A BEGINNER'S GUIDE TO BAYESIAN INFERENCE

RUI LUO (CASS-ATNF)

Co-learnium on 5 Dec. 2019

OUTLINE

- What is Bayesian inference?
  - Bayes' theorem
- Why use Bayesian?
- How to do Bayesian?
  - Sampling: MCMC, M-H algorithm, Nested sampling, etc
  - Available toolkits: MultiNest, emcee, etc
- Applications in astrophysics
- Consideration on the outlier detections

BAYESIAN INFERENCE

- Conditional Probability:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) \Rightarrow P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- Bayes' Theorem:

$$\underbrace{P(\boldsymbol{\Theta}|\mathbf{D},\mathbf{I})}_{\text{Posterior}} = \frac{\overbrace{P(\boldsymbol{\Theta}|\mathbf{I})}^{\text{Prior}}\,\overbrace{P(\mathbf{D}|\boldsymbol{\Theta},\mathbf{I})}^{\text{Likelihood}}}{\underbrace{P(\mathbf{D}|\mathbf{I})}_{\text{Evidence}}}$$

$\boldsymbol{\Theta}$: Unknown Parameters

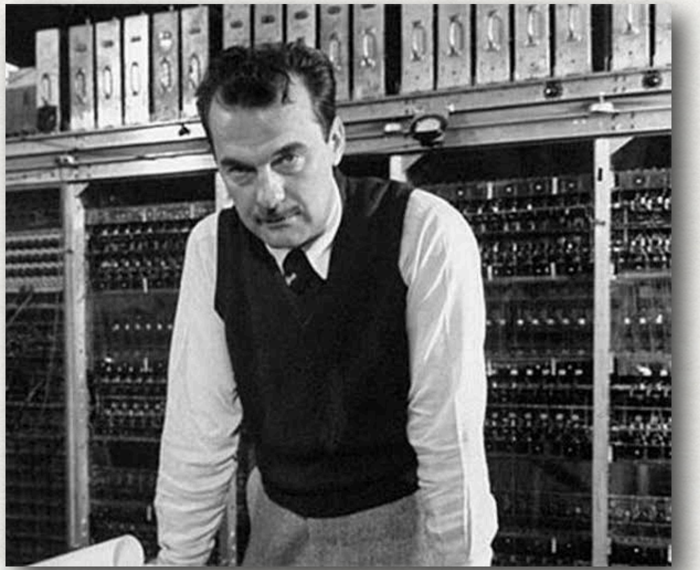
$\mathbf{D}$: Data

$\mathbf{I}$: Knowledge or Assumptions

HISTORY OF DEVELOPMENT

- A new probability theory to calculate limits on unknown parameters (Bayes & Price 1763)
- Theory of Probability: Bayesian Analysis (Jeffreys 1939)
- Markov Chain Monte Carlo (Metropolis et al. 1953)
- Metropolis-Hastings Algorithm for sampling (Hastings 1970)
- Probability Theory: The Logic of Science (Jaynes 2003)

Thomas Bayes (1701-1761)



Nicholas Metropolis (1915-1999)

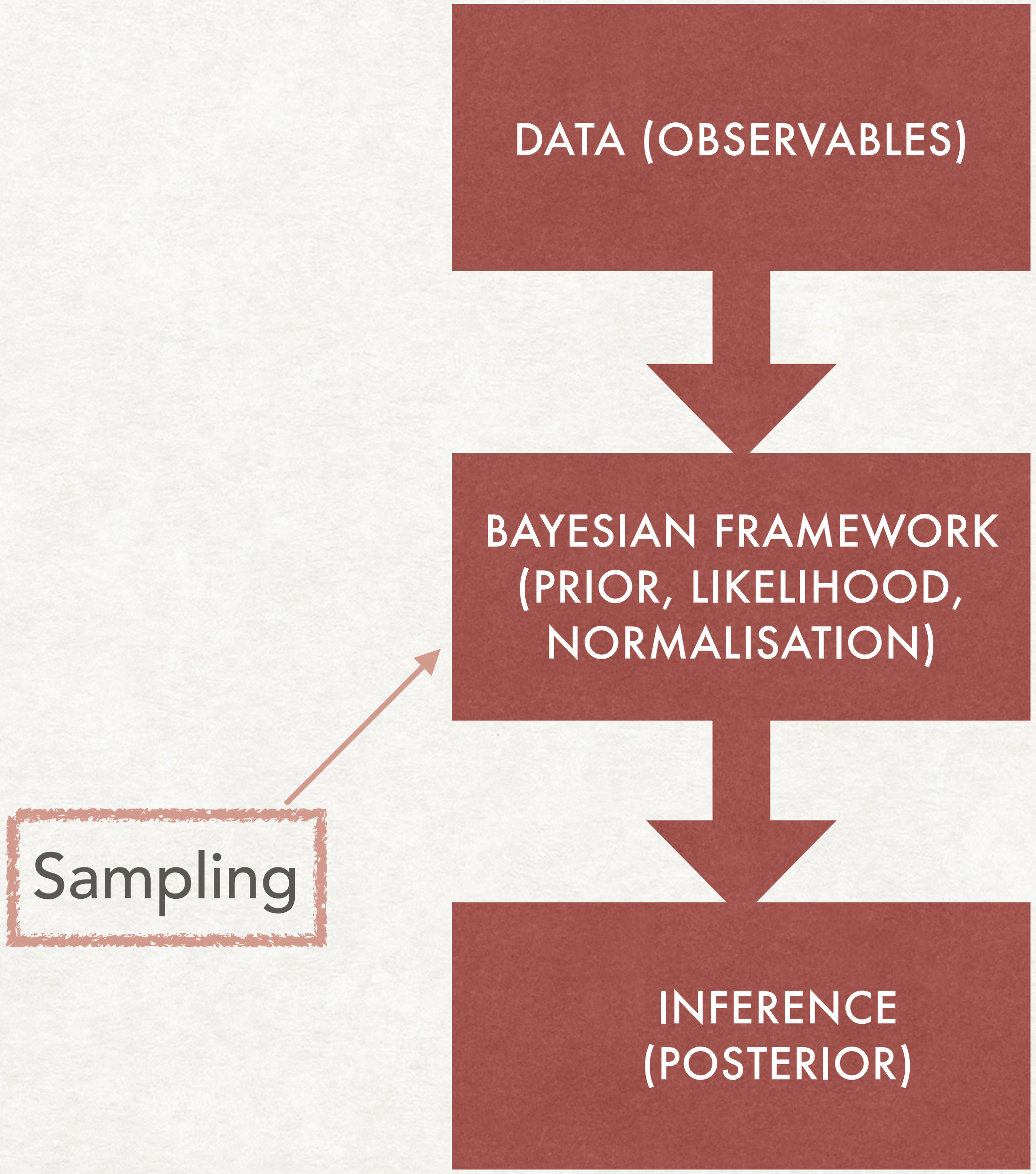
FREQUENTIST V.S. BAYESIAN

- Frequentist inference: draws conclusions from sample data by emphasising the **frequency** or proportion of the data.
- Bayesian inference: use **Bayes' Theorem** to infer the probability of hypothesis based on **priori information**.
- Different ways of thinking statistically.

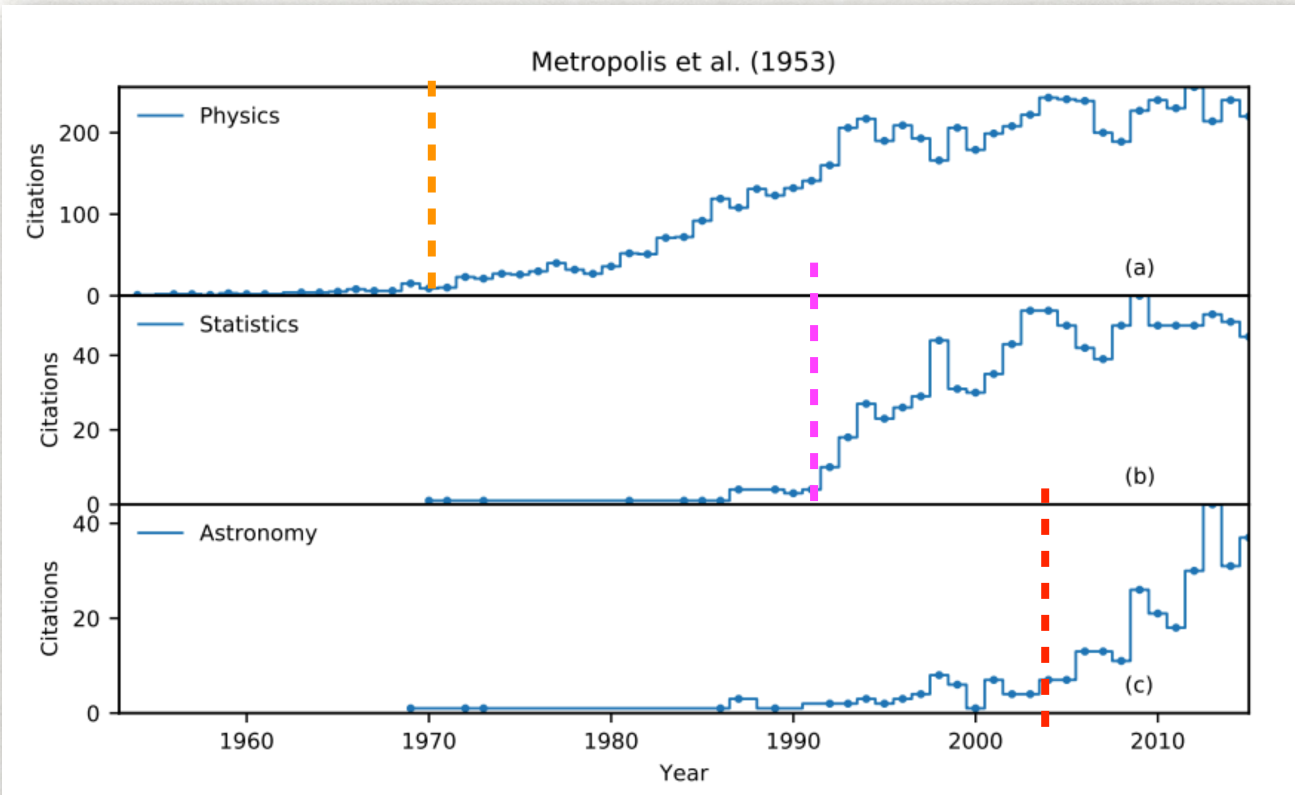
	Frequentist	Bayesian
Parameters	Fixed	Varied
Data	Varied	Fixed
Statistic	Frequency	Likelihood
Prior	No	Yes
Outcome	Estimates with error	Posterior distribution
Evaluation	Confidence	Credibility

WHY USE BAYESIAN?

- Problem with conditional probability
- Goal: **Estimation for parameters**
- Prior:
  - From knowledges or hypothesis
- Data:
  - From **only one trial (observation)**
  - With multiple-parameter space
- Likelihood
  - With unknown parameters
  - With coupled parameters

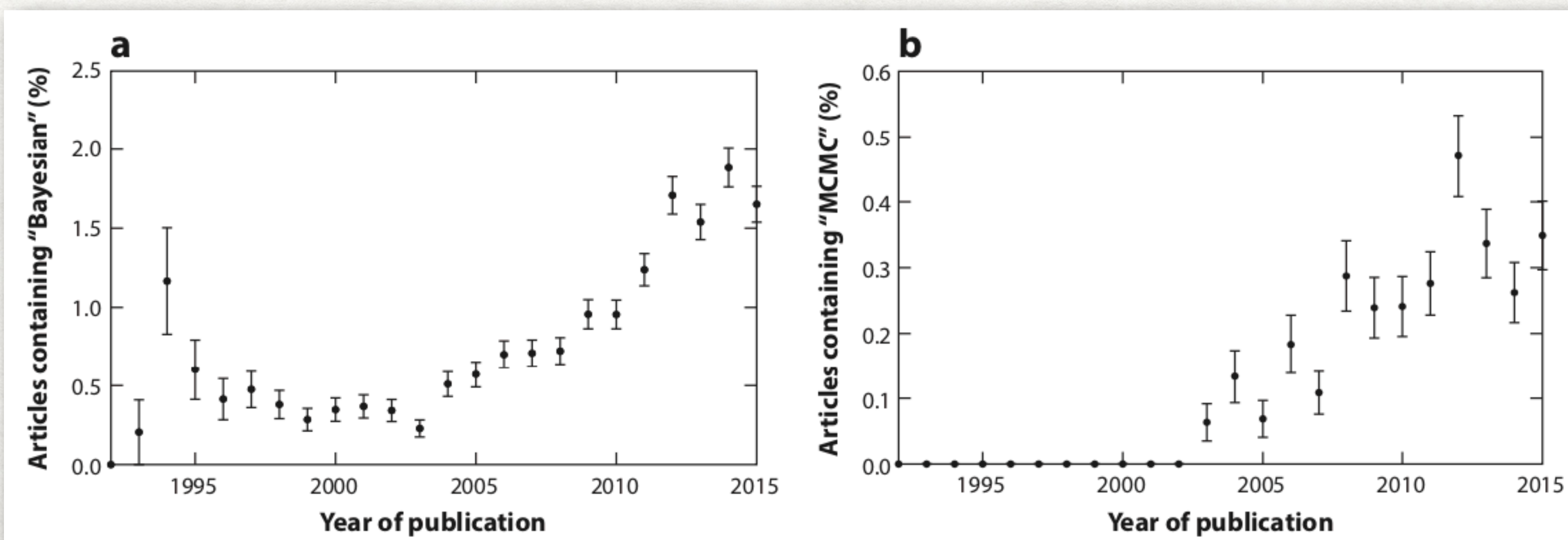


INCREASING APPLICATIONS

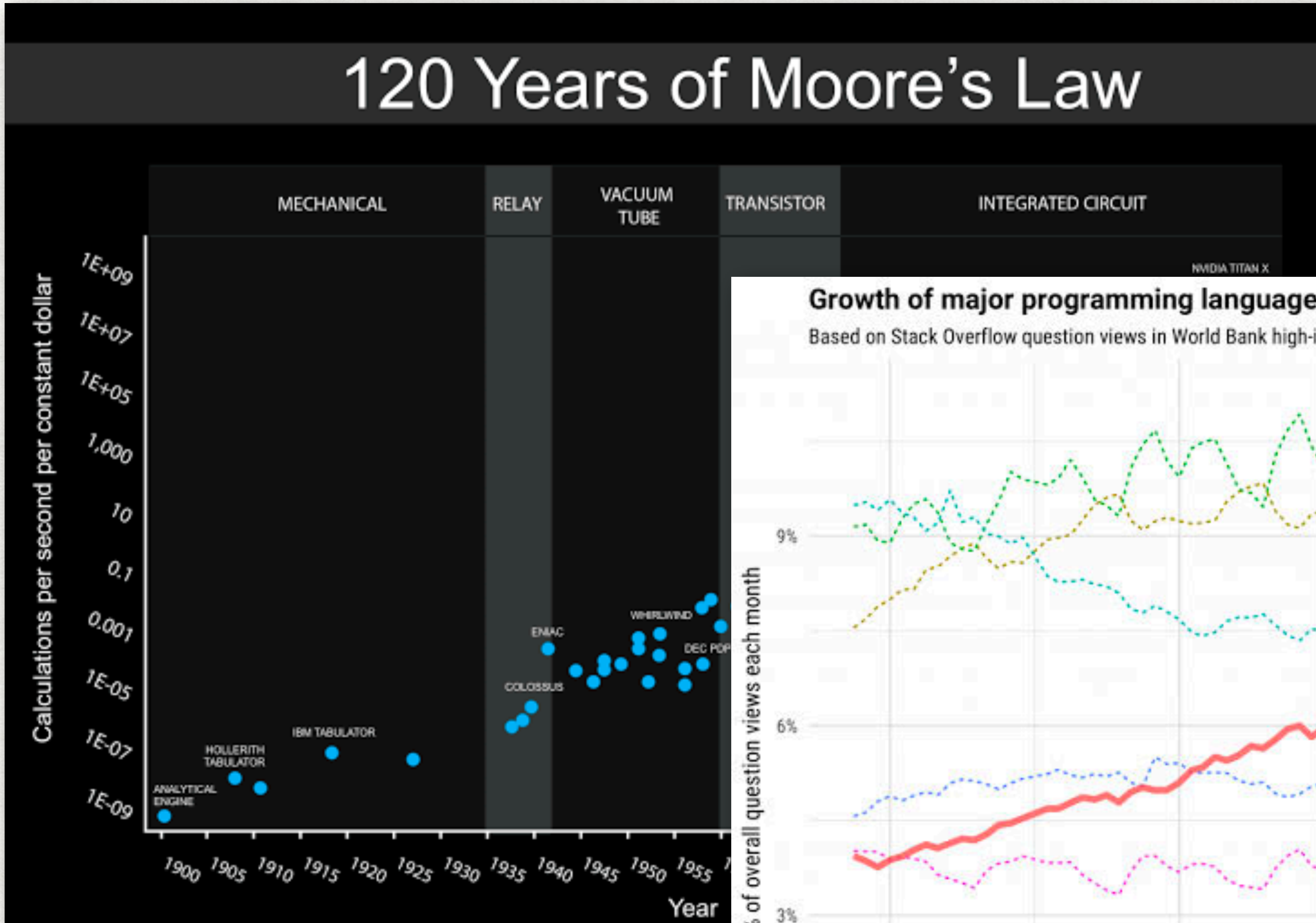


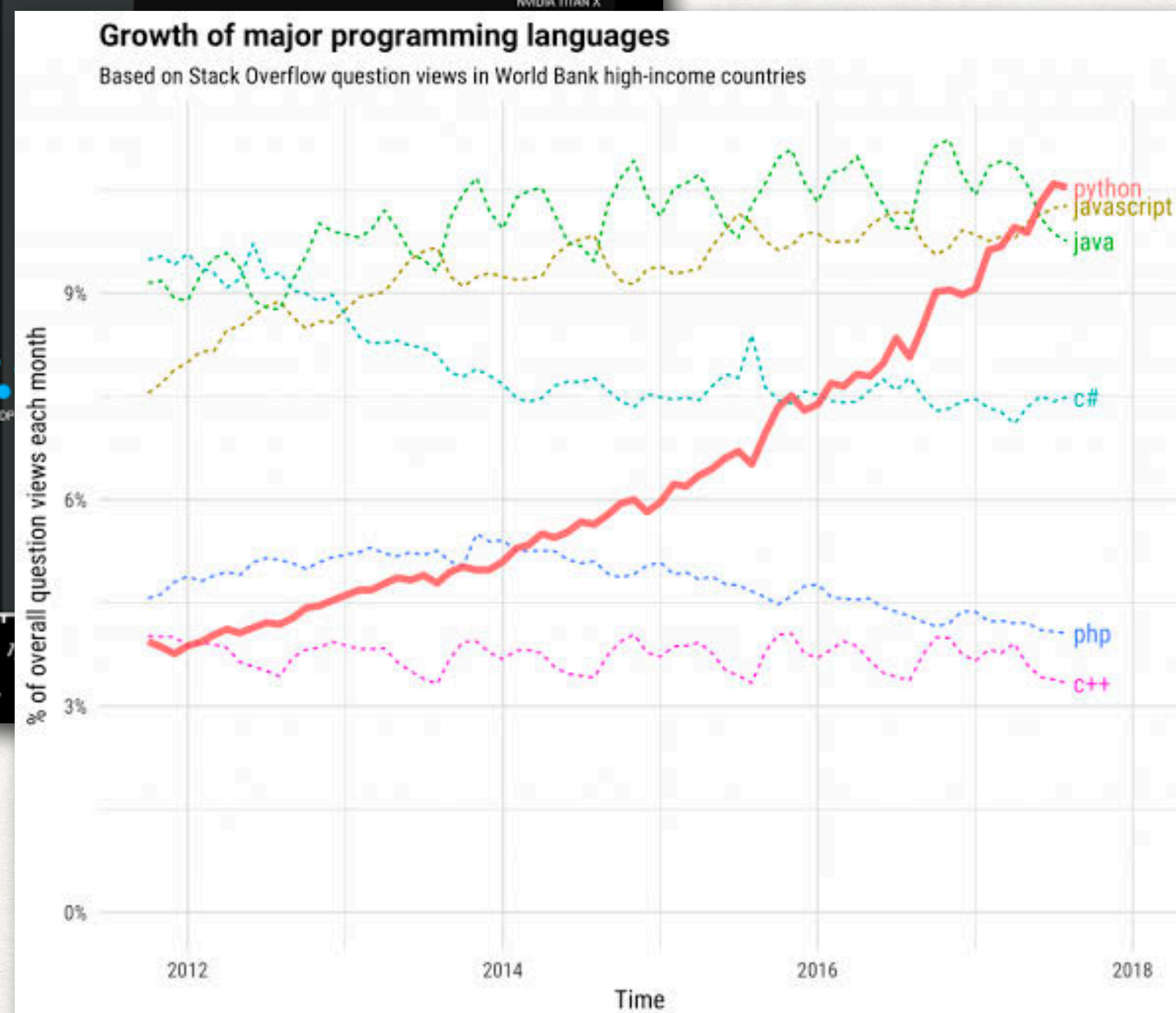
[PDF] Equation of State Calculations by Fast Computing Machines
https://bayes.wustl.edu › Manual › EquationOfState
by N Metropolis - 1953 - Cited by 40838 - Related articles
VOLUME 21, NUMBER 6. JUNE, **1953**. Equation of State Calculations by Fast Computing Machines. NICHOLAS **METROPOLIS**, ARIANNA W. ROSENBLUTH, ...

Sharma 2017, ARA&A



WHY IS MORE AND MORE POPULAR?





WHY TIME CONSUMING?

- Marginalised likelihood

$$P(\theta_{\mathrm{theo}}\,|\,D,I)=\frac{1}{P(D\,|\,I)}\int^{\forall\theta_i\in(\Theta\neq\theta_{\mathrm{theo}})}P(D\,|\,\Theta,I)P(\Theta\,|\,I)\mathrm{d}\theta_i$$

  - Normalisation is indispensable.

$$P(D\,|\,I)=\int^{\forall\theta_i\in\Theta}P(D\,|\,\Theta,I)P(\Theta\,|\,I)\mathrm{d}\theta_i$$

- Computational consumptions:
  - Integrations for N-dimensional parameters
  - Sampling for data in multiple parameter space

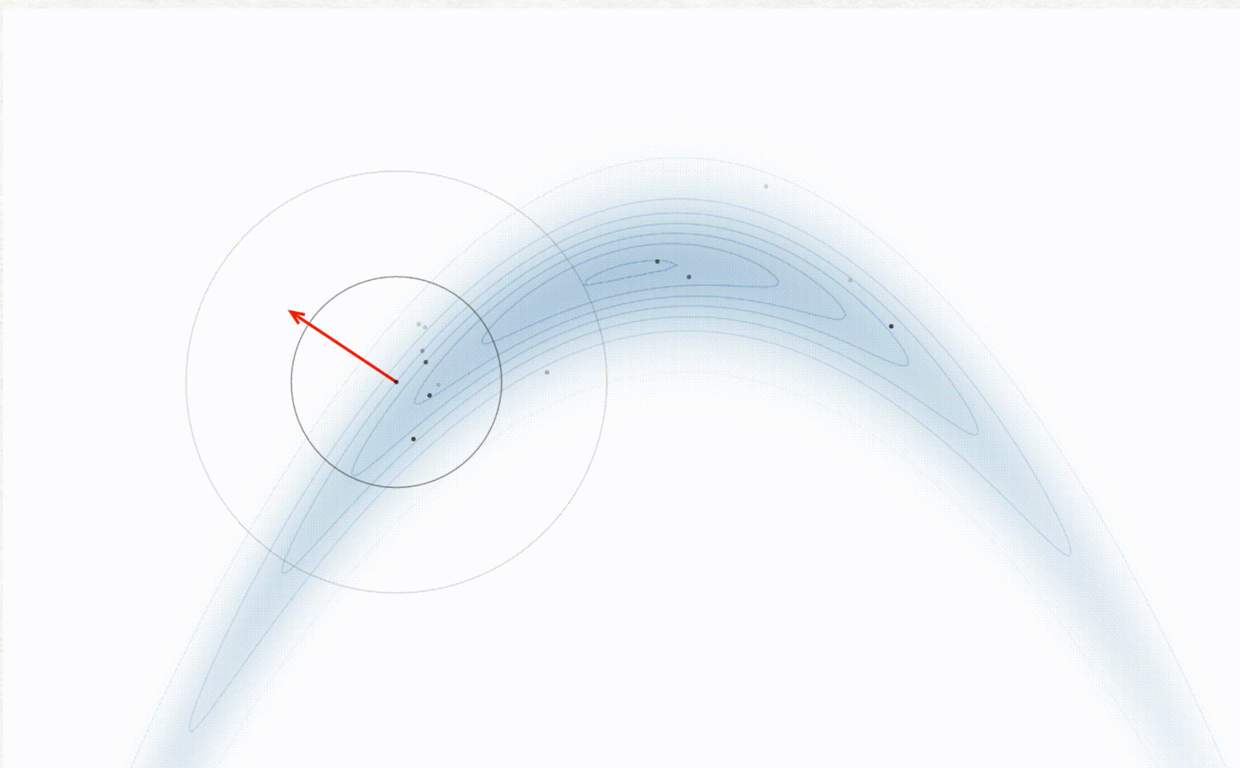
SAMPLING METHODS IN MONTE CARLO

- **Markov Chain Monte Carlo (MCMC)**
- Metropolis-Hastings Algorithm:
- Gibbs sampling
- Adaptive Metropolis
- **Ensemble and Affine Invariant Sampling**
- Monte Carlo Metropolis-Hastings
- Hamiltonian Monte Carlo
- Population Monte Carlo
- **Nested sampling**

Sharma 2017, ARA&A

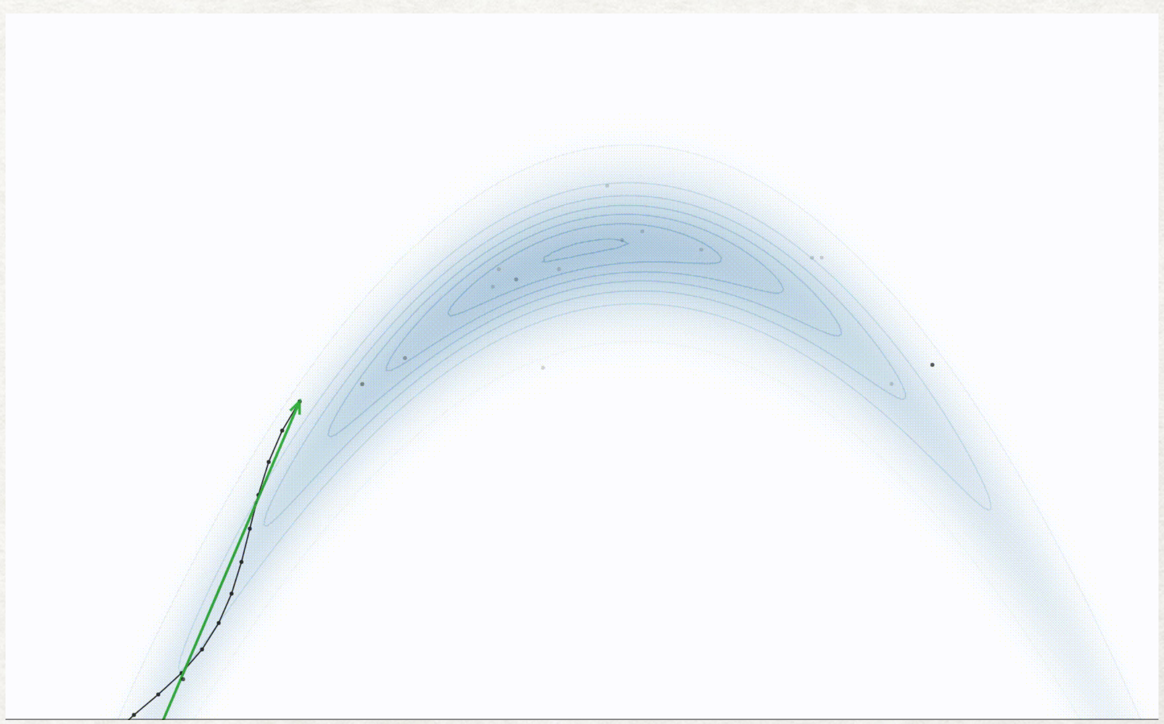
HOW SAMPLINGS WORK?

- Proposal distribution motions in each sampling



Random Walk MH

Hamiltonian MC



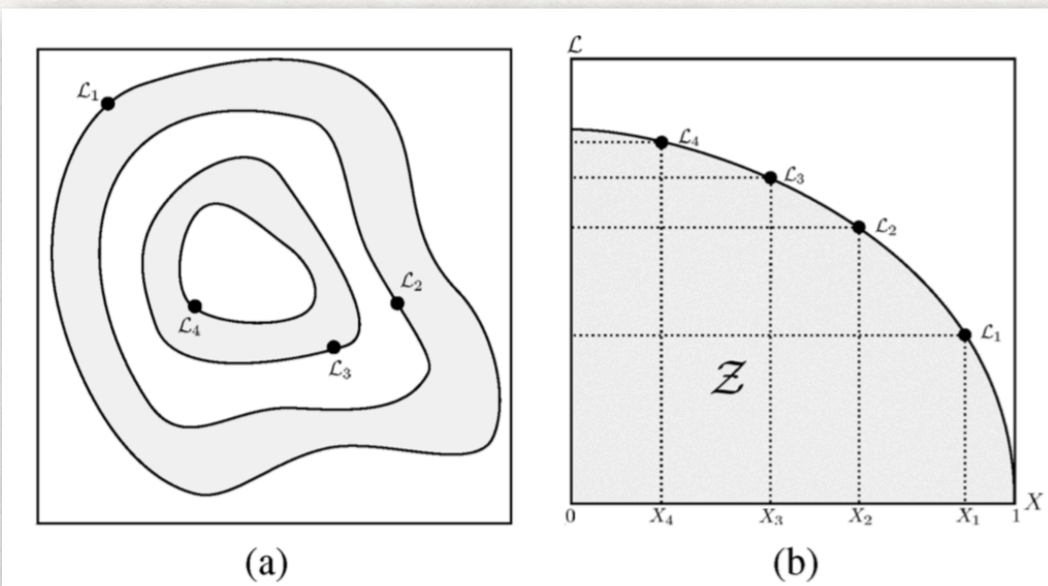
http://chi-feng.github.io/mcmc-demo/app.html

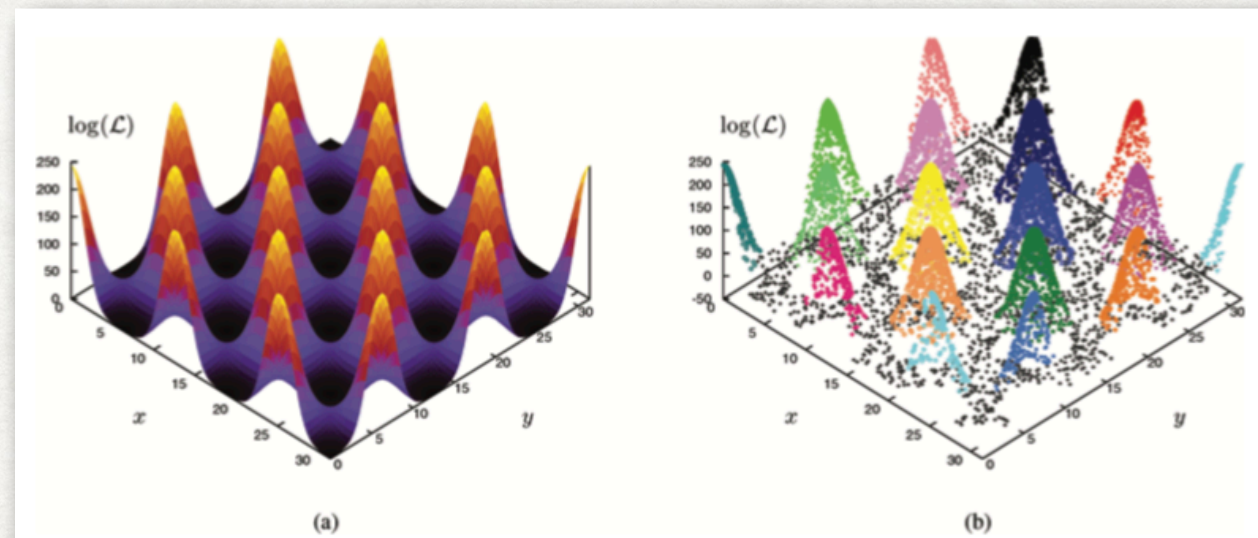
COMMON SAMPLERS

- MCMC:
  - **Emcee**
  - PyMC
  - Edward
  - TensorFlow Probability
  - PyStan
  - PyJAGS
  - …
- Nested sampling:
  - **MultiNest**
  - Nestle
  - CPNest
  - Dynesty
  - UltraNest
  - DNest4
  - PyPolyChord
  - …

RECOMMENDED TOOLKITS

- MultiNest: a Bayesian inference tool, written in Fortran and based on the **Nested Sampling** algorithm developed by John Skilling,





Feroz et al. 2009

  - PyMultiNest: http://johannesbuchner.github.io/PyMultiNest/

- emcee: **affine-invariant ensemble** sampler for **MCMC**

https://emcee.readthedocs.io/en/stable/#

emcee
The MCMC Hammer

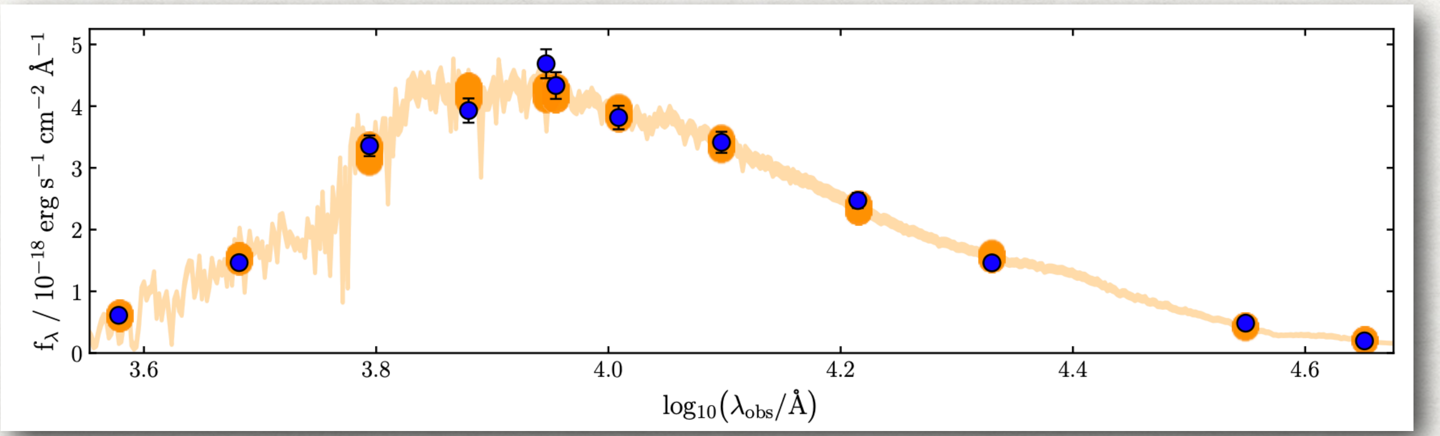
HOW TO DO BAYESIAN?

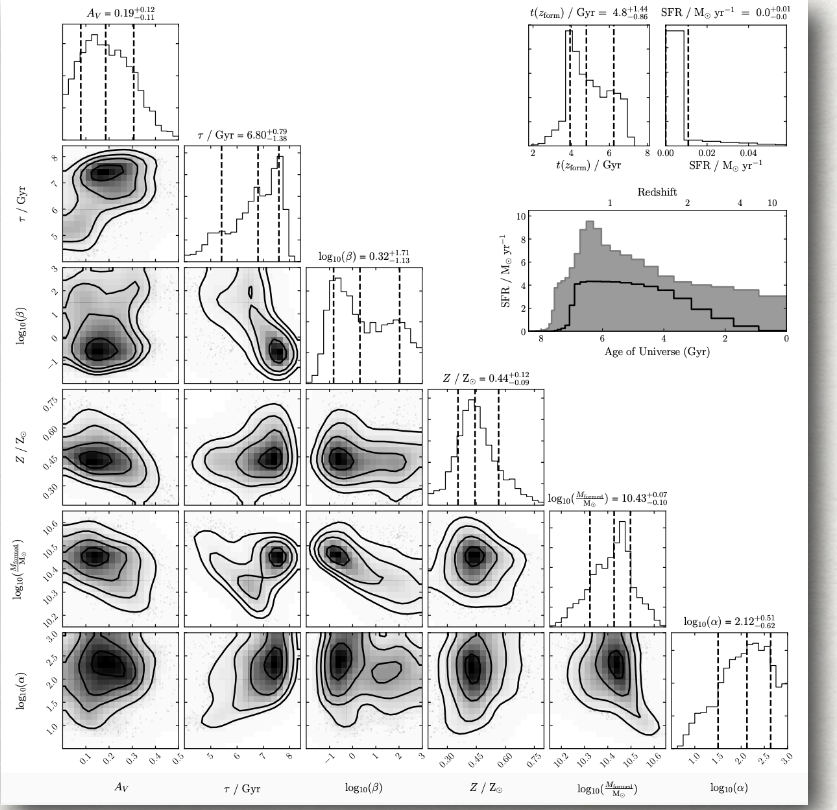
- Basic steps:
  - 1) Import the sampler package
  - 2) Define your prior function
  - 3) **Write log-likelihood function**
  - 4) Run sampler: do sampling
  - 5) Data analysis

```python
def LogLikelihood(self, cube):
    """
    The log likelihood function. This function has to be called "LogLike

    Args:
        cube (:class:`numpy.ndarray`): an array of parameter values.

    Returns:
        float: the log likelihood value.
    """

    # extract parameters
    m = cube[0]
    c = cube[1]

    # calculate the model
    model = self._model(x, m, c)

    # normalisation
    norm = -0.5*self._ndata*LN2PI - self._ndata*self._logsigma

    # chi-squared
    chisq = np.sum(((self._data - model)/(self._sigma))**2)

    return norm - 0.5*chisq
```

APPLICATIONS IN ASTROPHYSICS

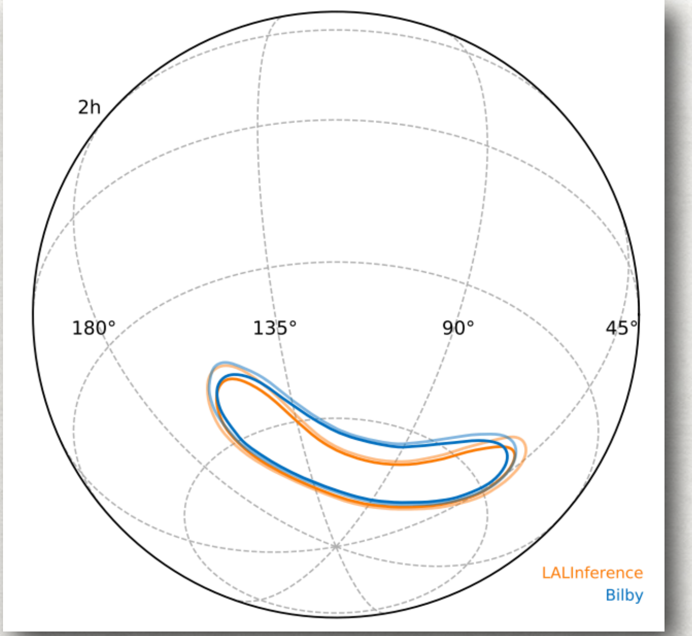
- Spectrum or SED fitting

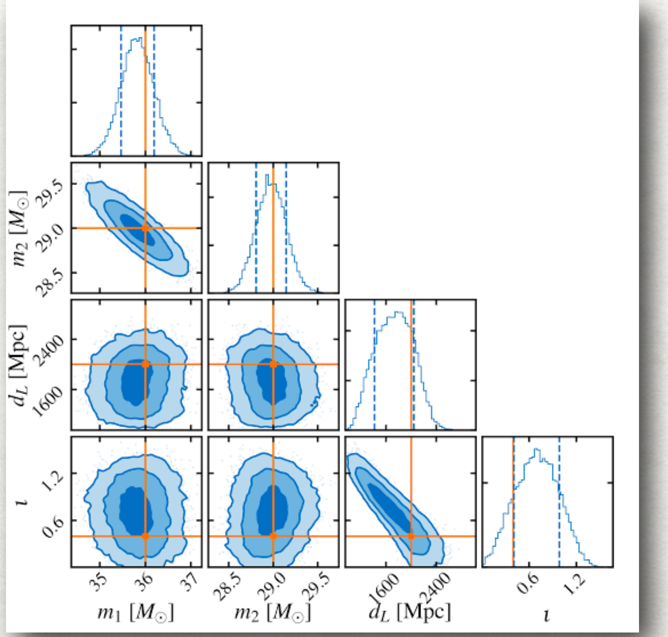




Using Bagpipes (Carnall et al. 2018)

- LIGO GW detection



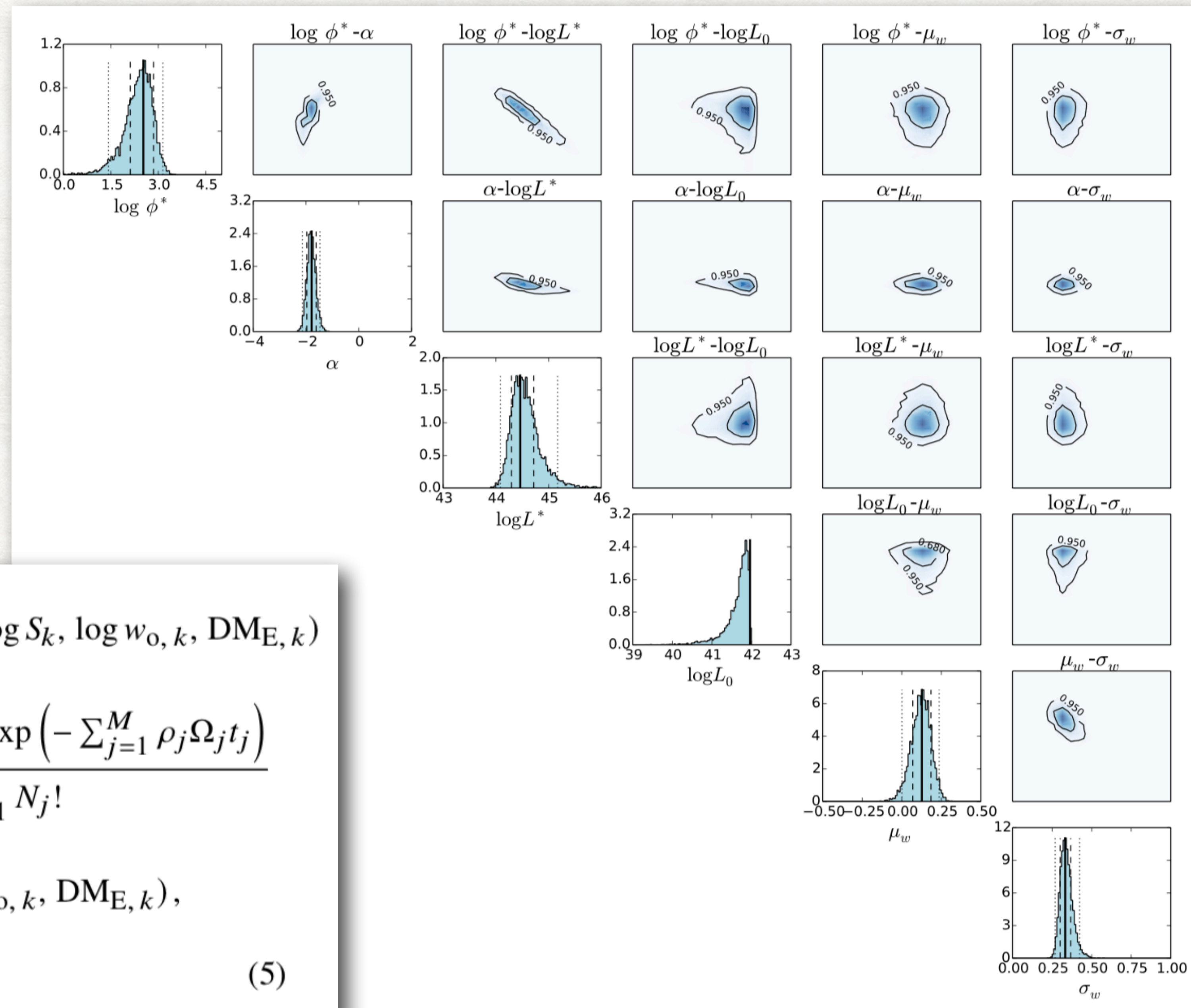


Using Bilby
(Ashton et al. 2018)

APPLICATIONS IN ASTROPHYSICS

- My work on FRB luminosity function

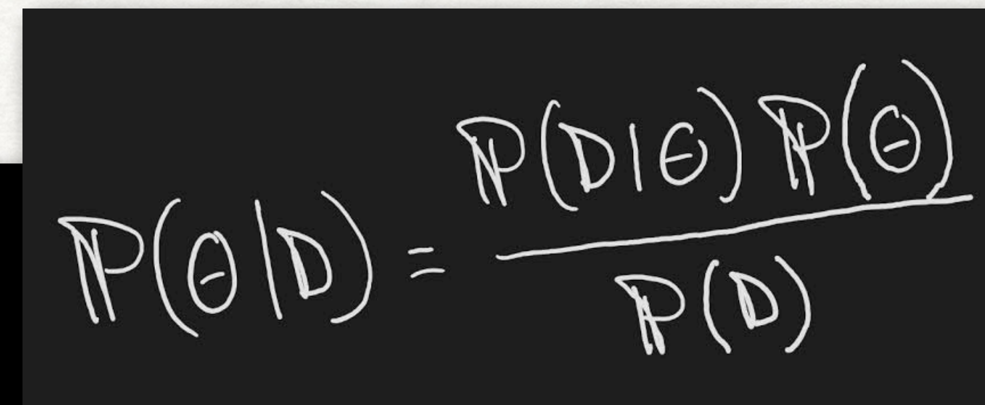
Luo et al. 2019a



$$
\begin{aligned}
\mathcal{L}(\log S, w_{\rm o}, N, \mathrm{DM_E}) &= \prod_{j=1}^{M} \mathcal{L}_i(N_i) \cdot \prod_{k=1}^{N} f(\log S_k,\, \log w_{\mathrm{o},\,k},\, \mathrm{DM}_{\mathrm{E},\,k}) \\
&= \frac{\prod_{j=1}^{M} (\rho_j \Omega_j t_j)^{N_j} \cdot \exp\left(-\sum_{j=1}^{M} \rho_j \Omega_j t_j\right)}{\prod_{j=1}^{M} N_j!} \\
&\quad \cdot \prod_{k=1}^{N} f(\log S_k,\, \log w_{\mathrm{o},\,k},\, \mathrm{DM}_{\mathrm{E},\,k}),
\end{aligned} \tag{5}
$$

A KEY TO OUTLIER DETECTION?

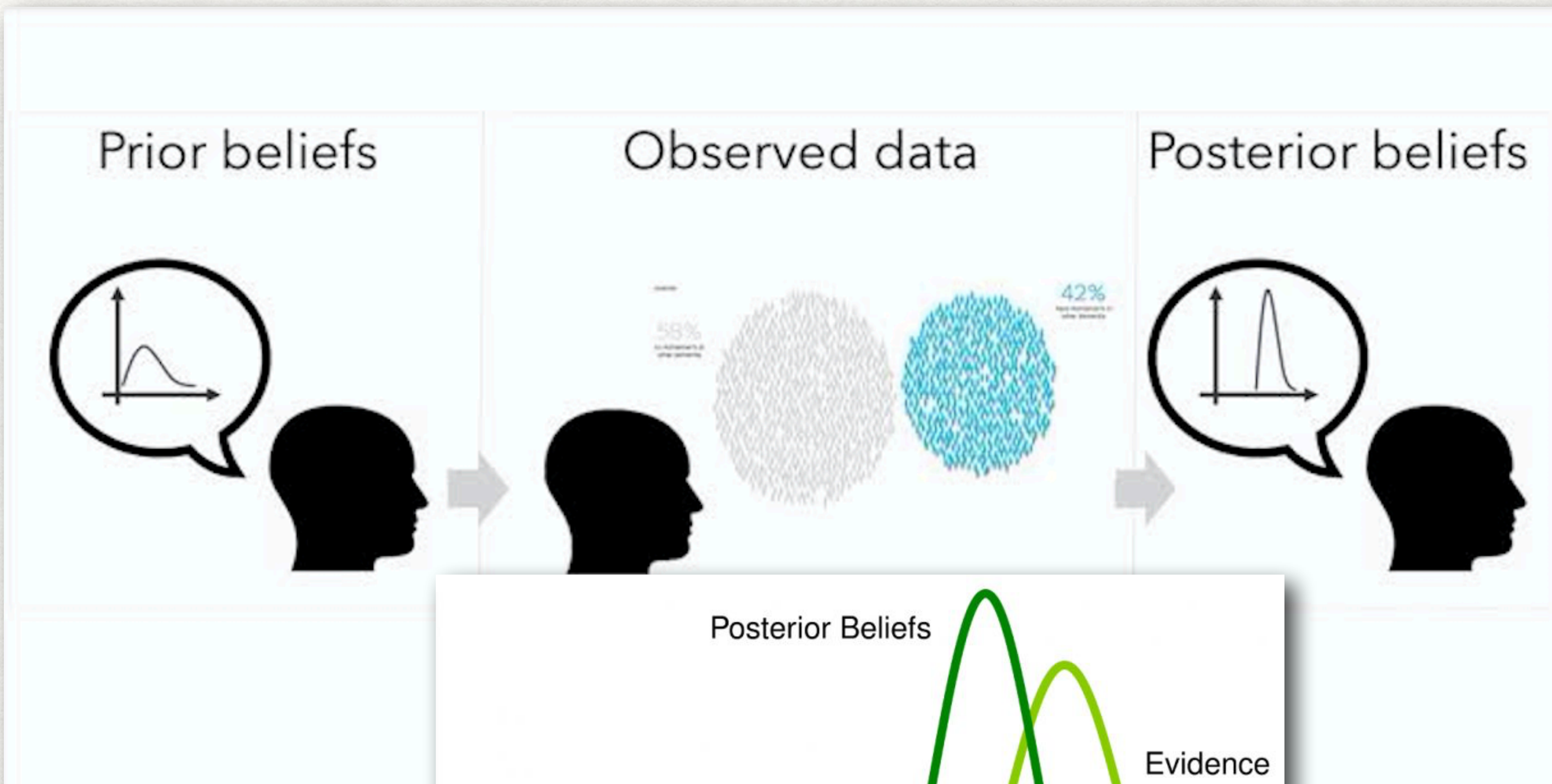
- Make the machines think like Bayesian?

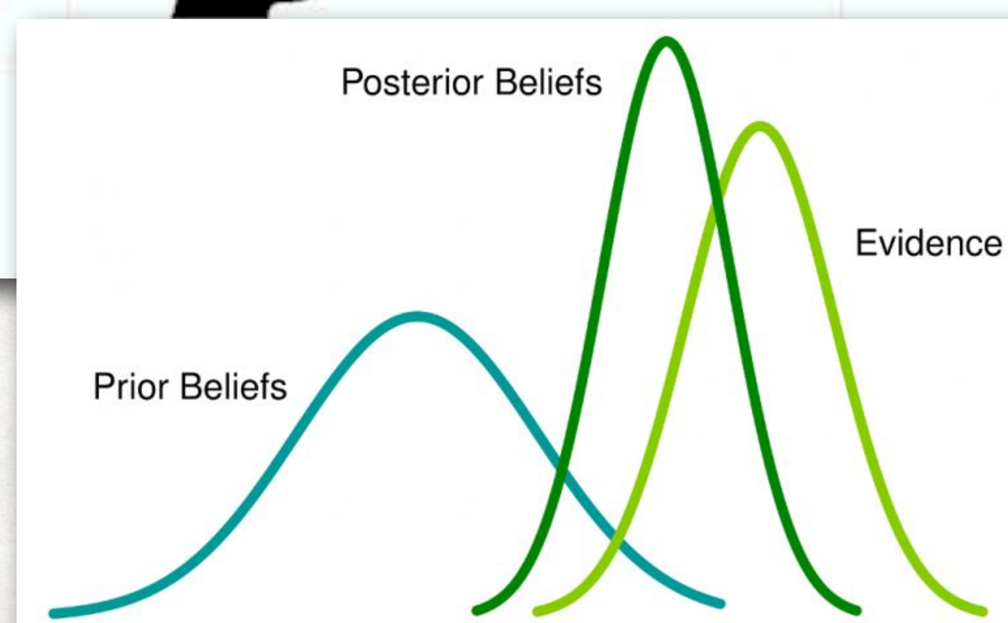


BELIEVE IT NOR NOT, IT IS A BAYESIAN QUESTION.

HOW BAYESIAN WORK FOR YOUR COGNITIONS

- Experience -> New Beliefs

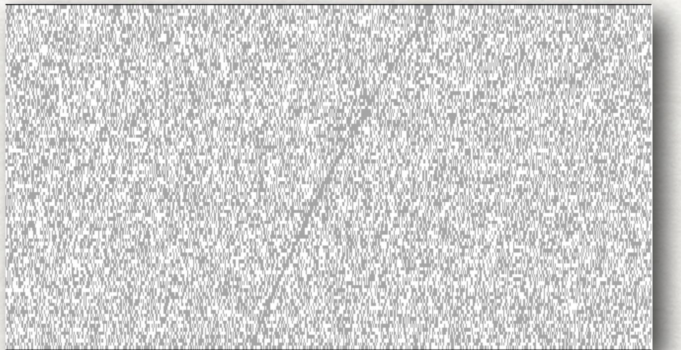


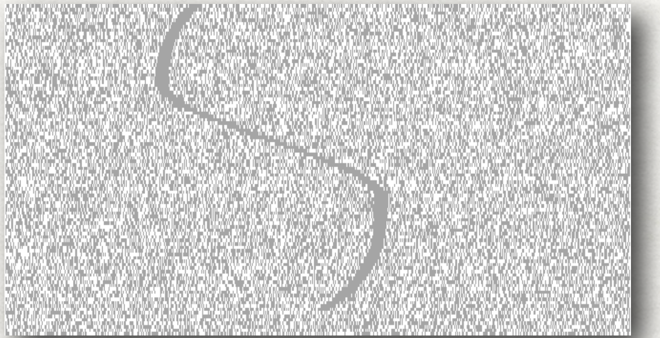


BAYESIAN NEURAL NETWORK?

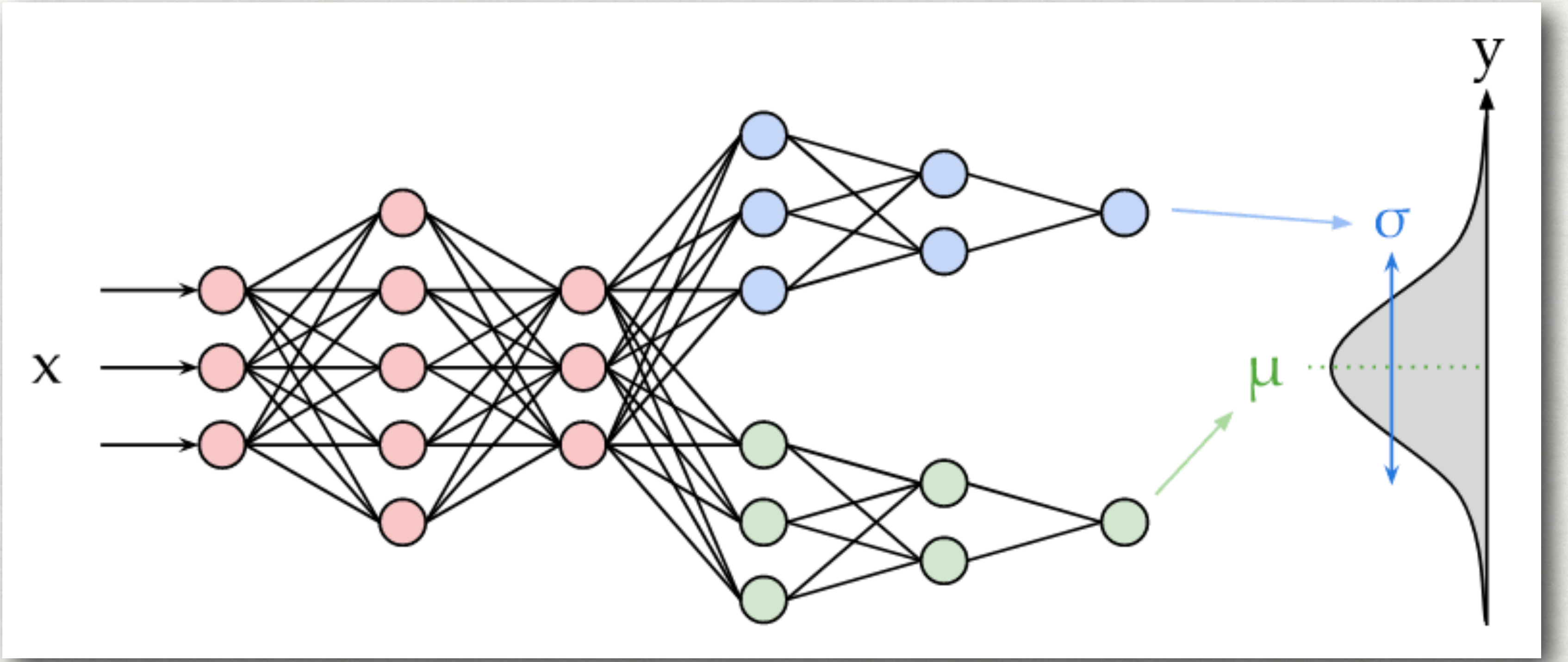
- Input: Data (Knowns and unknowns)
- Output: Probability of anomaly

FRB





Weird signal



OPEN QUESTIONS

- How to implement Bayesian inference in ML?
- **Is it computing limited for Bayesian in big data?**
- How to quantify anomaly in the probabilistic model?
- How to train the ML model? Supervised or Unsupervised?
- How to evaluate the performance?

THE END

Thank you for listening!