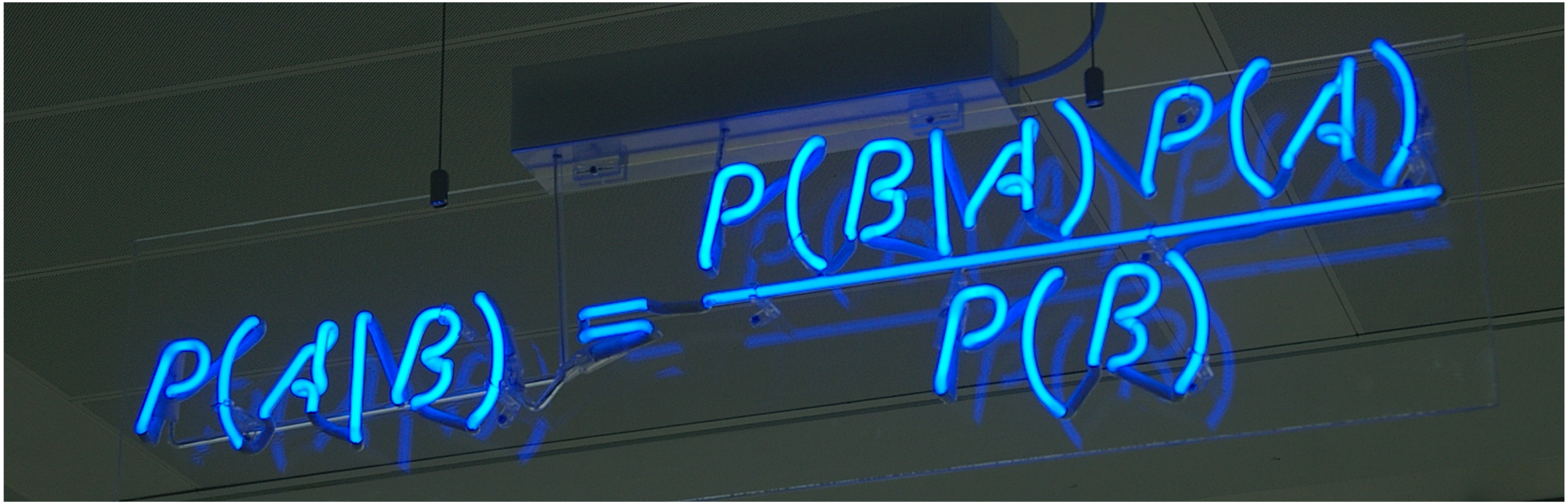
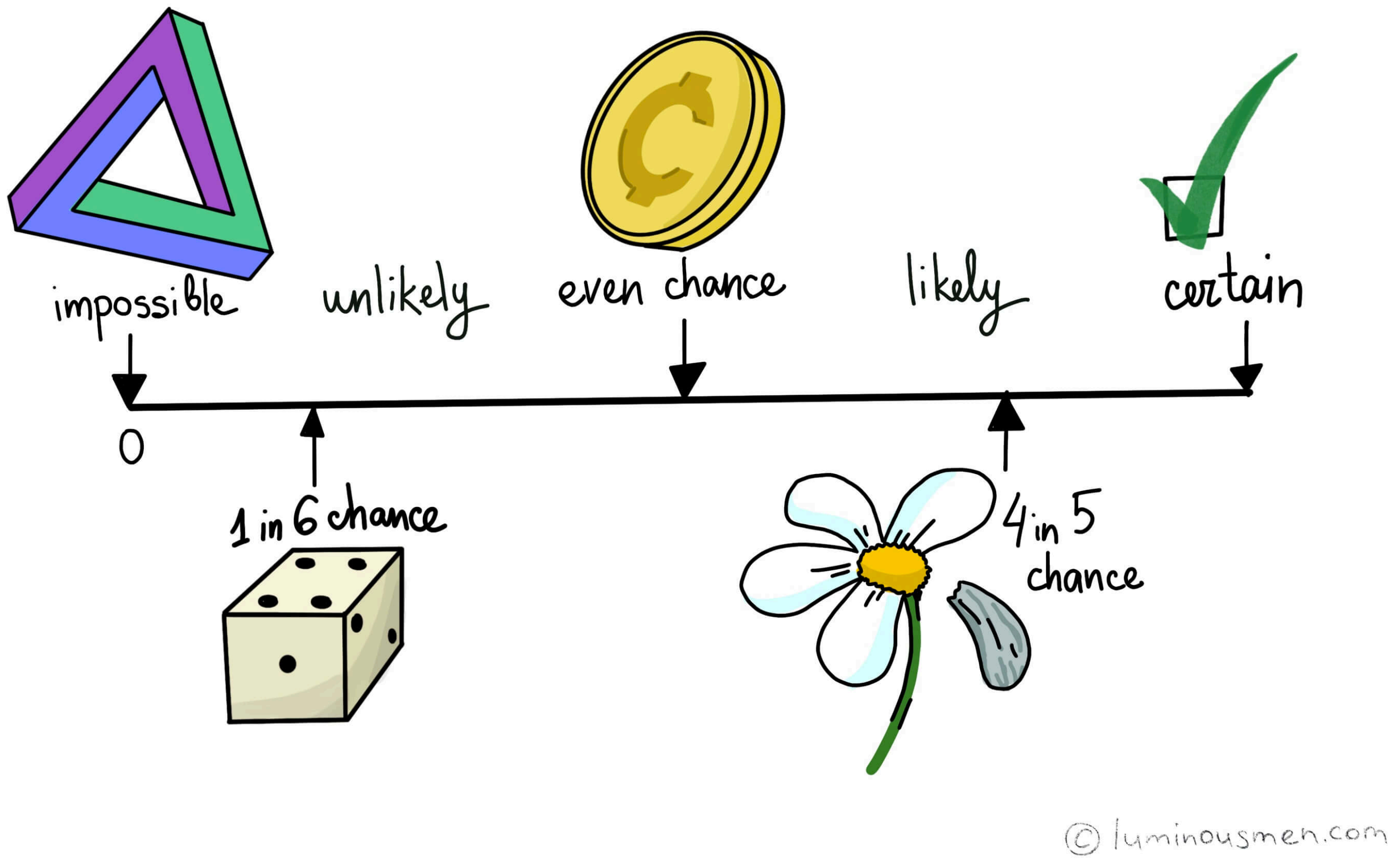
BAYESIAN STATISTICS FOR DUMMIES

AKA DREIA'S EXPERIENCE WITH UNDERSTANDING PROBABILITIES



HUGE CAVEAT:

Dreia is in no way an expert in this. She is merely summarizing what she has learned from a data science workshop. Please be kind to her.



PROBABILITY THEORY

PROBABILITY OF "A" HAPPENING

$P(A)$

PROBABILITY OF "B" HAPPENING

$P(B)$

PROBABILITY OF "A" AND "B" HAPPENING

$$P(A \cap B)$$

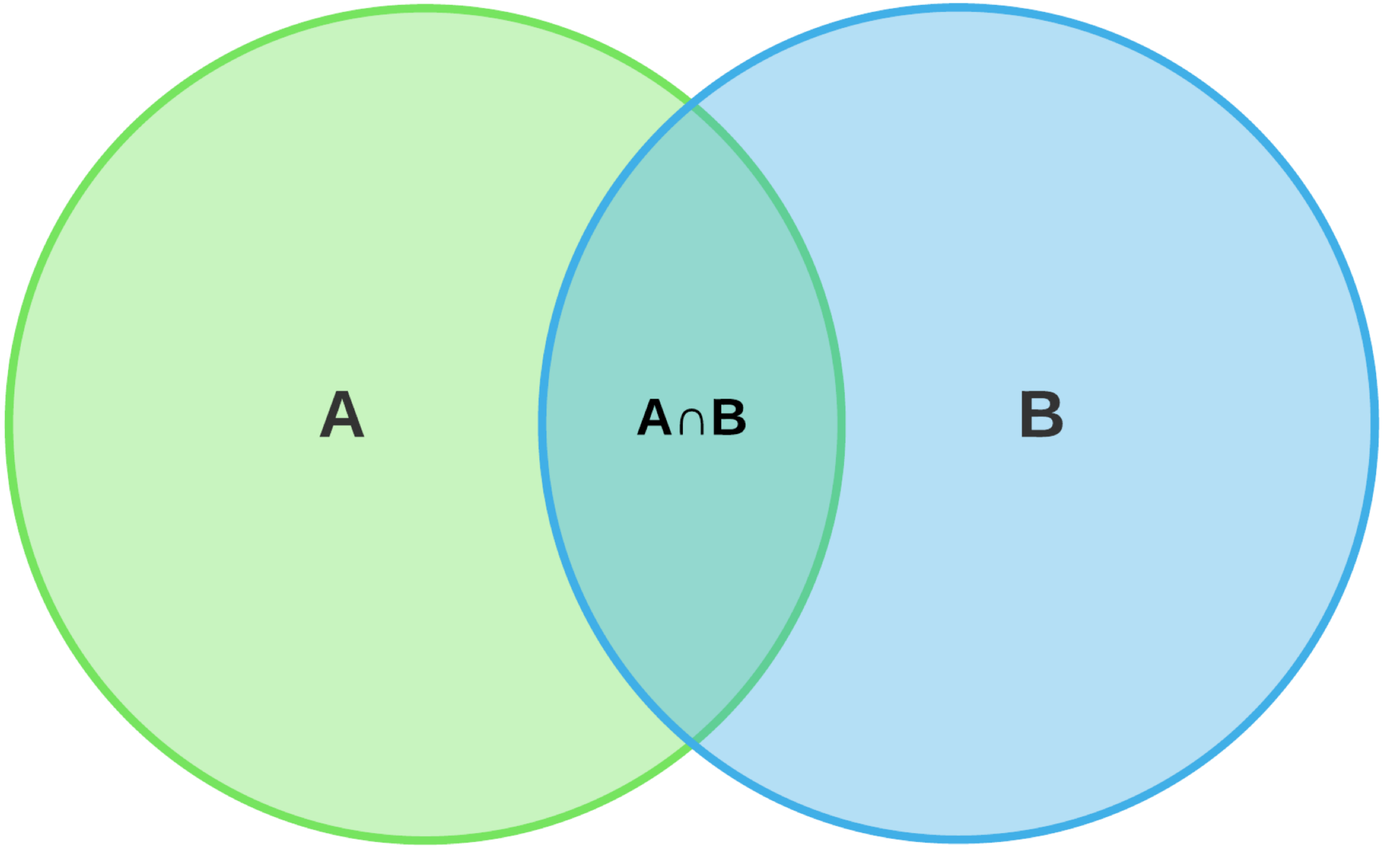
$$\rightarrow P(A)P(B)$$

PROBABILITY OF "A" AND "B" HAPPENING

$$P(A \cap B)$$

$$\rightarrow P(A|B)\,P(B)$$

PROBABILITY OF “B” AND “A” HAPPENING



PROBABILITY OF “B” AND “A” HAPPENING

$$P(B \cap A) = P(A \cap B)$$

$$P(B|A)P(A) = P(A|B)P(B)$$

- If B are parameters,
- And A is what we observe

$$B \rightarrow \theta, \quad A \rightarrow y$$

THEN

$$P(B|A)P(A) = P(A|B)P(B)$$

$$\downarrow$$

$$P(\theta|y)P(y) = P(y|\theta)P(\theta)$$

RE-ARRANGING THAT EQUATION…

Likelihood

Prior

Posterior

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

Evidence

BAYES' THEOREM!

Aka a seemingly smarter way of saying "the probability of SOME STUFF happening based on the probability of SOME OTHER STUFF."

RULES

$$\boxed{P(a|b) \geq 0}$$

$$\boxed{\int P(a|b)\ da = 1}$$

$$\int P(a|b)\ db = \text{WRONG!!!!!}$$

RULES

$$P(a|b) \geq 0$$

$$\int P(a|b) \; da = 1$$

$$\int P(a,c|b) \; da = P(c|b)$$

IF THERE ARE SEPARABLE (INDEPENDENT) DATA, THEN…

$$P(y|\theta) = \prod_{n} P(y_n|\theta)$$

RE-ARRANGING THAT EQUATION…

Likelihood

Prior

Posterior

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

Evidence

BAYES' THEOREM!

Aka a seemingly smarter way of saying "we're trying to figure out the probability of this NEW STUFF based on the probability of STUFF WE KNOW."

LIKELIHOOD FUNCTION

$$P(y|\theta)$$

Given my assumptions and set of parameters,
what is the probability distribution of the data?

Likelihood Principle
- All the information about the data is in the likelihood function

PRIOR

$$P(\theta)$$

How are the parameters distributed?

1 5

1 5

EVIDENCE

$$P(y)$$

Probability of data…??????? I don't really know

$$P(y|\theta)P(\theta)\ d\theta = P(y)$$

Hard to compute!

POSTERIOR (DISTRIBUTION FUNCTION)

$$P(\theta|y)$$

Given my observed data, what is the probability
Of measuring my parameters?

This is usually what we WANT in Astronomy!

SOME OTHER STUFF TO KNOW

- Products are hard! Summations are easier!!!
  - Take the log of things
- When the data is really good, then you don’t necessarily need the best prior
- Evidence is hard to compute but we can get away with it by doing MCMC

MARKOV CHAIN MONTE CARLO

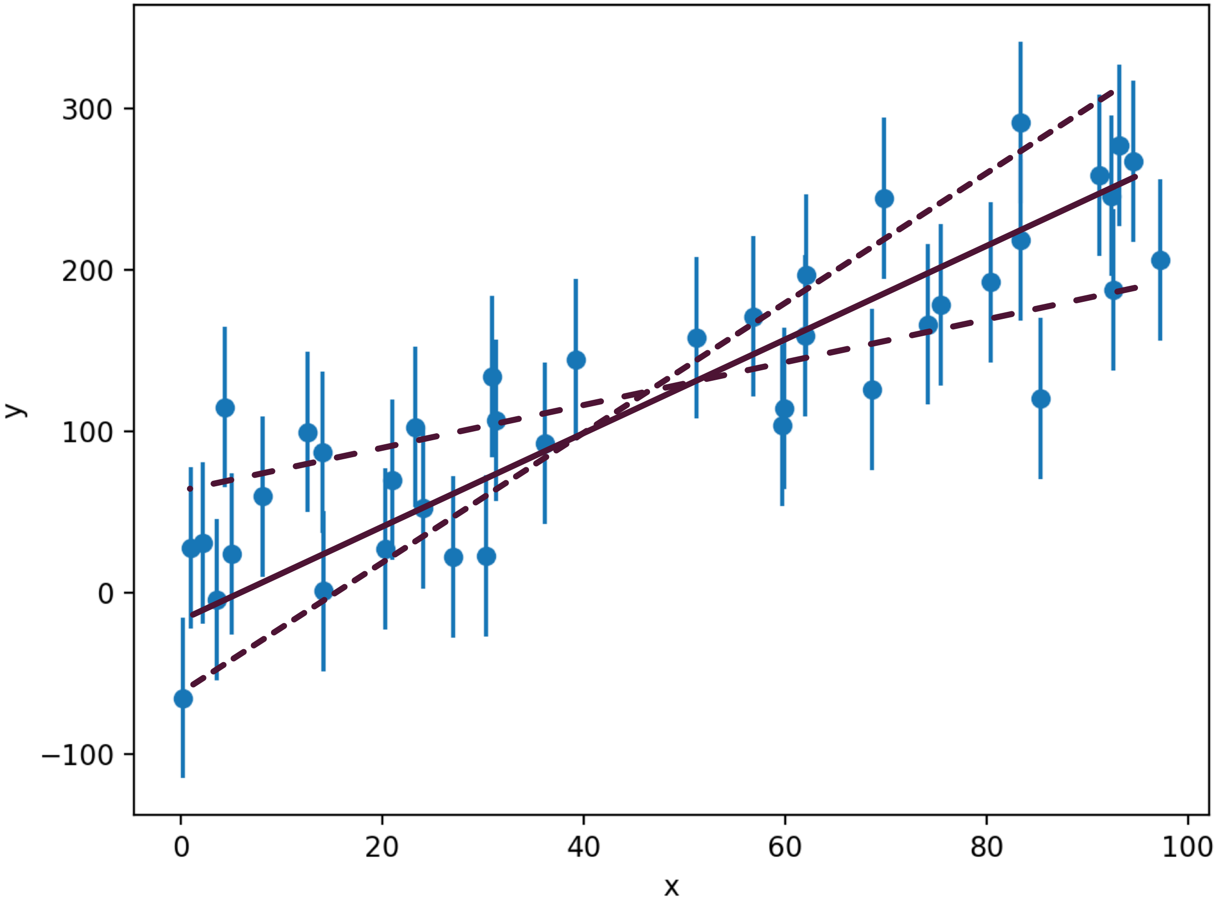
- Markov Chain
  - A mathematical sequence that's a stochastic process
  - The next element in the sequence only depends on the current element and not on the other position
  - http://setosa.io/ev/markov-chains/
- Monte Carlo
  - Randomized sample of parameters

A = walk, B = no walk

WHAT DO WE USE THIS FOR?

- To sample the posterior distribution of the parameter space
- For uncertainty estimation
- To visualize and marginalize over covariances between parameters
- To see how likely the model fits the data

APPLICATION: FITTING A LINE TO DATA



$y = m * x + b$

APPLICATION: FITTING A LINE TO DATA

- Assuming gaussian distributed scatter in the observations

$$p(y_i \mid m, b, x_i, \sigma_i) = \frac{1}{\sqrt{2\pi\sigma_{y_i}^2}} \exp - \frac{(y_i - m\,x_i - b)^2}{2\sigma_{y_i}^2}$$

- Assuming the observations are independent, then the probability of *all* the observations is the product of the individual probabilities **(LIKELIHOOD!)**

$$\mathcal{L} = \prod_i p(y_i \mid m, b, x_i, \sigma_i) = \prod_i \frac{1}{\sqrt{2\pi\sigma_{y_i}^2}} \exp - \frac{(y_i - m\,x_i - b)^2}{2\sigma_{y_i}^2}$$

APPLICATION: FITTING A LINE TO DATA

$$\mathcal{L} = \prod_i p(y_i \mid m, b, x_i, \sigma_i) = \prod_i \frac{1}{\sqrt{2\pi\sigma_{y_i}^2}} \exp - \frac{(y_i - m\,x_i - b)^2}{2\sigma_{y_i}^2}$$

- Products are hard! Also the big product is going to yield very tiny numbers so…

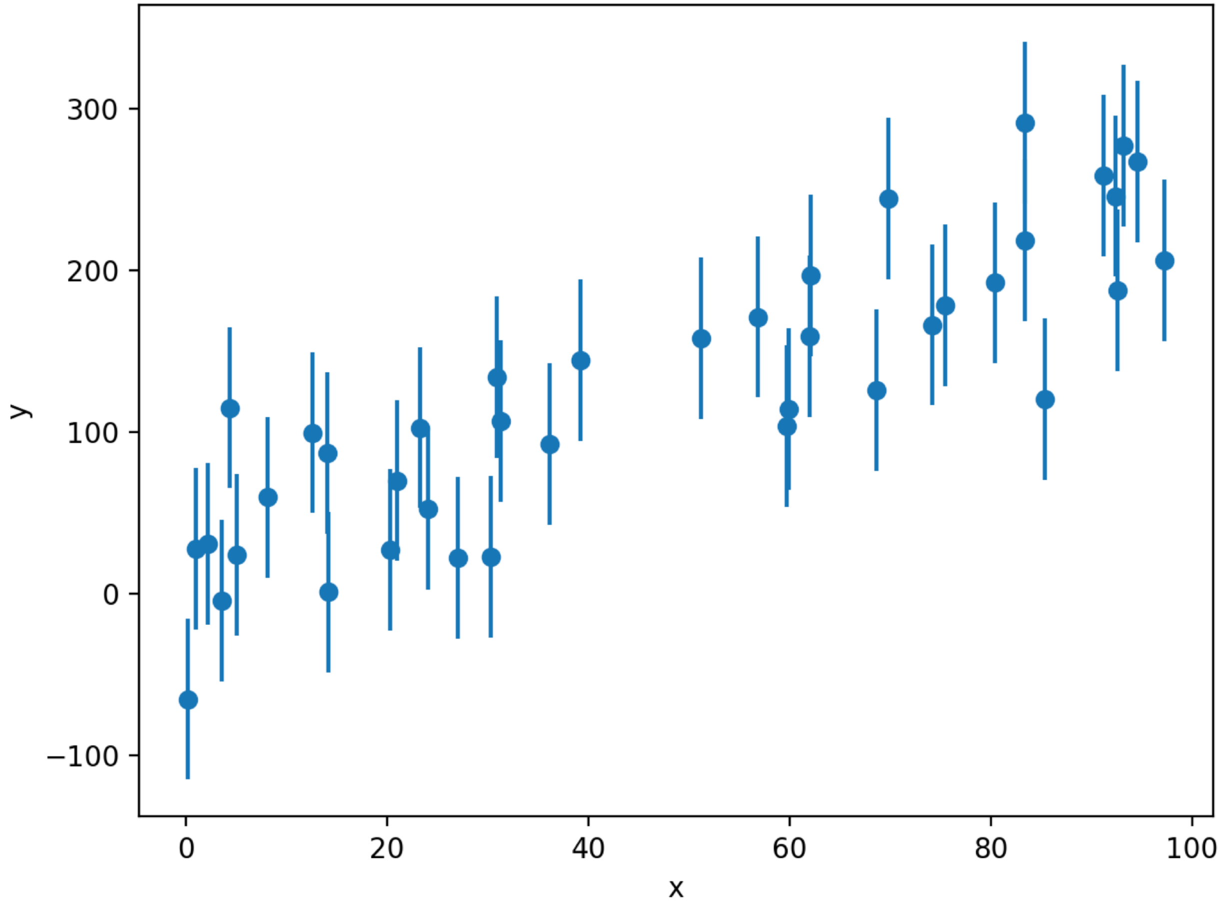
$$\log \mathcal{L} = \mathcal{K} - \sum_i \frac{(y_i - m\,x_i - b)^2}{2\sigma_{y_i}^2}$$

- And Kappa is just a constant:

$$-\frac{n}{2}\log 2\pi - \sum_i \log \sigma_{y_i}$$

APPLICATION: FITTING A LINE TO DATA

- Applying a uniform **PRIOR** for m and b:
  - P(m) = Uniform(0,300)
  - P(b) = Uniform(-100,100)



METROPOLIS-HASTINGS ALGORITHM

We compare to a random number whether or not we accept or reject the next position **(acceptance criterion),** which helps in exploring the **full posterior**

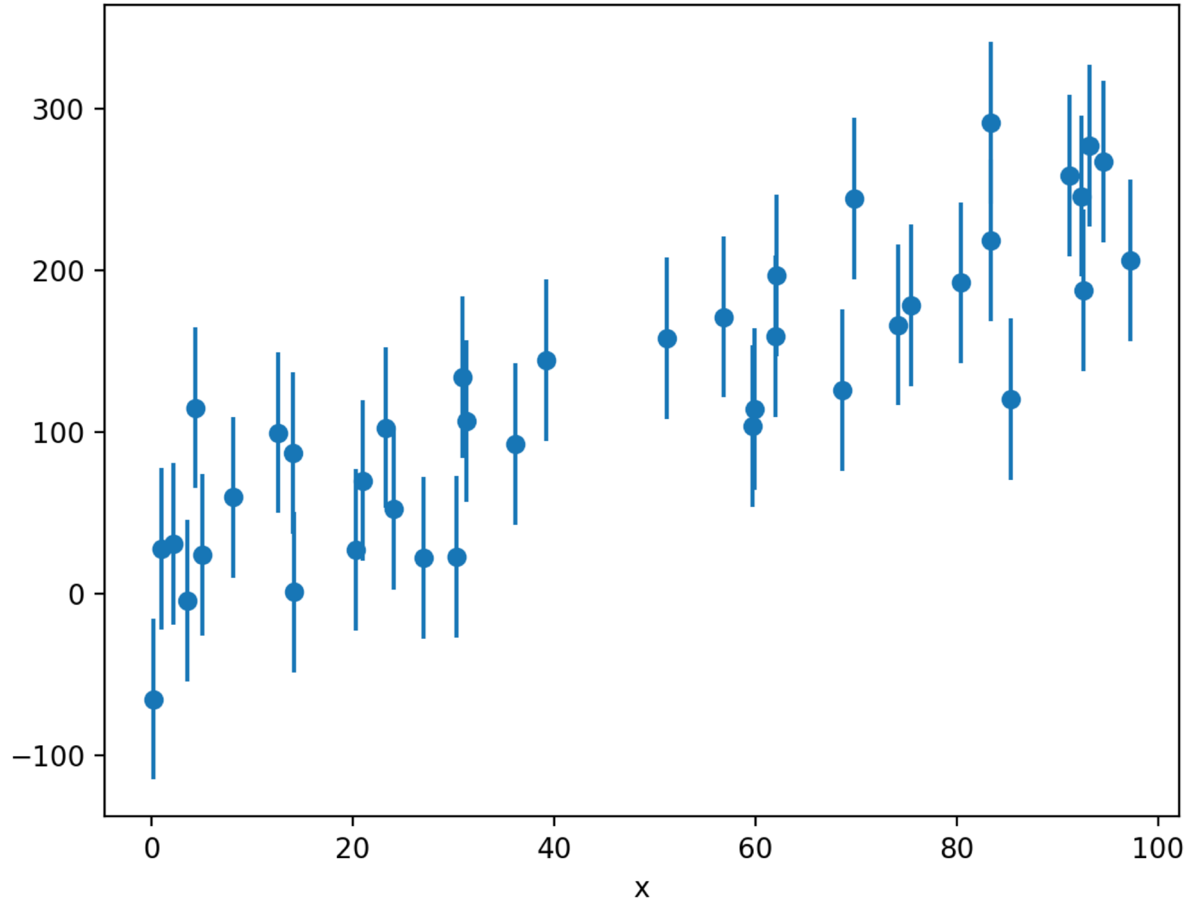
This is **not an optimization routine** which simply moves in the direction of greater probability

METROPOLIS-HASTINGS ALGORITHM

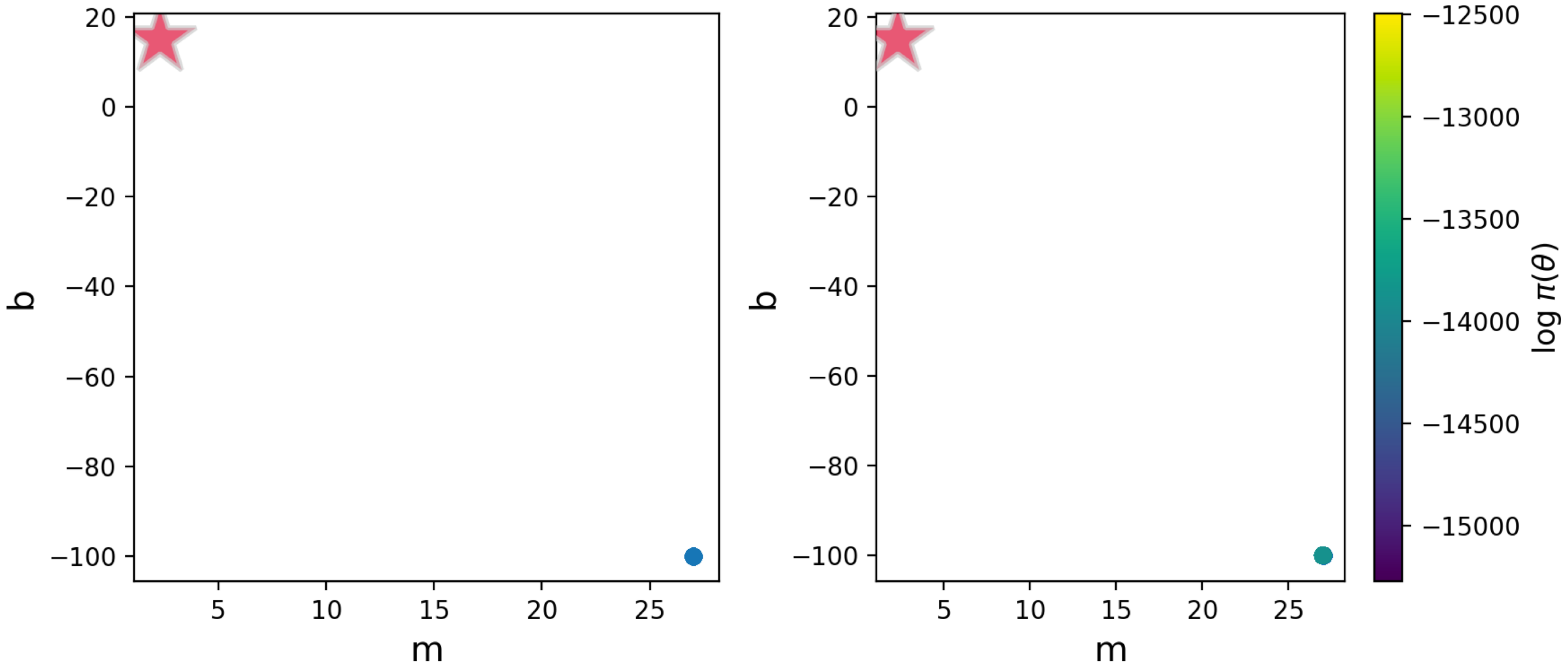
- pick some position θ_0 in the parameter space and calculate the posterior $P(\theta_0|x)$
- begin the chain
  - "propose" a move from the current position θ_i to a new position θ_{i+1}
  - calculate the posterior at θ_{i+1}, $P(\theta_{i+1}|x)$
  - draw a random number, R from a distribution that goes from 0 to 1
  - if the ratio $P(\theta_{i+1}|x)/P(\theta_i|x)$ is $>R$, "accept" the proposed move and advance the chain to θ_{i+1}
  - else "reject" the proposal and set $\theta_{i+1} = \theta_i$
  - repeat until chain is "finished"

STUFF TO THINK ABOUT WHEN DOING MCMC

- Initialize with parameters that make sense!
- How you propose to jump to the next θ is going to affect how you explore the posterior
  - i.e. we need $(\mu_m, \sigma_m), (\mu_b, \sigma_b)$

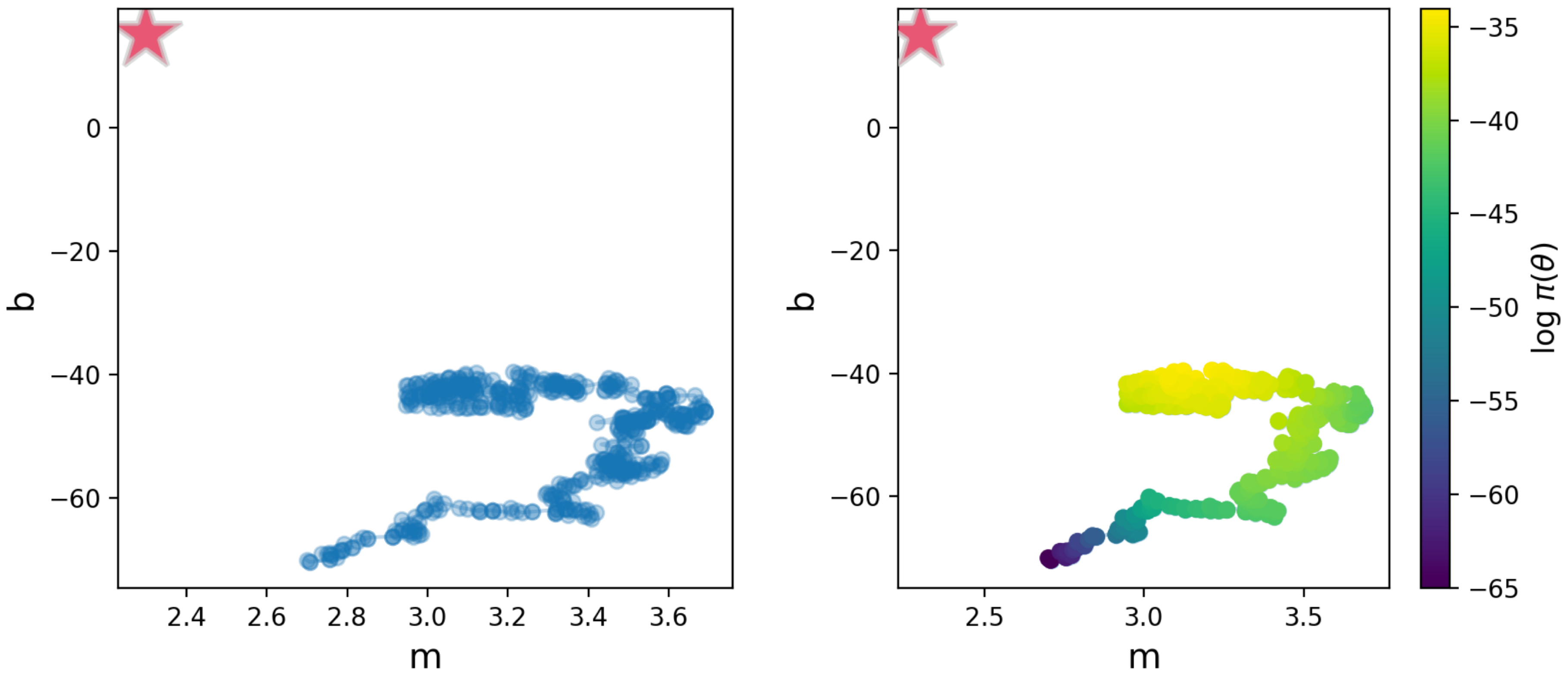


IF UR A DONGUS



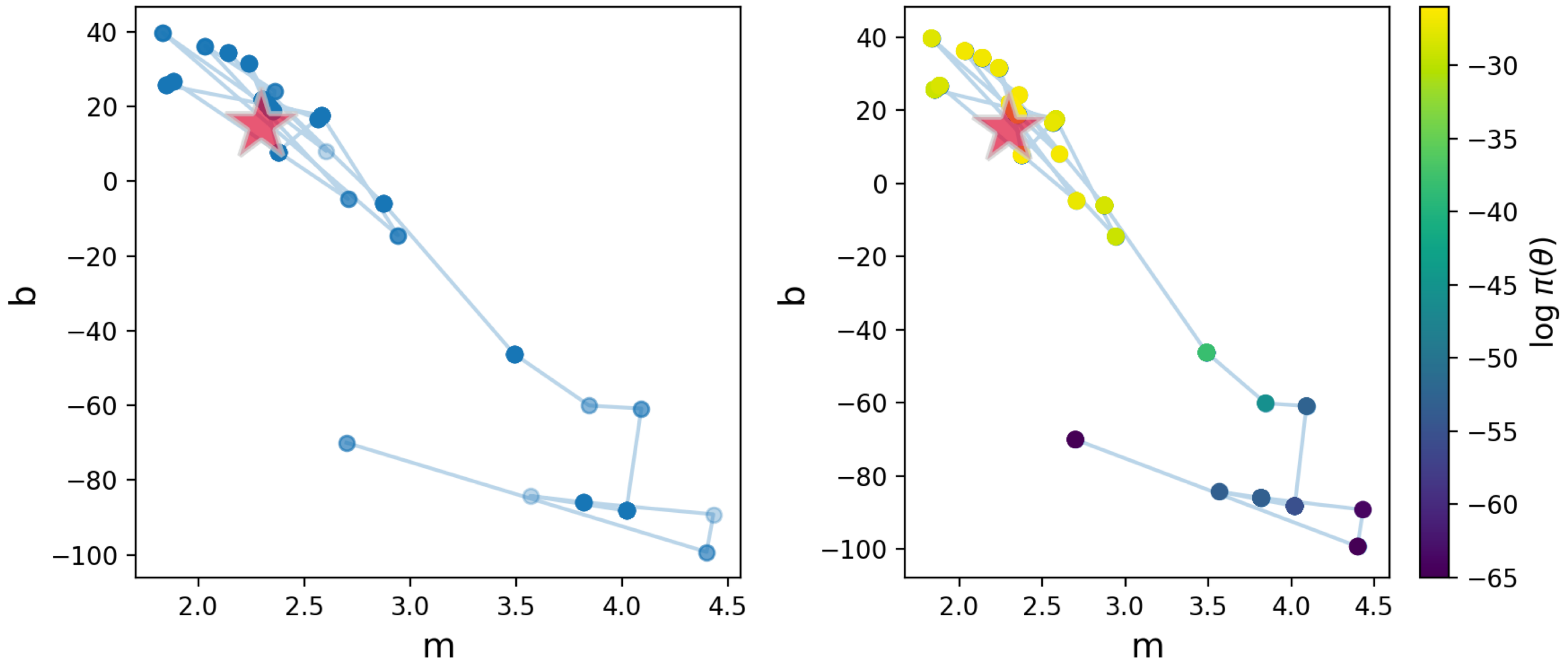
If you start very far from the truth, you'll never get there…

TOO SMALL JUMPS

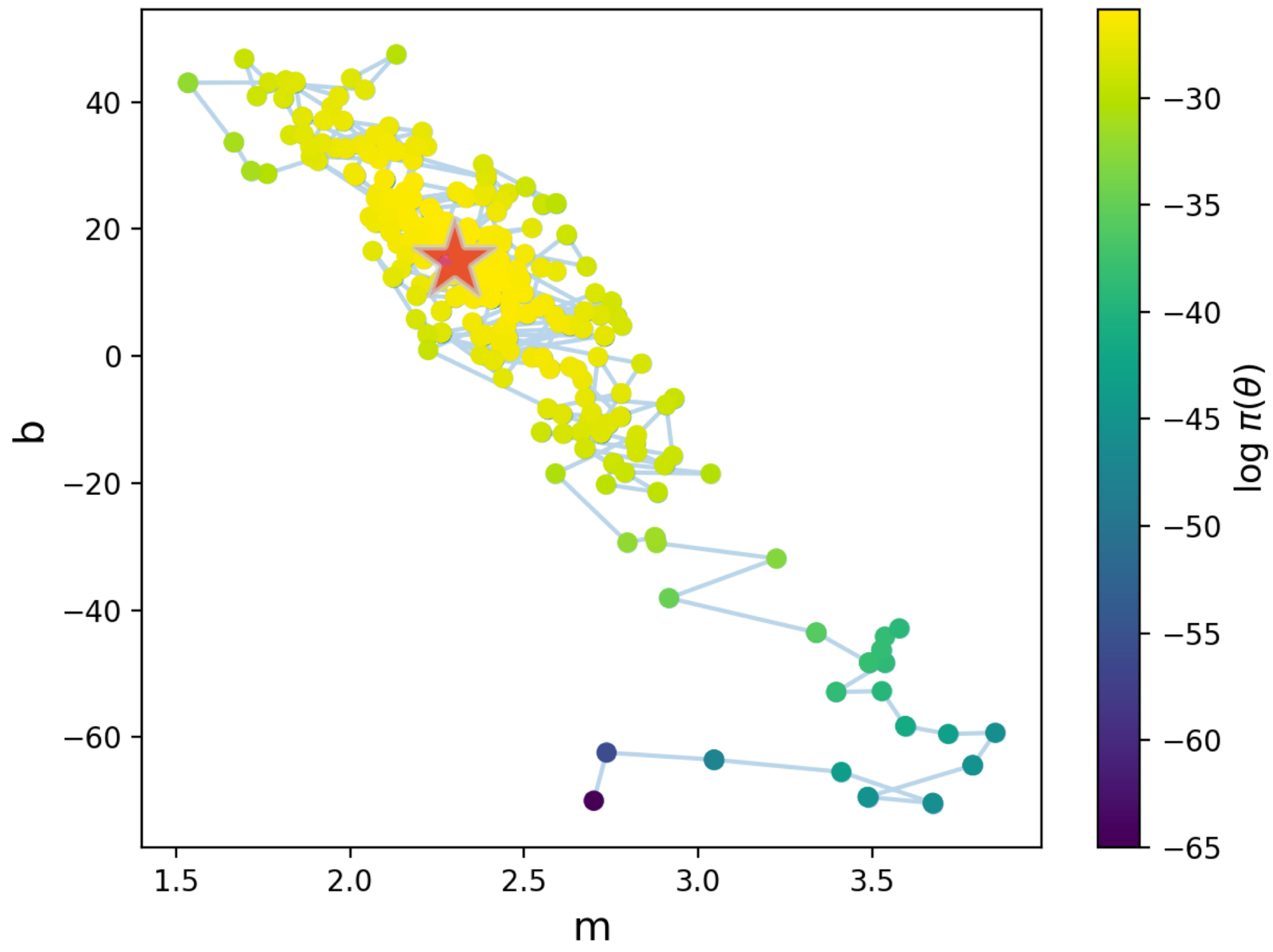


Requires many steps to explore the PDF

TOO BIG JUMPS

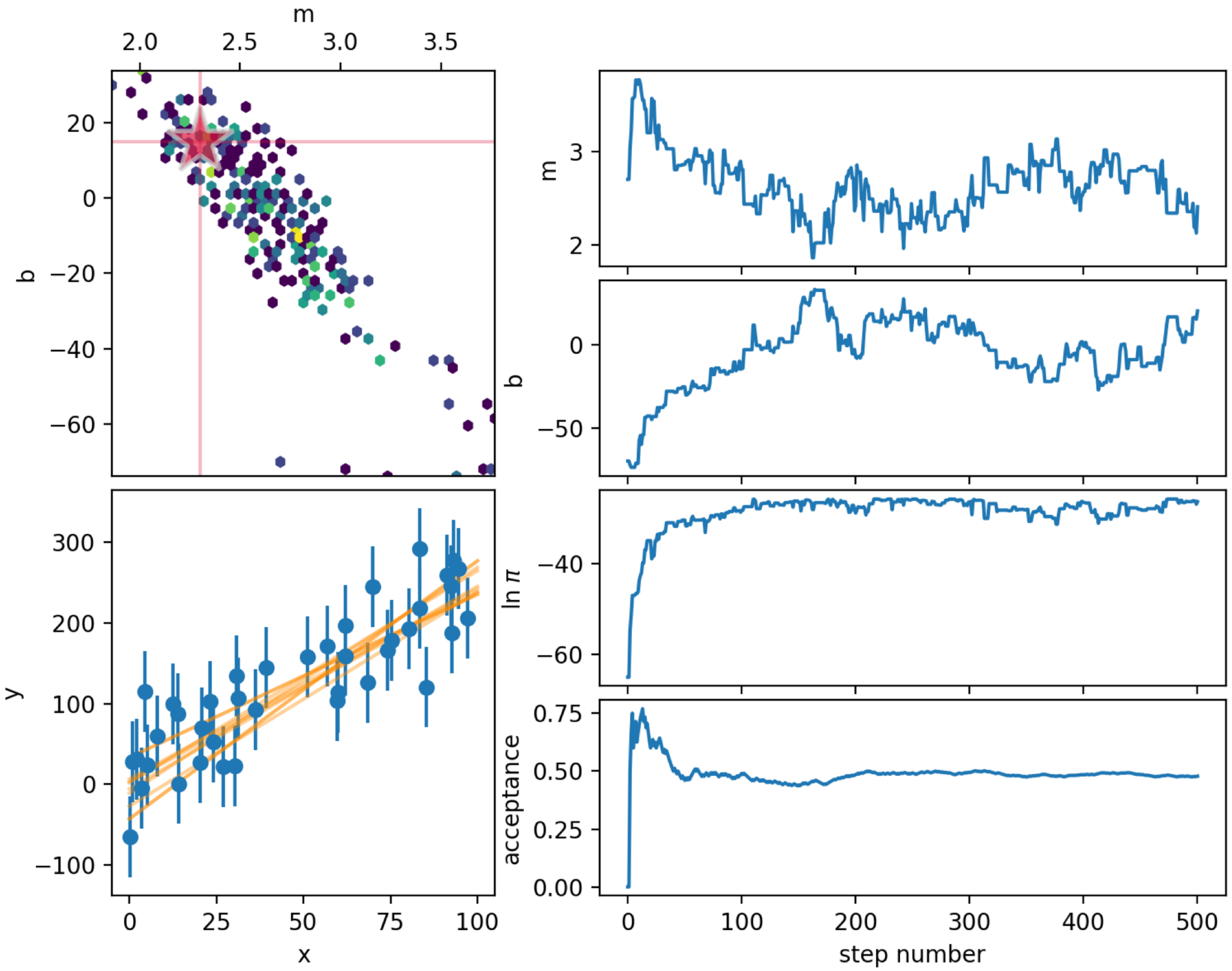


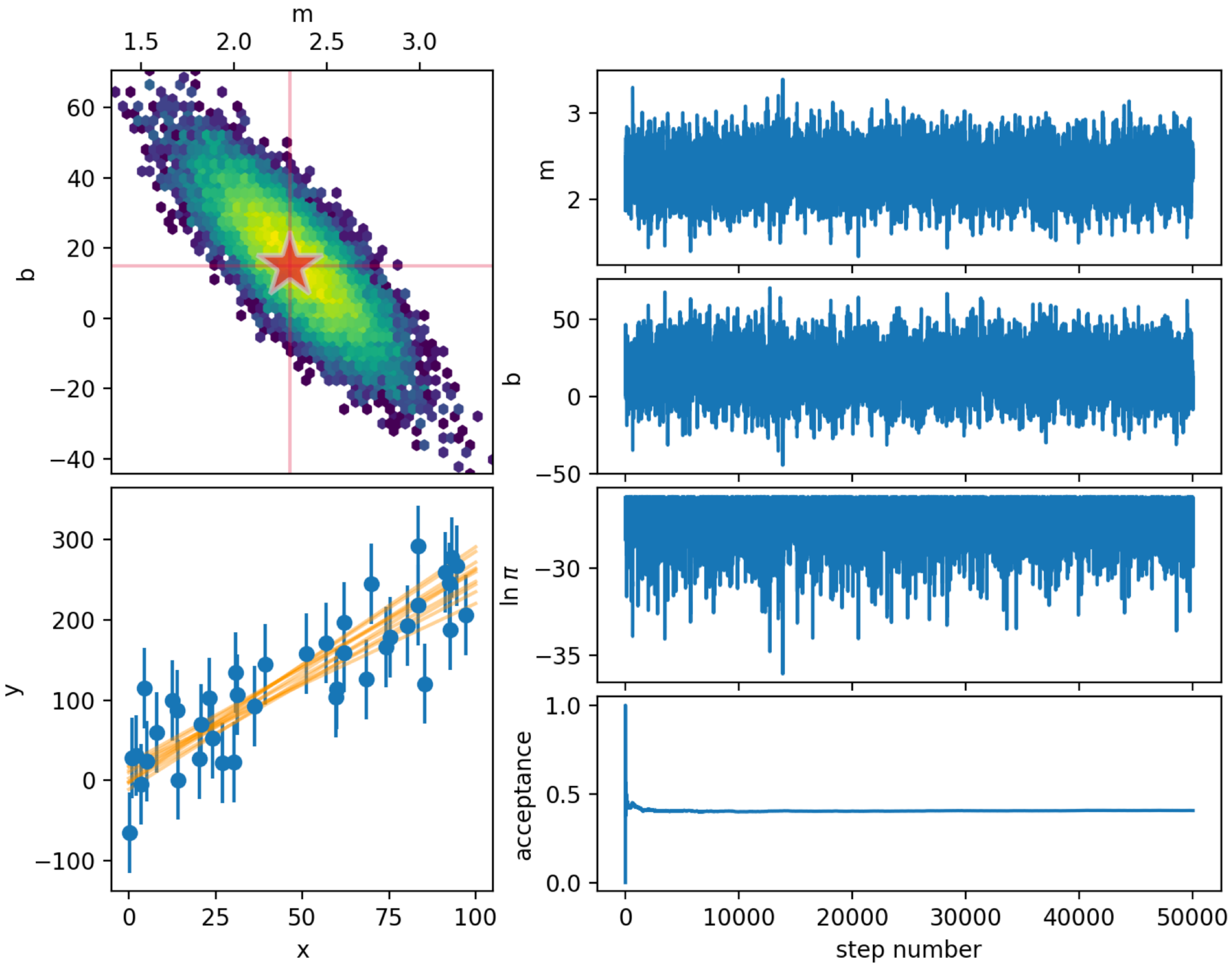
Will reject so many proposals so can't explore the full parameter space



VOILA!!!

MCMC USING 500 STEPS

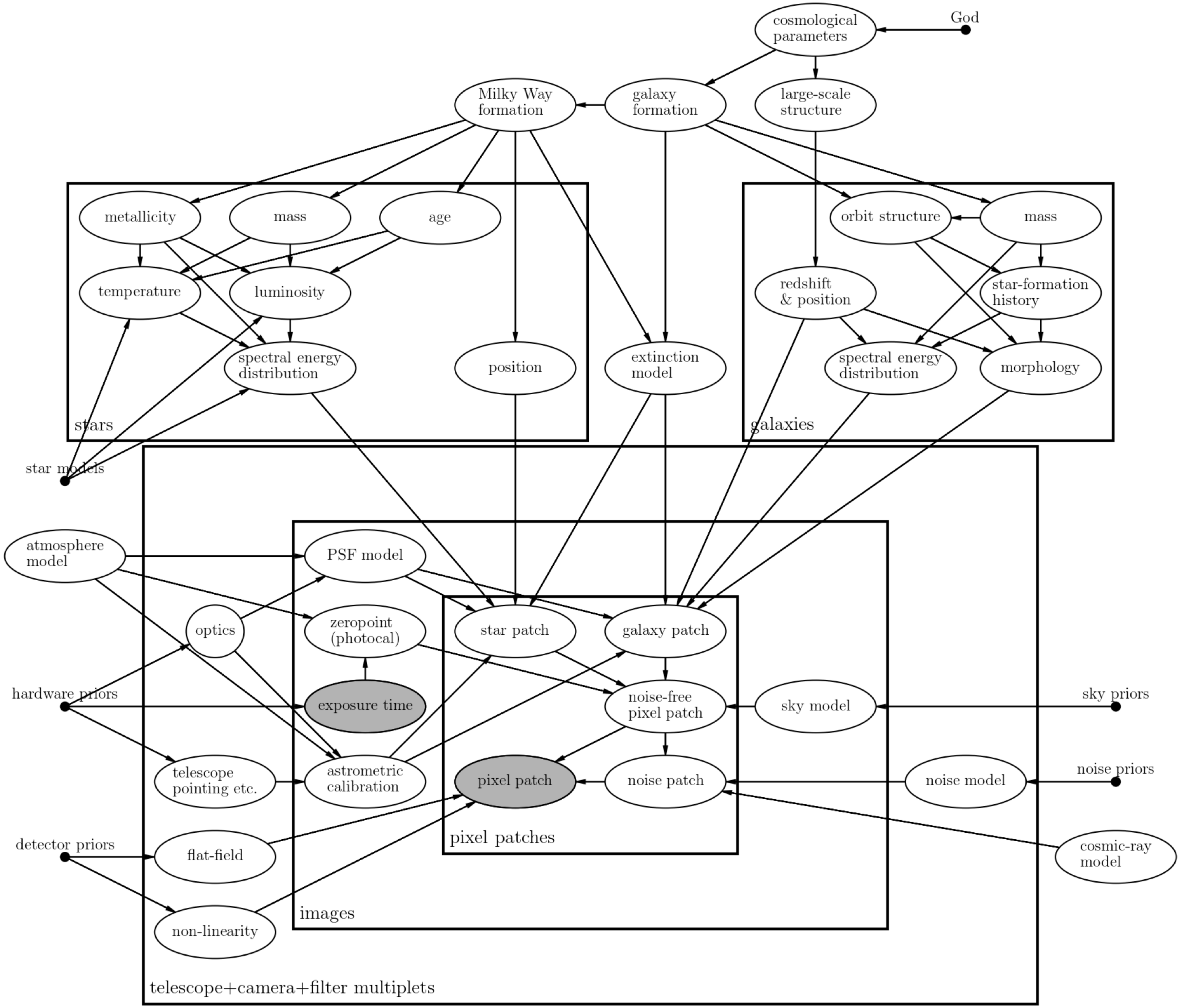




MCMC USING 5000 STEPS

PROBABILISTIC GRAPHICAL MODELS

PGM OF ASTRONOMY



RESOURCES

- Wikipedia
- https://github.com/LSSTC-DSFP/LSSTC-DSFP-Sessions