Markov modelling III:

behaviour of Markov chains

§11.3 – §11.6

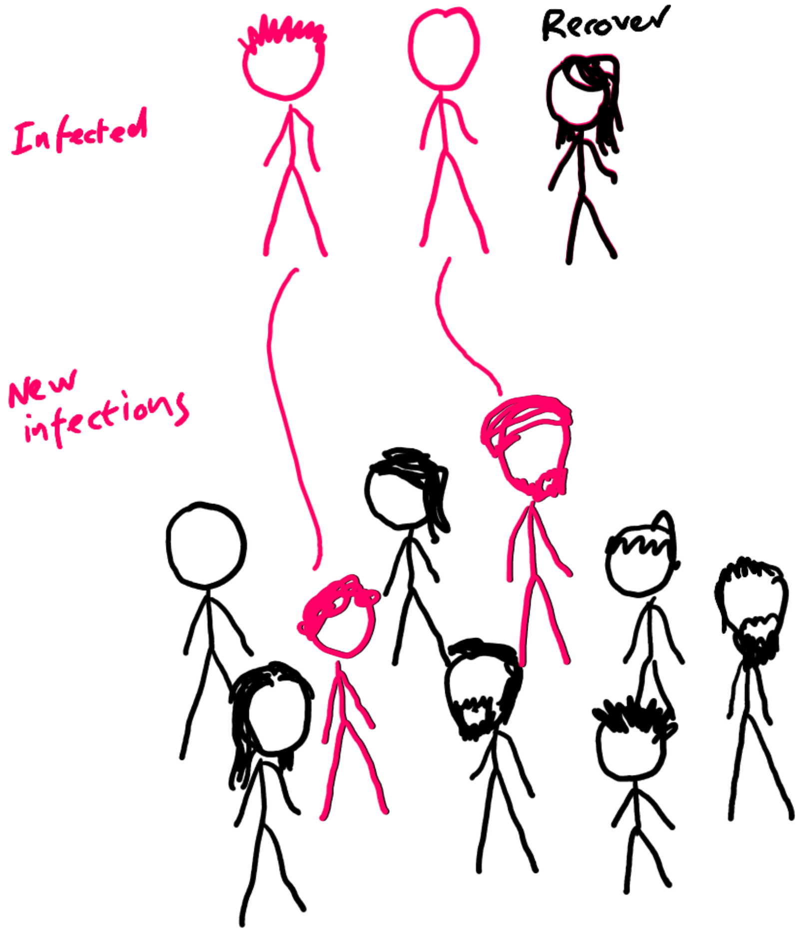
Example 11.1.2: dynamical system model of an epidemic

Let $X_n \in \mathbb{N}$ be the number of infected people on day n, and let it evolve according to

$$X_{n+1} = X_n - \mathrm{Recoveries}_n + \mathrm{Infections}_n$$

(We'll let the distributions of $\mathrm{Recoveries}_n$ and $\mathrm{Infections}_n$ depend only on X_n, making this a Markov model.)

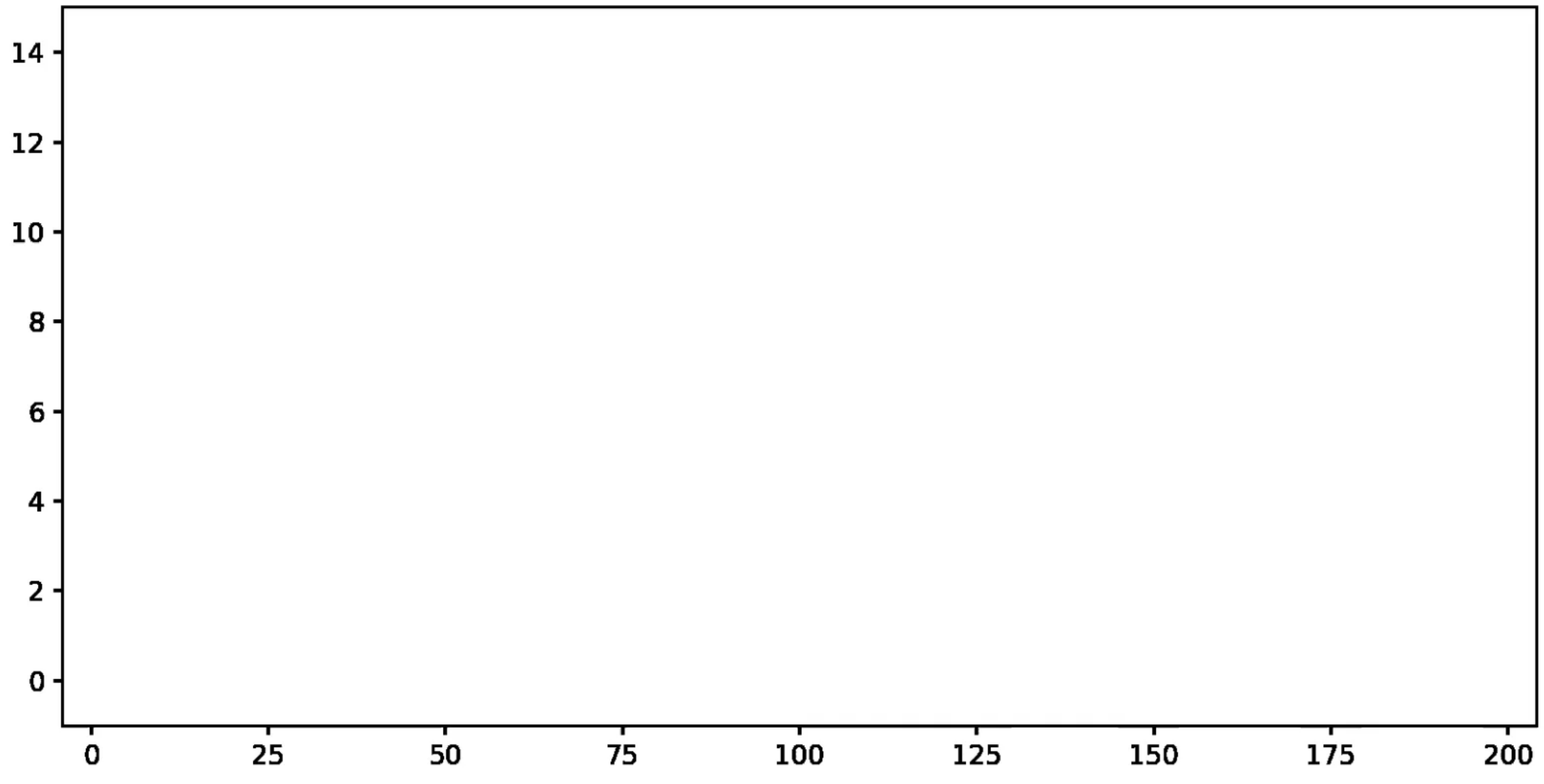
~~Day~~ ~~2~~ Day 1: # ~~infected~~ = ~~3~~ 3 + 2 - 1 = 4



Example 11.1.2: dynamical system model of an epidemic

Let $X_n \in \mathbb{N}$ be the number of infected people on day n

num. infected X_n
(5 simulation runs)



time n

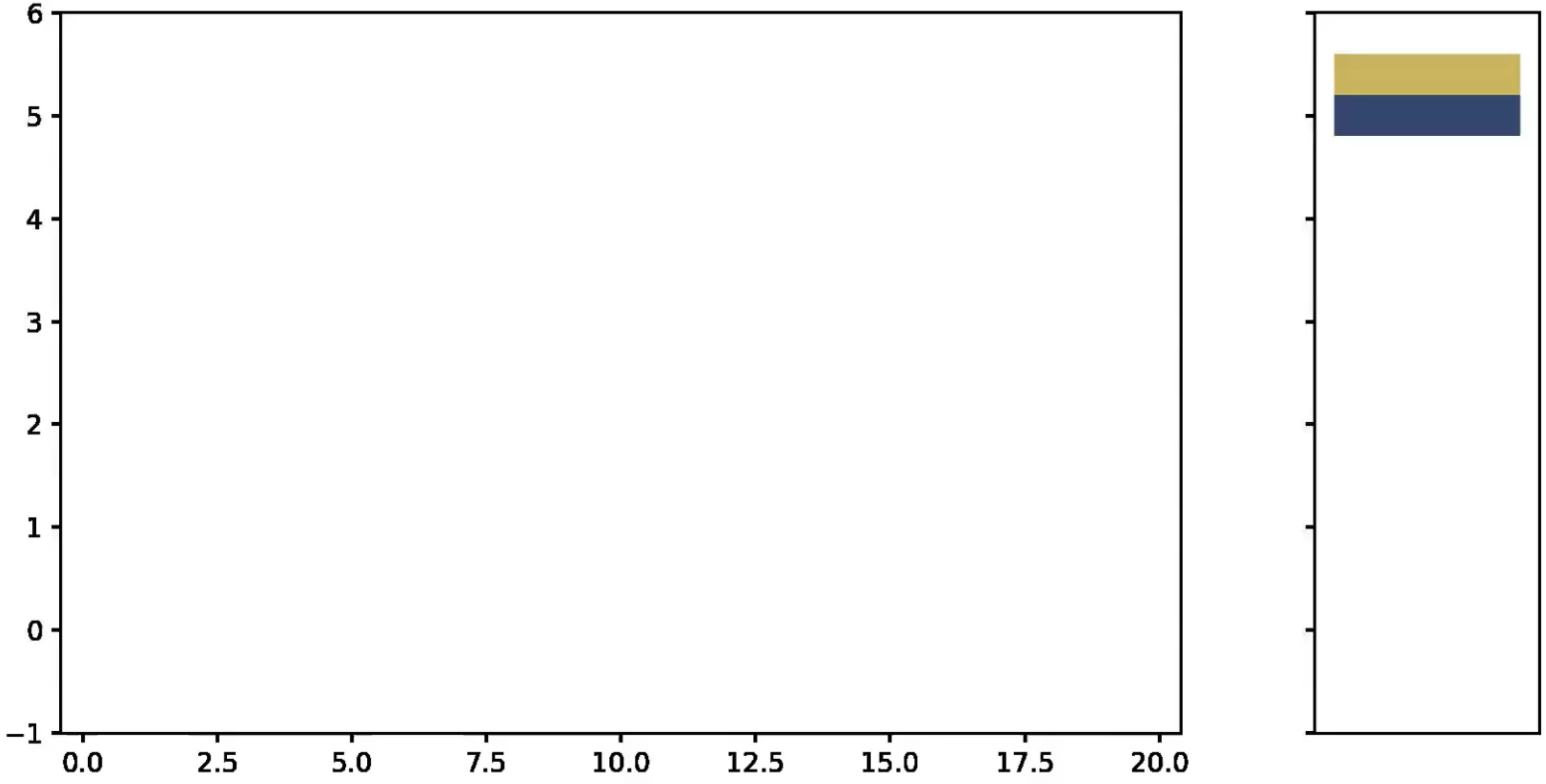
Example 11.1.3 dynamical system model of an online platform

Let $X_n \in \mathbb{N}$ be the number of users currently using an online platform at timestep n, and let it evolve according to

$$X_{n+1} = X_n + \text{Newusers}_n - \text{Departures}_n$$

(We'll let the distributions of Newusers_n and Departures_n depend only on X_n, making this a Markov model.)

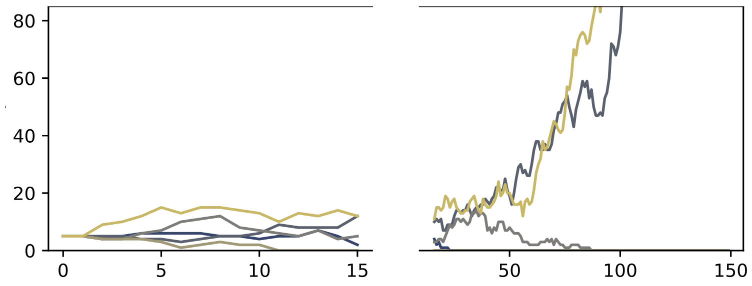
num. users X_n (2 simulation runs)



time n

Markov models can have very different behaviours, depending on the distribution of $(X_{n+1}|X_n)$.

EPIDEMIC MODEL

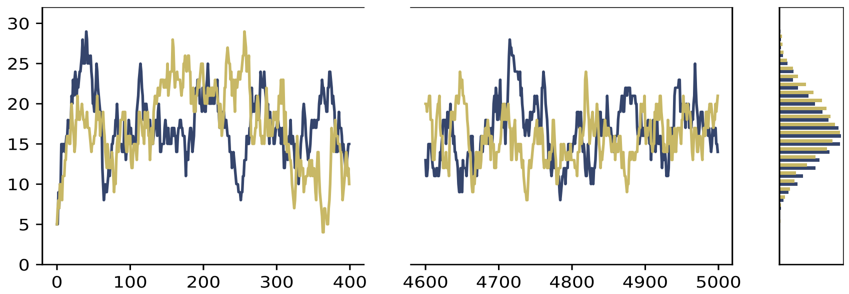


~~It might collapse or explode.~~

- ~~How likely is it that the epidemic dies out?~~
- ~~If it doesn't die out, what's the growth rate?~~

no longer on the syllabus

ACTIVE USERS MODEL



It might settle down to a stable **stationary distribution**.

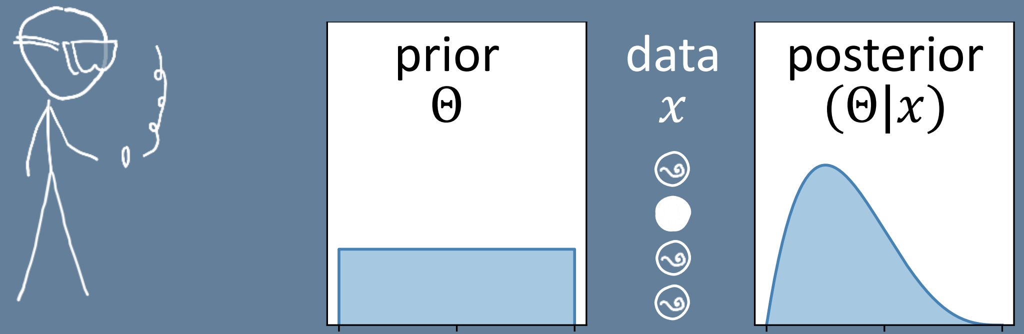
- How can we calculate this distribution?

Applications of Markov chains: better computational Bayes

Bayesian problem setup

Unknown parameter Θ, data model $(X|\Theta=\theta)$

We want the posterior belief about the parameter $(\Theta|X=x)$



Markov Chain Monte Carlo

1. Devise a Markov chain $\Theta_1 \to \Theta_2 \to \cdots$ whose stationary distribution is $(\Theta|X=x)$
2. Simulate it—run it until it's close to its stationary distribution—then sample it!

[see Part II Machine Learning and Bayesian Inference]

NOTATIONAL ASIDE

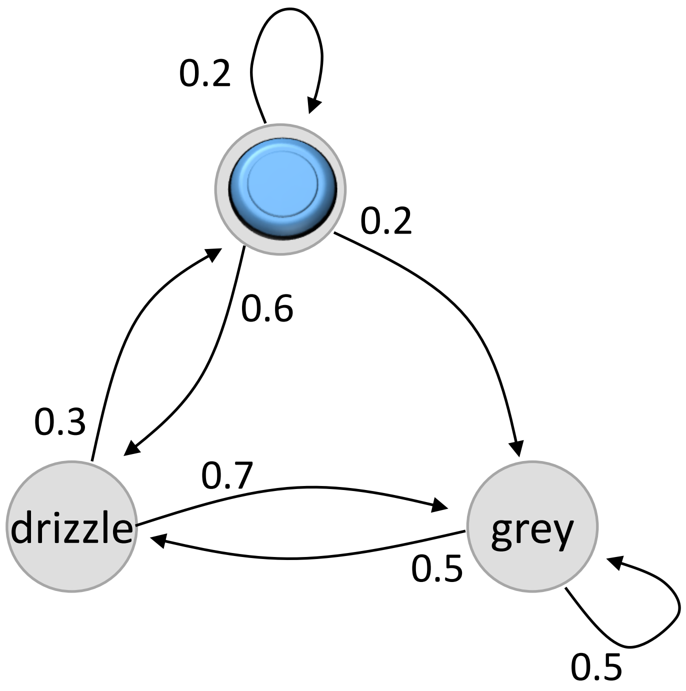
$X_1 \to X_2 \to X_3 \to$

The causal diagram for a Markov chain tells us that X_{n+1} is generated solely from X_n.
It's useful to have notation to say *how* it's generated.

STATE SPACE DIAGRAM

state space = {rain, drizzle, grey}

X_n = state on day n



TRANSITION PROBABILITY MATRIX

$$P = \begin{array}{r} \\ \text{rain} \\ \text{drizzle} \\ \text{grey} \end{array} \begin{array}{c} \begin{array}{ccc} \text{rain} & \text{drizzle} & \text{grey} \end{array} \\ \begin{bmatrix} .2 & .6 & .2 \\ .3 & 0 & .7 \\ 0 & .5 & .5 \end{bmatrix} \end{array}$$

useful sanity check: rows sum to 1

$$P_{ij} = \mathbb{P}\left(\begin{matrix} \text{next state} \\ \text{is } j \end{matrix} \,\middle|\, \begin{matrix} \text{in state} \\ i \end{matrix} \right)$$

$$X_{n+1} \sim \text{Cat}(P_{X_n \bullet})$$

row X_n of P

A distribution π over the state space is called a **stationary distribution** if

$$X_0 \sim \pi \quad \Rightarrow \quad X_1 \sim \pi$$

If π is a stationary distribution, and if $X_0 \sim \pi$, then by induction $X_n \sim \pi$ for all $n > 0$.

For Markov chains with a finite state space,

- $\pi \in \mathbb{R}^n_{\geq 0}$ where n is the number of states
- "$X_i \sim \pi$" means $\mathbb{P}(X_i = x) = \pi_x$ for all states x

If the state space is continuous, let π be a likelihood instead.

Assuming we start in a stat. dist π,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} 1_{X_i = x}\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(1_{X_i = x}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(X_i = x)$$

$$= \pi_x$$

long-run avg. fraction of time spent in state x

stationary distribution.

A distribution π over the state space is called a **stationary distribution** if

$$X_0 \sim \pi \quad \Rightarrow \quad X_1 \sim \pi$$

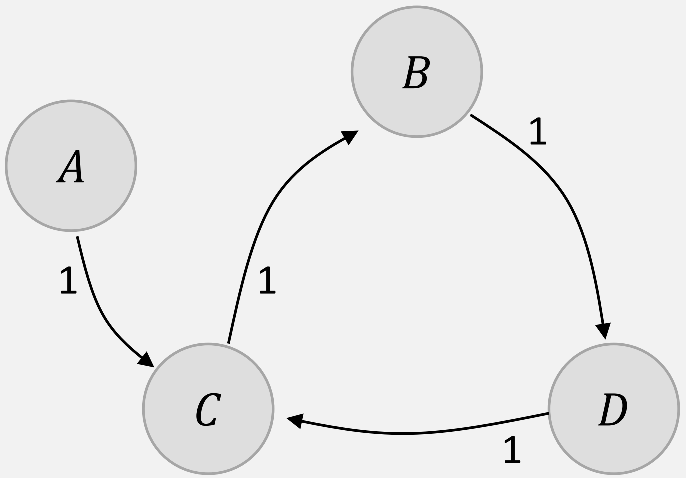
If π is a stationary distribution, and if $X_0 \sim \pi$, then by induction $X_n \sim \pi$ for all $n > 0$.

For Markov chains with a finite state space,

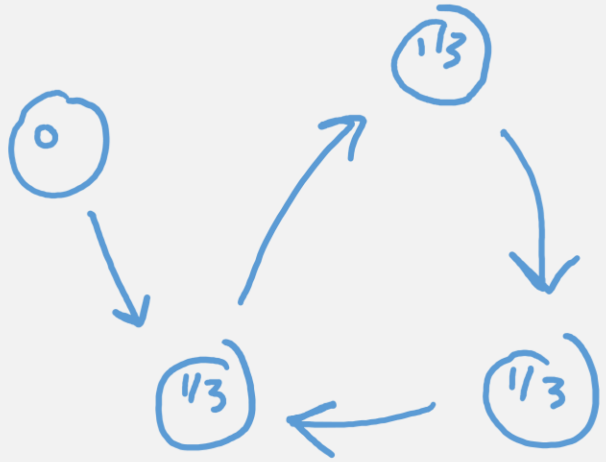
- $\pi \in \mathbb{R}^n_{\geq 0}$ where n is the number of states
- "$X_i \sim \pi$" means $\mathbb{P}(X_i = x) = \pi_x$ for all states x

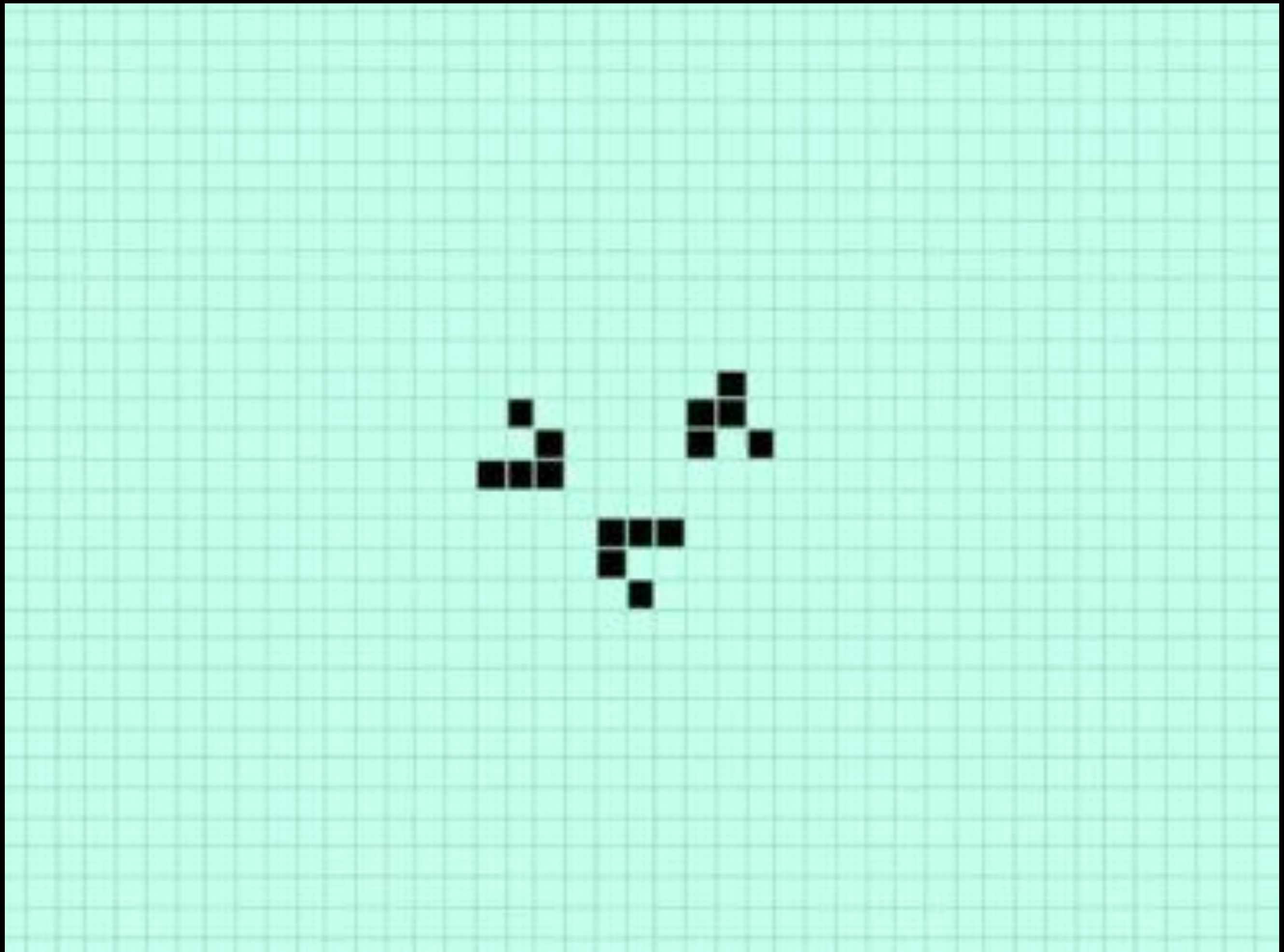
If the state space is continuous, let π be a likelihood instead.

QUESTION. What's the stationary distribution of this Markov chain?



Stat. Dist.

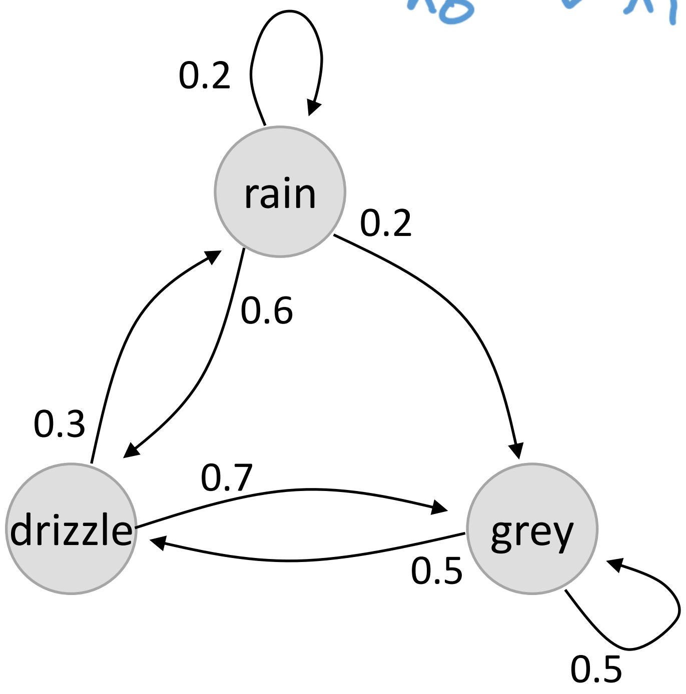




Example 11.4.1 (Stationary distribution)

Find the stationary distribution of Cambridge weather, generated from this Markov chain:

$X_0 \longrightarrow X_1 \longrightarrow \cdots$



let's suppose a stat. dist exists, call it π.

Suppose $X_0 \sim \pi$.

Then $X_1 \sim \pi$ by defn. of stationarity.

Also: $\underbrace{\mathbb{P}(X_1 = x)}_{\pi_x} = \sum_y \underbrace{\mathbb{P}(X_1 = x \mid X_0 = y)}_{P_{yx}} \underbrace{\mathbb{P}(X_0 = y)}_{\pi_y}$ by LOTP

$\Rightarrow \quad \pi_x = \sum_y \pi_y P_{yx}$ for all states x.

$x =$ rain: $\pi_{rain} = 0.2\,\pi_{rain} + 0.3\,\pi_{drizz}$

$x =$ drizzle: $\pi_{drizz} = 0.6\,\pi_{rain} \quad 0.5\,\pi_{grey}$

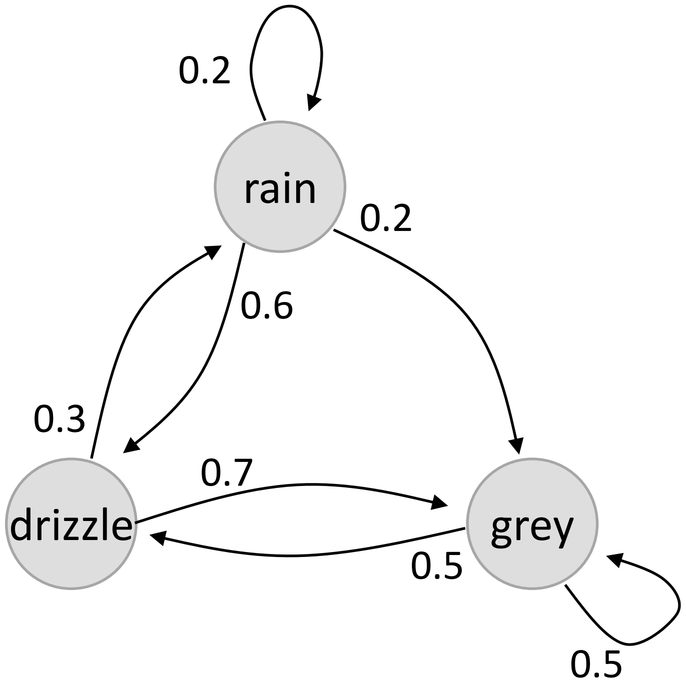
$x =$ grey: $\pi_{grey} = 0.2\,\pi_{rain} + 0.7\,\pi_{drizz} + 0.5\,\pi_{grey}$

Also: $\pi_{rain} + \pi_{drizz} + \pi_{grey} = 1$.

$\longrightarrow$ can solve for π.

Example 11.4.1 (Stationary distribution)

Find the stationary distribution of Cambridge weather, generated from this Markov chain:



In matrix notation,

$$\pi = \pi P \qquad \text{or equivalently } \pi(P-I)=0 \text{ or } (P-I)^{\mathsf{T}}\,\pi = 0$$

$$\pi \cdot 1 = 1$$

Or, putting these two together,

$$A\pi = b$$

$$\underbrace{\begin{bmatrix} (P-I)^{T} \\ \\ \text{---}1\text{---} \end{bmatrix}}_{\text{The A matrix}} \begin{bmatrix} | \\ \pi \\ | \end{bmatrix} = \underbrace{\begin{bmatrix} 0 \\ | \\ 0 \\ 1 \end{bmatrix}}_{\text{The b vector}}$$

```
# let states be rain=0, drizzle=1, grey=2
P = np.array([[.2,.6,.2], [.3,0,.7], [0,.5,.5]])
A = np.concatenate([(P-numpy.eye(3)).T, [[1,1,1]]])
π = np.linalg.lstsq(A, [0,0,0,1])[0]
```

- `np.linalg.lstsq(A,b)` seeks $\min_x |Ax-b|^2$. If $Ax=b$ can be solved, it will find a solution. It doesn't care about redundant equations.
- `np.linalg.solve(A,b)` solves $Ax=b$. It requires an exact system of equations, i.e. A square with no redundant equations.

Stationarity equations

If π is a stationary distribution, then it solves

$$\pi = \pi P, \qquad \pi \cdot 1 = 1$$

Conversely, if π is a distribution that solves $\pi = \pi P$ then π is a stationary distribution.

- There might be a unique solution
- There might be multiple solutions
- There might not even be any solutions (for ∞ state spaces)

For Markov Chain Monte Carlo, we aim to design a system that
- *has a unique stationary distribution*
- *settles down quickly to this distribution whatever state we start at*

Detailed balance equations

Lemma. If π is a vector that satisfies

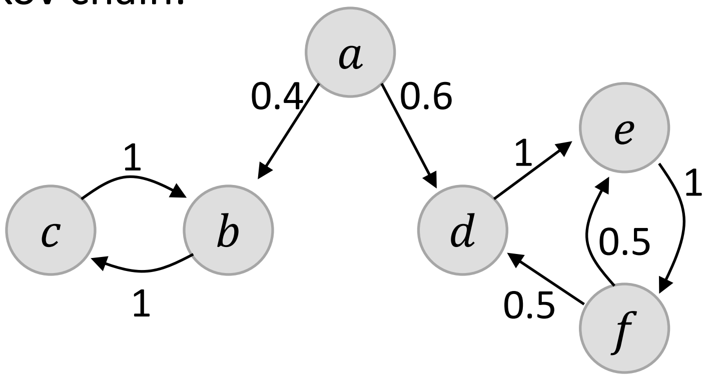
$$\pi_x P_{xy} = \pi_y P_{yx} \quad \text{for all } x, y$$

then π solves $\pi = \pi P$.

It doesn't hurt to try to solve detailed balance!
- If we're lucky, it tells us the stationary distribution
- If not, we just have to slog through solving $\pi = \pi P$

Example 11.4.2 (Stationary distribution)

Find all the stationary distributions of this Markov chain:



Let's solve the stationarity equations

$$\pi = \pi P \quad \text{and} \quad \pi \cdot 1 = 1$$

We get multiple solutions:

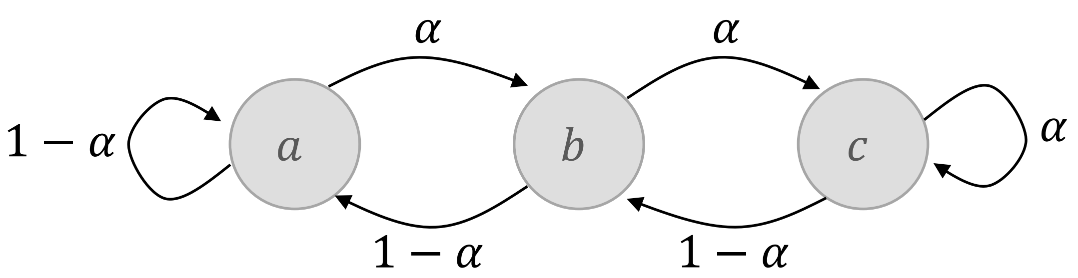
$$\begin{bmatrix}\pi_a\\ \pi_b\\ \pi_c\\ \pi_d\\ \pi_e\\ \pi_f\end{bmatrix} = p\begin{bmatrix}0\\ 0.5\\ 0.5\\ 0\\ 0\\ 0\end{bmatrix} + (1-p)\begin{bmatrix}0\\ 0\\ 0\\ 0.2\\ 0.4\\ 0.4\end{bmatrix}$$

QUESTION. Interpret these solutions in terms of the state space diagram.

[There are routines in numpy for getting all solutions. Recall IA NST maths / linear algebra.]

Example 11.4.4
(Stationary distribution via detailed balance)

Find the stationary distribution of this Markov chain:



let's see if we can solve detailed balance

$\forall x, y \quad \pi_x P_{xy} = \pi_y P_{yx}$

$x=a,\ y=b:$ $\quad \pi_a \alpha = \pi_b (1-\alpha)$

$x=b,\ y=a:$ $\quad \cancel{\pi_b(1-\alpha) = \pi_a \alpha}$

$x=a,\ y=c:$ $\quad \cancel{\pi_a 0 = \pi_c 0}$

$\vdots$

$x=a,\ y=a:$ $\quad \cancel{\pi_a(1-\alpha) = \pi_a(1-\alpha)}$

$x=b,\ y=c:$ $\quad \pi_b \alpha = \pi_c (1-\alpha)$

$$\pi_b = \pi_a \frac{\alpha}{1-\alpha}$$

$$\pi_c = \pi_b \frac{\alpha}{1-\alpha} = \pi_a \left(\frac{\alpha}{1-\alpha}\right)^2$$

Add in the constraint

$$\pi_a + \pi_b + \pi_c = 1$$

can solve for π.

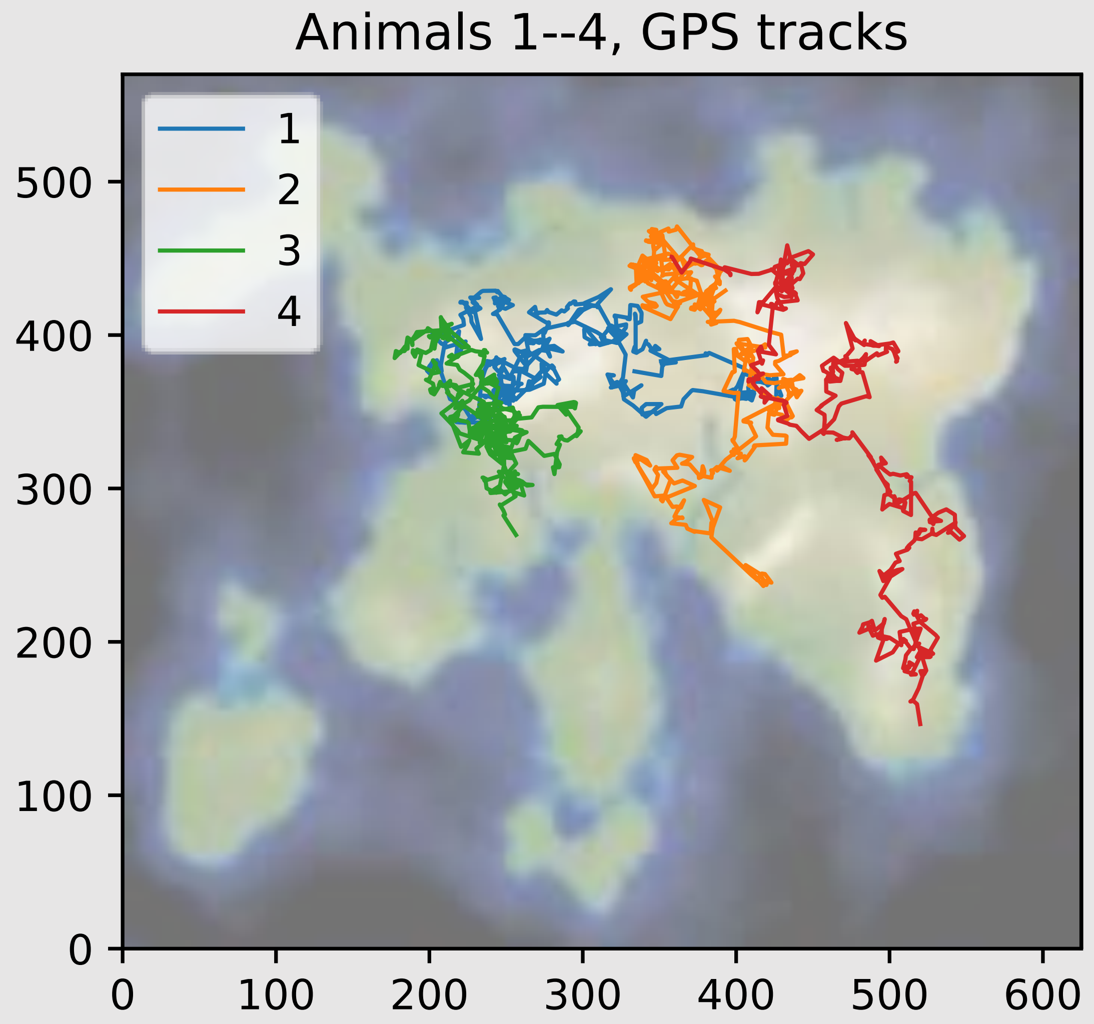
Challenge.

Data Stoat has gone missing!

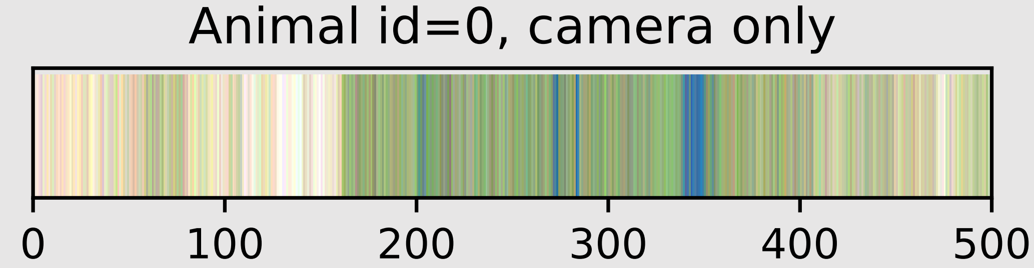
The GPS sensor that she normally carries has stopped working. But she still has a low-res camera with mobile uplink, so we know what sort of scenery she's in.

We also have full data from some of Data Stoat's friends, including Correlation Weasel and Bayes Ferret.

Can you help find Data Stoat?



Animals 1--4, GPS tracks
1
2
3
4



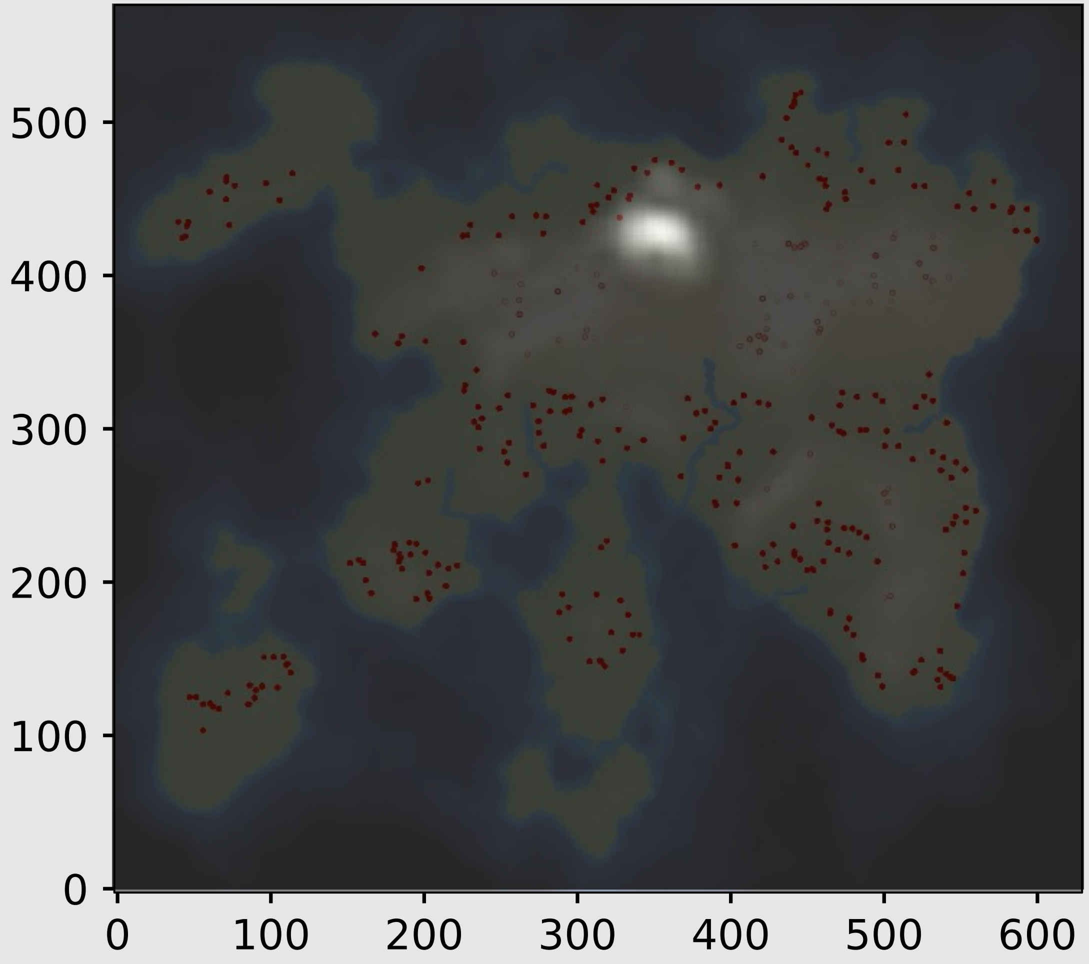
Animal id=0, camera only

Submit your answer as a heatmap.

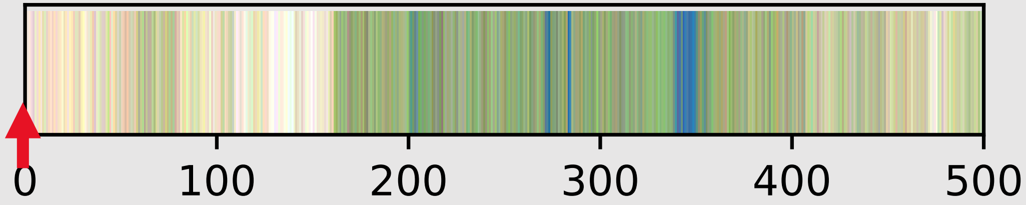
Your score will be the probability that you assign to Data Stoat's true location.

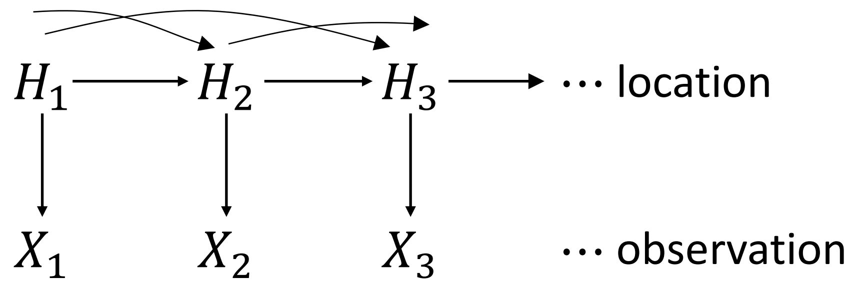
Winner will be announced on Christmas day.

Particle Filter to estimate Animal 0's location



Animal id=0, camera only





- invent models for H and $(X|H)$
- fit them
- choose between them

- computational Bayes (particle filter)

CONFIDENCE
- what's my spread of answers between runs?
- what's my holdout performance?

Using a probability model to describe data

Handling random variables and likelihood

- inventing probability models
- linear models
  - parameter identification
- causal diagrams
  - stationarity

Fitting a model to the dataset

Least squares
Brute force optimization

- Likelihood maximization
- Likelihood as a measure of fit

Do my conclusions generalize beyond the dataset?

Bayes's rule
Monte Carlo
Empirical distributions

- QUESTIONS—
  confidence, model choice
- APPROACHES—
  holdout, Bayesian, frequentist

THANK YOU!

Pick up a Data Stoat sticker

[on Handouts table tomorrow]

Office hours today

1–1.30pm in the cafe area