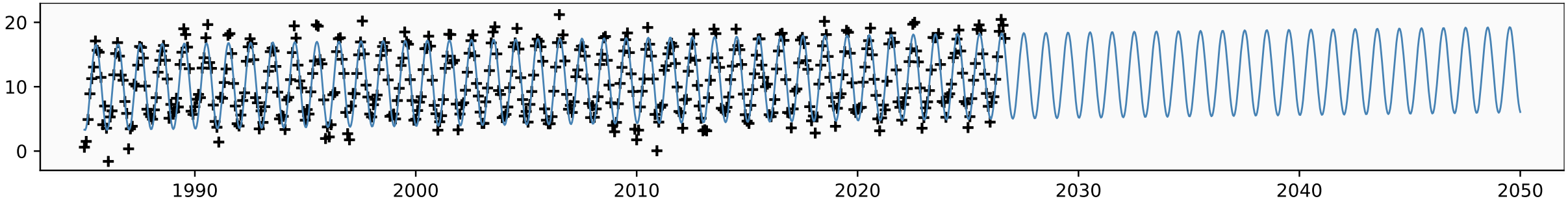
Consider the dataset of monthly average temperatures in Cambridge, since 1985.

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 $\hat{\gamma} = 0.042$ °C/year.

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



**THE TOTALITARIAN RULE**

Apply Bayes's rule to *all* the parameters and *all* the data:

$$\Pr(\text{all pars} \mid \text{all data})$$

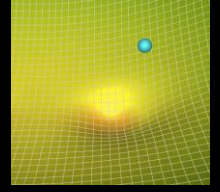


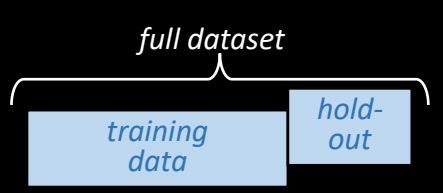
In problems with lots of datapoints, the likelihood is tiny. To avoid underflow, use the log-sum-exp trick (exercise 8.3.4)

## Climate confidence challenge.

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

Please submit your answer on Moodle.








| What can I say about ... | FITTING | HOLDOUT | BAYESIANISM |
|---|---|---|---|
| ... which model to choose? | MLE / prediction accuracy<br>Pick the model that best fits the data<br>⚠ overfitting | Pick the model that best fits *holdout data* |  |
| ... parameter values? | MLE / prediction accuracy<br>Pick the parameter values that give the best-fitting model | (a special case of model choice) | Treat unknown parameters as random variables<br>... and this will tell me how uncertain I should be |
| ... predicted outcomes? | | |  |

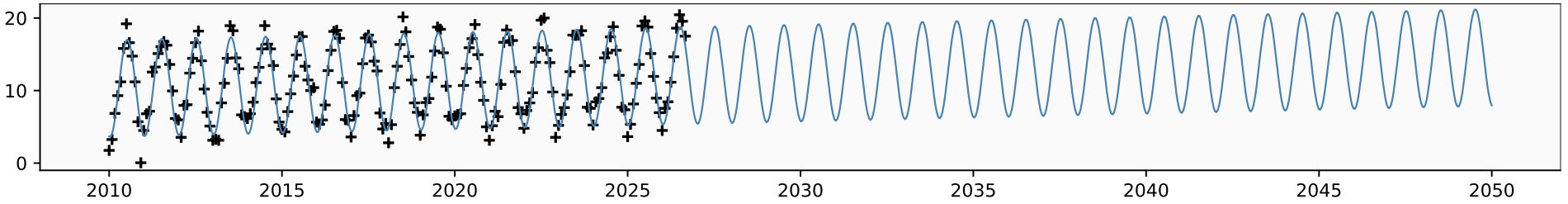
⚠ These methods only rank parameter values – *how much* should we trust the best answer?

# Choosing between models

Two scientists disagree about which climate model to use. Which model do you believe?

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

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



# THE BAYESIANIST LITANY



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

Q. What are we uncertain about?
*Varies from problem to problem*

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

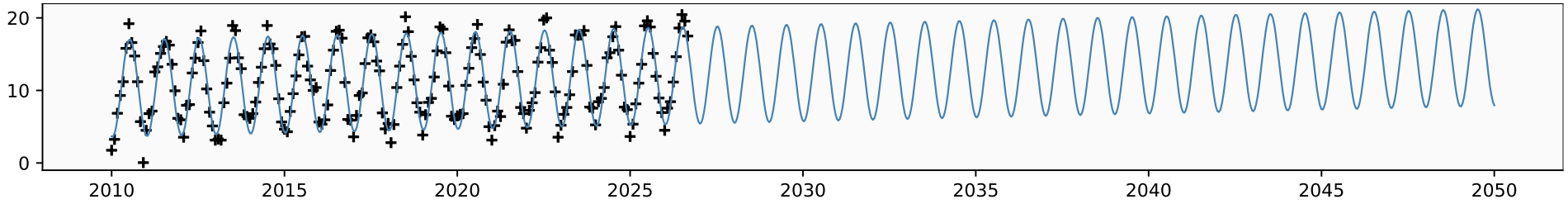
Q. What do we report?
*Answer: The posterior distribution of the quantity of interest*

Q. How do we find this?
*Answer: Using (totalitarian) Bayes's rule*

Two scientists disagree about which climate model to use. Which model do you believe?

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

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




What are we uncertain about?

How do we represent uncertainty?

Which model is correct and also about all 9 unknown pars.

With random variables, with prior distributions.

---

Let $M \in \{sci1, sci2\}$ say who is correct.

My prior: $\mathbb{P}(M = sci1) = 0.5$, $\mathbb{P}(M = sci2) = 0.5$

sci1 prior $Pr_{sci1}(\alpha, \beta, \phi, \gamma, \sigma)$

sci2 prior $Pr_{sci2}(\alpha', \beta', \phi', \sigma')$.

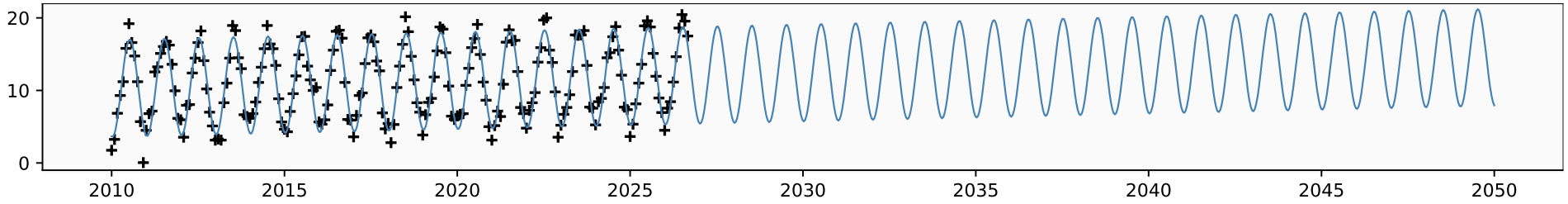
Data model:

$$Pr(data \mid m, \alpha, \beta, \phi, \gamma, \sigma, \alpha', \beta', \phi', \sigma') = \begin{cases} Pr_{sci1}(data \mid \alpha, \beta, \phi, \gamma, \sigma) & \text{if } m = sci1 \\ Pr_{sci2}(data \mid \alpha', \beta', \phi', \sigma') & \text{if } m = sci2. \end{cases}$$

Two scientists disagree about which climate model to use. Which model do you believe?

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

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




Write out the (totalitarian) prior:

$$\Pr(m, pars, pars') = \Pr_M(m)\,\Pr_{\text{sci1}}(pars)\,\Pr_{\text{sci2}}(pars')$$

I'm choosing independent priors

Write out the (totalitarian) data model:

$$\Pr(\text{data} \mid m, pars, pars') = \begin{cases} \Pr_{\text{sci1}}(\text{data} \mid pars) & \text{if } m = \text{sci1} \\ \Pr_{\text{sci2}}(\text{data} \mid pars') & \text{if } m = \text{sci2} \end{cases}$$

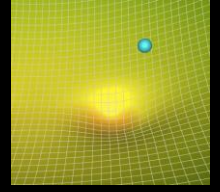
call M a "model switch" parameter

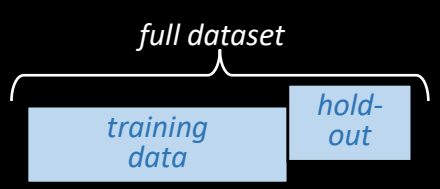
Get the posterior using Bayes's rule: mathematically or computationally

$$\Pr(m, pars, pars' \mid \text{data}) \propto \Pr(m, pars, pars')\,\Pr(\text{data} \mid m, pars, pars')$$

Report the parameters we're interested in (by marginalization):

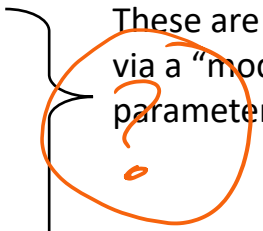
$$\Pr(m \mid \text{data}) = \text{<integrate out the other parameters>}$$







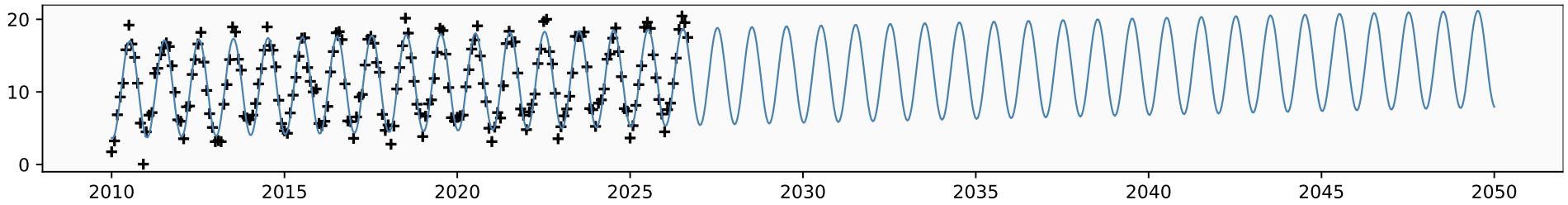

| What can I say about ... | FITTING | HOLDOUT | BAYESIANISM |
|---|---|---|---|
| ... which model to choose? | MLE / prediction accuracy<br>Pick the model that best fits the data<br>⚠ overfitting | Pick the model that best fits *holdout data* | These are the same thing, via a "model switch" parameter |
| ... parameter values? | MLE / prediction accuracy<br>Pick the parameter values that give the best-fitting model | (a special case of model choice) | Treat unknown parameters as random variables<br>... and this will tell me how uncertain I should be |
| ... predicted outcomes? | ⚠ These methods only rank parameter values – *how much* should we trust the best answer? | | |





What do you believe about the temperature in July 2050?

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




According to this model, temperature in July 2050 will be

$$T = \alpha + \beta \sin\big(2\pi(t_{\text{new}}+\phi)\big) + \gamma(t_{\text{new}}-2000) + N(0,\sigma^2) \quad \text{where} \quad t_{\text{new}} = 2050.54$$

$+\sigma E$ where $E \sim N(0,1)$

What are we uncertain about?

We're uncertain about the exact values of the parameters $\alpha, \beta, \phi, \gamma, \sigma$
We're also uncertain about what the month's noise $E$ will be.

To see the posterior distribution of $T$, we can take lots of samples of $(T|\text{data})$ and plot a histogram.
$T$ is a function of $pars$ and of $E$. So we need samples of $(pars, E \mid \text{data})$

The data is obviously informative about $pars$. But what does the data tell us about $E$?

nothing!

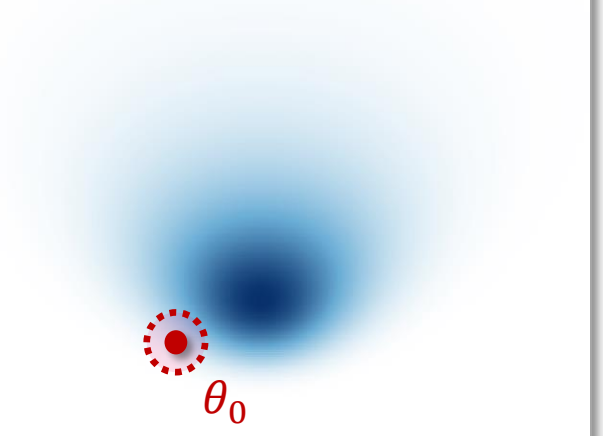
# We can approximate "sample from the posterior distribution" by drawing from the weighted samples from Computational Bayes.

⚠ It's a noisy procedure! If there's too much noise in your problem, plotting these weighted points can help understand why.

MATHEMATICAL BAYES

The true posterior distribution

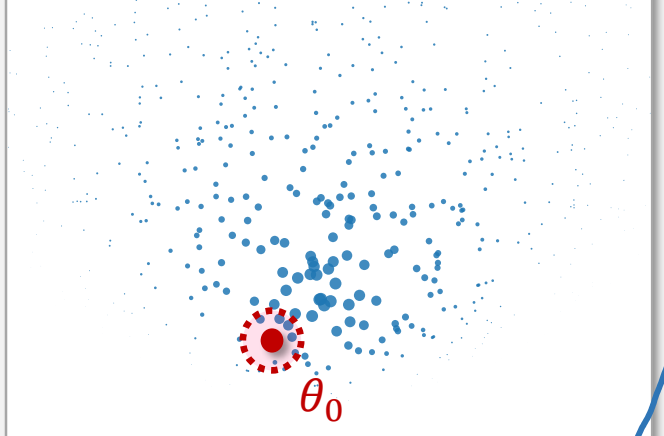
$\Pr_\Theta(\theta|\text{data})$




COMPUTATIONAL BAYES

Sample points $\theta_i$ from the prior $\Pr_\Theta$

Weight them by $w_i \propto \Pr(\text{data}|\theta_i)$




SETUP PHASE: prepare a bag $\{(\theta_i, w_i),\ i=1 \cdots \#\text{samples}\}$

USAGE PHASE Draw from our bag. Let $\hat{\Theta}$ be a value we draw.

If we sample randomly from the posterior distribution, what's the chance of getting a value in a small region around $\theta_0$?

$$\mathbb{P}(\Theta \in \text{region} \mid \text{data}) = \int_{\theta \in \text{region}} \Pr_\Theta(\theta|\text{data})\, d\theta$$

$$\approx \Pr_\Theta(\theta_0|\text{data}) \times \text{area}$$ assuming the function is continuous

$$\propto \Pr_\Theta(\theta_0)\Pr(\text{data}|\theta_0) \times \text{area}$$ by Bayes's rule

Pick one of the $\theta_i$ samples randomly, weighted by $w_i$. What's the chance of getting a value in a small region around $\theta$?

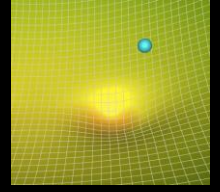
$$\mathbb{P}(\hat{\Theta} \in \text{region} \mid \text{data}) = \text{total weight in region}$$

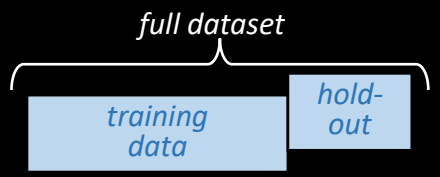
$$= \#\text{points in region} \times \text{avg weight in reg.}$$

$$\propto \Pr_\Theta(\theta_0) \times \text{area} \qquad \approx \Pr(\text{data}|\theta_0)$$

because of how we generate our $\theta_i$

because of how we weight them

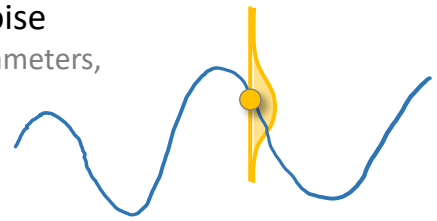


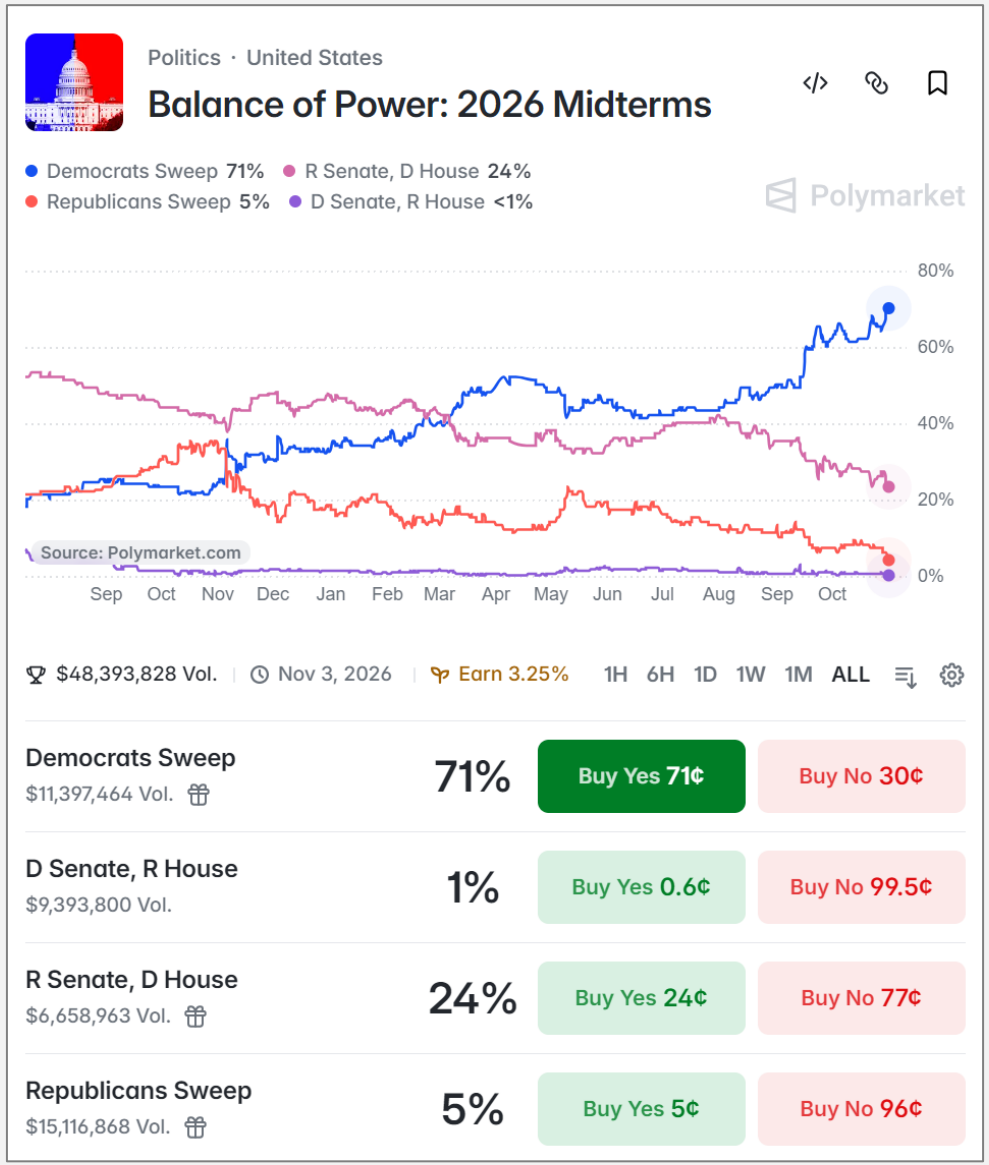





| What can I say about ... | FITTING | HOLDOUT | BAYESIANISM |
|---|---|---|---|
| ... which model to choose? | MLE / prediction accuracy<br>Pick the model that best fits the data<br>⚠ overfitting | Pick the model that best fits *holdout data* | These are the same thing, via a "model switch" parameter |
| ... parameter values? | MLE / prediction accuracy<br>Pick the parameter values that give the best-fitting model | (a special case of model choice) | Treat unknown parameters as random variables<br>... and this will tell me how uncertain I should be |
| ... predicted outcomes? | A prediction is a readout function of unknown parameters and of noise<br>... so use your best guess for the parameters, and your data-model for the noise | | A prediction is a readout function of unknown parameters and of noise<br>... so take account of the uncertainty in both. |

⚠ These methods only rank parameter values – *how much* should we trust the best answer?






**My model:**

I believe there is a "national sentiment" that describes swing voters, voter turnout, etc. This differs from one election to the next, and it's hard to predict. There's also uncertainty because of chance events, in each race.

**What am I uncertain about?**

- national sentiment, call it $\Theta$
- the winner of the house, $H \in \{\text{dem}, \text{rep}\}$ (1, 0)
- the winner of the senate, $S \in \{\text{dem}, \text{rep}\}$

Treat them all as random variables.

Here's a detailed model that's calibrated to the current prediction market:

$$\alpha = -3.1$$
$$\beta = -1.1$$
$$\tau = 1.61$$

$$\Theta \sim N(0, \tau^2)$$

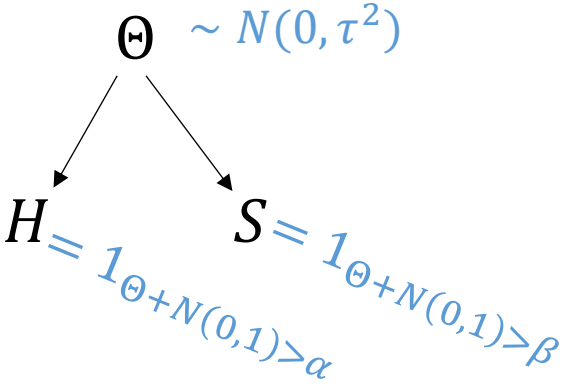
$$H = 1_{\Theta + N(0,1) > \alpha}$$

$$S = 1_{\Theta + N(0,1) > \beta}$$

epistemic: my uncertainty about the system, which reduces the more data I get

aleatoric: the intrinsic irreducible noise associated with an observation

NOTE: there's no Bayesian updating yet — this is merely a Bayesian model that we might propose, before we update it in the light of data.




*Interpretation of the "confidence ribbon":*

- *Both $H$ and $S$ involve comparing $\Theta + N(0,1)$ to a threshold*
- *By linearity of the Gaussian, $\Theta + N(0,1) \sim N(0,1+\tau^2)$*
- *Rule of thumb: a 95% confidence interval for $N(\mu,\sigma^2)$ is $\mu \pm 1.96\sigma$*
- *So, the confidence ribbon in this plot shows a 95% confidence interval for $\Theta + N(0,1)$, which we can then compare to $\alpha$ and $\beta$*

total uncertainty
(95% confidence)
$\pm 1.96\sqrt{1+\tau^2}$

