Introduction to statistical inference

Aarya Patil
LSST-Discovery Alliance Catalyst Fellow,
Max-Planck-Institut für Astronomie

Quick Review of Day 1

Bayesian and Frequentist Statistics

probability of a data set given the null hypothesis

purely driven by the data

prior information

probability of a hypothesis given a particular data set

p-value

probability of obtaining another data set at least as extreme as the one collected

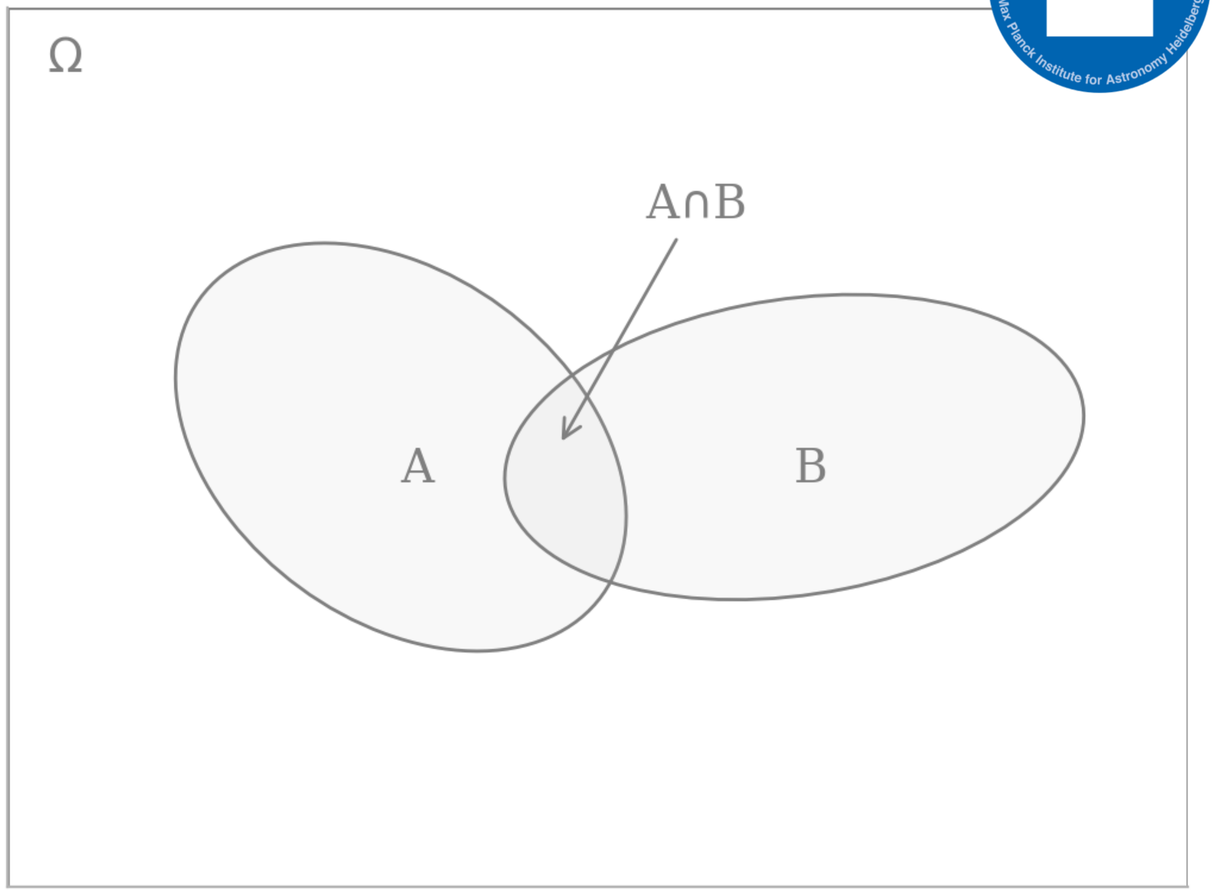
Fornacon-Wood et al. 2022

Bayesian Statistics

Bayes' theorem

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

$$[P(B) \neq 0]$$



$$P(A, B \mid \Omega) = P(A \mid B, \Omega)\, P(B \mid \Omega)$$
$$= P(B \mid A, \Omega)\, P(A \mid \Omega)$$

Data modeling

1. Parameter estimation

2. Model comparison

3. Prediction

Zero-parameter models

Result D	**Is Model M true?** M denotes if a person has COVID-19	
	Yes	No
positive	true positive $P(\mathrm{D} \mid \mathrm{M})$	false positive $P(\mathrm{D} \mid \mathrm{M}')$
negative	false negative $P(\mathrm{D}' \mid \mathrm{M})$	true negative $P(\mathrm{D}' \mid \mathrm{M}')$

Hypothesis testing

$$P(\mathrm{M} \mid \mathrm{D}) = \frac{1}{1 + \frac{1}{R}}$$

$$R = \frac{P(\mathrm{D} \mid \mathrm{M})\, P(\mathrm{M})}{P(\mathrm{D} \mid \mathrm{M}')\, P(\mathrm{M}')}$$

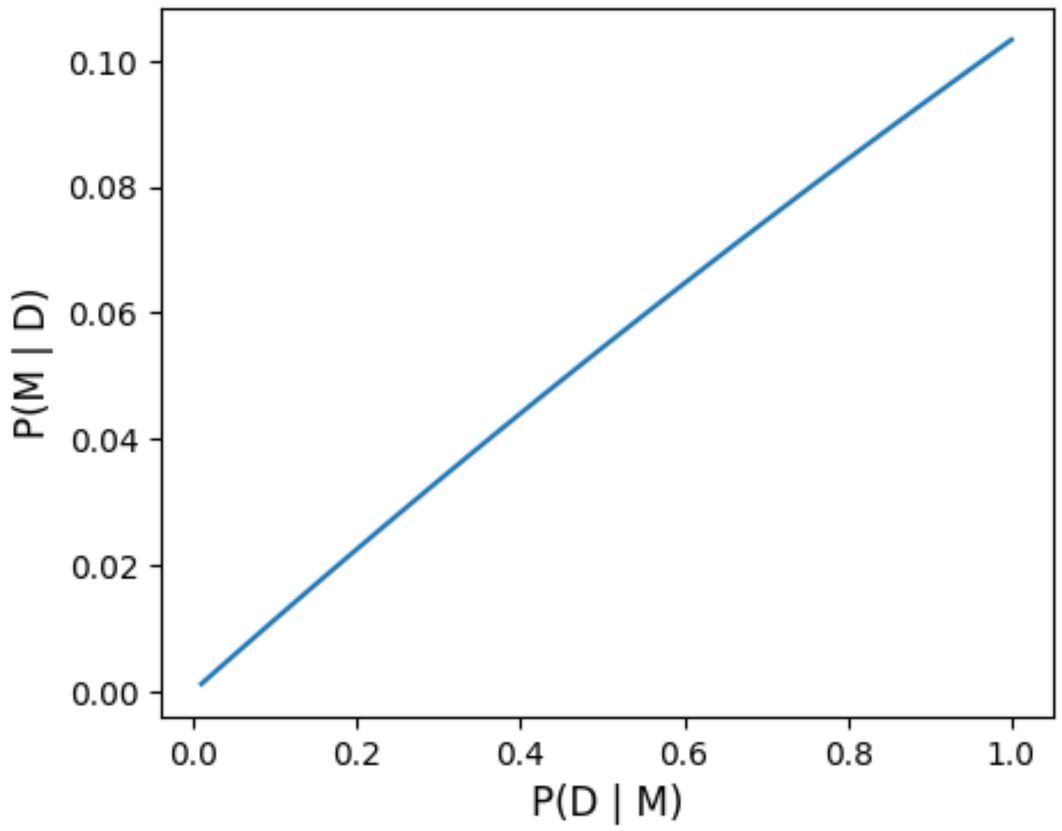
“Posterior odds ratio”

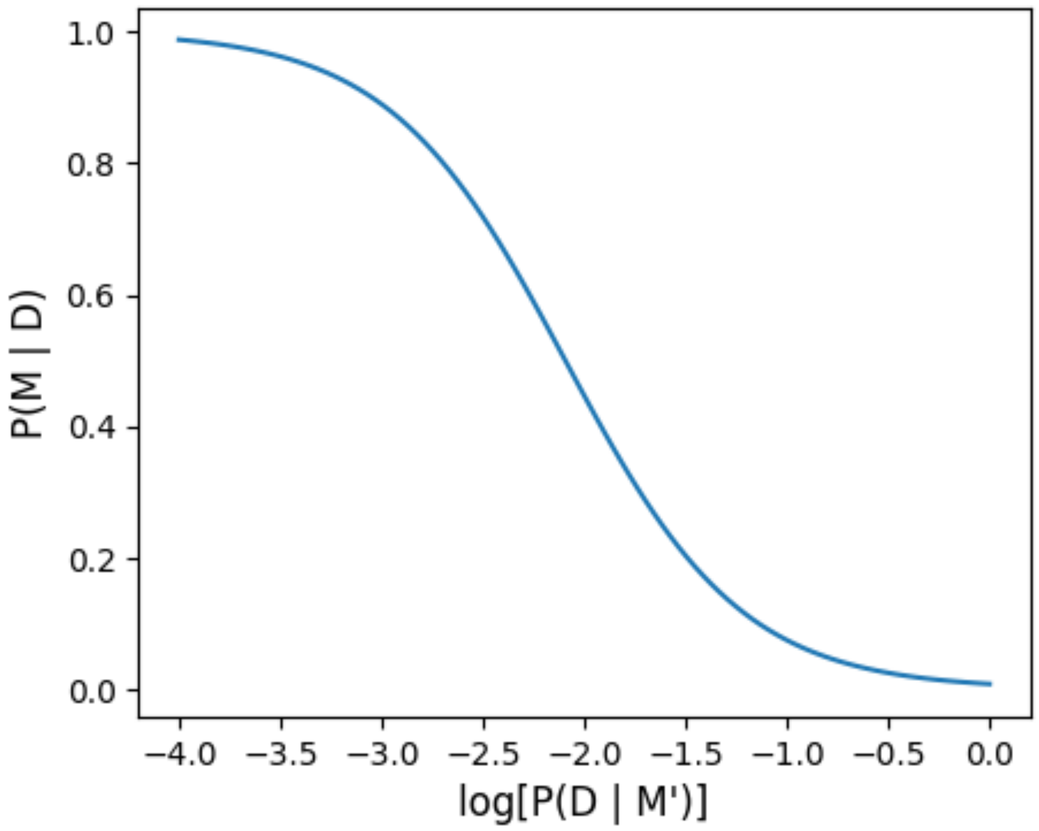
$$P(\mathrm{M} \mid \mathrm{D}) = \frac{P(\mathrm{D} \mid \mathrm{M})\, P(\mathrm{M})}{P(\mathrm{D})}$$

Exercise:
Derive using Bayes’ theorem

Thinking in terms of frequencies

$P(\mathrm{D} \mid \mathrm{M}) = 0.9 \quad P(\mathrm{D} \mid \mathrm{M}') = 0.07 \quad P(\mathrm{M}) = 0.008$





Hypothesis testing

$$P(\mathrm{M} \mid \mathrm{D}) = \frac{1}{1 + \frac{1}{R}}$$

$$[P(\mathrm{M}) = P(\mathrm{M}')]$$

$$\boxed{R = \frac{P(\mathrm{D} \mid \mathrm{M})\, P(\mathrm{M})}{P(\mathrm{D} \mid \mathrm{M}')\, P(\mathrm{M}')}}$$

$$\boxed{BF = \frac{P(\mathrm{D} \mid \mathrm{M})}{P(\mathrm{D} \mid \mathrm{M}')}}$$

“Posterior odds ratio”

“Bayes factor”

Hypothesis testing

$$P(\mathrm{M}_1 \mid \mathrm{D}) = \frac{P(\mathrm{D} \mid \mathrm{M}_1)P(\mathrm{M}_1)}{\sum_i P(\mathrm{D} \mid \mathrm{M}_i)P(\mathrm{M}_i)}$$

$$\boxed{R = \frac{P(\mathrm{D} \mid \mathrm{M}_1)\, P(\mathrm{M}_1)}{P(\mathrm{D} \mid \mathrm{M}_2)\, P(\mathrm{M}_2)}}$$

“Posterior odds ratio”

$$\boxed{BF = \frac{P(\mathrm{D} \mid \mathrm{M}_1)}{P(\mathrm{D} \mid \mathrm{M}_2)}}$$

“Bayes factor”

Parametric models

Posterior

Likelihood

Prior

$$P(\boldsymbol{\Theta} \mid \mathbf{D}, \mathrm{M}) = \frac{P(\mathbf{D} \mid \boldsymbol{\Theta}, \mathrm{M})\, P(\boldsymbol{\Theta} \mid \mathrm{M})}{P(\mathbf{D} \mid \mathrm{M})}$$

Evidence

Why is sampling from $P(\mathbf{x})$ difficult?

Let's assume we can evaluate a function $P^*(\mathbf{x})$ such that

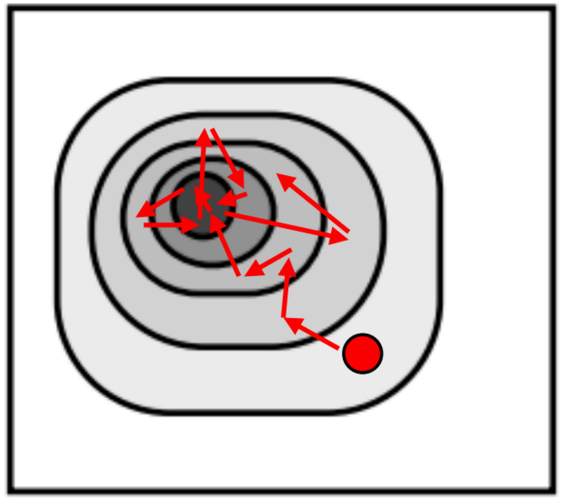
$$P(\mathbf{x}) = P^*(\mathbf{x})/Z$$

Difficulties:

1. Normalizing constant $\quad Z = \int d^N\mathbf{x}\, P^*(\mathbf{x})$
2. High-dimensional spaces

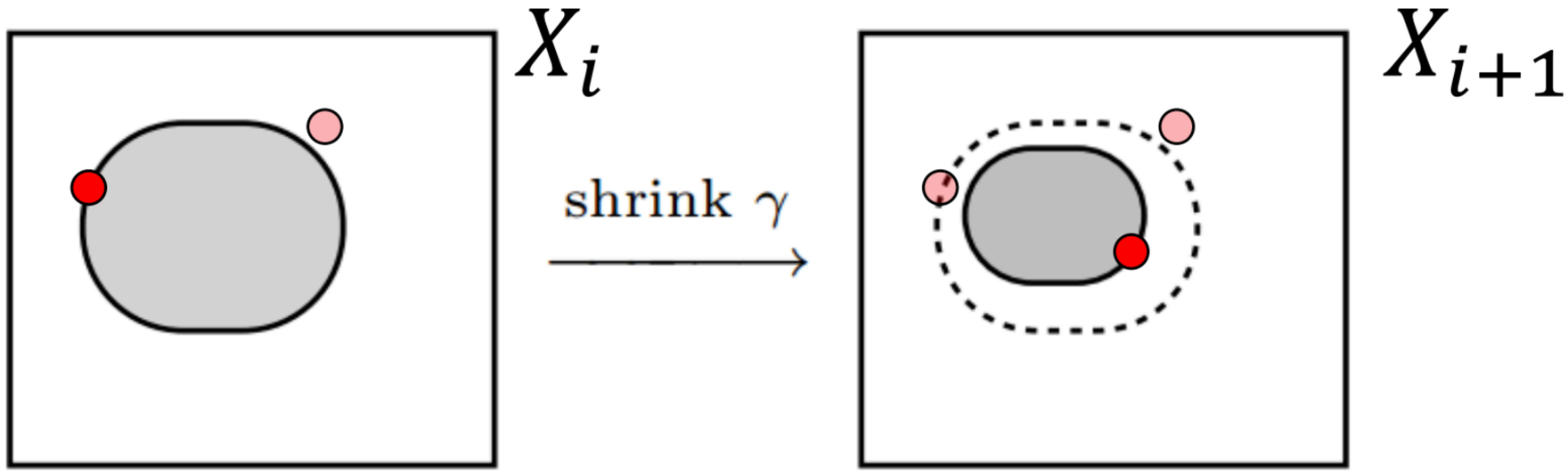
MacKay 1998

Sampling a distribution

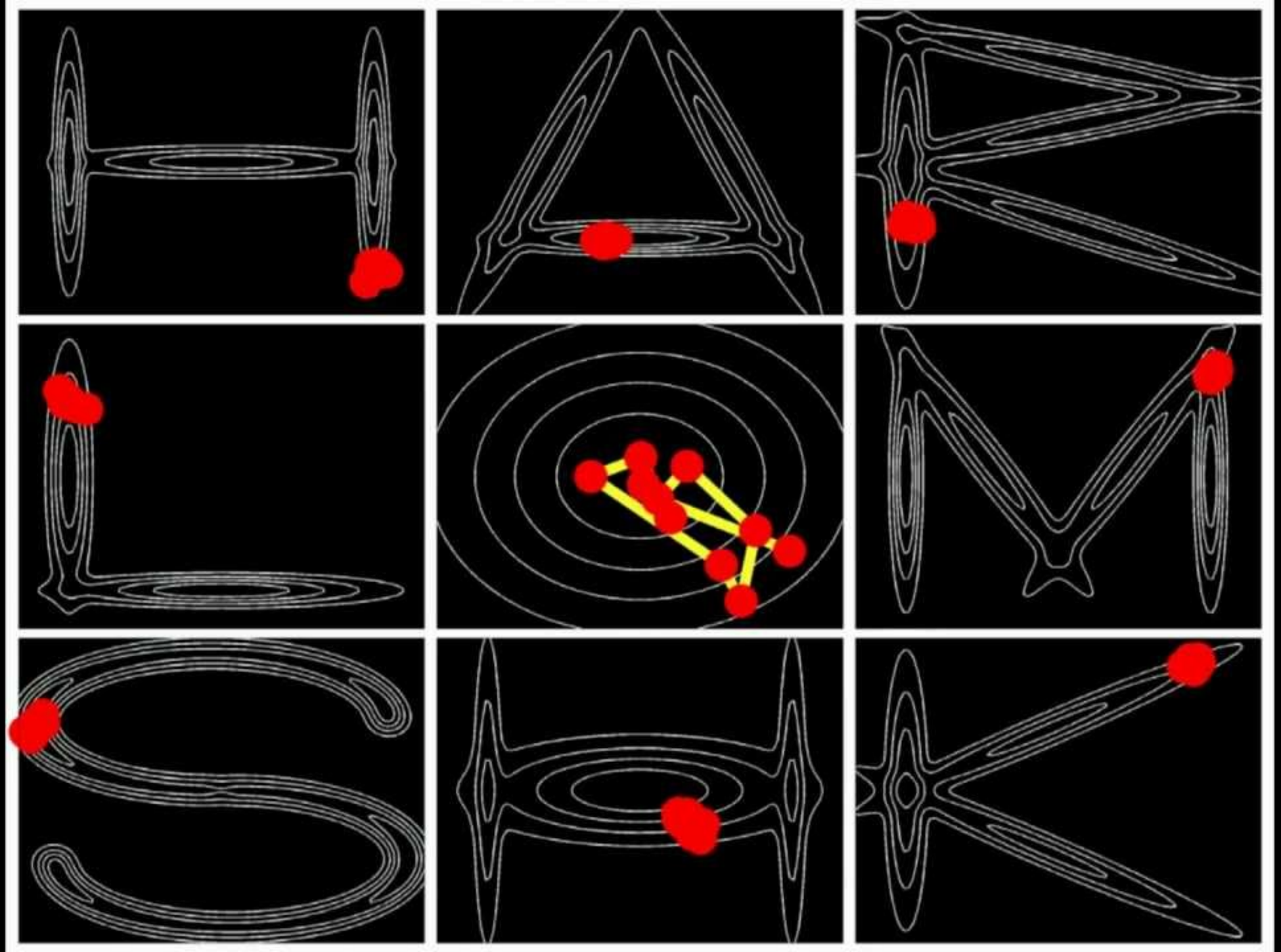


Markov chain Monte Carlo (MCMC)
solving a difficult problem once

Nested Sampling
solving an easier problem several times



Adapted from 2021 talk by Josh Speagle



Prior

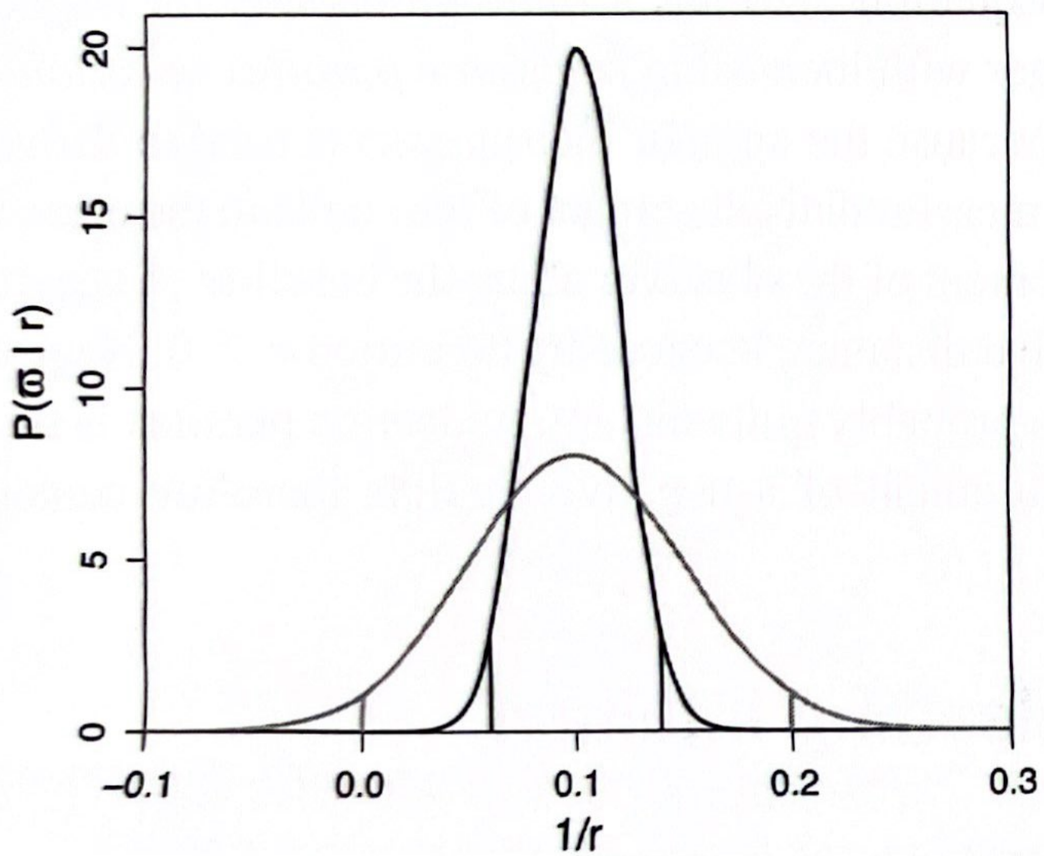
Parallax example

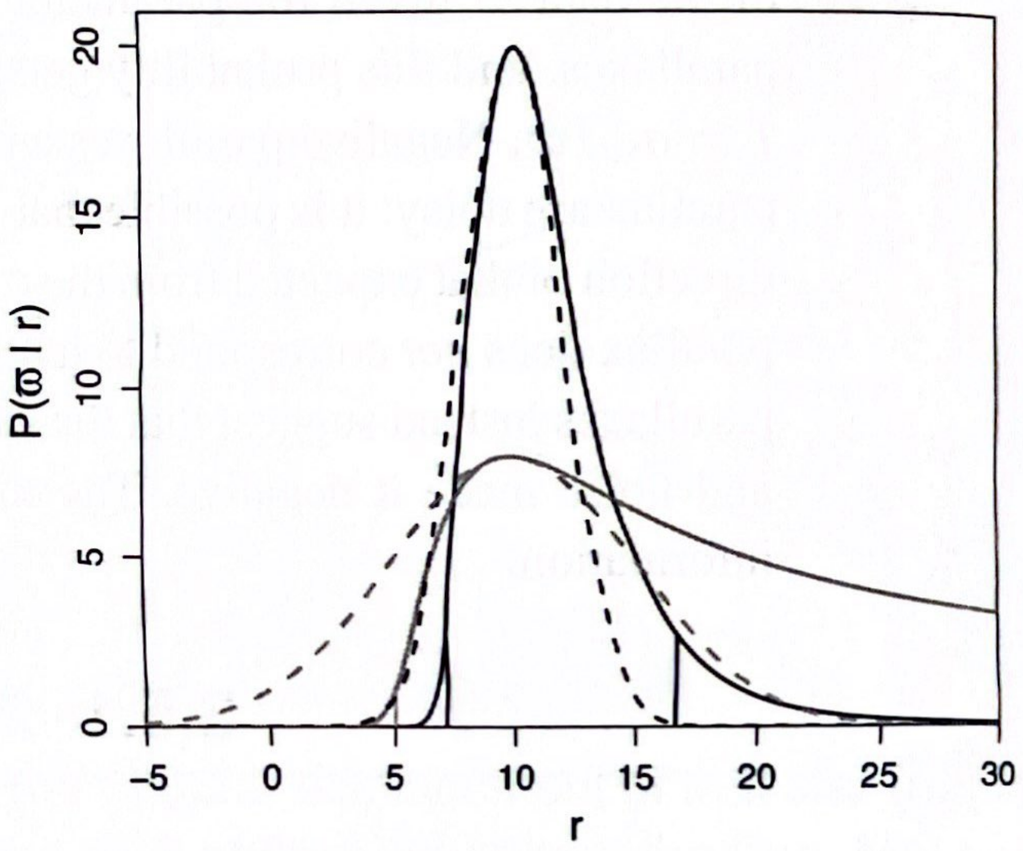
$$r\ [\mathrm{kpc}] = \frac{1}{\overline{\omega}}\ [\mathrm{mas}]$$

$$P(\overline{\omega} \mid r) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(\overline{\omega} - \frac{1}{r}\right)^2\right]$$

Bailer-Jones 2017
Eadie et al. 2023

Prior



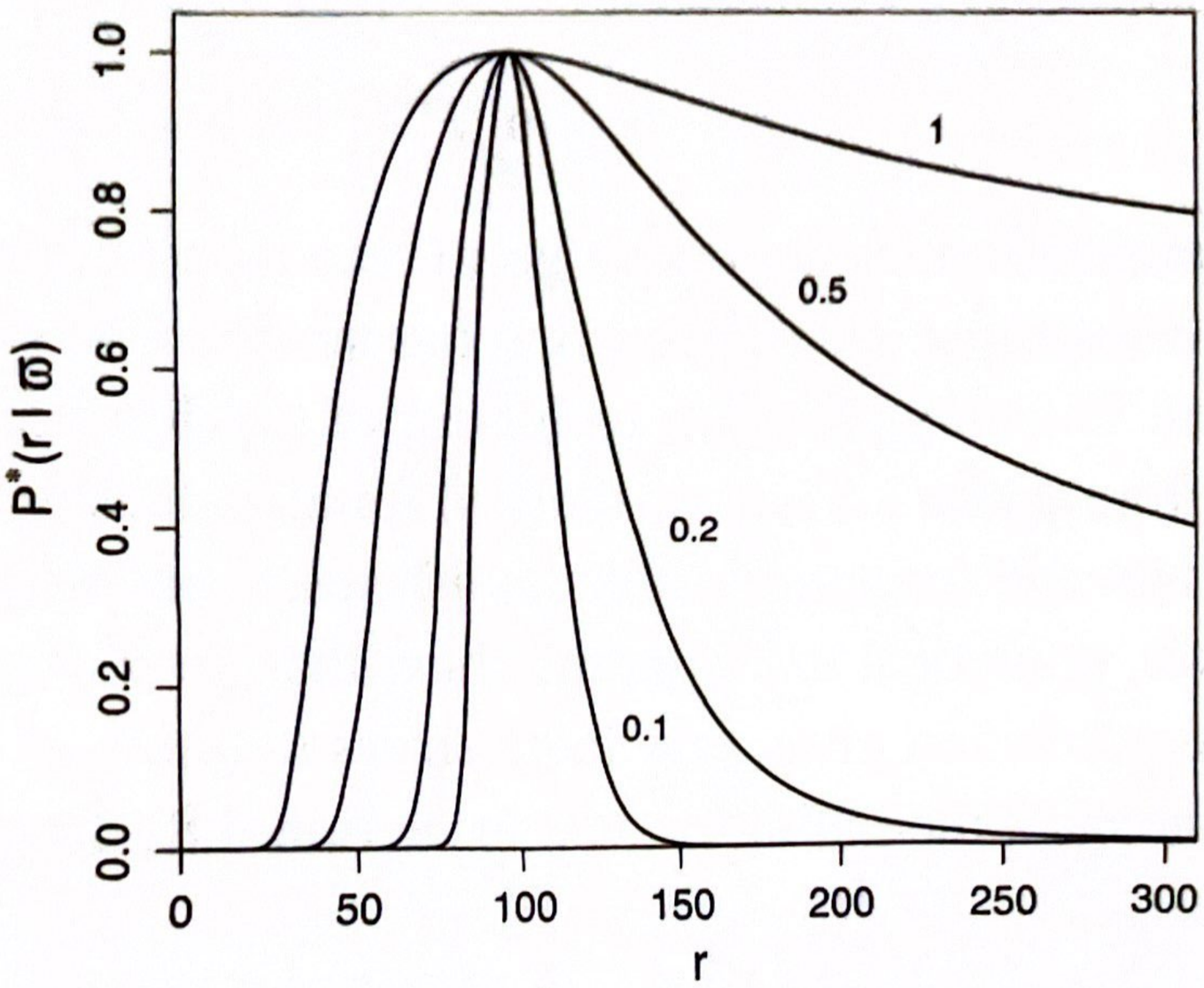


Prior

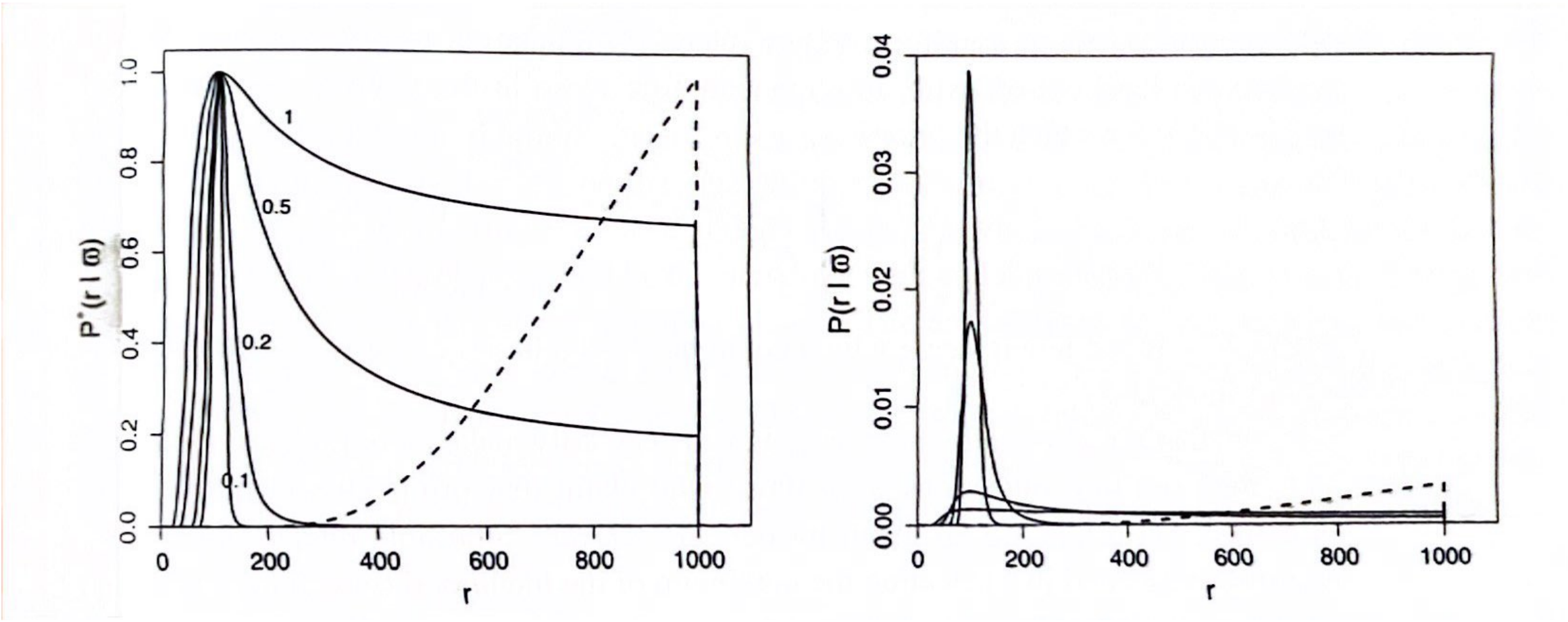
“How can science depend on assumptions?”

“How can I know the prior if I haven’t yet measured any distances?”

Improper uniform prior



Proper uniform prior



Conjugate priors

Example

Occurrence of objects in a large population e.g. brown dwarfs among all stars

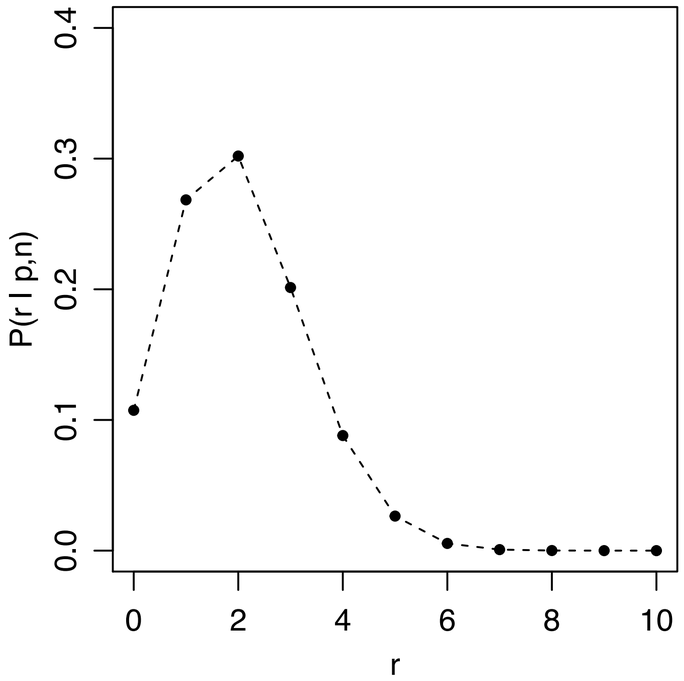
Survey just n objects of which *r* are brown dwarfs

What is the true population fraction *p* of brown dwarfs?

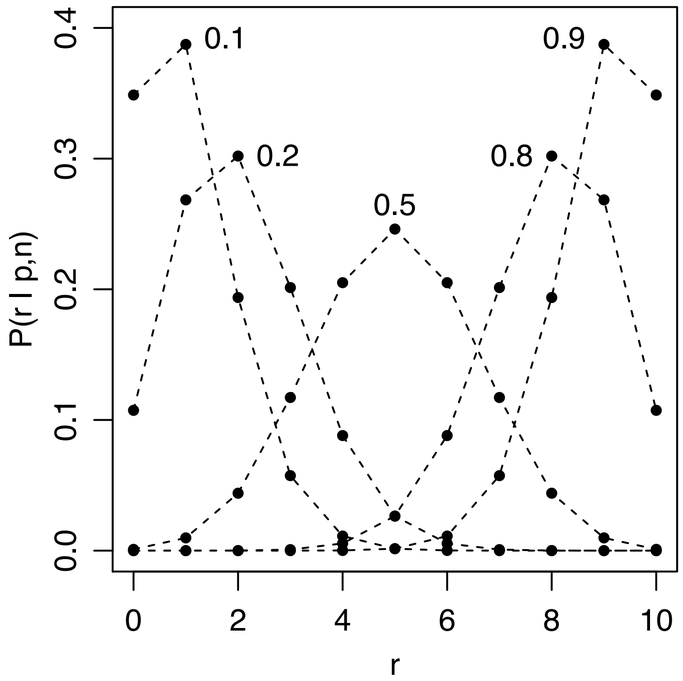
Binomial distribution

$$P(r \mid p, n) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} \quad \text{where} \quad n \geq 0,\ r \geq 0,\ 0 \leq p \leq 1$$

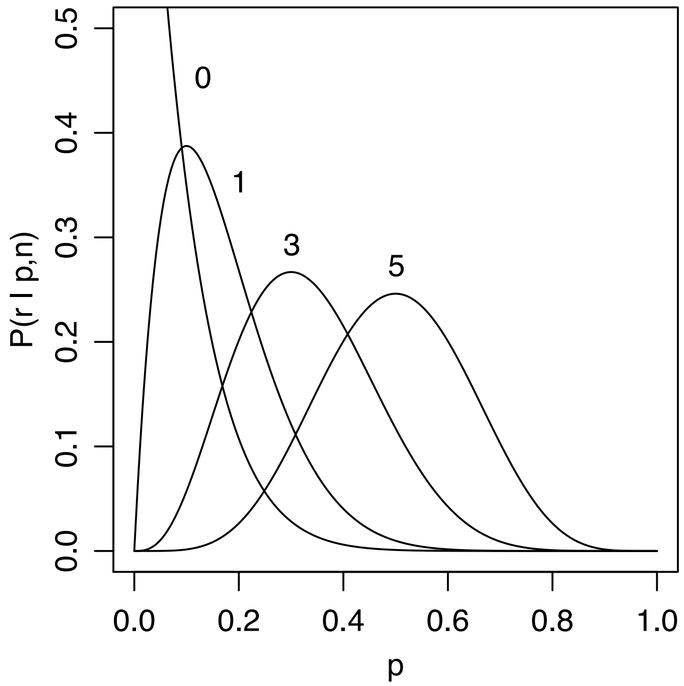
n=10, p=0.2



n=10, p=(0.1,0.2,0.5,0.8,0.9)



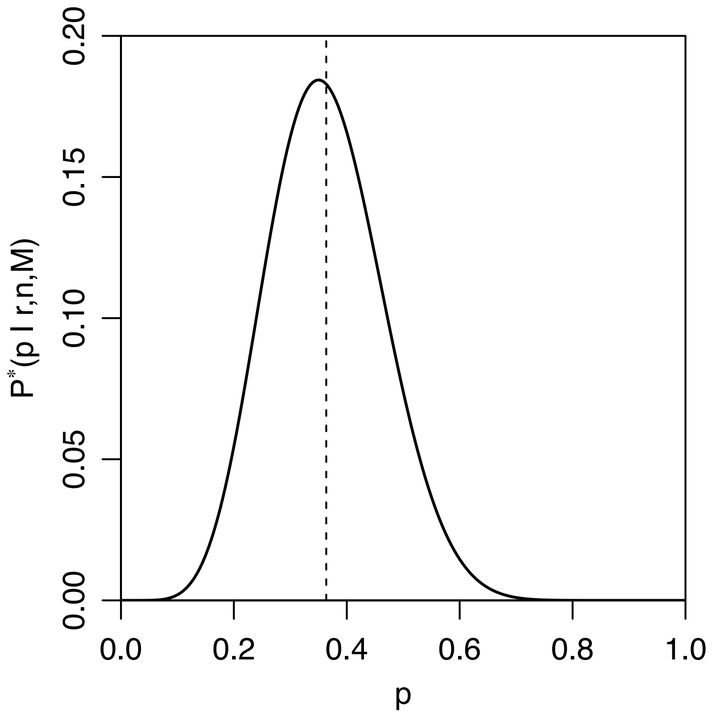
not a PDF in p
n=10, r=(0,1,3,5)

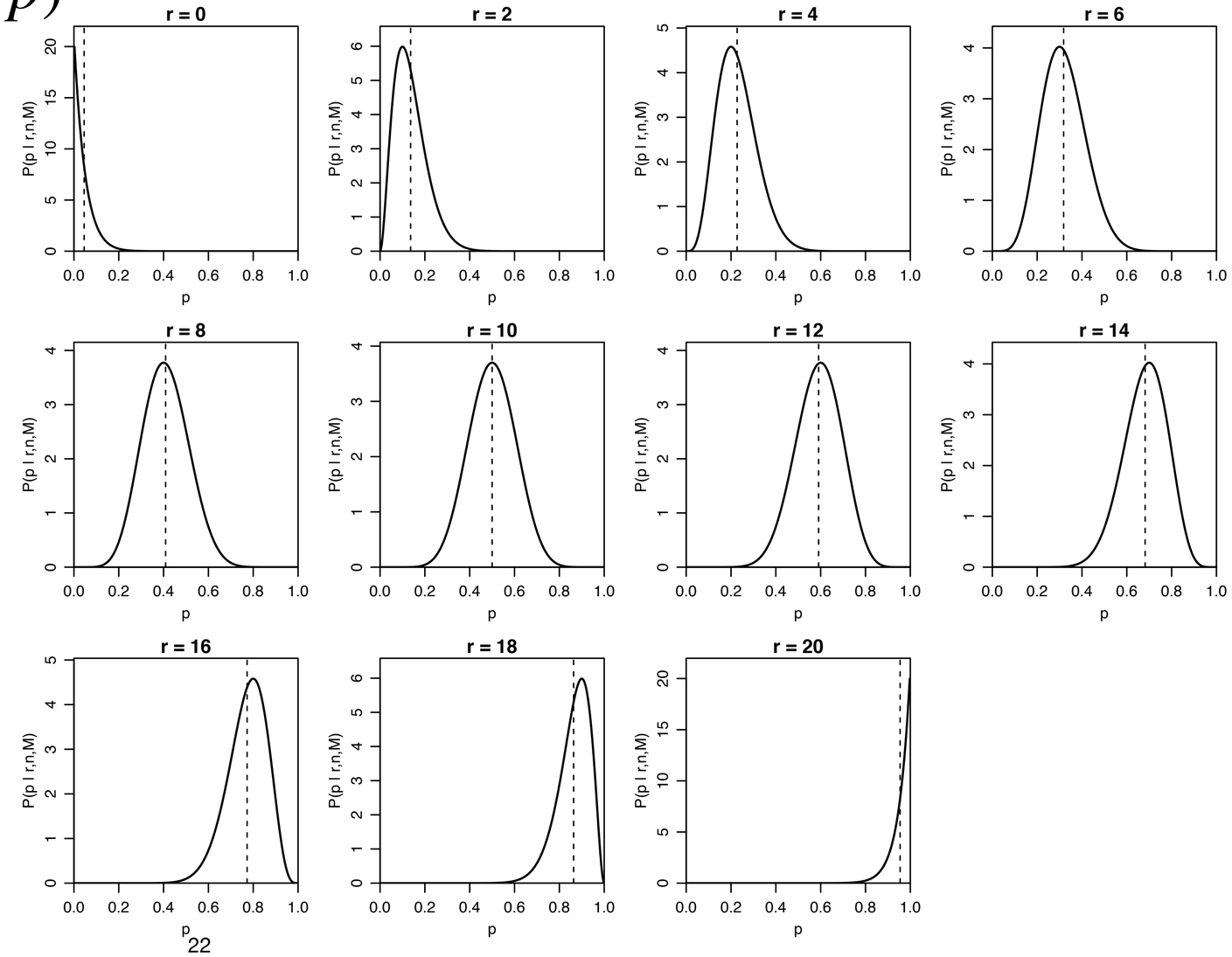


Posterior with uniform prior

$$P^*(p \mid r, n) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}$$

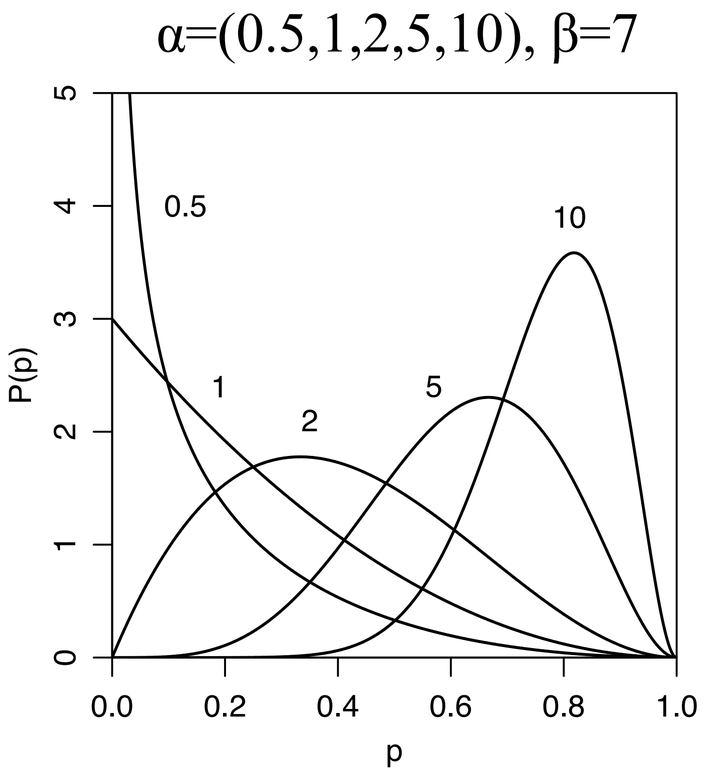
n=20, r=7, uniform prior
unnormalized posterior

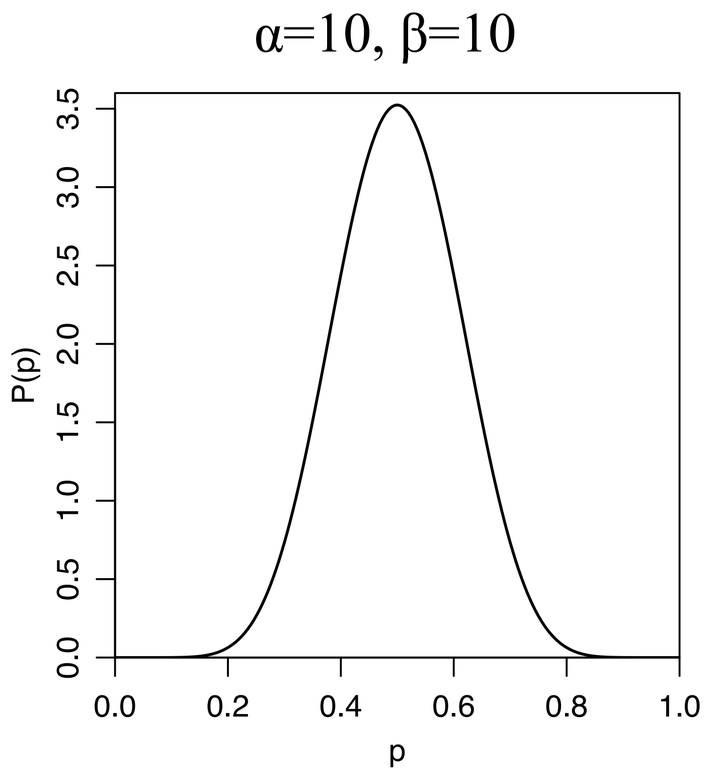




Beta distribution prior

$$P(p) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1}(1-p)^{\beta-1} \quad \text{where} \quad \alpha > 0, \beta > 0, 0 \leq p \leq 1$$

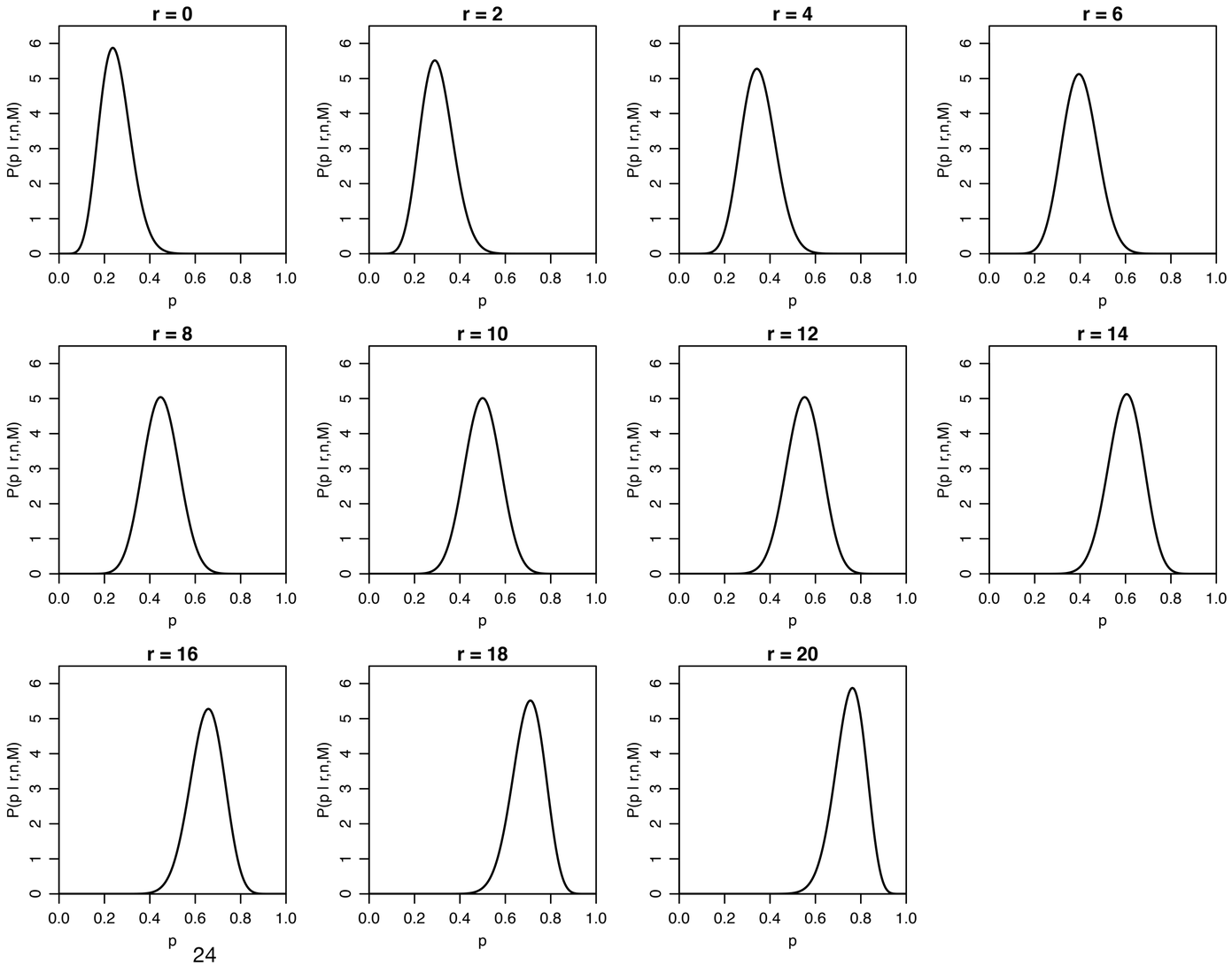




Posterior with beta prior

$$P^*(p \mid r, n) = p^{r+\alpha-1}(1-p)^{n-r+\beta-1}$$

α=10, β=10

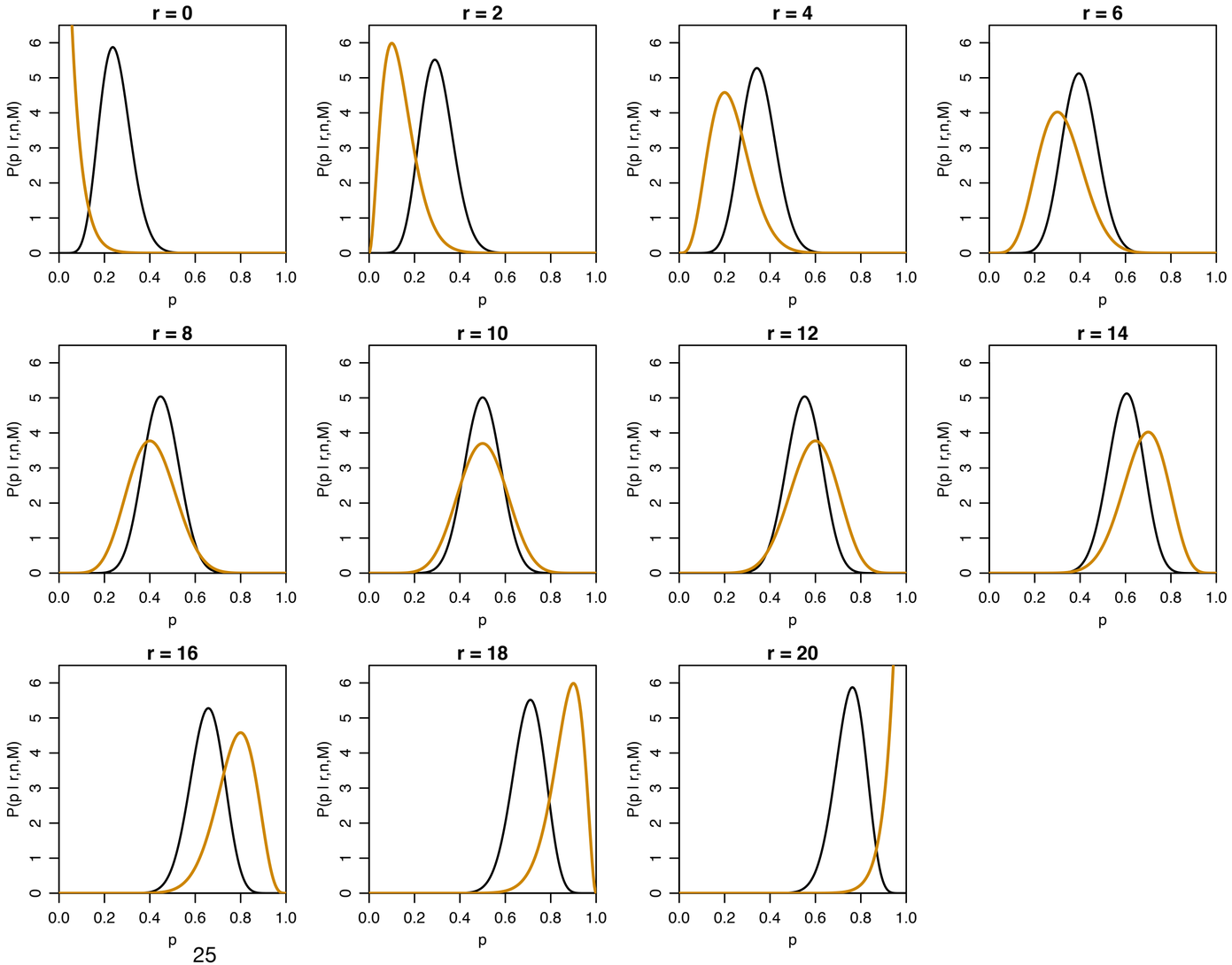


Posterior with beta and uniform priors

$$P^*(p \mid r, n) = p^{r+\alpha-1}(1-p)^{n-r+\beta-1}$$

α=10, β=10

α=1, β=1 (uniform)

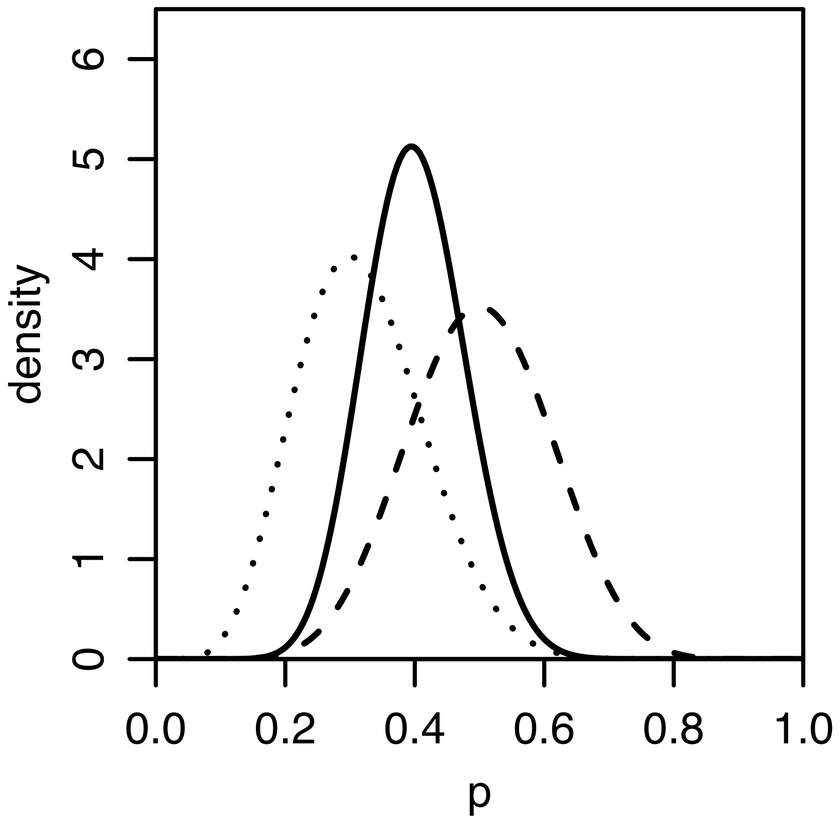


Posterior with beta prior

$$P^*(p \mid r, n) = p^{r+\alpha-1}(1-p)^{n-r+\beta-1}$$

$$P^*(p) = p^{\alpha-1}(1-p)^{\beta-1}$$

$$P^*(r \mid p, n) = p^{r}(1-p)^{n-r}$$



(beta) prior is *conjugate* for this (binomial) likelihood

Parametric models

$$\underbrace{P(\boldsymbol{\Theta} \mid \mathbf{D}, \mathrm{M})}_{\text{Posterior}} = \frac{\overbrace{P(\mathbf{D} \mid \boldsymbol{\Theta}, \mathrm{M})}^{\text{Likelihood}} \; \overbrace{P(\boldsymbol{\Theta} \mid \mathrm{M})}^{\text{Prior}}}{\underbrace{P(\mathbf{D} \mid \mathrm{M})}_{\text{Evidence}}}$$

Parametric models

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

$$\equiv \int_{\Omega_{\boldsymbol{\Theta}}} L(\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})d\boldsymbol{\Theta}$$

Model Comparison

$$P(\boldsymbol{\Theta} \mid \mathbf{D}, \mathrm{M}) = \frac{P(\mathbf{D} \mid \boldsymbol{\Theta}, \mathrm{M})\, P(\boldsymbol{\Theta} \mid \mathrm{M})}{P(\mathbf{D} \mid \mathrm{M})}$$

$$P(\mathbf{D} \mid \mathrm{M}) = \int P(\mathbf{D} \mid \boldsymbol{\Theta}, \mathrm{M})\, P(\boldsymbol{\Theta} \mid \mathrm{M})\, d\boldsymbol{\Theta}$$

“marginal likelihood” or “evidence”

Model Comparison

$$P(\boldsymbol{\Theta} \mid \mathbf{D}, \mathrm{M}) = \frac{P(\mathbf{D} \mid \boldsymbol{\Theta}, \mathrm{M})\, P(\boldsymbol{\Theta} \mid \mathrm{M})}{P(\mathbf{D} \mid \mathrm{M})}$$

$$P(\mathbf{D} \mid \mathrm{M}) = \int \underbrace{P(\mathbf{D} \mid \boldsymbol{\Theta}, \mathrm{M})}_{L(\boldsymbol{\Theta})}\, \underbrace{P(\boldsymbol{\Theta} \mid \mathrm{M})}_{\pi(\boldsymbol{\Theta})}\, d\boldsymbol{\Theta}$$

"marginal likelihood" or "evidence"

Model Comparison

$$P(\boldsymbol{\Theta} \mid \mathbf{D}, \mathrm{M}) = \frac{P(\mathbf{D} \mid \boldsymbol{\Theta}, \mathrm{M})\, P(\boldsymbol{\Theta} \mid \mathrm{M})}{P(\mathbf{D} \mid \mathrm{M})}$$

$$P(\mathbf{D} \mid \mathrm{M}) = \int P(\mathbf{D} \mid \boldsymbol{\Theta}, \mathrm{M})\, P(\boldsymbol{\Theta} \mid \mathrm{M})\, d\boldsymbol{\Theta}$$

"marginal likelihood" or "evidence"

$$BF_{12} = \frac{P(\mathbf{D} \mid \mathrm{M}_1)}{P(\mathbf{D} \mid \mathrm{M}_2)}$$

"Bayes factor"

$$R = \frac{P(\mathbf{D} \mid \mathrm{M}_1)\, P(\mathrm{M}_1)}{P(\mathbf{D} \mid \mathrm{M}_2)\, P(\mathrm{M}_2)}$$

"Posterior odds ratio"

Model Comparison

$$\text{AIC} = 2p - 2\log L(\hat{\boldsymbol{\Theta}})$$

$$\text{BIC} = p\log(n) - 2\log L(\hat{\boldsymbol{\Theta}})$$

p is the number of parameters in the model

n is the number of observations

$\hat{\boldsymbol{\Theta}}$ is the MLE of $\boldsymbol{\Theta}$

Model Comparison

$$\mathrm{AIC} = 2p - 2\log L(\hat{\boldsymbol{\Theta}})$$

$$\mathrm{BIC} = p\log(n) - 2\log L(\hat{\boldsymbol{\Theta}})$$

We can approximate the log marginal likelihood for model M as

$$P(\mathbf{D} \mid \mathrm{M}) \approx \log L(\hat{\boldsymbol{\Theta}}) - \frac{p}{2}\log n$$

Simulation-based inference

$$\underbrace{P(\boldsymbol{\Theta} \mid \mathbf{D}, \mathrm{M})}_{\text{Posterior}} = \frac{\overbrace{P(\mathbf{D} \mid \boldsymbol{\Theta}, \mathrm{M})}^{\text{Likelihood}} \; \overbrace{P(\boldsymbol{\Theta} \mid \mathrm{M})}^{\text{Prior}}}{\underbrace{P(\mathbf{D} \mid \mathrm{M})}_{\text{Evidence}}}$$

What is simulation-based inference?

$$\underbrace{P(\boldsymbol{\Theta} \mid \mathbf{D}, \mathrm{M})}_{\text{Posterior}} = \frac{\overbrace{P(\mathbf{D} \mid \boldsymbol{\Theta}, \mathrm{M})}^{\text{Likelihood}} \; \overbrace{P(\boldsymbol{\Theta} \mid \mathrm{M})}^{\text{Prior}}}{\underbrace{P(\mathbf{D} \mid \mathrm{M})}_{\text{Evidence}}}$$

Standard approach: χ^2 minimisation

Simulation-based (likelihood-free) inference

$$\underbrace{P(\boldsymbol{\Theta} \mid \mathbf{D}, \mathrm{M})}_{\text{Posterior}} = \frac{\overbrace{P(\mathbf{D} \mid \boldsymbol{\Theta}, \mathrm{M})}^{\text{Likelihood}} \; \overbrace{P(\boldsymbol{\Theta} \mid \mathrm{M})}^{\text{Prior}}}{\underbrace{P(\mathbf{D} \mid \mathrm{M})}_{\text{Evidence}}}$$

Approximate Bayesian Computation

Tavaré et al. 1997

Simulation-based (likelihood-free) inference

$$\underbrace{P(\boldsymbol{\Theta} \mid \mathbf{D}, \mathrm{M})}_{\text{Posterior}} = \frac{\overbrace{P(\mathbf{D} \mid \boldsymbol{\Theta}, \mathrm{M})}^{\text{Likelihood}} \; \overbrace{P(\boldsymbol{\Theta} \mid \mathrm{M})}^{\text{Prior}}}{\underbrace{P(\mathbf{D} \mid \mathrm{M})}_{\text{Evidence}}}$$

Approximate Bayesian Computation

$$\epsilon \to 0$$

computationally expensive

Tavaré et al. 1997

Simulation-based (likelihood-free) inference

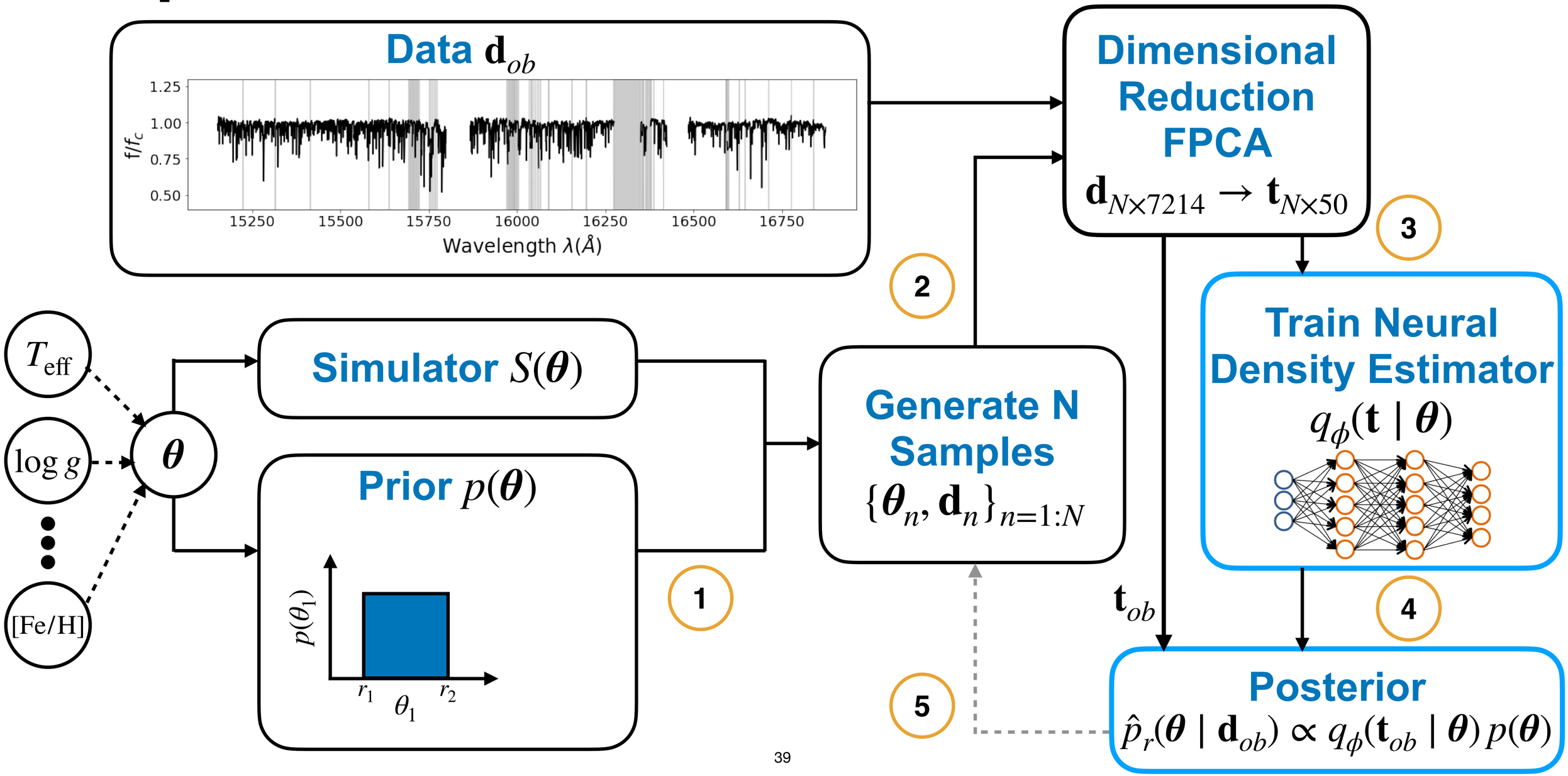
$$\underbrace{P(\boldsymbol{\Theta} \,|\, \mathbf{D}, \mathrm{M})}_{\text{Posterior}} = \frac{\overbrace{P(\mathbf{D} \mid \boldsymbol{\Theta}, \mathrm{M})}^{\text{Likelihood}} \; \overbrace{P(\boldsymbol{\Theta} \mid \mathrm{M})}^{\text{Prior}}}{\underbrace{P(\mathbf{D} \mid \mathrm{M})}_{\text{Evidence}}}$$

Neural Posterior/Likelihood/Ratio Estimation

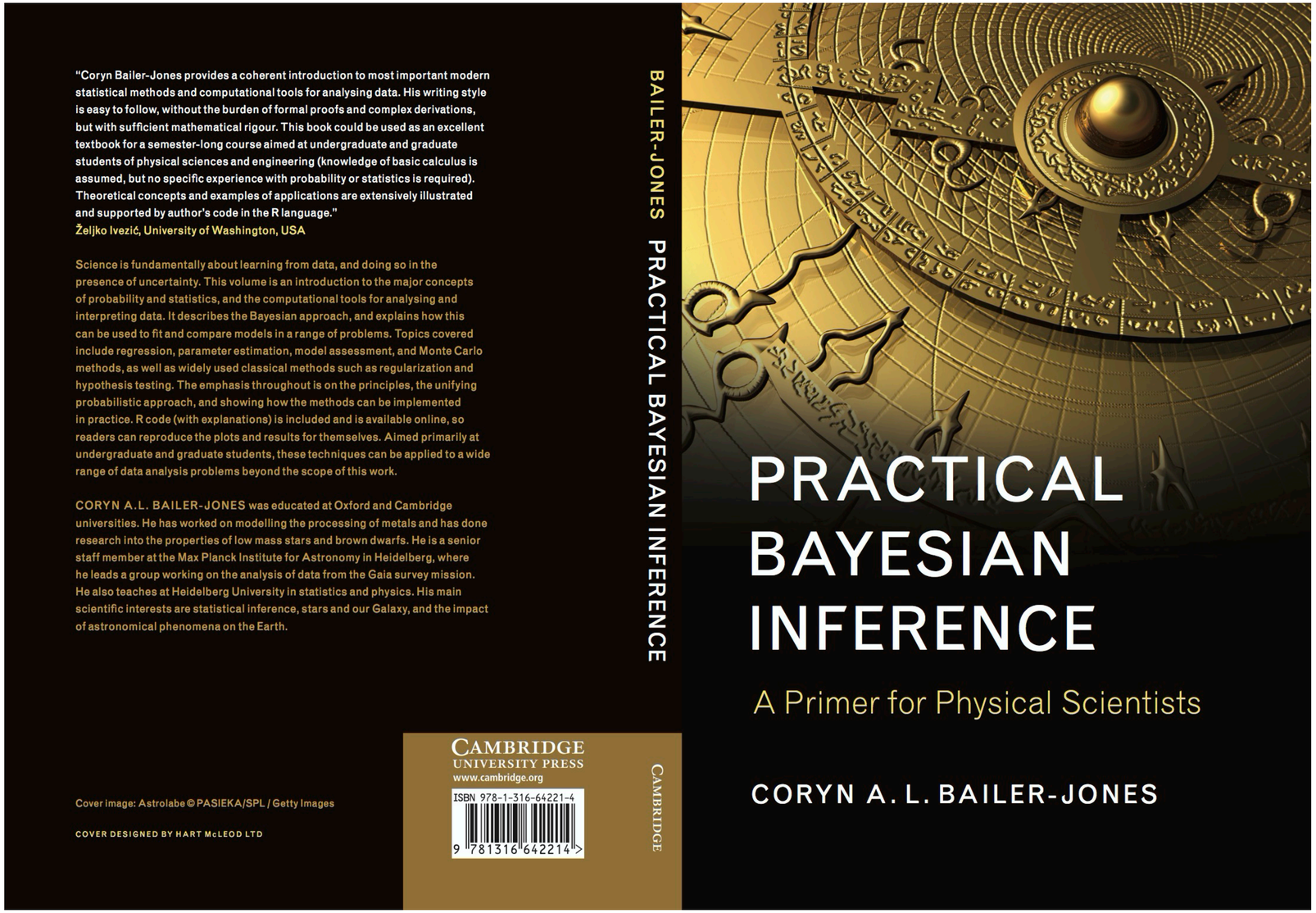
ϵ free?

e.g., Papamakarios et al. 2019

Sequential Neural Likelihood



Simulation-based (likelihood-free) inference

https://github.com/sbi-dev/sbi/

Source of a lot of covered material