Reverend Thomas Bayes, 1701–1761

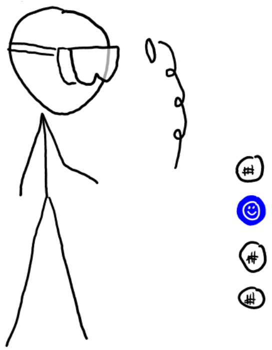
Bayes's rule for random variables

For any pair of random variables (X, Y)

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

Bayesianism

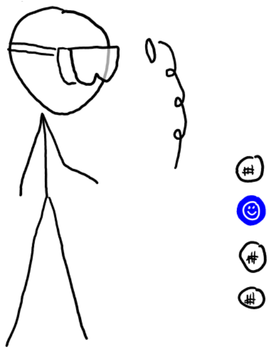
Whenever there's an unknown parameter, you should express your uncertainty about it by treating it as a random variable.



I tossed four coins and got one head.
What is it reasonable to infer about the probability of heads (call it θ)?

- "The maximum likelihood estimator is $\hat{\theta} = 25\%$, thus the true probability of heads is 25%"
  (hence if I tossed millions 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!
- ???

Bayesianists represent their uncertainty about an unknown parameter by using a random variable.



probability of heads, unknown

Θ We use a random variable here to express our beliefs about Θ.

We might choose $\Theta \sim U[0,1]$ to express ignorance about Θ.

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

number of heads from 4 coin tosses

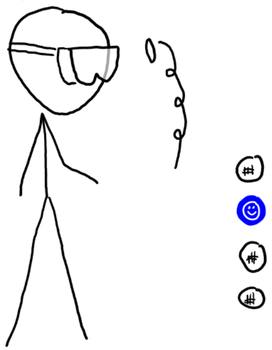
$\Pr_\Theta(\theta)$ is called the prior.

It expresses our beliefs prior to having seen this data.

$\Pr_\Theta(\theta | X = 1)$ is called the posterior.

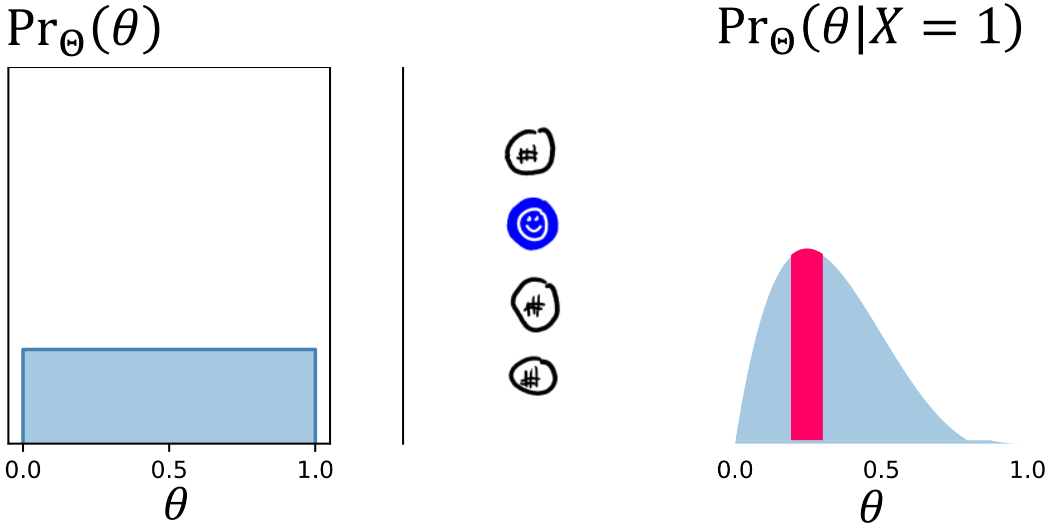
It expresses our beliefs about Θ *in the light of the data.*

By using random variables for unknown quantities, we can reason about confidence.



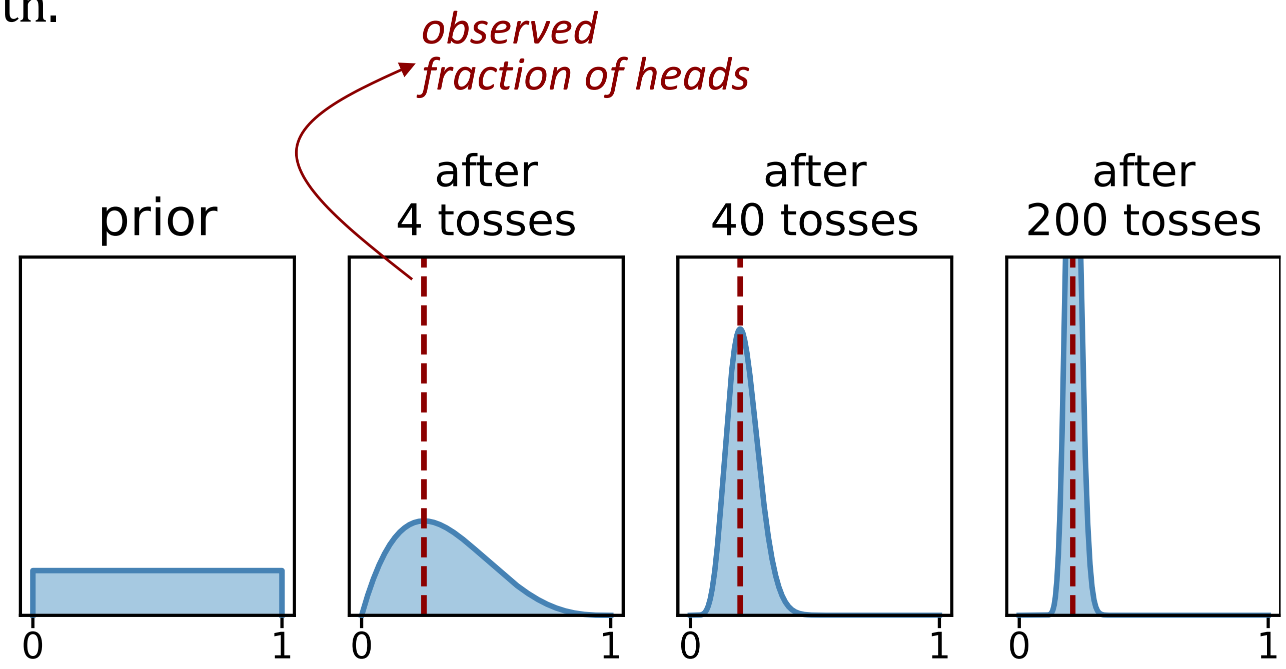
$$\Theta \sim U[0,1]$$

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

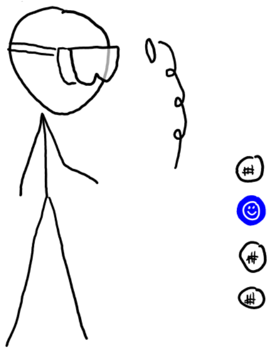


This Bayesianist approach lets us say something justifiable *and* useful: for example, "$\mathbb{P}(\Theta \in [.2, .3] \mid \text{data}) = 21\%$".

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



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



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

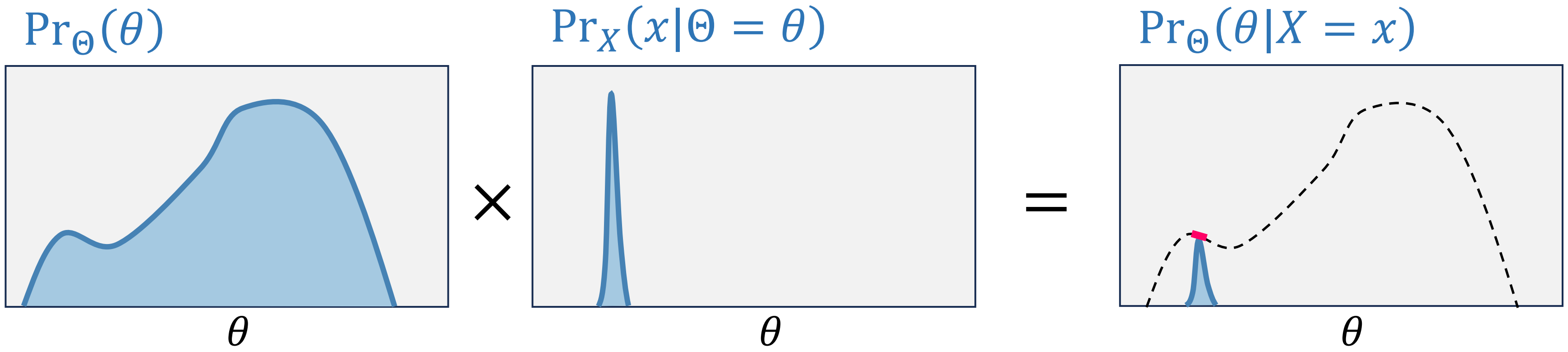
$X \sim \text{Bin}(4, \Theta)$

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.

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

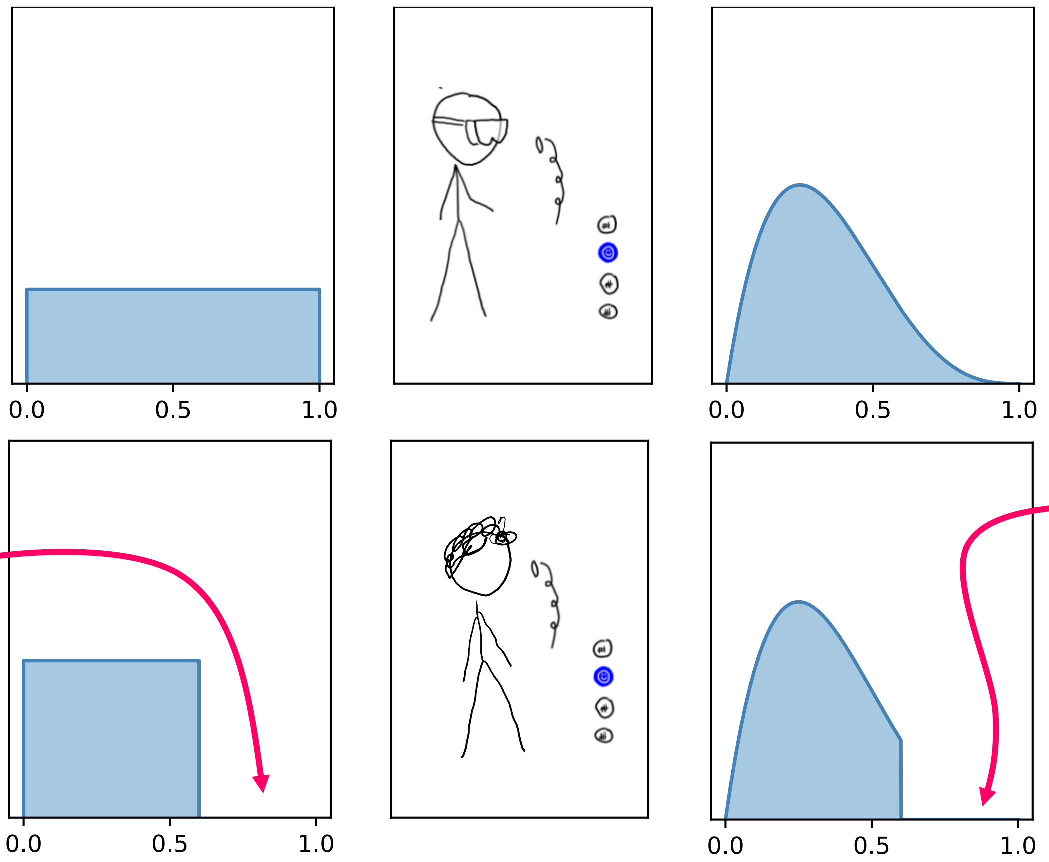
$$\Pr_{\Theta}(\theta|X=x) \;=\; \kappa \Pr_{\Theta}(\theta) \Pr_{X}(x|\Theta=\theta)$$



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

$\Pr_\Theta(\theta)$

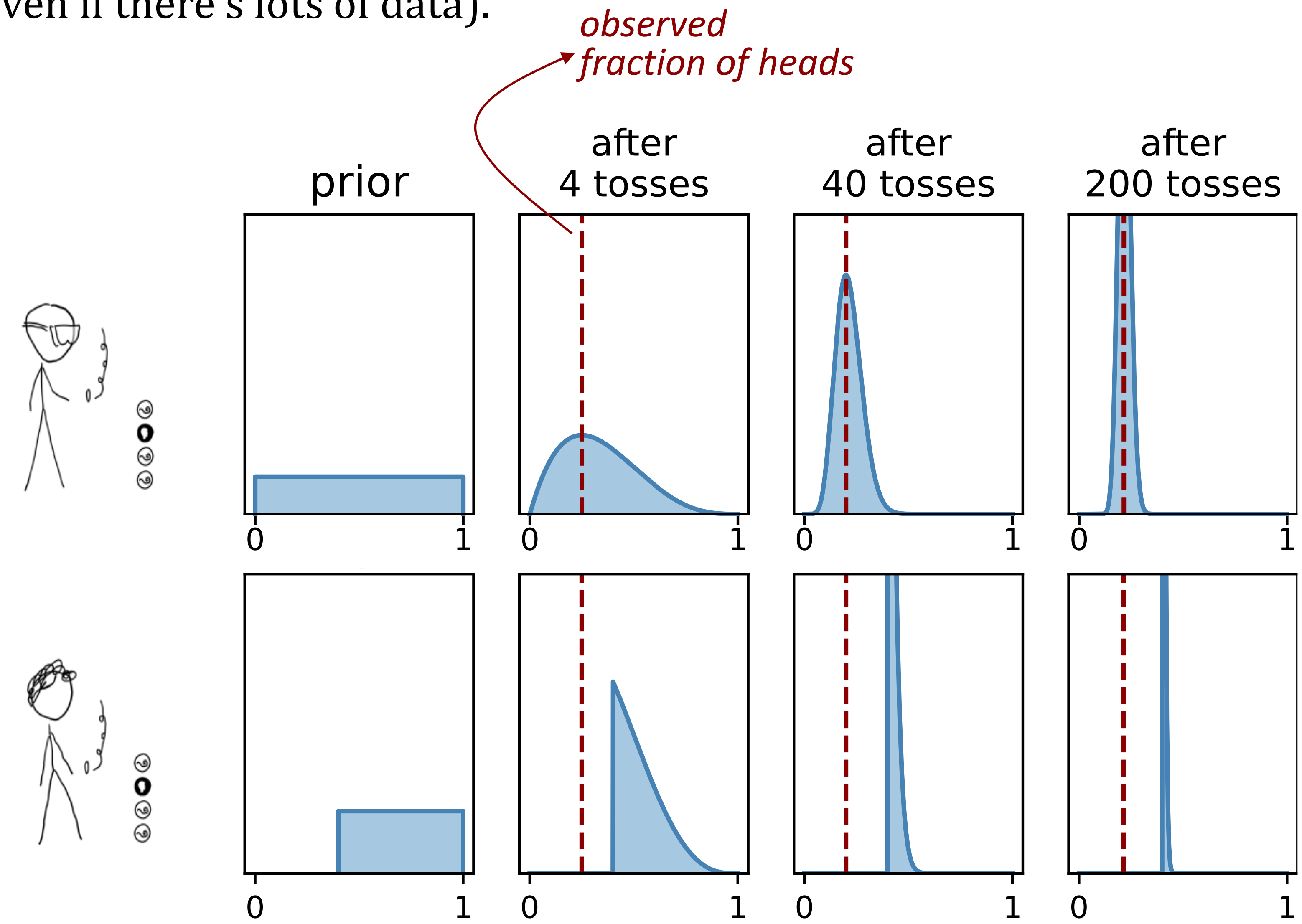
$\Pr_\Theta(\theta \mid X = x) \;= \kappa \Pr_\Theta(\theta)\,\Pr_X(x \mid \Theta = \theta)$



Preconception that $\theta > 0.6$ is impossible

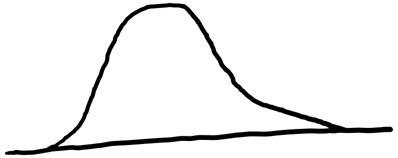
The preconception is unshakeable

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

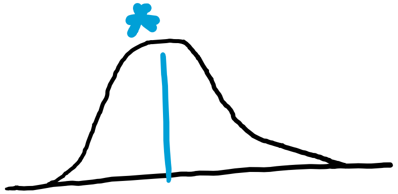


Prior distribution for Θ

Posterior distribution for Θ



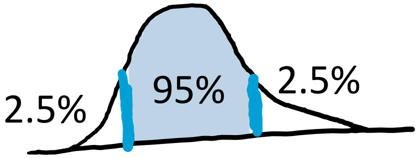
QUESTION.
How should we report the posterior distribution?



We could report the *posterior mean.*

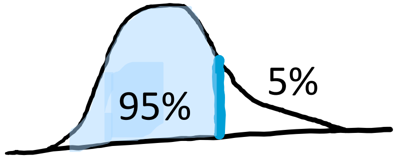
We could report the point with highest likelihood, the *MAP* or *maximum a-posteriori* estimate.

Example (Laplace smoothing).
We counted x successful outcomes from n trials.
Using the model $X \sim \mathrm{Bin}(n, \Theta)$, and the prior $\Theta \sim U[0,1]$, the posterior mean of Θ is $(x+1)/(n+2)$.



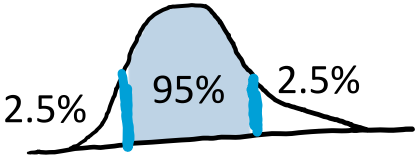
We could report a *95% confidence interval* [lo,hi] such that

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



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

$$\mathrm{lo} = -\infty$$
$$\mathbb{P}(\Theta > \mathrm{hi} \mid \mathrm{data}) = 5\%$$



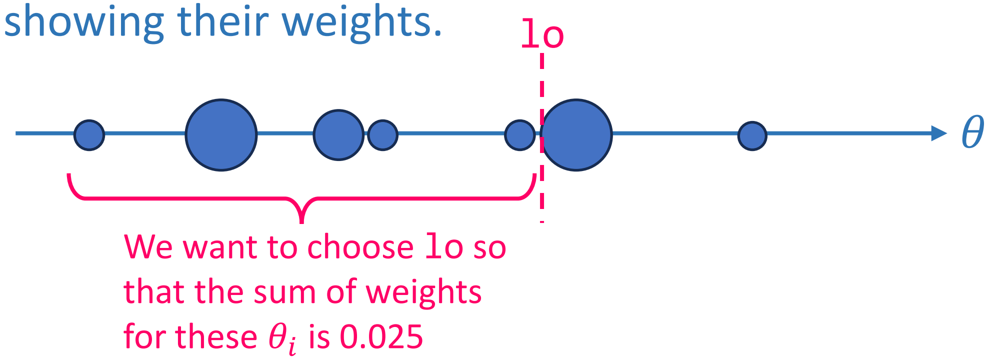
We could report a *95% confidence interval* [lo,hi] such that

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

(though this only really works well for continuous Θ,
as for discrete Θ we might not be able to hit those probabilities exactly)

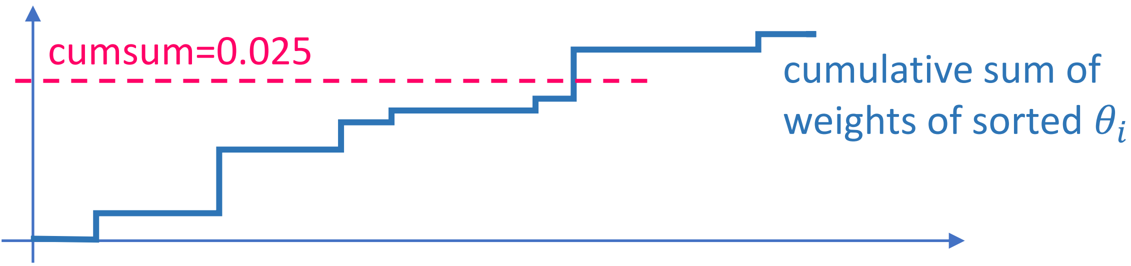
How can we compute lo and hi?

Consider a plot of the θ_i, showing their weights.



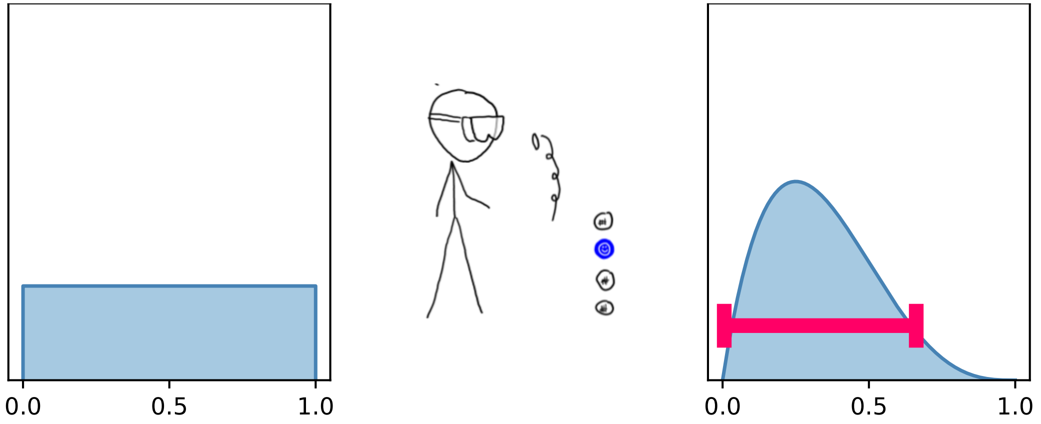
Via the computational Bayes estimate:

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

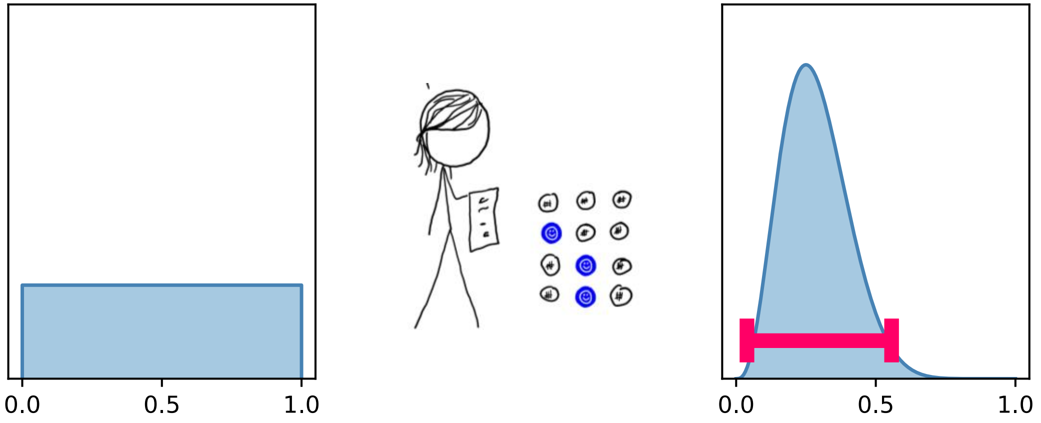



```
θsamp, w = …
i = np.argsort(θsamp)
θsamp, w = θsamp[i], w[i]
F = np.cumsum(w)
lo = θsamp[F<0.025][-1]
```

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



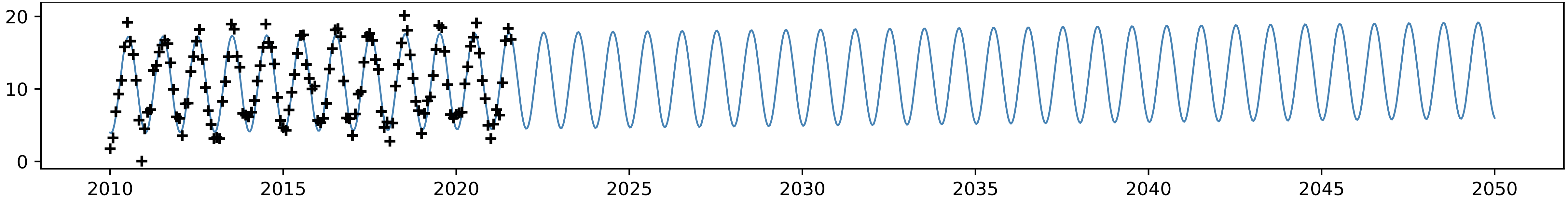
I estimate the probability of heads is 25%, and my 95% confidence interval is [3%, 72%]



I estimate the probability of heads is 25%, and my 95% confidence interval is [12%, 51%]

Consider the dataset of monthly average temperatures in Cambridge.

Proposed model: $\text{Temp} \sim \alpha + \beta \sin\big(2\pi(t+\phi)\big) + \gamma(t-2000) + N(0,\sigma^2)$



If we fit this model we get the maximum likelihood estimate $\hat{\gamma} = 0.027$ °C/year.

How **confident** are we about this value?

Climate confidence challenge.

Find a 95% confidence interval for γ,
for Cambridge from 1985 to the present.
(Use your own priors for the unknowns.)

Please submit your answer on Moodle
by Monday 6 November

§8.2 Asking the right question

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

Q. What don't we know?

Q. How do we represent unknowns?
Answer: As random variables, with a prior.

Q. What do we report?
Answer: The posterior distribution of the quantity of interest.

Q. How do we find this?
Answer: Using Bayes's rule.

Exercise 8.3.3 (Bayesian classification)

There are two types of expense claims, legitimate and fraudulent.
The legitimate claim sizes are $\sim \mathrm{Exp}(\lambda_L)$ and the fraudulent ones are $\sim \mathrm{Exp}(\lambda_F)$ where $\lambda_L = 0.1$ and $\lambda_F = 0.02$.
In my prior experience, 99% of claims I've seen are legitimate.
A new claim comes in, for an amount £x. Is it likely to be fraudulent?

What are we uncertain about?

whether the new claim is fraudulent

How do we represent uncertainty?

$$\text{Let } \Theta = \begin{cases} \ell & \text{if the new claim is legitimate} \\ f & \text{if it's fraudulent} \end{cases}$$

What is my prior?

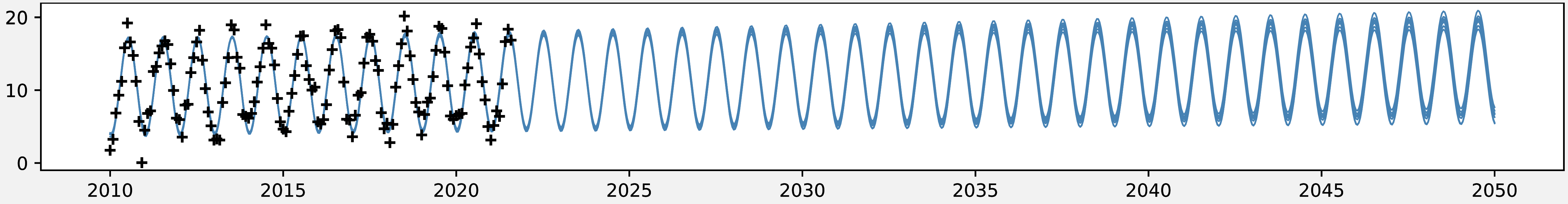
$\Pr_\Theta(\ell) = 0.99$ and $\Pr_\Theta(f) = 0.01$

What is the posterior I want to report?

$\Pr_\Theta(f \mid x)$
i.e. $\mathbb{P}(\Theta = f \mid x)$

Exercise.
Calculate $\mathbb{P}(\Theta = f \mid x)$.
(See lecture notes for solution.)

How should we express uncertainty about *predictions*?



I've fitted the model: $\text{Temp} \sim \alpha + \beta \sin\big(2\pi(t+\phi)\big) + \gamma(t-2000) + N(0,\sigma^2)$

I predict the temperature in January 2050 is $\text{pred}(2050) = \alpha + \beta \sin\big(2\pi(2050+\phi)\big) + 50\gamma$.

How confident am I about this prediction?

What are we uncertain about? The unknown parameters $\alpha, \beta, \phi, \gamma, \sigma$

How do we represent uncertainty? Treat the unknowns as random variables.

Concretely, we'll generate M samples $(\alpha_i, \beta_i, \phi_i, \gamma_i, \sigma_i)$, i=1,…,M, from our chosen prior, then compute weights w_i.

What do I want to report? The posterior distribution of pred(2050).

Each sample of the parameters gives a different prediction, call it $\text{pred}_i(2050)$.

Each sample also has an associated weight. Use these weights to find a confidence interval for pred(2050).

Why is this the right way to compute a confidence interval for a prediction?

(* non-examinable)

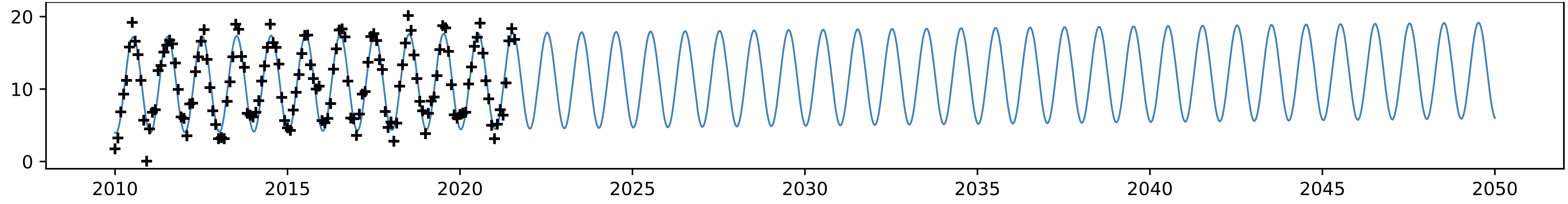
Let $h(\alpha,\beta,\varphi,\gamma,\sigma) = 1_{\text{pred}(2050\,;\,\alpha,\beta,\varphi,\gamma)\,\leq\,\text{lo}}$

$$\begin{aligned}
\mathbb{P}(\text{pred}(2050;\alpha,\beta,\varphi,\gamma) \leq \text{lo}) &= \mathbb{E}\, 1_{\text{pred}(2050\,;\,\alpha,\beta,\varphi,\gamma)\,\leq\,\text{lo}} && \text{since } \mathbb{E}1_{X\in A} = \mathbb{P}(X \in A) \\
&= \mathbb{E}\, h(\alpha,\beta,\varphi,\gamma,\sigma) && \text{by definition of } h \\
&\approx \sum_{i=1}^{n} w_i\, h(\alpha_i,\beta_i,\phi_i,\gamma_i,\sigma_i) && \text{by Computational Bayes} \\
&= \sum_{i=1}^{n} w_i\, \text{pred}_i(2050) && \text{where } \text{pred}_i \text{ is the prediction from the } i\text{th parameter sample}
\end{aligned}$$

How should we choose between two models?

Modeller 1: $\text{Temp} \sim \alpha + \beta \sin\big(2\pi(t+\phi)\big) + \gamma(t-2000) + N(0,\sigma^2)$

Modeller 2: $\text{Temp} \sim \alpha' + \beta' \sin\big(2\pi(t+\phi')\big) + N(0,\sigma'^2)$



What are we uncertain about? Which model is correct (and also all nine unknown parameters)

How do we represent uncertainty? With random variables.

Let M be a random variable saying which model is correct, M=1 or M=2. Invent a prior for it.

$$\Pr(\text{data} \mid \text{params}) = \Pr(\text{temp}_1,\dots,\text{temp}_n \mid M=m, \alpha,\beta,\varphi,\gamma,\sigma,\alpha',\beta',\varphi',\sigma') = \begin{cases} \cdots & \text{if } m=1 \\ \cdots & \text{if } m=2 \end{cases}$$

What do I want to report? The posterior distribution of M given the data. In other words, $\mathbb{P}(M=1 \mid \text{data})$.