Derived parameters with specified distributions

Maximum entropy prior choices

Will Handley
wh260@cam.ac.uk

Astrophysics Group
Cavendish Laboratory
University of Cambridge

16th May 2018

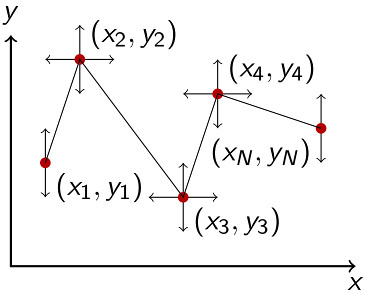
Bayesian inference

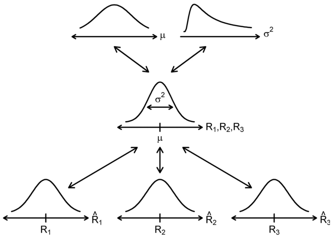
- Model parameters x describing data D:

$$P(x|D) = \frac{P(D|x)P(x)}{P(D)}$$

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

- Need Prior distribution $P(x)$
- Chosen to reflect initial knowledge, without data
- Harder to do with modern inference techniques:
  - Non-parametric (model-independent) reconstructions
  - Hierarchical models





Prior construction

The principle of maximum entropy

- We may wish to construct a prior "assuming the least information"
- One way to quantify this is using the Shannon entropy:

$$H(\Omega) = \sum_{E\in\Omega} P(E)\log\frac{1}{P(E)}$$

- Shannon information

$$\mathcal{I}(E) = \log\frac{1}{P(E)}, \qquad H = \langle \log \mathcal{I}(E)\rangle_{E\in\Omega}$$

- We construct priors by minimising H, subject to knowledge constraints

Maximum entropy prior examples

- Known mean μ and variance $\sigma \Rightarrow$ Gaussian:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

- Known mean x_0 and positive $x > 0 \Rightarrow$ Exponential:

$$P(x) = \frac{1}{x_0} \exp\left[-x/x_0\right]$$

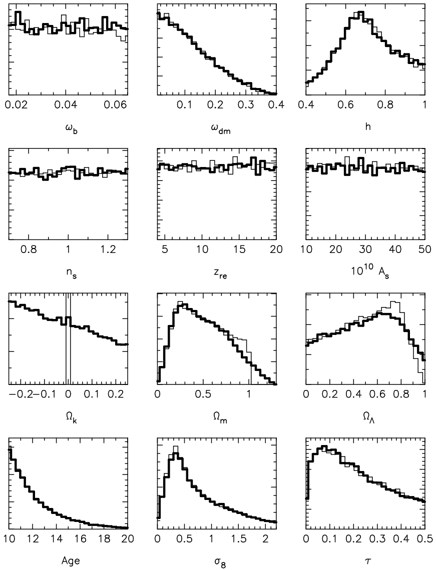
- Positive $x > 0 \Rightarrow$ Logarithmic (improper):

$$P(x) \propto 1/x$$

- Nothing $\Rightarrow$ Uniform (improper):

$$P(x) \propto 1$$

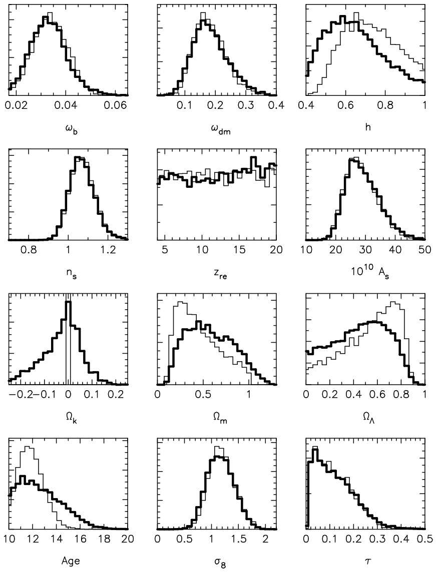
- Consider constraint on Hubble parameter h ($H_0 = 100h\frac{\mathrm{kms}^{-1}}{\mathrm{Mpc}}$)



The importance of plotting priors

VSA cosmological parameters (astro-ph:0212497)

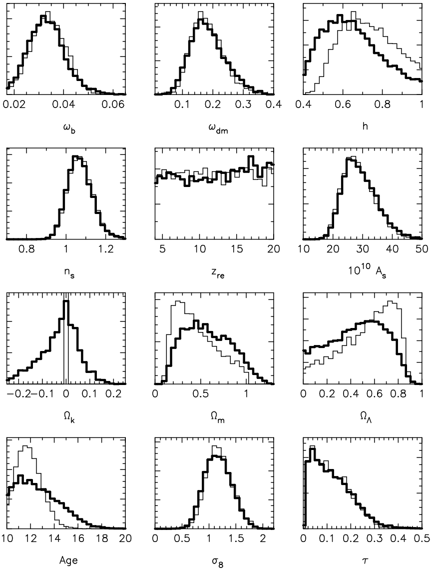
- Consider constraint on Hubble parameter h ($H_0 = 100h\frac{\mathrm{kms}^{-1}}{\mathrm{Mpc}}$)



The importance of plotting priors

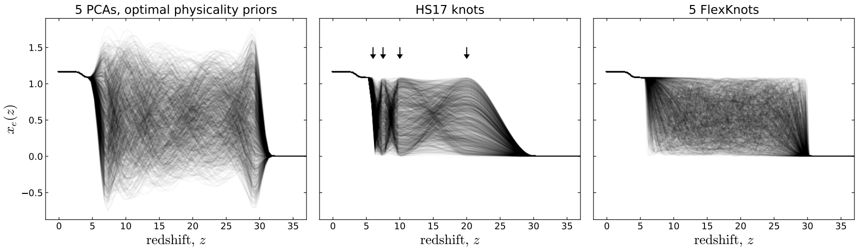
VSA cosmological parameters (astro-ph:0212497)

- Consider constraint on Hubble parameter h ($H_0 = 100h\frac{\mathrm{kms}^{-1}}{\mathrm{Mpc}}$)
- h-constraint *gets worse* with data
- Lesson: It is essential to plot priors and posteriors together.
- Particularly relevent for new data with weak constraints (e.g. EoR)



Non-parametric reconstructions

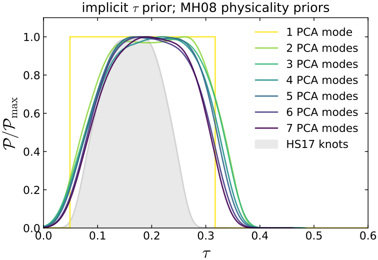
Example: cosmic reionisation history from CMB (Millea & Bouchet 1804.08476)

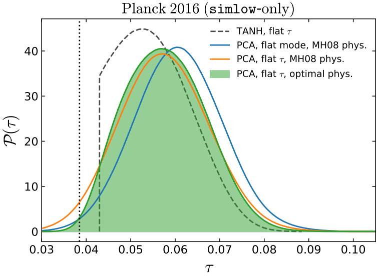


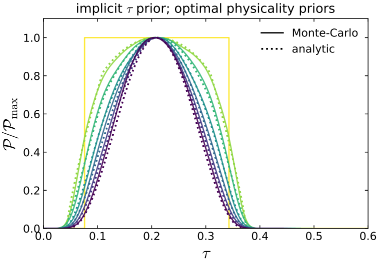
- Aim to reconstruct reionisation history $x_\mathrm{e}(z)$ from Planck data
- Model-independent/non-parametric
- Optical depth $\tau = \int \frac{n_\mathrm{H}(z)(1+z)^2}{H(z)} x_\mathrm{e}(z) dz$
- Reconstruction introduces non-trivial prior on derived parameter τ

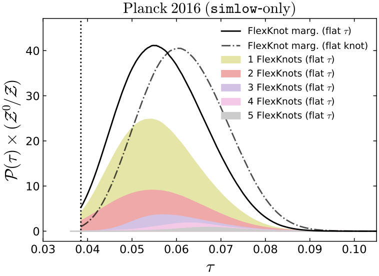
Tau prior and posterior

Example: cosmic reionisation history from CMB (Millea & Bouchet 1804.08476)



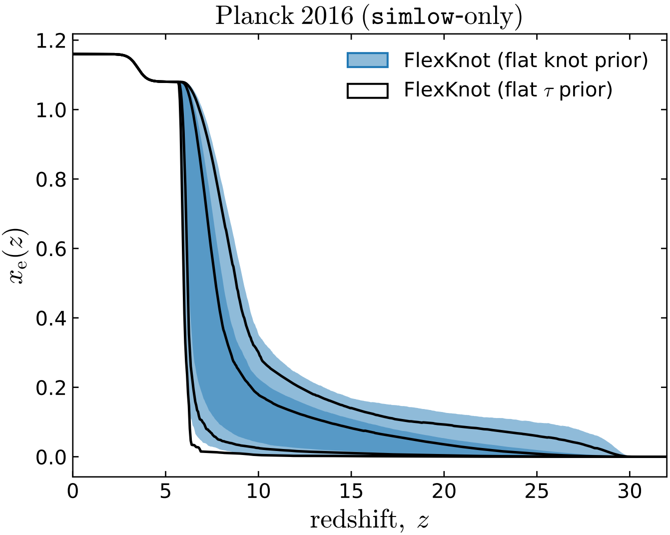






Reionisation posterior

Example: cosmic reionisation history from CMB (Millea & Bouchet 1804.08476)



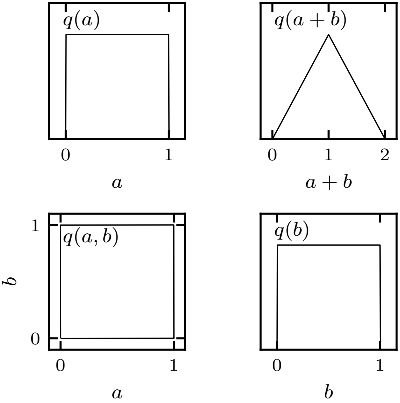
Derived parameter priors

Simplified example

- Uniform distribution $q(a, b)$
- $\Rightarrow$ triangular distribution on $a + b$.
- Remove this effect by dividing out this distribution:

$$p(a, b) = \frac{q(a, b)}{q(a + b)}$$

Uniform distribution on a and b

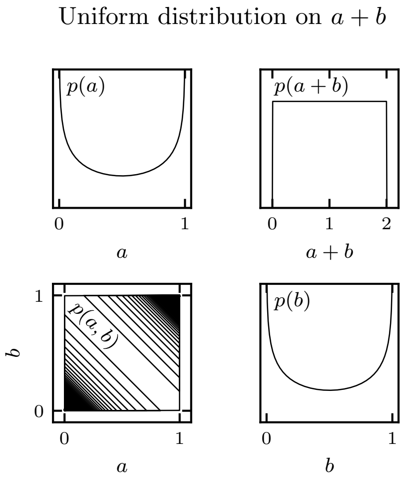


Derived parameter priors

Simplified example

- Uniform distribution $q(a,b)$
- $\Rightarrow$ triangular distribution on $a+b$.
- Remove this effect by dividing out this distribution:

$$p(a,b) = \frac{q(a,b)}{q(a+b)}$$



General result: This is maximum entropy

Handley & Millea 1804.08143

Theorem

If one has a distribution on parameters x with probability density function $q(x)$ along with a derived parameter f defined by a function $f = f(x)$, then the maximum entropy distribution $p(x)$ relative to $q(x)$ satisfying the constraint that f is distributed with probability density function to $r(f)$ is:

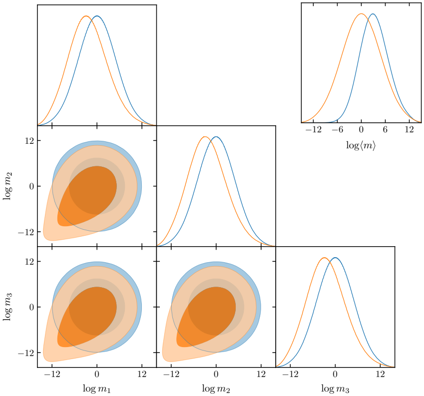
$$p(x) = \frac{q(x)r(f(x))}{P(f(x)|q)},$$

where $P(f|q)$ is the probability density for the distribution induced by q on $f = f(x)$.

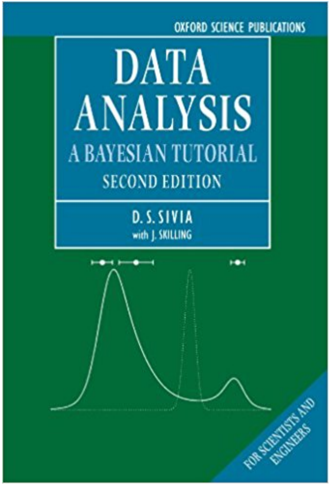
Derived parameter priors

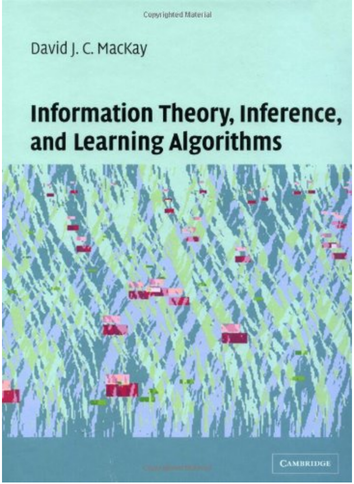
Neutrino example

- Initial spherical log-gaussian q
- $\Rightarrow$ non-trivial shifted distribution on mass sum $m_1 + m_2 + m_3$
- Apply maxent prior forcing this distribution back to center
- Creates heavy tail previously ruled out by q.



Further reading





- Data analysis: A Bayesian Tutorial (Sivia & Skilling)
- Information Theory, Inference and Learning Algorithms (Mackay)