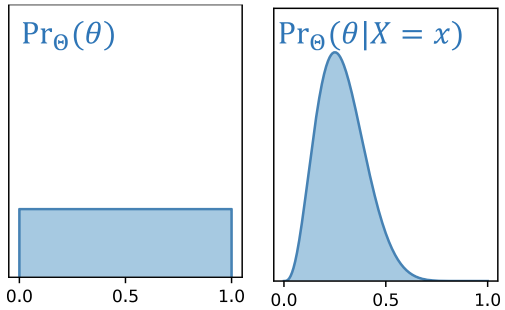
$$= \kappa \Pr_\Theta(\theta) \Pr_X(x|\Theta = \theta)$$

$$= \kappa' \theta^x (1-\theta)^{n-x}$$

Gather all non-θ terms into a constant

The constant κ' must be chosen so that this is a valid distribution i.e. $\int_\theta \kappa' \theta^x (1-\theta)^{n-x} \, d\theta = 1$

3. Report the distribution of $(\Theta|X = x)$, for example by plotting its likelihood $\Pr_\Theta(\theta|X = x)$



COMPUTATIONAL BAYESIAN METHODS

$$\Pr_X(x) = \int_{\theta'} \Pr_\Theta(\theta') \Pr_X(x \mid \Theta = \theta')\, d\theta'$$

It's useful to be able to generate samples from the posterior distribution $(\Theta | X = x)$.

For example, we could generate samples $\theta_1, \dots, \theta_n$ and then plot a histogram of their values.

The maths version of Bayes's rule isn't any help for this.

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

$$= \kappa \Pr_\Theta(\theta)\,\Pr_X(x|\Theta=\theta)$$

$$= \kappa'\,\theta^x (1-\theta)^{n-x}$$

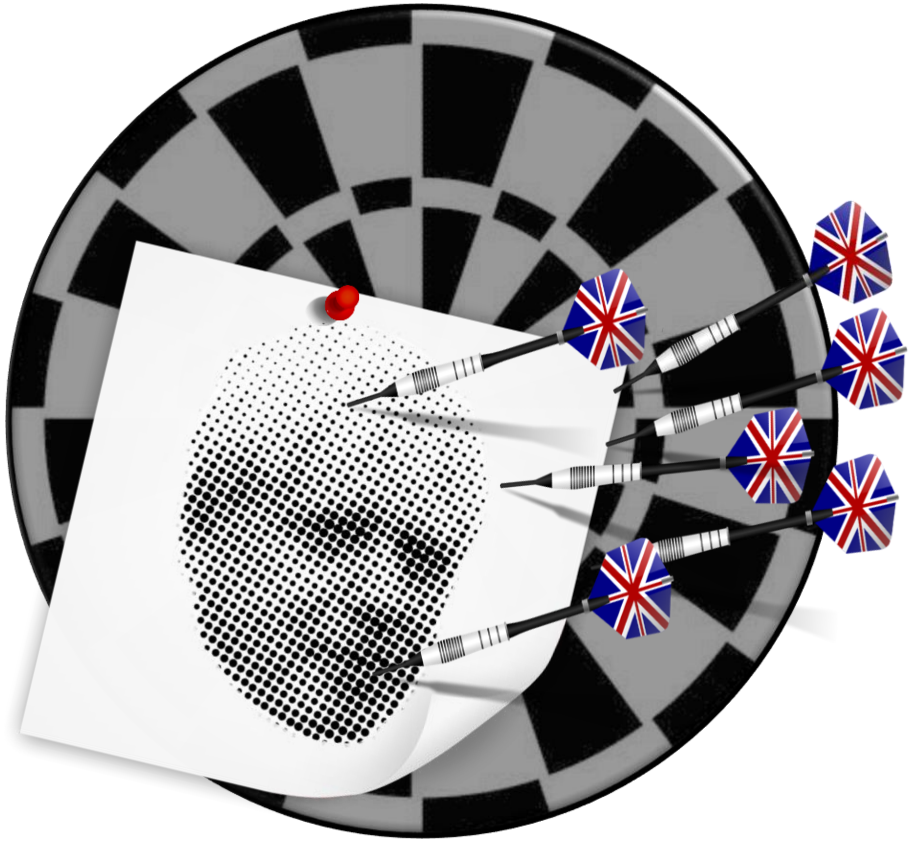
The constant κ' must be chosen so that this is a valid distribution i.e. $\int_\theta \kappa'\,\theta^x(1-\theta)^{n-x}\,d\theta = 1$

$$\frac{\Pr_\Theta(\theta) \Pr_X(x \mid \Theta = \theta)}{\int_{\theta'} \Pr_\Theta(\theta') \Pr_X(x \mid \Theta = \theta')\, d\theta'}$$

This integral is usually impossible to solve. And even if we could solve it, how do we sample from this distribution?

SECTION 6.1. COMPUTATIONAL METHODS

What's the chance that a randomly thrown dart will hit the mystery object A?



Let X be the location of a randomly thrown dart, and let $x_1, \dots, x_n$ be some throws.

The probability of hitting A is

$$\mathbb{P}(X \in A) \approx \frac{1}{n}\sum_{i=1}^{n} 1_{x_i \in A}$$

```
# Let X ~ N(μ = 1, σ = 3). What is ℙ(X > 5)?
x = np.random.normal(loc=1, scale=3, size=10000)
i = (x > 5)
np.mean(i)
```

Expectation

For a real-valued random variable X

$$\mathbb{E}X = \begin{cases} \sum_x x \Pr_X(x)\,, & \text{if } X \text{ is discrete} \\ \int_x x \Pr_X(x)\, dx\,, & \text{if } X \text{ is continuous} \end{cases}$$

Law of the Unconscious Statistician

For a random variable X and a real-valued function h

$$\mathbb{E}h(X) = \begin{cases} \sum_x h(x) \Pr_X(x), & \text{if } X \text{ is discrete} \\ \int_x h(x) \Pr_X(x)\, dx, & \text{if } X \text{ is continuous} \end{cases}$$

Law of the Unconscious Statistician

For a random variable X and a real-valued function h

$$\mathbb{E}h(X) = \begin{cases} \sum_x h(x) \Pr_X(x), & \text{if } X \text{ is discrete} \\ \int_x h(x) \Pr_X(x)\, dx, & \text{if } X \text{ is continuous} \end{cases}$$

Monte Carlo integration

$$\mathbb{E}h(X) \approx \frac{1}{n} \sum_{i=1}^{n} h(x_i)$$

where $x_1, \dots, x_n$ is a sample drawn from X

Let $h(x) = 1_{x \in A} = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$

By Monte Carlo,

$$\mathbb{E}\, h(X) \approx \frac{1}{n} \sum_{i=1}^{n} h(x_i)$$

$$\frac{1}{n} \sum_i 1_{x_i \in A}$$

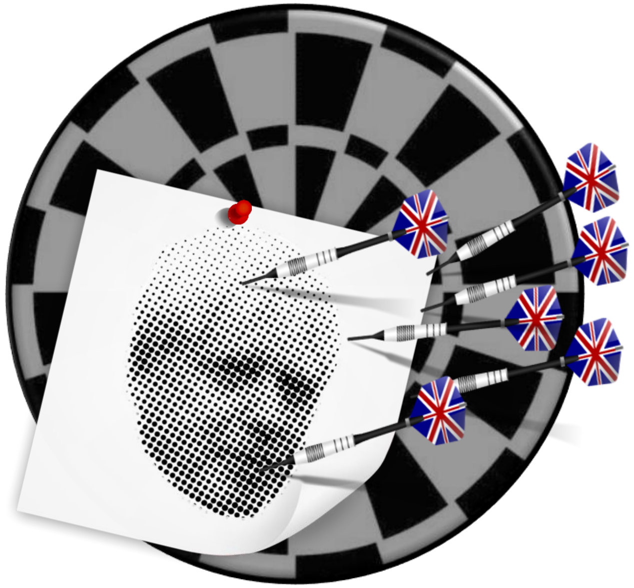
$$Y = h(X) = 1_{X \in A}$$

$$\begin{aligned} \mathbb{E} Y &= 0 \times \mathbb{P}(Y=0) + 1 \times \mathbb{P}(Y=1) \\ &= \mathbb{P}(Y=1) \\ &= \mathbb{P}(1_{X \in A} = 1) \\ &= \mathbb{P}(X \in A) \end{aligned}$$

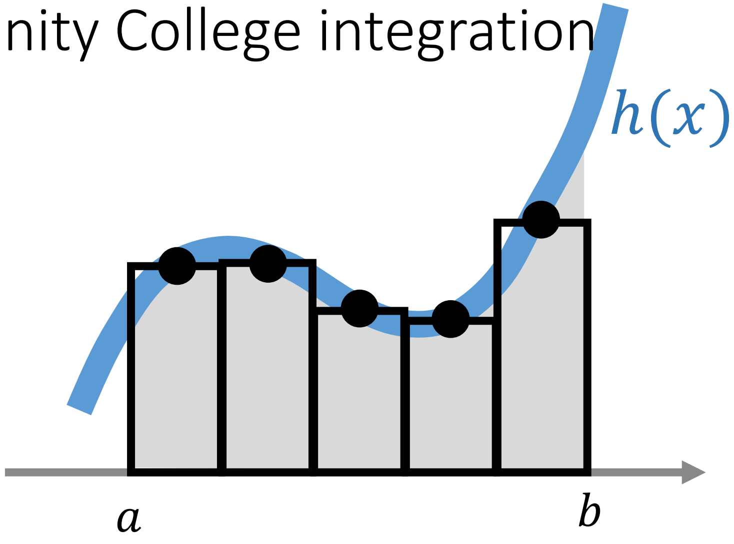
Let X be the location of a randomly thrown dart, and let $x_1, \dots, x_n$ be some throws.

The probability of hitting A is

$$\mathbb{P}(X \in A) \approx \frac{1}{n} \sum_{i=1}^{n} 1_{x_i \in A}$$



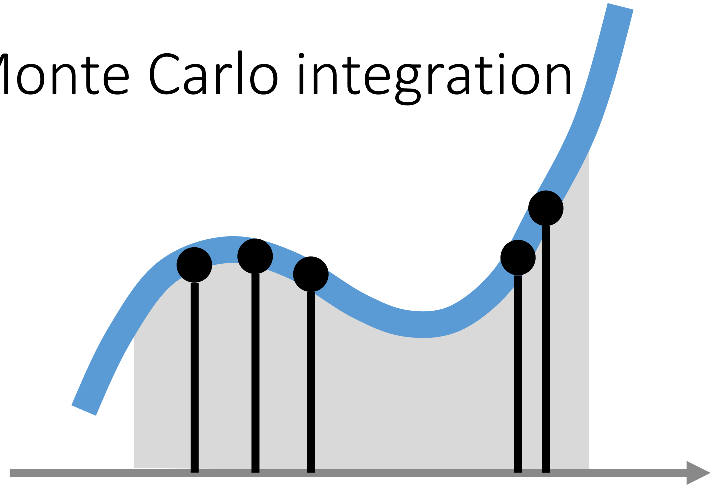
Trinity College integration



$$\int_{x=a}^{b} h(x)\,dx \approx \sum_{i=1}^{n} h(x_i)\,\frac{b-a}{n}$$

where x_i is the midpoint of interval i

Monte Carlo integration



Let's instead approximate this integral using Monte Carlo. Let $X \sim U[a,b]$.

By Monte Carlo,

$$\underbrace{\mathbb{E}h(X)}_{\int_{x=a}^{b} h(x)\,\Pr_X(x)\,dx} \approx \frac{1}{n}\sum_{i=1}^{n} h(x_i) \quad \text{where } x_1, \ldots, x_n \text{ sampled from } X$$

$$\int_{x=a}^{b} h(x)\,\Pr_X(x)\,dx = \int_{x=a}^{b} h(x)\,\frac{1}{b-a}\,dx$$

Thus,

$$\int_{x=a}^{b} h(x)\,dx \approx \frac{b-a}{n}\sum_{i=1}^{n} h(x_i)$$

COMPUTATIONAL METHODS

- If we want $\mathbb{E}h(X)$ but the maths is too complicated, we can approximate it using $x_1, \ldots, x_n$ sampled from X
- This formula for expectation also tells us how to estimate probabilities, since $\mathbb{P}(X \in A) = \mathbb{E}1_{X \in A}$
- For computational Bayes, we need something a bit fancier: *weighted samples*

One way to do

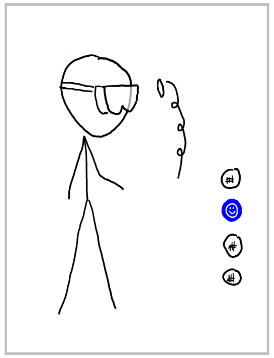
SECTION 6.2. COMPUTATIONAL BAYES

0. Write out the likelihood of the dataset $\Pr_X(x|\Theta = \theta)$
1. Invent a prior distribution for Θ and generate a sample $(\theta_1, \dots, \theta_n)$ from it
2. Compute weights $w_i = \Pr_X(x|\Theta = \theta_i)$, then rescale weights to sum to one
3. Reason about $(\Theta|X = x)$ indirectly, using $\mathbb{E}[h(\Theta)|X = x] \approx \Sigma_i w_i h(\theta_i)$

MATHS.

0. Write $\Pr_X(x|\Theta=\theta)$
1. Write $\Pr_\Theta(\theta)$
2. Use Bayes's rule to get $\Pr_\Theta(\theta|X=x)$
3. Use it.

Example



I got $x = 1$ head out of $n = 4$ coin tosses. I propose the probability model $X \sim \mathrm{Bin}(n, \Theta)$. I don't know Θ, so I'll treat it as a random variable, $\Theta \sim U[0,1]$.

Plot a histogram of the posterior distribution of Θ.

Likelihood of the dataset: $X \sim \mathrm{Bin}(n, \theta)$ $\qquad \Pr_X(x|\Theta = \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$

Invent a prior distribution for Θ and generate a sample $(\theta_1, \ldots, \theta_n)$ from it:

```
θsamp = np.random.uniform(0,1, size=1000)
```

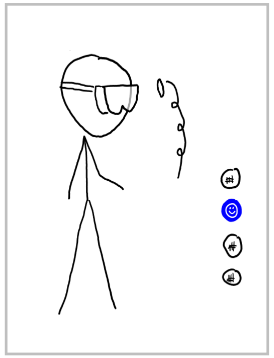
Compute weights $w_i = \Pr_X(x|\Theta = \theta_i)$, then rescale weights to sum to one:

```
w = 4 * θsamp**1 * (1-θsamp)**3
w = w / np.sum(w)
```

Reason about $(\Theta|X = x)$ indirectly, using

$$\mathbb{E}[h(\Theta)|X = x] \approx \Sigma_i w_i h(\theta_i)$$

Example



I got $x = 1$ head out of $n = 4$ coin tosses. I propose the probability model $X \sim \text{Bin}(n, \Theta)$. I don't know Θ, so I'll treat it as a random variable, $\Theta \sim U[0,1]$.

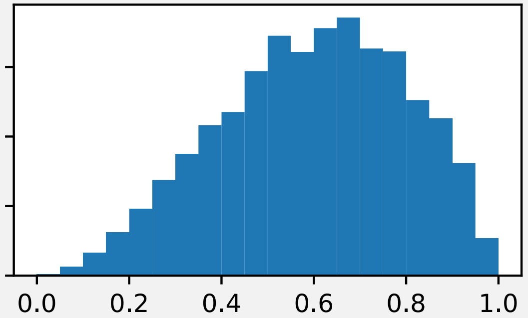
Plot a histogram of the posterior distribution of Θ.

For each bin, I want to plot a bar of height $\mathbb{P}(\Theta \in \text{bin} \mid \text{data})$

$$
\begin{aligned}
\mathbb{P}(\Theta \in \text{bin} \mid \text{data}) &= \mathbb{E}\left(1_{\Theta \in \text{bin}} \mid \text{data}\right) \\
&= \mathbb{E}\left(h(\Theta) \mid \text{data}\right) \quad \text{where } h(\theta) = 1_{\theta \in \text{bin}} \\
&\approx \sum_{i=1}^{n} w_i \, 1_{\theta_i \in \text{bin}} \\
&= \sum_{i:\, \theta_i \in \text{bin}} w_i
\end{aligned}
$$

for each bin, sum up the weights of the θ-samples in that bin.

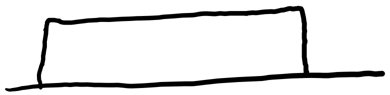
```
plt.hist(θsamp, weights=w)
```



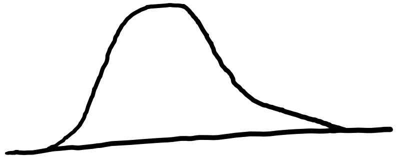
Reason about $(\Theta | X = x)$ indirectly, using
$\mathbb{E}[h(\Theta) | X = x] \approx \Sigma_i w_i h(\theta_i)$

SECTION 8.3. BAYESIAN READOUTS

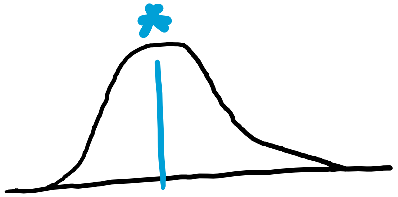
Prior distribution for Θ



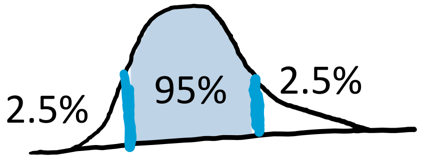
Posterior distribution for Θ



How should we report this distribution?

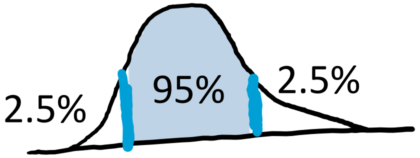


We could report the point with highest likelihood, the *MAP* or *maximum a-posteriori* estimate



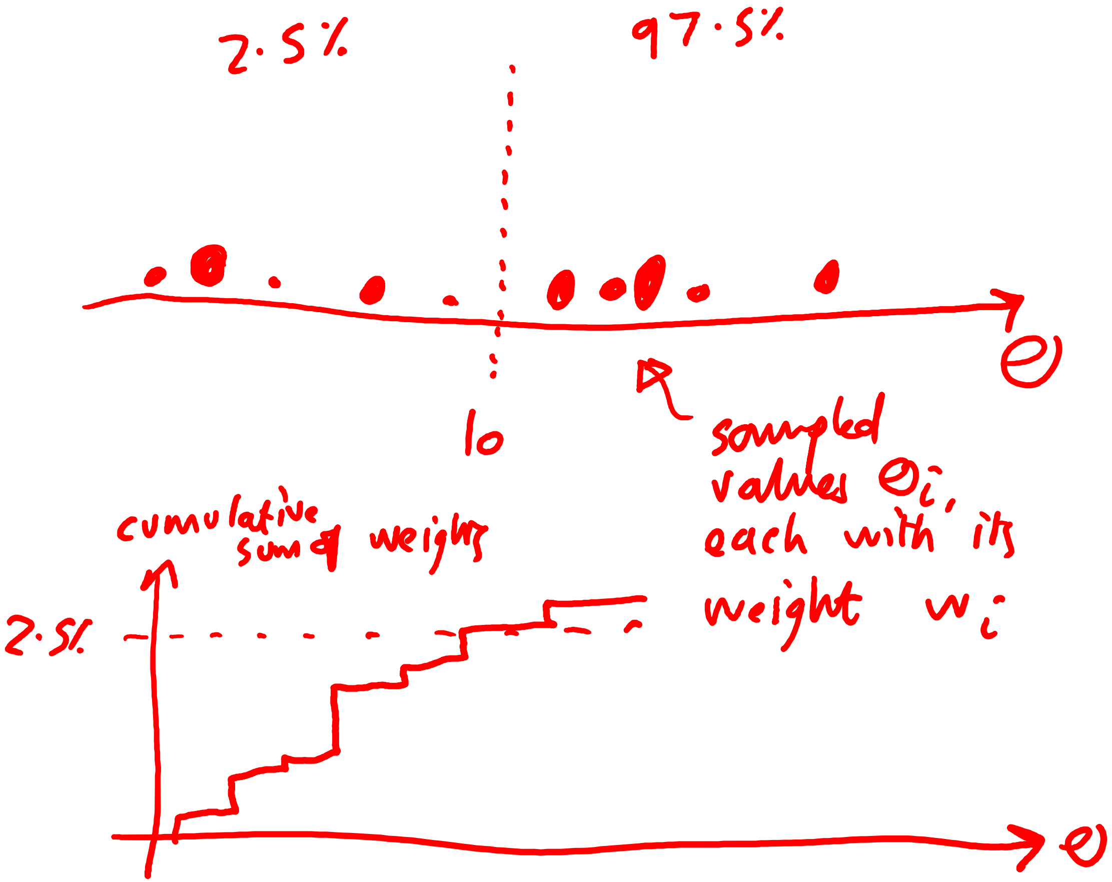
We could report a *95% confidence interval* $[\mathtt{lo},\mathtt{hi}]$ such that

$$\mathbb{P}(\Theta < \mathtt{lo} \mid \text{data}) = 2.5\%$$
$$\mathbb{P}(\Theta > \mathtt{hi} \mid \text{data}) = 2.5\%$$



We could report a *95% confidence interval* [lo,hi] such that

$$\mathbb{P}(\Theta < \texttt{lo} \mid \text{data}) = 2.5\%$$
$$\mathbb{P}(\Theta > \texttt{hi} \mid \text{data}) = 2.5\%$$



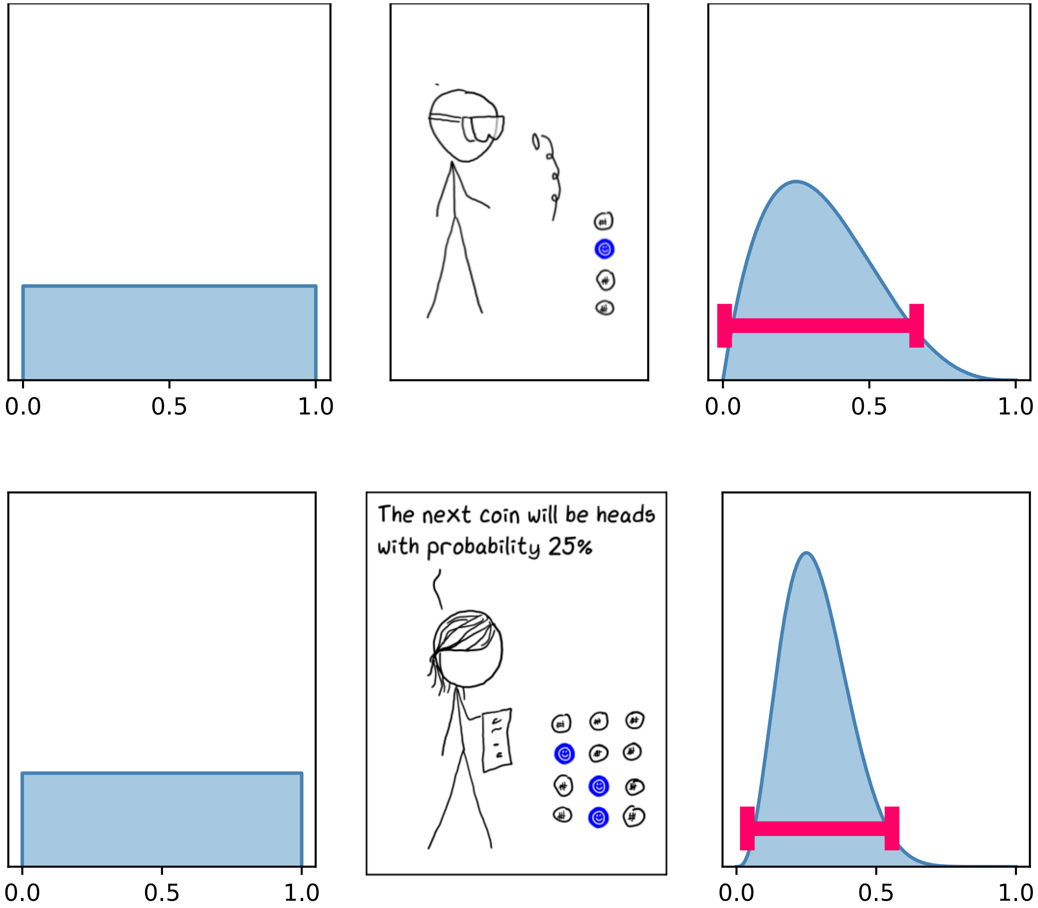
How can we compute lo and hi?

Via the computational Bayes estimates:

$$\mathbb{P}(\Theta < \texttt{lo} \mid \text{data}) \approx \sum_i w_i \, 1_{\theta_i < \texttt{lo}}$$

$$\mathbb{P}(\Theta > \texttt{hi} \mid \text{data}) \approx \sum_i w_i 1_{\theta_i > \texttt{hi}}$$

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

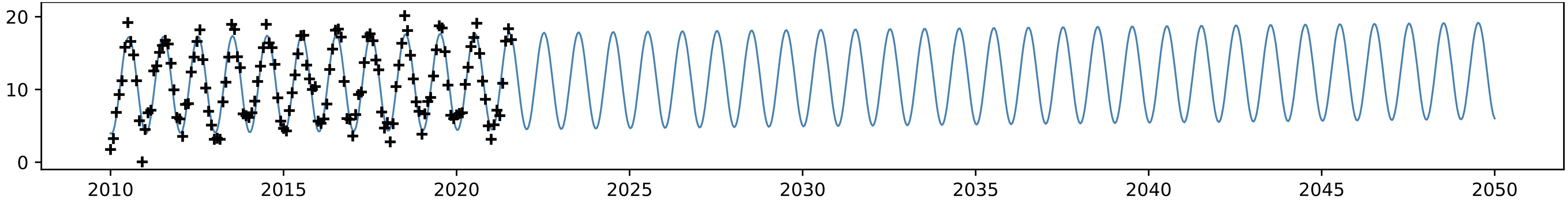


I estimate the probability of heads is 25%, and my confidence interval is [3%, 72%]

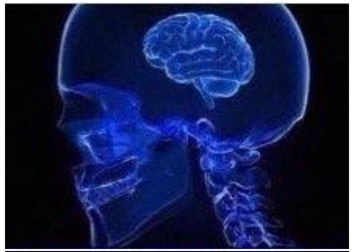
I estimate the probability of heads is 25%, and my confidence interval is [12%, 51%]

SECTION 8.5. BAYESIAN PREDICTION

I proposed the probability model: $Y \sim \alpha + \beta \sin\big(2\pi(t+\phi)\big) + \gamma t + N(0,\sigma^2)$



What will be the temperature in $t^* = \text{Jan } 2050$?



I'll fit the model using maximum likelihood, and report my estimated value

$$\hat{P} = \hat{\alpha} + \hat{\beta} \sin\Big(2\pi\big(t^* + \hat{\phi}\big)\Big) + \hat{\gamma} t^*$$

The actual observed temperature will actually have noise. I'll report my estimated distribution

$$Y^* \sim N(\hat{P}, \hat{\sigma}^2)$$

I'm not even certain about $\hat{P}$, because I'm not certain about my parameter estimates.
I should report my uncertainty about $P = \alpha + \beta \sin\big(2\pi(t^* + \phi)\big) + \gamma t^*$.

QUESTION. How should I compute and report my uncertainty about P?

I'm not even certain about $\hat{P}$, because I'm not certain about my parameter estimates.
I should report my uncertainty about $P = \alpha + \beta \sin\big(2\pi(t^* + \phi)\big) + \gamma t^*$.

CONFIDENCE INTERVALS FOR PREDICTIONS

1. The fundamental tenet of Bayesianism is that we should represent our parameter uncertainty by treating our parameters as random variables.

2. The parameters I'm uncertain about are $\theta = (\alpha, \beta, \gamma, \phi, \sigma)$. I shall use a random variable Θ, taking values in $\mathbb{R}^5$, to represent this uncertainty.

3. What distribution should I use for Θ? The Bayesianist view is that it's entirely up to me what prior I choose to use, and that I must choose my prior without looking at the data.
   I might choose for example $\alpha \sim N(10,1^2)$°C if I'm confident that α should be around 10;
   I might choose $\alpha \sim N(10,8^2)$°C if I'm uncertain.

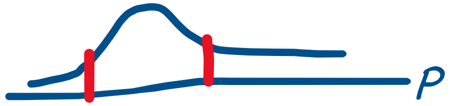
4. For Computation Bayes, we first generate a large number of possible parameters $\theta_1, \dots, \theta_m$ from the prior distribution, then we compute a weight w_i for every θ_i, $i \in \{1, \dots, m\}$

5. For every one of these parameter choices θ_i, there's a corresponding value for P, call them $p_1, \dots, p_m$

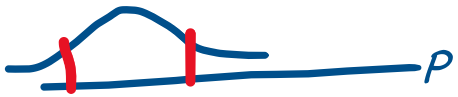
6. I thus have a collection of possible values for P, each with an associated weight. I can use these weights to find a confidence interval for P.

I'm not even certain about $\hat{P}$, because I'm not certain about my parameter estimates.
I should report my uncertainty about $P = \alpha + \beta \sin\big(2\pi(t^* + \phi)\big) + \gamma t^*$.

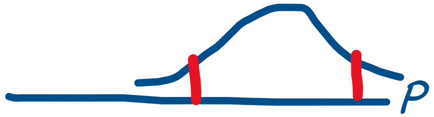
At $t^* =$ Jan 2050, I get this confidence interval:



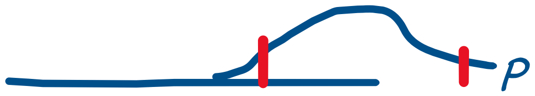
At $t^* =$ Feb 2050, I get this confidence interval:



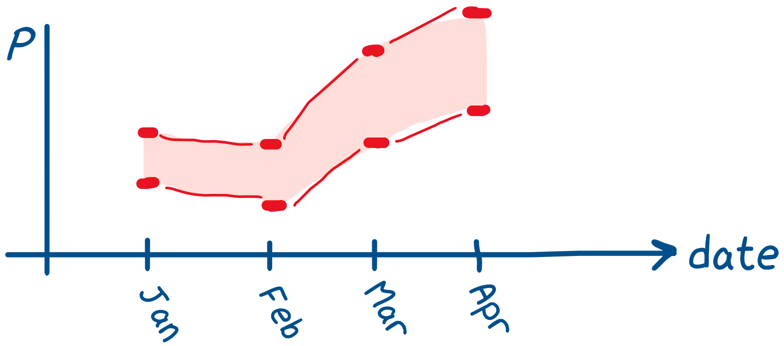
At $t^* =$ Mar 2050, I get this confidence interval:



At $t^* =$ Apr 2050, I get this confidence interval:



I can show all my confidence intervals as a *ribbon plot:*



I'm not even certain about $\hat{P}$, because I'm not certain about my parameter estimates.
I should report my uncertainty about $P = \alpha + \beta \sin\big(2\pi(t^* + \phi)\big) + \gamma t^*$.

TIP. Use all the unknowns when you apply Bayes's rule, even if you're only interested in some of them.

S FOR PREDICTIONS

t of Bayesianism is that we should represent our parameter uncertainty by rs as random variables.

2. The parameters I'm uncertain about are $\theta = (\alpha, \beta, \gamma, \phi, \sigma)$. I shall use a random variable Θ, taking values in $\mathbb{R}^5$, to represent this uncertainty.

3. What distribution should I use for Θ? The Bayesianist view is that it's entirely up to me what prior I choose to use, and that I must choose my prior without looking at the data.
   I might choose for example $\alpha \sim N(10, 1^2)$°C if I'm confident that α should be around 10;
   I might choose $\alpha \sim N(10, 8^2)$°C if I'm uncertain.

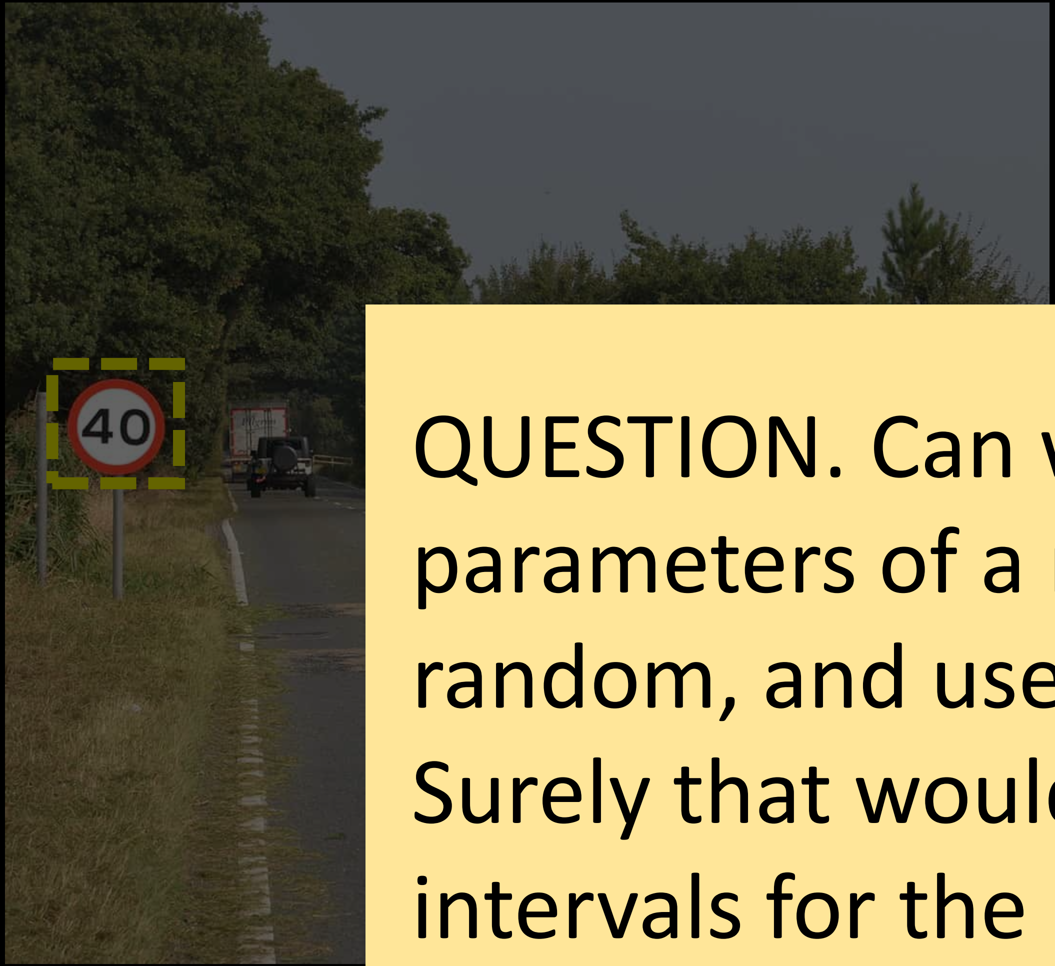
4. For Computation Bayes, we first generate a large number of possible parameters $\theta_1, \ldots, \theta_m$ from the prior distribution, then we compute a weight w_i for every θ_i, $i \in \{1, \ldots, m\}$

5. For every one of these parameter choices θ_i, there's a corresponding value for P, call them $p_1, \ldots, p_m$

6. I thus have a collection of possible values for P weights to find a confidence interval for P.

TIP. In large datasets, you can run into underflow problems when you compute the likelihood and when you rescale weights. See the *log-sum-exp* trick (exercise 8.2.4 in notes).

I'm not even certain about $\hat{P}$, because I'm not certain about my parameter estimates.
I should report my uncertainty about $P = \alpha + \beta \sin\big(2\pi(t^* + \phi)\big) + \gamma t^*$.



QUESTION. Can we just take the parameters of a neural network to be random, and use Bayesian methods? Surely that would give us confidence intervals for the probability!

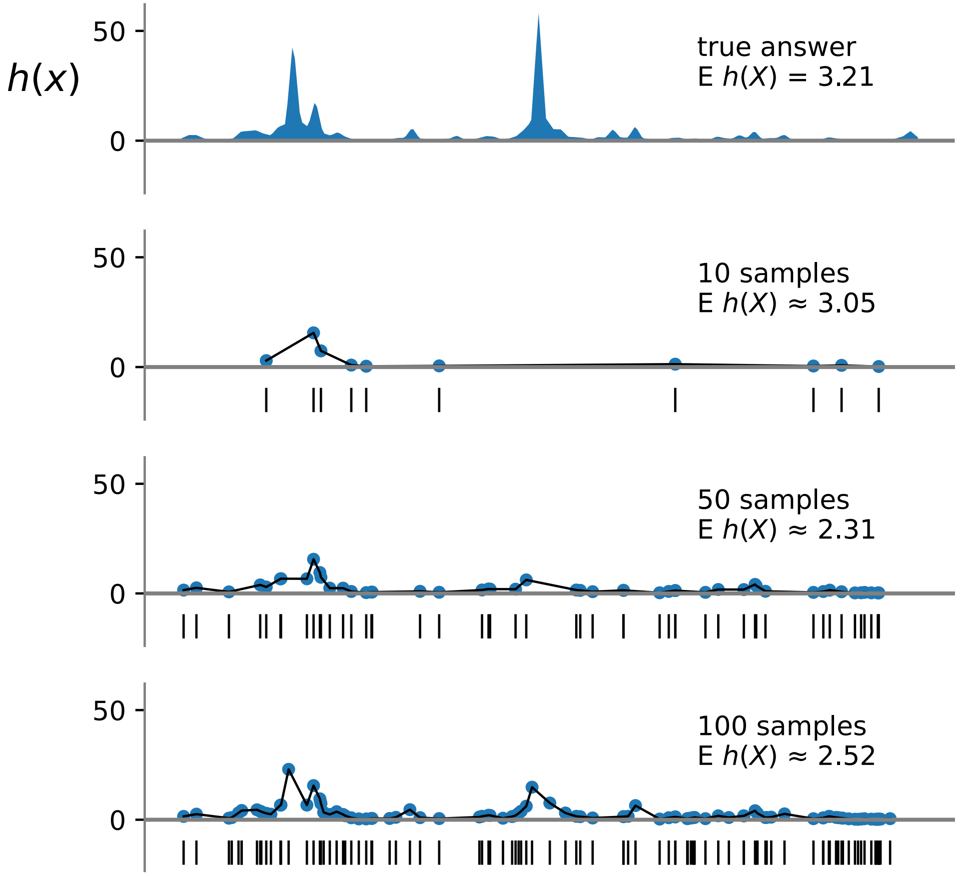
0mph speed
probability 98%.

works tell us
es, but they don't
r *confidence.*

No one has worked out how to extract confidences from neural networks. But, in Bayesian statistics, we do know how to …

Let X be a random variable, let h be a real-valued function. Then

$$\mathbb{E}h(X) \approx \frac{1}{n}\sum_{i=1}^{n} h(x_i) \quad \text{where } x_1, \dots, x_n \text{ is a sample drawn from } X$$



Example: $X \sim U[0,1]$, spiky h function

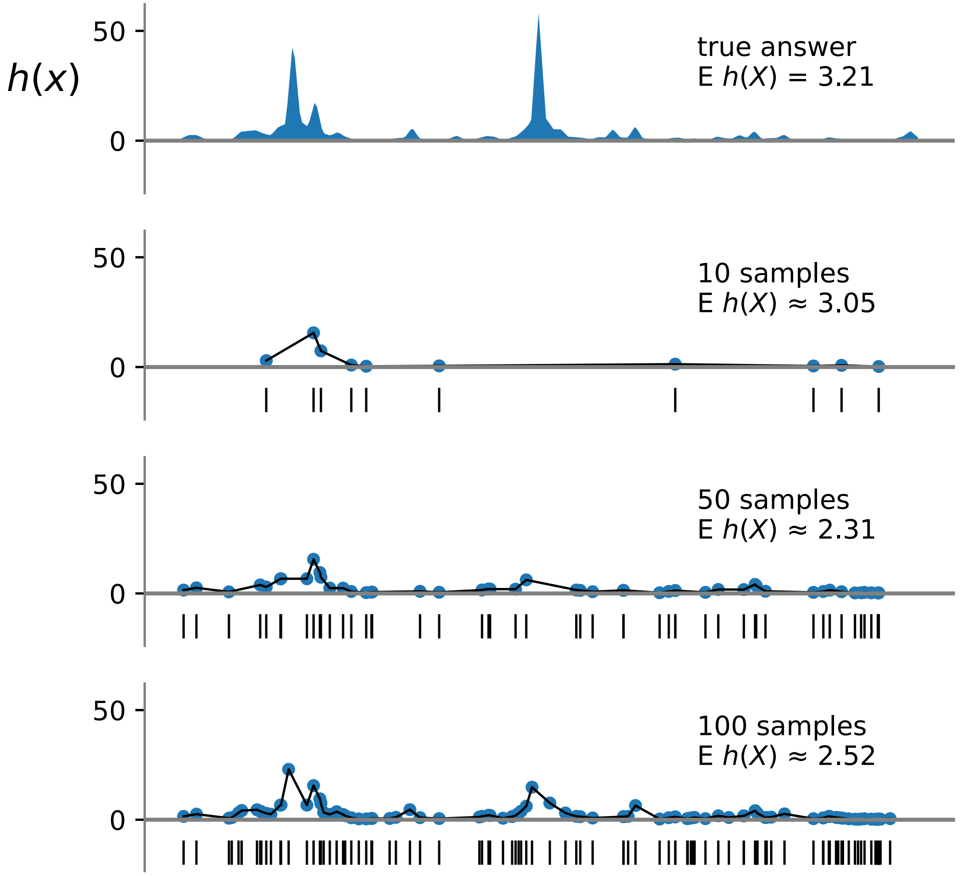
It may take very large n
to get a good approximation

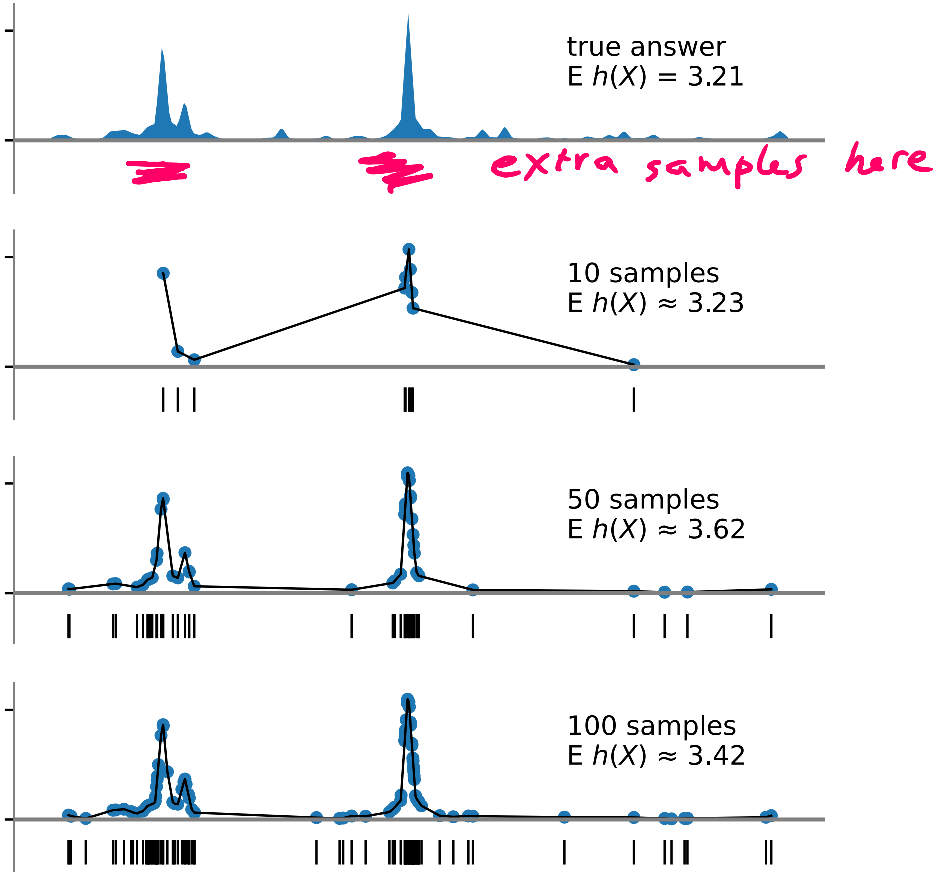
SECTION 6.3. IMPORTANCE SAMPLING

Let X be a random variable, let h be a real-valued function. Then

$$\mathbb{E}h(X) \approx \frac{1}{n}\sum_{i=1}^{n} h(x_i) \quad \text{where } x_1, \dots, x_n \text{ is a sample drawn from } X$$

Can we speed things up with biased sampling?





Let X be a random variable, let h be a real-valued function. Then

$$\mathbb{E}h(X) \approx \frac{1}{n}\sum_{i=1}^{n} h(x_i) \quad \text{where } x_1, \dots, x_n \text{ is a sample drawn from } X$$

Importance sampling

Let X be a random variable, let h be a real-valued function, and let $\tilde{X}$ be any distribution. Then

$$\mathbb{E}h(X) \approx \frac{1}{n}\sum_{i=1}^{n} h(x_i)\frac{\Pr_X(x_i)}{\Pr_{\tilde{X}}(x_i)} \quad \text{where } x_1, \dots, x_n \text{ is a sample drawn from } \tilde{X}$$

the "sampling distribution"

correction for biased sampling

This works for *any* sampling distribution $\tilde{X}$.
But it will only be useful if we choose a sensible sampling distribution!

Importance sampling

Let X be a random variable, let h be a real-valued function, and let $\tilde{X}$ be any distribution. Then

$$\mathbb{E}h(X) \approx \frac{1}{n}\sum_{i=1}^{n} h(x_i)\frac{\Pr_X(x_i)}{\Pr_{\tilde{X}}(x_i)} \quad \text{where } x_1, \dots, x_n \text{ is a sample drawn from } \tilde{X}$$

Why does this work?

$$\mathbb{E}\,h(X) = \int_x h(x)\Pr_X(x)\,dx \qquad \text{by definition of expectation}$$

$$= \int_x \underbrace{h(x)\frac{\Pr_X(x)}{\Pr_{\tilde{X}}(x)}}_{g(x)} \underbrace{\Pr_{\tilde{X}}(x)}\,dx$$

$$\approx \frac{1}{n}\sum_{i=1}^{n} g(x_i) \quad \text{where } x_1, \dots, x_n \text{ sampled from } \tilde{X}. \qquad \text{by Monte Carlo}$$

$$= \frac{1}{n}\sum_{i} h(x_i)\frac{\Pr_X(x_i)}{\Pr_{\tilde{X}}(x_i)} \qquad \text{by definition of } g.$$

Importance sampling

Let X be a random variable, let h be a real-valued function, and let $\tilde{X}$ be any distribution. Then

$$\mathbb{E}h(X) \approx \frac{1}{n}\sum_{i=1}^{n} h(x_i)\frac{\Pr_X(x_i)}{\Pr_{\tilde{X}}(x_i)}$$ where $x_1, \ldots, x_n$ is a sample drawn from $\tilde{X}$

Computational Bayes is based on importance sampling. It's based on using samples from the prior distribution (Θ) to get estimates for things derived from the posterior distribution $(\Theta|\text{data})$.

Correction factor: $$\frac{\Pr_{(\Theta|\text{data})}(\theta_i)}{\Pr_\Theta(\theta_i)} = \frac{\kappa \Pr_\Theta(\theta_i)\Pr(\text{data}|\theta_i)}{\Pr_\Theta(\theta_i)} = \kappa \Pr(\text{data}|\theta_i)$$ by Bayes's rule

To see how to estimate κ, see notes section 6.2.