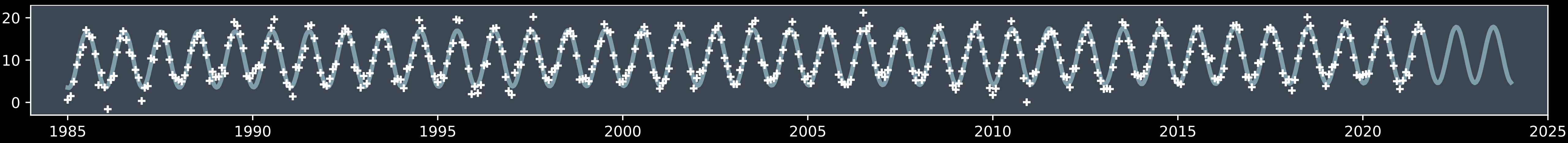
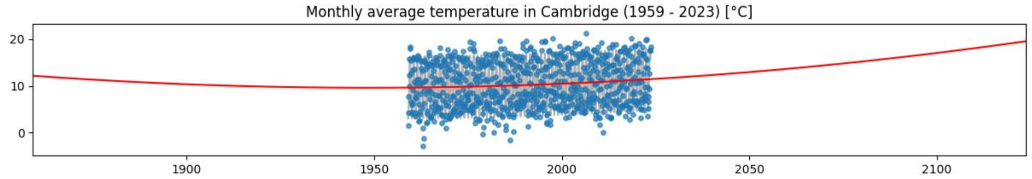
Climate challenge

- ❖ What is the rate of temperature increase at Cambridge?
- ❖ Are temperatures increasing at a constant rate, or has the increase accelerated?
- ❖ How do the results compare across the whole of the UK?

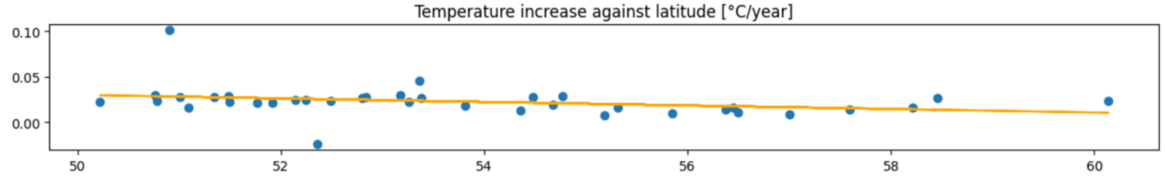
Your task is to answer these questions using appropriate linear models, and to produce elegant plots to communicate your findings.



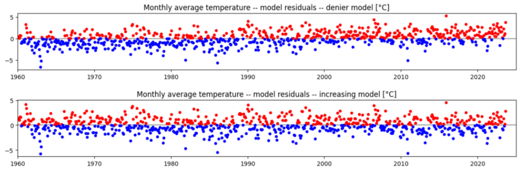
Anant Gupta (Fitzwilliam)



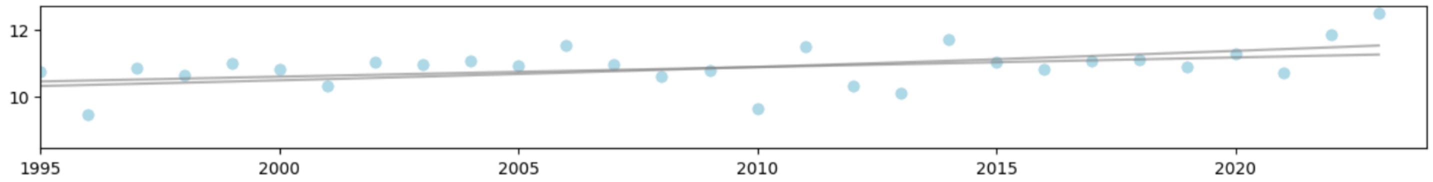
Jing Xuan Tan (Hughes Hall)



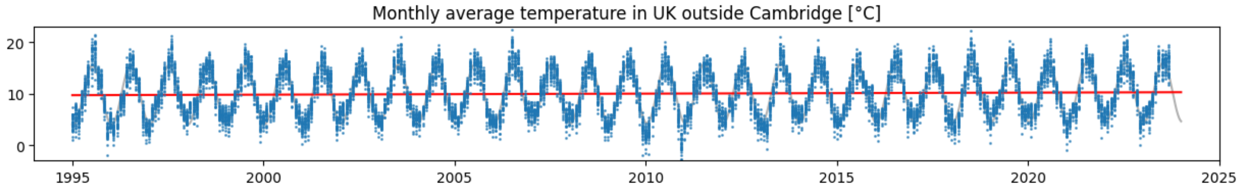
Joel Robinson (Emmanuel)



Paul D'Souza (Robinson)



Wei Chuen Sin (Hughes Hall)



Q1. What is the rate of temperature increase in Cambridge?

Fit the model:

$$\text{Temp} \approx \alpha + \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t) + \gamma t$$

```
X = np.column_stack([np.sin(2*π*df.t), np.cos(2*π*df.t), df.t])
model = sklearn.linear_model.LinearRegression()
model.fit(X, df.temp)
α,(β1,β2,γ) = (model.intercept_, model.coef_)
γ
```

- 0.028°C per year [1959 to present]
- 0.025°C per year [???]

Always start by saying exactly what data you're working with.

It's a great idea to run sensitivity analyses. Is my answer robust, if I look at a different subset of the data?

Q2. Are temperatures increasing at a constant rate?

To see if there's a sign of nonlinearity, fit the model:

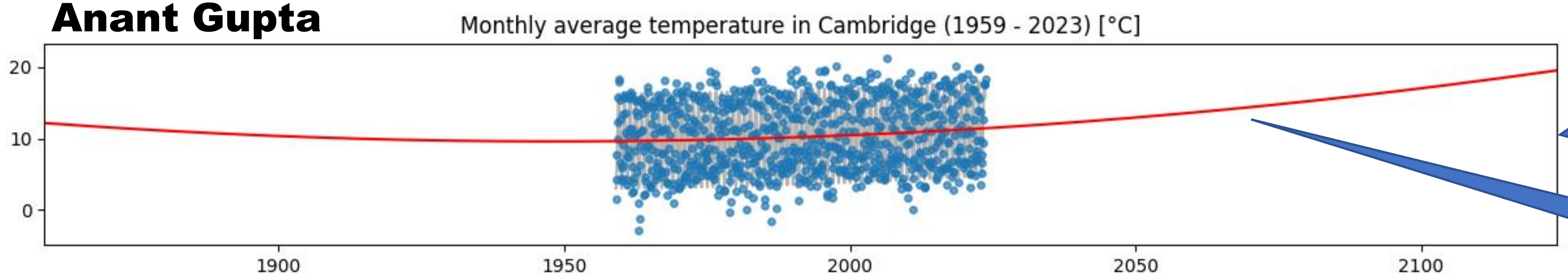
$$\text{Temp} \approx \alpha + \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t) + \gamma t + \delta t^2$$

Conclusion: $\delta = 0.00032$

Report a few significant figures, rather than "$\delta = 0.000$"

Joel Robinson. This change is very small so may be insignificant. Assuming that the change *is* significant, we can conclude that the temperature change is increasing and accelerating. However since we have no data for values past 2024 it would be unwise to try to extrapolate what future temperature values may be from this model.

Anant Gupta



Report your fitted model in meaningful units (e.g. impact on predicted response, not just raw coefficients)

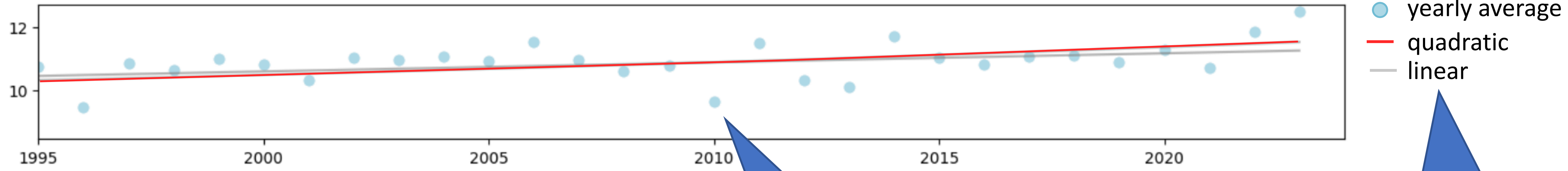
Show the context

Q2. Are temperatures increasing at a constant rate?

To see if there's a sign of nonlinearity, fit the model:

$$\text{Temp} \approx \alpha + \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t) + \gamma t + \delta t^2$$

Paul D'Souza



Make it easy to compare models

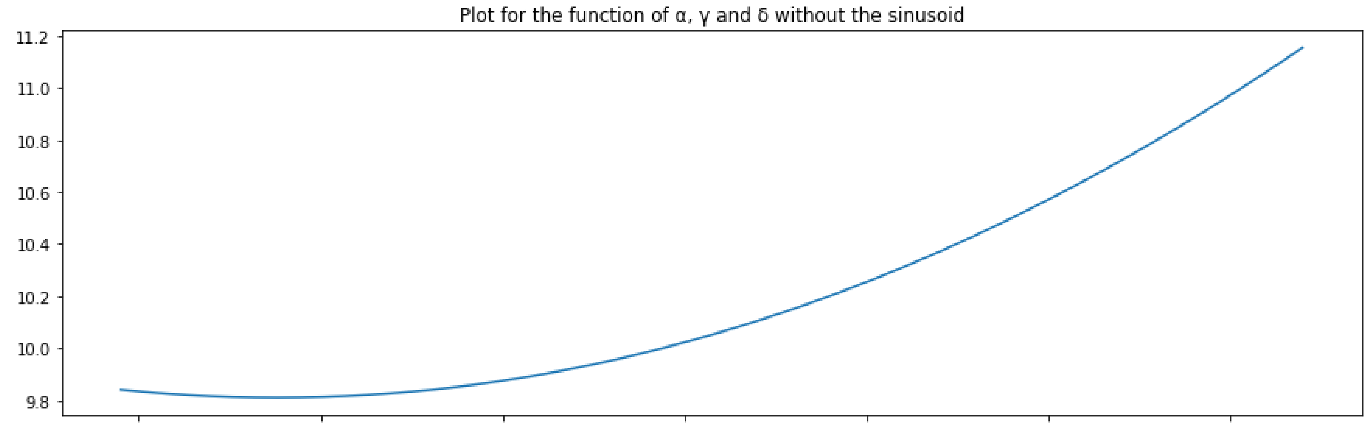
Make it easy to compare models to the data.

… but it's a sin to waste data! Use *all* the data to fit your model (if your model is expressive enough to use it)

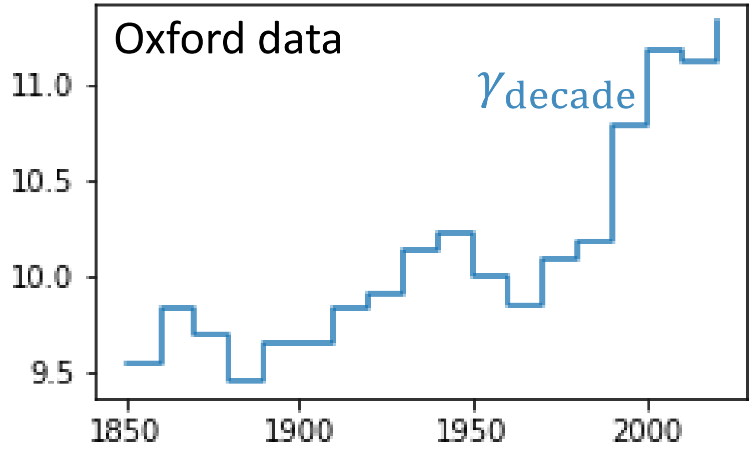
PRECONCEIVED BELIEFS

OPEN TO ANY EXPLANATION

$$\text{temp} \approx \alpha + \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t) + \gamma t + \delta t^2$$



$$\text{temp} \approx \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t) + \gamma_{\text{decade}}$$



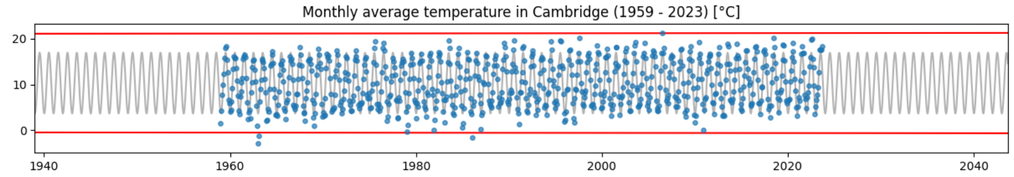
Q2. Are temperatures increasing at a constant rate?

What about other models for non-linearity?

Wei Chuen Sin
"from climate science, we know that temperature is rising at an exponential rate"

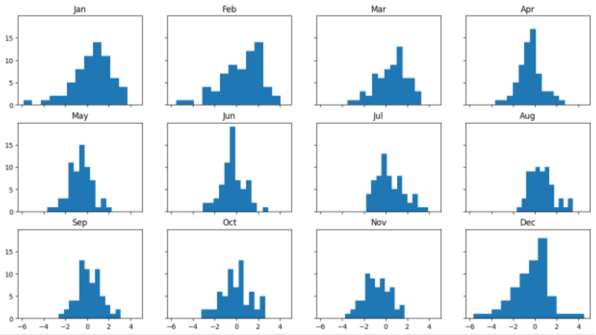
Anant Gupta
$\text{Temp} \approx \alpha + \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t) + \gamma e^{\delta(t-\varepsilon)}$ (needs scipy.optimize.fmin)

Anant Gupta
$\text{Temp} \approx \alpha + (\beta_1 + \gamma_1 t) \sin(2\pi t) + (\beta_2 + \gamma_2 t) \cos(2\pi t)$



Paul D'Souza
"What if the periodic part isn't a pure sinusoid?"
→ let's look at yearly averages instead of the full data

Anant Gupta
"The residuals are too low in Jan/Feb/Mar, too high for the rest of the year, so the sinusoid isn't a great fit."

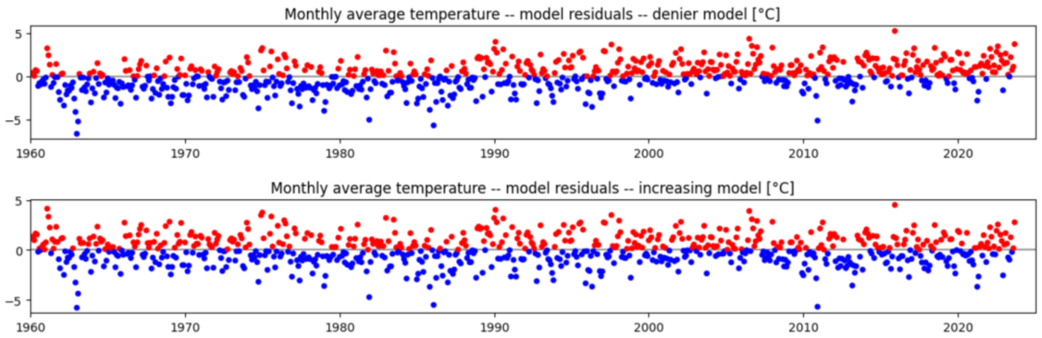


How should we compare models?

Joel Robinson

The mean square error, $n^{-1}\sum_{i=1}^{n}(y_i - \mathrm{pred}_i)^2$, measures how well a model fits.

"It seemed that our model better fitted weather station readings from the North of the UK; the mean residuals squared value was smaller for Bradford, Tiree and Armagh than for Oxford, Cambridge and Heathrow. This may suggest that the north is experiencing climate change at a faster rate than the south and is therefore more suited to a quadratic model."



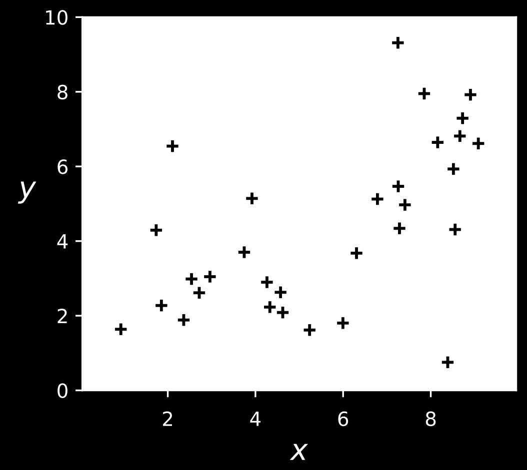
MSE = 2.40 for the no-change model

MSE = 2.10 for the linear-increase model

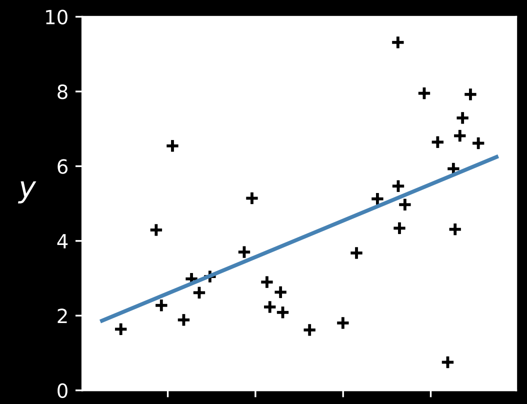
Paul D'Souza

`sklearn.linear_model.LinearRegression.score`

This measures R^2, which is a transformed version of MSE.



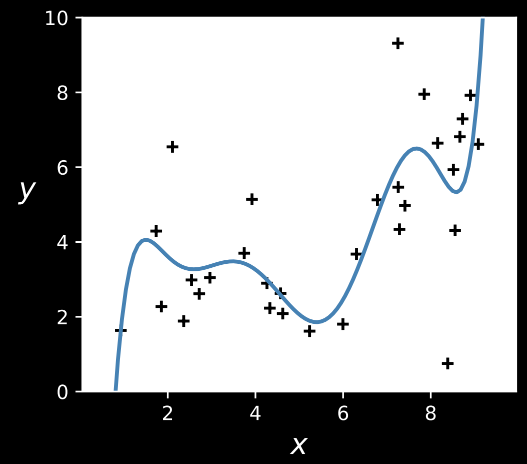
dataset of (x_i, y_i) pairs



Model A:

$$Y_i \sim 1.62 + 0.49\, x_i + \text{Normal}(0, 2.39^2)$$

MSE large



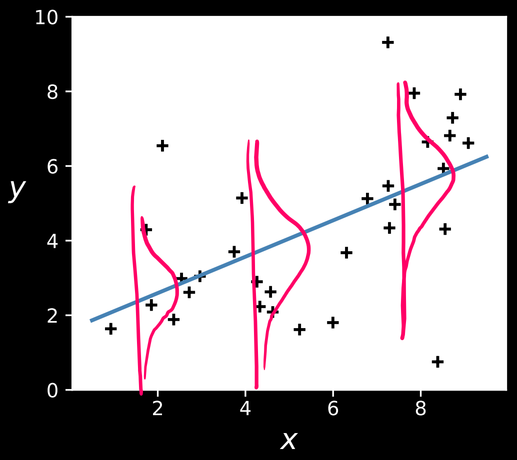
Model B:

$$Y_i \sim -38.5 + 95.7\, x_i - 84.8\, x_i^2 + 38.3\, x_i^3 - 9.5\, x_i^4 + 1.3\, x_i^5 - 0.09\, x_i^6 + 0.003\, x_i^7 + \text{Normal}(0, 0.31^2)$$

MSE small

This model doesn't just predict a value for y.

It predicts a distribution Y, at every x.

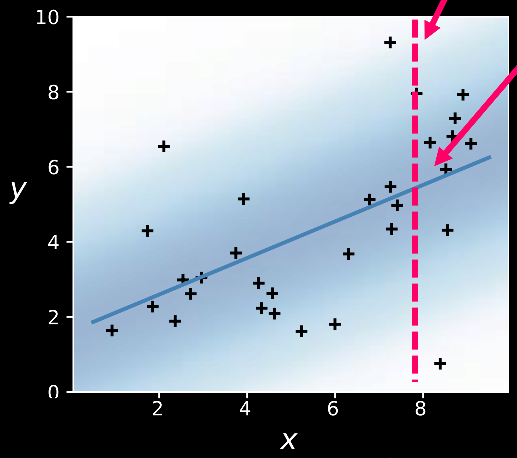


Model A:

$$Y_i \sim 1.62 + 0.49\, x_i + \text{Normal}(0, 2.39^2)$$

Area of low likelihood

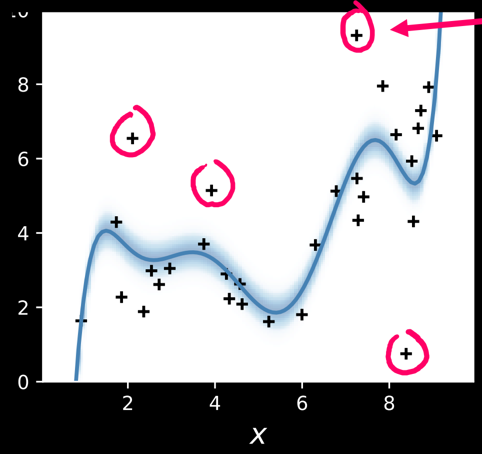
Area of high likelihood



Model A:

$$Y_i \sim 1.62 + 0.49\, x_i + \text{Normal}(0, 2.39^2)$$

These points are very unlikely to have been generated by this model



Model B:

$$Y_i \sim -38.5 + 95.7\, x_i - 84.8\, x_i^2 + 38.3\, x_i^3 - 9.5\, x_i^4 + 1.3\, x_i^5 - 0.09\, x_i^6 + 0.003\, x_i^7 + \text{Normal}(0, 0.31^2)$$

There are several datapoints y_i where model B says "The likelihood of this y_i is vanishingly small." But these y_i did appear in the dataset. So model B is a bad explanation.

MODEL EVALUATION AND COMPARISON

After we fit a model, how do we decide if it's a good fit?

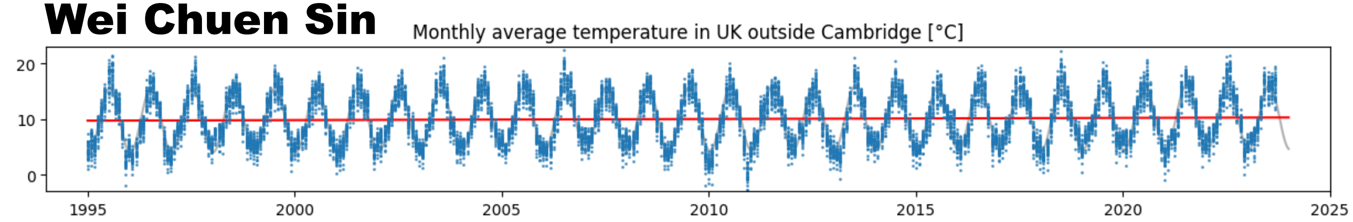
1. Evaluate the ~~mean square error~~ log likelihood of the dataset
2. Plot the ~~residuals~~ log likelihood of each datapoint, and look for systematic patterns.

Q3. How do the results compare across the UK?

We could model the entire dataset as

$$\text{Temp} \approx \alpha + \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t) + \gamma t$$

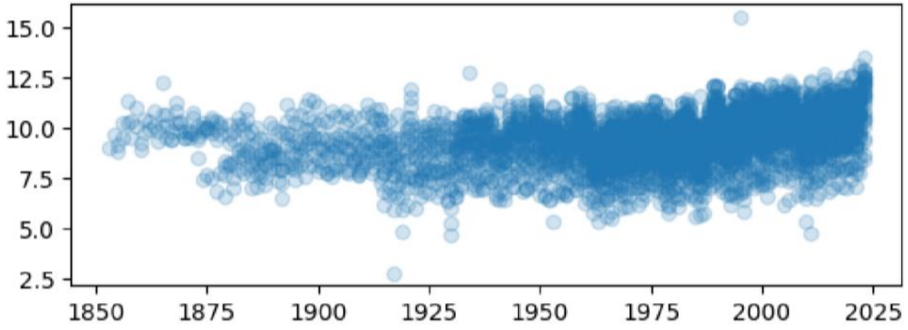
Wei Chuen Sin



(not the entire dataset, but only the subset for which all stations are present!)

It's a really useful sanity check to show the "disposition" of the entire dataset.

Paul D'Souza



Q3. How do the results compare across the UK?

We could model each station individually:

for s in stations:

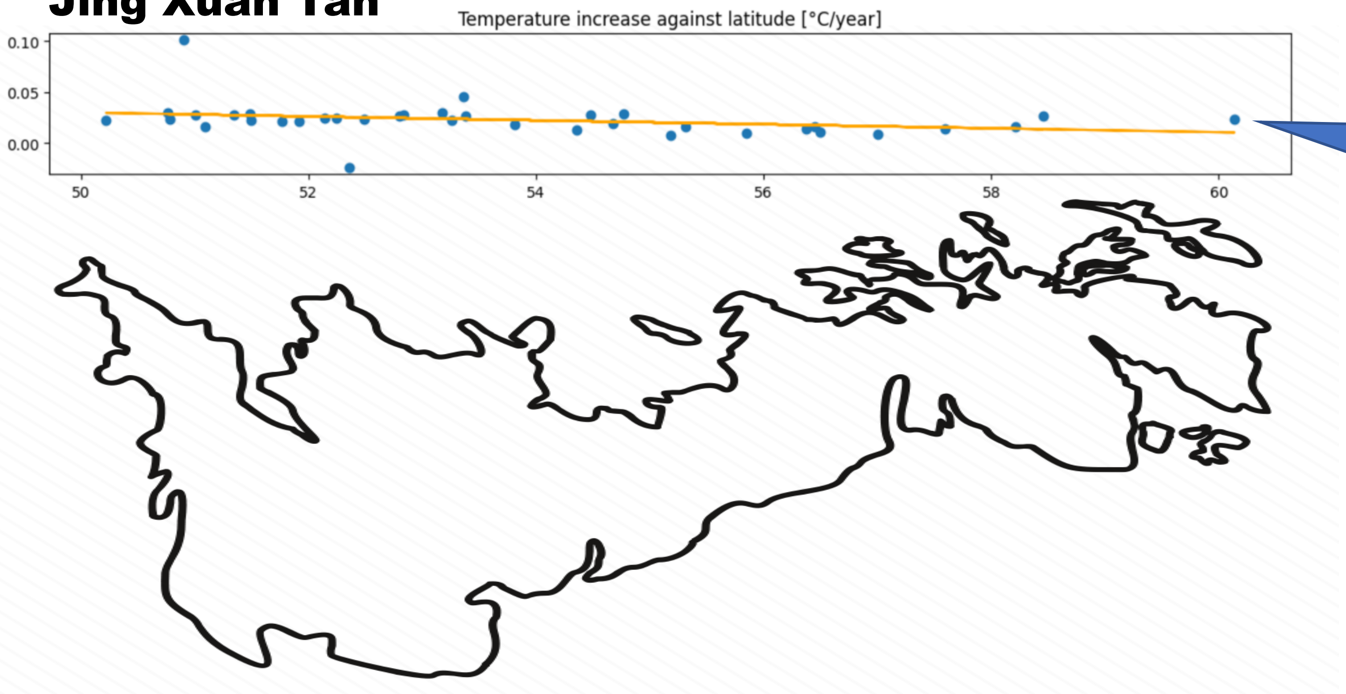
model data from station s as $\mathrm{Temp} \sim \alpha + \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t) + \gamma t + N(0, \sigma^2)$

Or, use one-hot coding to extract per-station coefficients:

$$\mathrm{Temp} \approx \alpha_{\mathrm{station}} + \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t) + \gamma_{\mathrm{station}} t$$

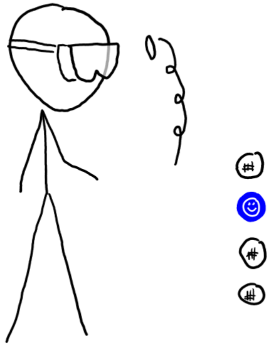
IMHO it's always cleaner to build a single model for your entire dataset.

Jing Xuan Tan



It's very powerful to be able to extract coefficients and plot them all together.

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



probability of
heads, unknown

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

$\downarrow$

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

number of heads
from 4 coin tosses

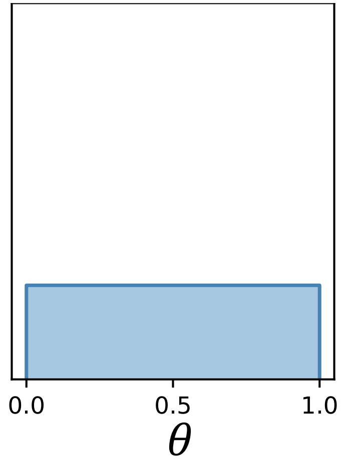
prior belief
$\text{Pr}_\Theta(\theta)$

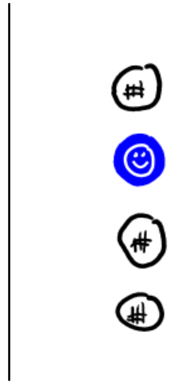
+

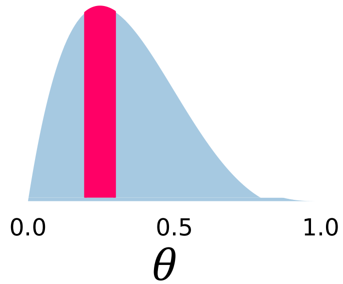
data
x

→

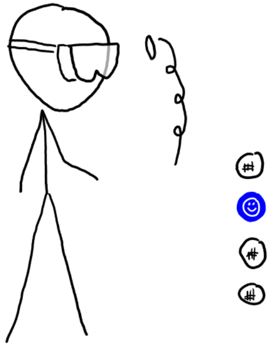
posterior belief
$\text{Pr}_\Theta(\theta | X = x)$







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



probability of heads, unknown

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

↓

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

number of heads from 4 coin tosses

0. First write out our probability model for the data $\Pr_X(x|\Theta=\theta)$
1. Write out $\Pr_\Theta(\theta)$
2. Use the formula $\Pr_\Theta(\theta|X=x) = \kappa \Pr_\Theta(\theta) \Pr_X(x|\Theta=\theta)$ then find κ to make this integrate to 1

This lets us calculate probabilities:

$$\mathbb{P}(\Theta \in \text{range}|X=x) = \int_{\theta \in \text{range}} \Pr_\Theta(\theta|X=x)\, d\theta$$

Exercise.

Consider the pair of random variables (Θ, X) where

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

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

$\Pr_\Theta(\theta) =$ 1 for $\theta \in [0,1]$

$\Pr_X(x|\Theta = \theta) =$ $\binom{n}{x} \theta^x (1-\theta)^{n-x} = 4\,\theta\,(1-\theta)^3$ for $n=4, x=1$

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

↑ a function of θ

$= \kappa \times 1 \times 4\,\theta\,(1-\theta)^3$

$= \kappa'\,\theta\,(1-\theta)^3$ κ' amalgamates non-θ terms.

$\int_0^1 \kappa'\,\theta(1-\theta)^3\,d\theta = 1 \Rightarrow \kappa' = \dfrac{1}{\int_0^1 \theta(1-\theta)^3\,d\theta}.$

Exercise.

Consider the pair of random variables (Θ, X) where

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

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

Beta

Probability density function	
Notation	$\text{Beta}(\alpha, \beta)$
PDF	$\dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$ where $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and Γ is the Gamma function.

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

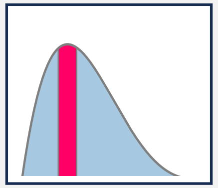
$$= \kappa\, \theta\, (1-\theta)^3 = \kappa\, B(\alpha,\beta)\, \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} \quad \text{where } \alpha = 2, \ \beta = 4$$

this is a standard pdf

so this constant must be 1 (otherwise this pdf wouldn't integrate to 1 w.r.t. θ)

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

What is $\mathbb{P}(\Theta \in [.2,3] \mid X = 1)$?



```
D = scipy.stats.beta(a=2,b=4)
D.cdf(.3) - D.cdf(.2)
```

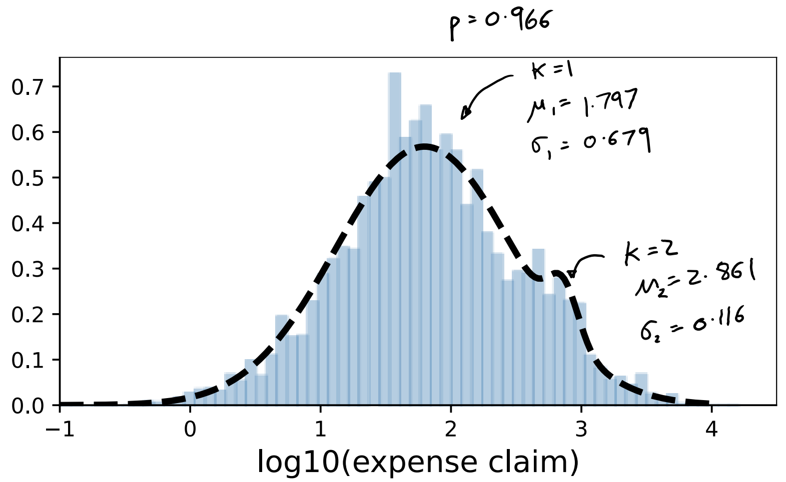
Exercise 5.2.3 (classification)

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,

$$C = \begin{cases} 1 & \text{with prob } p \\ 2 & \text{with prob } 1-p \end{cases}$$

$$Y \sim \text{Normal}(\mu_C, \sigma_C^2)$$

Find the probability that a claim amount £5000 belongs to the component $c = 2$.



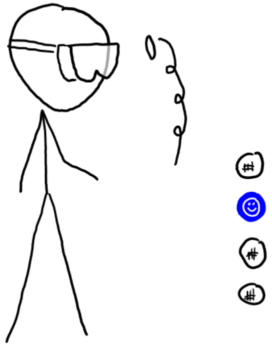
$$\Pr_C(c) =$$

$$\Pr_Y(y|C = c) =$$

$$\Pr_C(c|Y = y) = \kappa \Pr_C(c) \Pr_Y(y|C = c)$$

Exercise.

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



probability of heads, unknown

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

↓

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

number of heads from 4 coin tosses

0. First write out our probability model for the data $\Pr_X(x|\Theta=\theta)$

1. Write out $\Pr_\Theta(\theta)$

2. Use the formula $\Pr_\Theta(\theta|X=x) = \kappa \Pr_\Theta(\theta)\Pr_X(x|\Theta=\theta)$ then find κ to make this integrate to 1

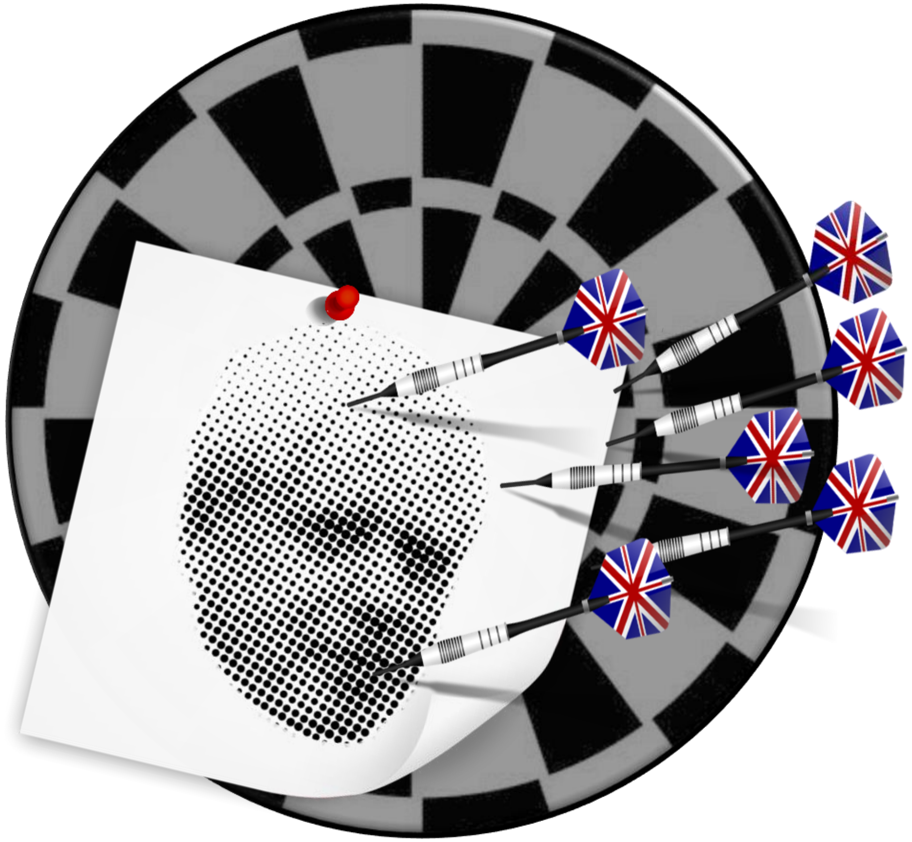
… but these are usually intractable

This lets us calculate probabilities:

$$\mathbb{P}(\Theta \in \text{range}|X=x) = \int_{\theta \in \text{range}} \Pr_\Theta(\theta|X=x)\, d\theta$$

§6. 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

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

$$1_{x_i \in A} = \begin{cases} 1 & \text{if } x_i \in A \\ 0 & \text{else} \end{cases}$$

```
# 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)    10,000 Booleans
np.mean(i)
```

typecast bool to int.

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

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

If we want to know the average properties of a rich random variable (random images, random texts), we have to use real-valued property readout functions h(X) so that we can take averages.

Monte Carlo integration

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

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

let $h(x) = 1_{x \in A}$

By Monte Carlo, $x_1 \dots x_n$ sampled from X

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

$$\frac{1}{n} \sum_{i} 1_{x_i \in A}$$

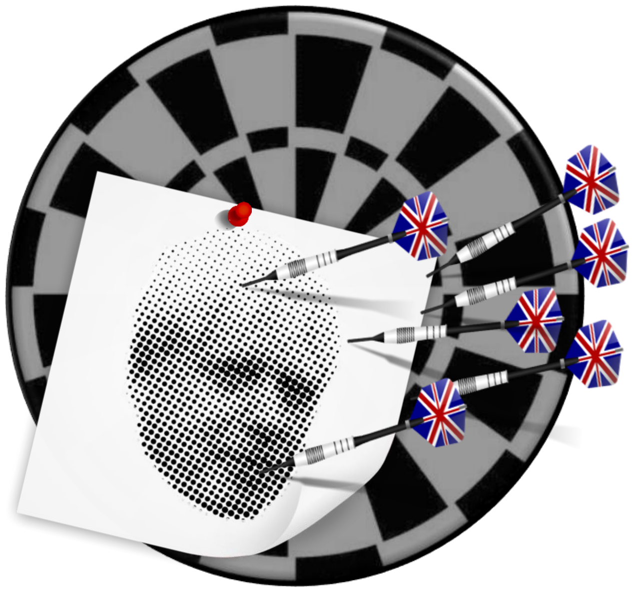
let $Y = h(X)$

$$\mathbb{E} Y = 0 \times \mathbb{P}(Y=0) + 1 \times \mathbb{P}(Y=1)$$
$$= \mathbb{P}(Y=1)$$
$$= \mathbb{P}(h(X)=1)$$
$$= \mathbb{P}(1_{X \in A} = 1)$$
$$= \mathbb{P}(X \in A)$$

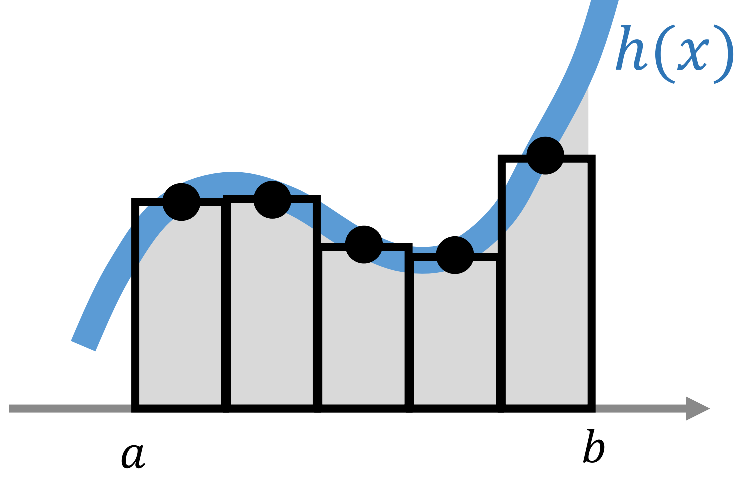
Let X be the location of a randomly thrown dart, and let $x_1, \dots, x_n$ be some throws.

The probability of hitting A is

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



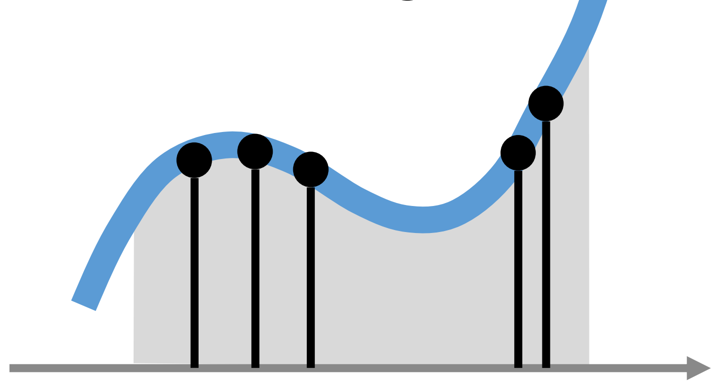
Trinity College integration



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

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,

$$\underbrace{\mathbb{E}h(X)}_{} \approx \frac{1}{n}\sum_{i=1}^{n} h(x_i) \quad \text{where } x_1, \dots, x_n \text{ sampled from } X$$

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

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
- The approximation for $\mathbb{E}h(X)$ 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*