UNIVERSITY OF CAMBRIDGE

Lecture 14
Applications of Bayesian Analysis

Lecturer: **Dr Dominic Anstey** (da401)

UNIVERSITY OF CAMBRIDGE

Overview

Conjugate Priors
- Use case in radiometer calibration

Likelihood Reweighting
- Use case in RFI removal

Analytic Marginalisation of Linear Parameters
- Use case in 21cm beam modelling

Marginal Statistics and Joint Analysis
- Use case in early universe constraints

UNIVERSITY OF CAMBRIDGE

Conjugate Priors

In cases where speed of computation is critical, it may be desirable for the full posterior to have a closed-form analytical expression, making sampling unnecessary

$$\mathcal{L}\Pi = \mathcal{P}\mathcal{Z}$$

In order to have an analytic posterior, the functional form of the likelihood and priors must be such that their product takes an analytical form

In general, these must be determined on a case-by-case basis, but there are many established cases

UNIVERSITY OF CAMBRIDGE

Conjugate Priors

Many typically used probability distributions fall into defined distribution families of a certain type:

- e.g. - normal
  - multivariate normal
  - uniform
  - gamma distribution

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

distinguished by parameter values

For certain likelihoods, there exists a family of probability distributions that, when used as a prior, result in a **posterior of the same family as the prior**, but with different parameter values. This is called the **conjugate prior** of that likelihood.

UNIVERSITY OF CAMBRIDGE

Conjugate Priors

The parameters that describe the details of the prior and posterior distributions are referred to as **hyperparameters**.

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

Using the conjugate prior of the chosen likelihood makes the posterior analytic, allowing it to be evaluated without sampling

UNIVERSITY OF CAMBRIDGE

Normal Distribution Example

Consider a likelihood that is normal, with a known variance

$$\mathcal{L}(\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{\mathcal{D}-\mu}{\sigma}\right)^2}$$

Assume also a normal distribution prior

$$\Pi = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2}$$

Prior hyperparameters: $\{\mu_0, \sigma_0\}$

If this is the conjugate prior, the posterior should also be normal

$$\mathcal{P} \propto e^{-\frac{1}{2}\left(\frac{\mu-\alpha}{\beta}\right)^2} \qquad \{\alpha, \beta\}$$

UNIVERSITY OF CAMBRIDGE

Normal Distribution Example

$$\mathcal{P} \propto e^{-\frac{1}{2}\left(\frac{\mu-\alpha}{\beta}\right)^2} \qquad \mathcal{L}\Pi = \mathcal{P}\mathcal{Z}$$

$$\mathcal{P} \propto e^{-\frac{1}{2}\left(\frac{\mathcal{D}-\mu}{\sigma}\right)^2} e^{-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2}$$

Expand

$$\mathcal{P} \propto e^{-\frac{1}{2}\left[\frac{\mathcal{D}^2}{\sigma^2}+\frac{\mu^2}{\sigma^2}-\frac{2\mathcal{D}\mu}{\sigma^2}+\frac{\mu^2}{\sigma_0^2}+\frac{\mu_0^2}{\sigma_0^2}-\frac{2\mu\mu_0}{\sigma_0^2}\right]}$$

Group terms

$$\mathcal{P} \propto e^{-\frac{1}{2}\left[\mu^2\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_0^2}\right)-2\mu\left(\frac{\mathcal{D}}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)+\left(\frac{\mathcal{D}^2}{\sigma^2}+\frac{\mu_0^2}{\sigma_0^2}\right)\right]}$$

$$\alpha = \frac{1}{\frac{1}{\sigma^2}+\frac{1}{\sigma_0^2}}\left(\frac{\mathcal{D}}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right) \qquad \beta^2 = \frac{1}{\frac{1}{\sigma^2}+\frac{1}{\sigma_0^2}}$$

UNIVERSITY OF CAMBRIDGE

Other Conjugate Pairs

Likelihood:	Conjugate prior:
- Normal with known mean →	- Inverse gamma distribution
- Normal (both mean and variance as parameters) →	- Normal-inverse gamma distribution
- Multivariate normal with known covariance →	- Multivariate normal
- Multivariate normal →	- Normal-inverse-Wishart distribution
- Uniform →	- Pareto distribution
- Exponential distribution →	- Gamma distribution
- etc.	- etc.

https://en.wikipedia.org/wiki/Conjugate_prior

UNIVERSITY OF CAMBRIDGE

Use Case – Radiometer Calibration

The output from a radiometer must be amplified , which also adds noise and reflections

Have to calibrate all these effects to recover the original temperature from the measured power

$$P_{\mathrm{cal}} = g_{\mathrm{sys}} \left[T_{\mathrm{cal}} \left(1 - |\Gamma_{\mathrm{cal}}|^2\right) \left| \frac{\sqrt{1-|\Gamma_{\mathrm{rec}}|^2}}{1-\Gamma_{\mathrm{cal}}\Gamma_{\mathrm{rec}}} \right|^2 + T_{\mathrm{unc}} |\Gamma_{\mathrm{cal}}|^2 \left| \frac{\sqrt{1-|\Gamma_{\mathrm{rec}}|^2}}{1-\Gamma_{\mathrm{cal}}\Gamma_{\mathrm{rec}}} \right|^2 + T_{\mathrm{cos}} \,\mathrm{Re} \left(\Gamma_{\mathrm{cal}} \frac{\sqrt{1-|\Gamma_{\mathrm{rec}}|^2}}{1-\Gamma_{\mathrm{cal}}\Gamma_{\mathrm{rec}}} \right) + T_{\mathrm{sin}} \,\mathrm{Im} \left(\Gamma_{\mathrm{cal}} \frac{\sqrt{1-|\Gamma_{\mathrm{rec}}|^2}}{1-\Gamma_{\mathrm{cal}}\Gamma_{\mathrm{rec}}} \right) + T_0 \right]$$

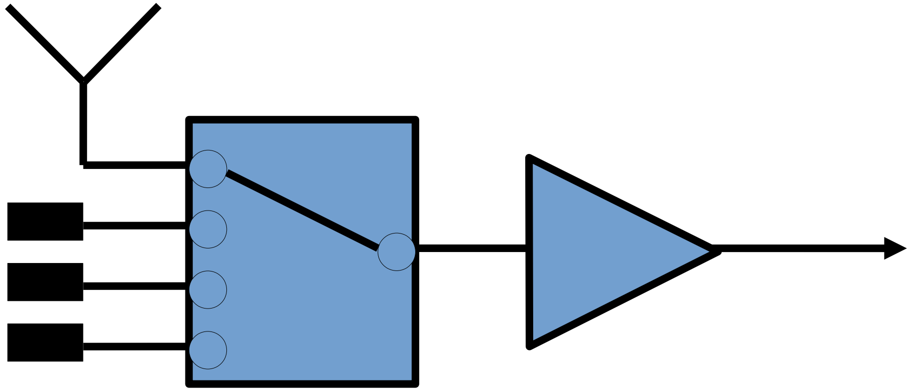
Need to solve for
many unknowns

From Roque et al. (2021)

UNIVERSITY OF CAMBRIDGE

Dicke Switches

From Dicke (1946)



$$\{X_{\mathrm{unc}}, X_{\cos}, X_{\sin}, X_{\mathrm{NS}}, X_{\mathrm{L}}\}$$

Functions of the reflection coefficients

$$T_{\mathrm{cal}} = X_{\mathrm{unc}}T_{\mathrm{unc}} + X_{\cos}T_{\cos} + X_{\sin}T_{\sin} + X_{\mathrm{NS}}T_{\mathrm{NS}} + X_{\mathrm{L}}T_{\mathrm{L}}$$

$$\theta = \{T_{\mathrm{unc}}, T_{\cos}, T_{\sin}, T_{\mathrm{NS}}, T_{\mathrm{L}}\}$$

$$\log \mathcal{L} = -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\left(\frac{T_{\mathrm{cal}} - \underline{X}\theta}{\sigma}\right)^2$$

For a known calibrator, can fit for theta

From Roque et al. (2021)

UNIVERSITY OF CAMBRIDGE

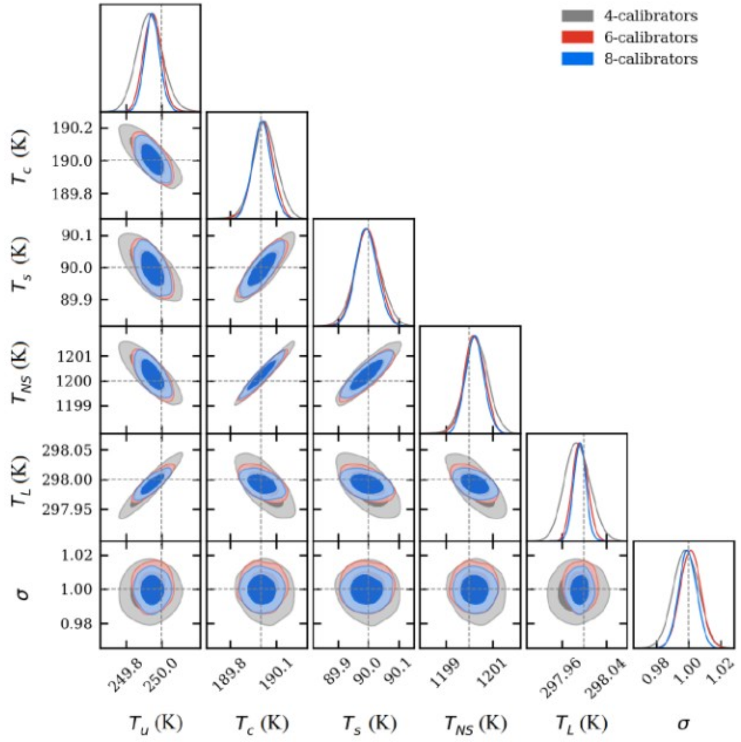
Loop Conjugate Priors

In many cases, radiometers have limits on storage and data transmission rates (e.g. due to remote locations)
Need the calibration to be performed on-site, at a rate matching the data collection

Normal likelihood with unknown mean and variance → Normal-inverse gamma distribution

$$\begin{tikzcd}
\Pi_0\left(\theta\right) \arrow[d] & & \\
\mathcal{L}\left(\theta\right)\Pi_i\left(\theta\right) \arrow[r, bend left] & \mathcal{P}_i\left(\theta\right) \arrow[l, bend left] \arrow[r] & \mathcal{P}_{\text{final}}\left(\theta\right)
\end{tikzcd}$$

Loop over all calibrators



From Roque et al. (2021)

UNIVERSITY OF CAMBRIDGE

Likelihood Reweighting

The speed of the likelihood evaluation is the limiting factor in computation time for most Bayesian algorithms

If the required likelihood is slow to evaluate, it can impede or prevent the analysis

The bulk of likelihood calls are used for exploring the parameter volume in order to isolate the region of high posterior probability

Consider the possibility of using a faster likelihood to locate the posterior peak first

UNIVERSITY OF CAMBRIDGE

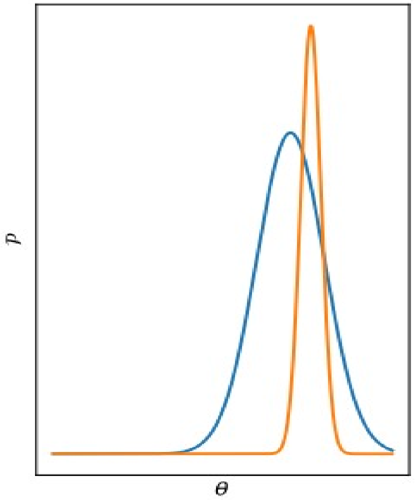
Slow and Fast Likelihoods

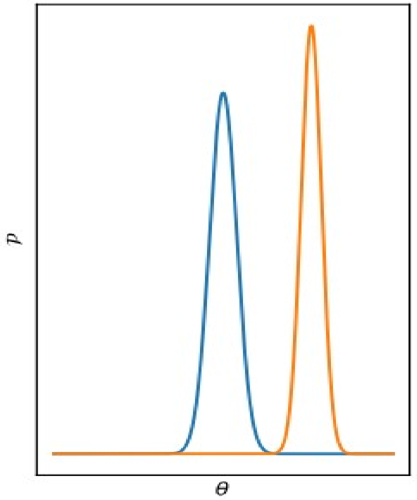
Assume 2 likelihoods, one of which is much faster to evaluate, can be defined, subject to the conditions:

- Both likelihoods take the same parameters with the same priors

- Both likelihoods have a posterior peak in the same location of parameter space

✓





$$\log \mathcal{L}_{\mathrm{S}}\left(\theta\right) = \sum_{ij} -\frac{1}{2}\log\left(2\pi\sigma_{\mathrm{n}}^{2}\right) - \frac{1}{2}\left(\frac{\mathcal{D}_{ij} - \mathcal{M}_{ij}\left(\theta\right)}{\sigma_{\mathrm{n}}}\right)^{2}$$

$$\log \mathcal{L}_{\mathrm{F}}\left(\theta\right) = \sum_{i} -\frac{1}{2}\log\left(2\pi\sigma_{\mathrm{n}}^{2}\right) - \frac{1}{2}\left(\frac{\widetilde{D}_{i} - \widetilde{M}_{i}\left(\theta\right)}{\sigma_{\mathrm{n}}}\right)^{2}$$

Reweighting Relation

$$\mathcal{P}_{\mathrm{S}}\left(\theta|\mathcal{D},\mathcal{M}_{\mathrm{S}}\right)=\frac{\mathcal{L}_{\mathrm{S}}\left(\mathcal{D}|\theta,\mathcal{M}_{\mathrm{S}}\right)\pi\left(\theta\right)}{\mathcal{Z}_{\mathrm{S}}}$$

$$\mathcal{P}_{\mathrm{F}}\left(\theta|\mathcal{D},\mathcal{M}_{\mathrm{F}}\right)=\frac{\mathcal{L}_{\mathrm{F}}\left(\mathcal{D}|\theta,\mathcal{M}_{\mathrm{F}}\right)\pi\left(\theta\right)}{\mathcal{Z}_{\mathrm{F}}}$$

$$\pi\left(\theta\right)=\frac{\mathcal{Z}_{\mathrm{F}}\mathcal{P}_{\mathrm{F}}\left(\theta|\mathcal{D},\mathcal{M}_{\mathrm{F}}\right)}{\mathcal{L}_{\mathrm{F}}\left(\mathcal{D}|\theta,\mathcal{M}_{\mathrm{F}}\right)}$$

$$\mathcal{P}_{\mathrm{S}}\left(\theta|\mathcal{D},\mathcal{M}_{\mathrm{S}}\right)=\mathcal{P}_{\mathrm{F}}\left(\theta|\mathcal{D},\mathcal{M}_{\mathrm{F}}\right)\frac{\mathcal{L}_{\mathrm{S}}\left(\mathcal{D}|\theta,\mathcal{M}_{\mathrm{S}}\right)}{\mathcal{L}_{\mathrm{F}}\left(\mathcal{D}|\theta,\mathcal{M}_{\mathrm{F}}\right)}\frac{\mathcal{Z}_{\mathrm{F}}}{\mathcal{Z}_{\mathrm{S}}}$$

UNIVERSITY OF CAMBRIDGE

Reweighting Relation

Process:

- Define a comparable fast likelihood with the same parameters and similar posterior to the slow

- Perform a model fit to acquire a representative set of samples from the posterior

- For each posterior sample, evaluate the fast and slow likelihoods

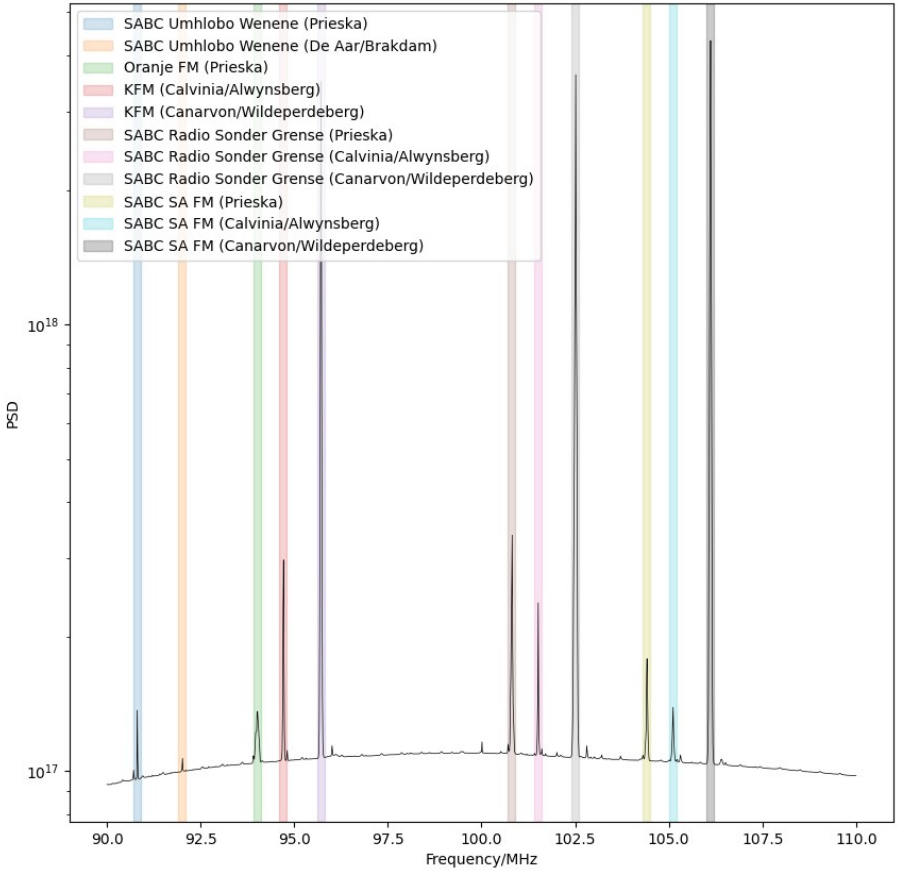
- Multiply the posterior sample weights by the reweight factor to convert them to samples of the slow posterior

- Slow likelihood only need to be evaluated for the few posterior samples

$$w\left(\theta\right) = \frac{\mathcal{L}_{\mathrm{S}}\left(\mathcal{D}|\theta, \mathcal{M}_{\mathrm{S}}\right)}{\mathcal{L}_{\mathrm{F}}\left(\mathcal{D}|\theta, \mathcal{M}_{\mathrm{F}}\right)}$$

Use Case – RFI Excision

Radio Frequency Interference: Contamination of data by radio emissions from local sources

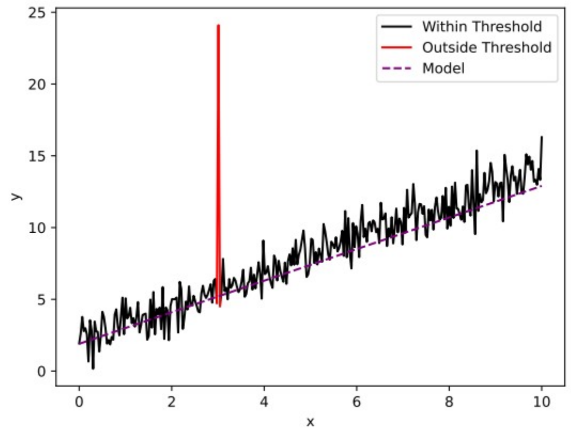


Needs flagging and removing from data

Removing entire channels of transient RFI results in the loss of useful information

UNIVERSITY OF CAMBRIDGE

Bayesian Anomaly Flagging



$$\log \mathcal{L}_{ij} = -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\left(\frac{\mathcal{D}_{ij} - \mathcal{M}_{ij}(\boldsymbol{\theta})}{\sigma}\right)^2$$

$$\log \mathcal{L} = \sum_{ij} \begin{cases} \log \mathcal{L}_{ij} + \log(1-p) & \text{if } \log \mathcal{L}_{ij} + \log(1-p) > \log p - \log \Delta \\ \log(p) - \log \Delta & \text{otherwise} \end{cases}$$

Probability a point is contaminated (p)

Approximate scale of contamination (Δ)

From Leeney et al. (2023)

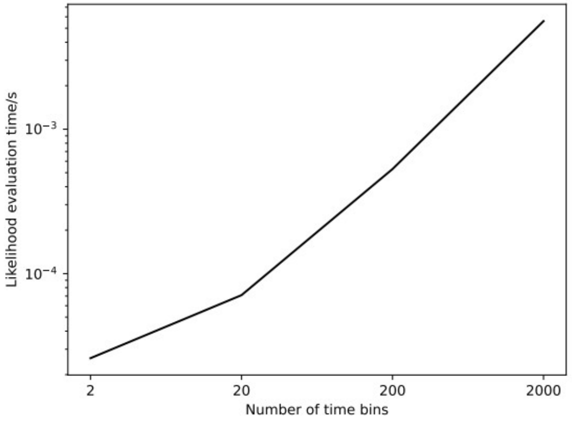
UNIVERSITY OF CAMBRIDGE

Use Case – RFI Excision

$$\log \mathcal{L} = \sum_{ij} \begin{cases} \log \mathcal{L}_{ij} + \log(1+p) & \text{if} \quad \log \mathcal{L}_{ij} + \log(1-p) > \log p - \log \Delta \\ \log(p) - \log \Delta & \text{otherwise} \end{cases}$$

$$\log \mathcal{L}_{ij} = -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\left(\frac{\mathcal{D}_{ij} - \mathcal{M}_{ij}(\boldsymbol{\theta})}{\sigma}\right)^2$$

Full likelihood – Slow to evaluate



$$\log \mathcal{L} = \sum_{i} \begin{cases} \log \mathcal{L}_{i} + \log(1+p) & \text{if} \quad \log \mathcal{L}_{i} + \log(1-p) > \log p - \log \Delta \\ \log(p) - \log \Delta & \text{otherwise} \end{cases}$$

$$\log \mathcal{L}_{i} = -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\left(\frac{\frac{1}{N_t}\sum_j \mathcal{D}_{ij} - \frac{1}{N_t}\sum_j \mathcal{M}_{ij}(\boldsymbol{\theta})}{\sigma}\right)^2$$

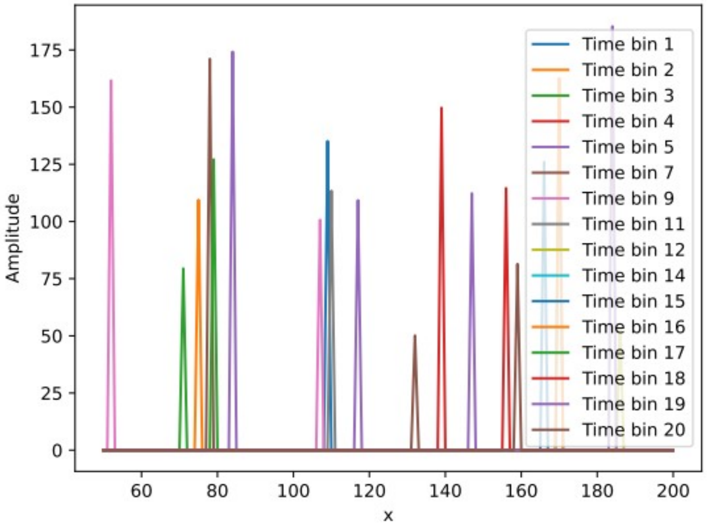
Averaged likelihood – Quick to evaluate

UNIVERSITY OF CAMBRIDGE

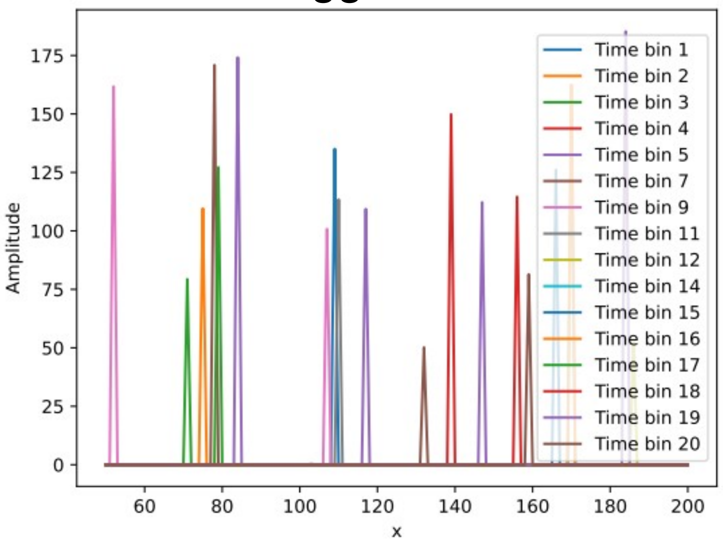
Use Case – RFI Excision

Test on simulated 21cm data with known RFI added

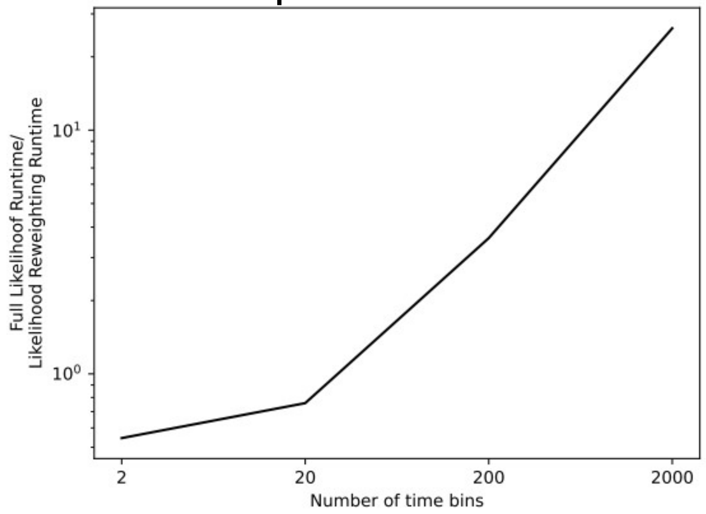
True



Flagged



Speed Increase



UNIVERSITY OF CAMBRIDGE

Marginalisation

Consider the circumstance where the model has too many nuisance parameters to viable fit in a reasonable time frame

Nuisance parameters can be marginalised

$$\mathrm{P}\left(\theta_{\mathcal{M}}|\mathcal{D},\mathcal{M}\right) = \mathrm{P}\left(\theta_{\mathcal{S}},\theta_{\mathcal{N}}|\mathcal{D},\mathcal{M}\right)$$

$$\mathrm{P}\left(A\right) = \sum_{i} \mathrm{P}\left(A, B_i\right)$$

$$\mathrm{P}\left(\theta_{\mathcal{S}}|\mathcal{D},\mathcal{M}\right) = \int \mathrm{P}\left(\theta_{\mathcal{S}},\theta_{\mathcal{N}}|\mathcal{D},\mathcal{M}\right) \mathrm{d}\theta_{\mathcal{N}}$$

UNIVERSITY OF CAMBRIDGE

Marginal Likelihoods

Marginalisation can also be performed at the likelihood level

$$\mathcal{P}\left(\theta_{\mathcal{S}}, \theta_{\mathcal{N}}\right) = \frac{\mathcal{L}\left(\theta_{\mathcal{S}}, \theta_{\mathcal{N}}\right) \Pi\left(\theta_{\mathcal{S}}\right) \Pi\left(\theta_{\mathcal{N}}\right)}{\mathcal{Z}}$$

$$\mathcal{P}_{\mathrm{eff}}\left(\theta_{\mathcal{S}}\right) = \int \mathcal{P}\left(\theta_{\mathcal{S}}, \theta_{\mathcal{N}}\right) \mathrm{d}\theta_{\mathcal{N}}$$

$$\mathcal{P}_{\mathrm{eff}}\left(\theta_{\mathcal{S}}\right) = \frac{\Pi\left(\theta_{\mathcal{S}}\right) \int \mathcal{L}\left(\theta_{\mathcal{S}}, \theta_{\mathcal{N}}\right) \Pi\left(\theta_{\mathcal{N}}\right) \mathrm{d}\theta_{\mathcal{N}}}{\mathcal{Z}} = \frac{\Pi\left(\theta_{\mathcal{S}}\right) \mathcal{L}_{\mathrm{eff}}\left(\theta_{\mathcal{S}}\right)}{\mathcal{Z}}$$

$$\mathcal{L}_{\mathrm{eff}}\left(\theta_{\mathcal{S}}\right) = \int \mathcal{L}\left(\theta_{\mathcal{S}}, \theta_{\mathcal{N}}\right) \Pi\left(\theta_{\mathcal{N}}\right) \mathrm{d}\theta_{\mathcal{N}}$$

Performing an integral over the parameter space in the likelihood is normally too slow

UNIVERSITY OF CAMBRIDGE

Analytic Marginalisation

Consider the case where the model is linear in the nuisance parameters

$$\mathcal{M}\left(\theta_{\mathrm{S}}, \theta_{\mathcal{N}}\right) = \underline{\underline{A}}\left(\theta_{\mathcal{S}}\right)\underline{\theta_{\mathcal{N}}} = \underline{\underline{A}}\,\underline{\theta}$$

Assume a Gaussian likelihood

$$\log \mathcal{L}\left(\theta_{\mathcal{S}}, \theta_{\mathcal{N}}\right) = -\frac{1}{2}\log\left|\left(2\pi\right)^{n}\underline{\underline{C}}\right| - \frac{1}{2}\left(\mathcal{D} - \mathcal{M}\left(\theta_{\mathcal{S}}, \theta_{\mathcal{N}}\right)\right)^{\mathrm{T}}\underline{\underline{C}}^{-1}\left(\mathcal{D} - \mathcal{M}\left(\theta_{\mathcal{S}}, \theta_{\mathcal{N}}\right)\right)$$

$$\log \mathcal{L} = -\frac{1}{2}\log\left|\left(2\pi\right)^{n}\underline{\underline{C}}\right| - \frac{1}{2}\left(\underline{\mathcal{D}} - \underline{\underline{A}}\,\underline{\theta}\right)^{\mathrm{T}}\underline{\underline{C}}^{-1}\left(\underline{\mathcal{D}} - \underline{\underline{A}}\,\underline{\theta}\right)$$

UNIVERSITY OF CAMBRIDGE

Uniform Priors

$$\mathcal{L}_{\text{eff}}\left(\theta_{\mathcal{S}}\right) = \int \mathcal{L}\left(\theta_{\mathcal{S}}, \theta_{\mathcal{N}}\right) \Pi\left(\theta_{\mathcal{N}}\right) \mathrm{d}\theta_{\mathcal{N}}$$

Assume for now the priors are uniform

$$\log \mathcal{L} = -\frac{1}{2}\log\left|\left(2\pi\right)^n \underline{\underline{C}}\right| - \frac{1}{2}\left(\underline{\mathcal{D}} - \underline{\underline{A}}\,\underline{\theta}\right)^{\mathrm{T}} \underline{\underline{C}}^{-1}\left(\underline{\mathcal{D}} - \underline{\underline{A}}\,\underline{\theta}\right)$$

Expand out:

$$\log \mathcal{L} = -\frac{1}{2}\log\left|\left(2\pi\right)^n \underline{\underline{C}}\right| - \frac{1}{2}\underline{\mathcal{D}}^{\mathrm{T}}\underline{\underline{C}}^{-1}\underline{\mathcal{D}} + \frac{1}{2}\underline{\mathcal{D}}^{\mathrm{T}}\underline{\underline{C}}^{-1}\underline{\underline{A}}\,\underline{\theta} + \frac{1}{2}\underline{\theta}^{\mathrm{T}}\underline{\underline{A}}^{\mathrm{T}}\underline{\underline{C}}^{-1}\underline{\mathcal{D}} - \frac{1}{2}\underline{\theta}^{\mathrm{T}}\underline{\underline{A}}^{\mathrm{T}}\underline{\underline{C}}^{-1}\underline{\underline{A}}\,\underline{\theta}$$

Make the substitutions:

$$\underline{\mu} = \underline{\underline{\Sigma}}\,\underline{\underline{A}}^T\underline{\underline{C}}^{-1}\underline{\mathcal{D}} \qquad \underline{\underline{\Sigma}}^{-1} = \underline{\underline{A}}^{\mathrm{T}}\underline{\underline{C}}^{-1}\underline{\underline{A}}$$

$$\log \mathcal{L} = -\frac{1}{2}\log\left|\left(2\pi\right)^n \underline{\underline{C}}\right| - \frac{1}{2}\underline{\mathcal{D}}^{\mathrm{T}}\underline{\underline{C}}^{-1}\underline{\mathcal{D}} + \frac{1}{2}\underline{\mu}^{\mathrm{T}}\underline{\underline{\Sigma}}^{-1}\underline{\theta} + \frac{1}{2}\underline{\theta}^{\mathrm{T}}\underline{\underline{\Sigma}}^{-1}\underline{\mu} - \frac{1}{2}\underline{\theta}^{\mathrm{T}}\underline{\underline{\Sigma}}^{-1}\underline{\theta} + \frac{1}{2}\underline{\mu}^{\mathrm{T}}\underline{\underline{\Sigma}}^{-1}\underline{\mu} - \frac{1}{2}\underline{\mu}^{\mathrm{T}}\underline{\underline{\Sigma}}^{-1}\underline{\mu}$$

UNIVERSITY OF CAMBRIDGE

Uniform Priors

$$\log \mathcal{L} = -\frac{1}{2}\log|(2\pi)^n \underline{\underline{C}}| - \frac{1}{2}\underline{\mathcal{D}}^{\mathrm{T}}\underline{\underline{C}}^{-1}\underline{\mathcal{D}} + \frac{1}{2}\underline{\mu}^{\mathrm{T}}\underline{\underline{\Sigma}}^{-1}\underline{\mu} - \frac{1}{2}\left(\underline{\mu} - \underline{\theta}\right)^{\mathrm{T}}\underline{\underline{\Sigma}}^{-1}\left(\underline{\mu} - \underline{\theta}\right)$$

$$\mathcal{L}_{\mathrm{eff}}\left(\theta_{\mathcal{S}}\right) = \int \mathcal{L}\left(\theta_{\mathcal{S}}, \theta_{\mathcal{N}}\right)\Pi\left(\theta_{\mathcal{N}}\right)\mathrm{d}\theta_{\mathcal{N}}$$

All dependence on the linear nuisance parameters has been collected in this Gaussian term

$$\log \mathcal{L}_{\mathrm{eff}} = -\frac{1}{2}\log|(2\pi)^n \underline{\underline{C}}| - \frac{1}{2}\underline{\mathcal{D}}^{\mathrm{T}}\underline{\underline{C}}^{-1}\underline{\mathcal{D}} + \frac{1}{2}\underline{\mu}^{\mathrm{T}}\underline{\underline{\Sigma}}^{-1}\underline{\mu} + \frac{1}{2}\log|(2\pi)^k \underline{\underline{\Sigma}}|$$

If your likelihood is Gaussian and nuisance parameters are linear with uniform priors, they can be marginalised out analytically in the likelihood and do not need to be fit for

UNIVERSITY OF CAMBRIDGE

Non-Uniform Priors

Analytic marginalisation can also be performed for Gaussian priors

$$\Pi\left(\underline{\theta}\right) = \frac{1}{\left|2\pi\Lambda\right|^{\frac{1}{2}}} e^{-\frac{1}{2}\left(\underline{\theta}-\underline{\xi}\right)^{\mathrm{T}}\underline{\underline{\Lambda}}^{-1}\left(\underline{\theta}-\underline{\xi}\right)}$$

With the further substitutions:

$$\underline{\underline{\Omega}}^{-1} = \underline{\underline{\Sigma}}^{-1} + \underline{\underline{\Lambda}}^{-1}$$

$$\underline{\underline{\Omega}}^{-1}\underline{\omega} = \underline{\underline{\Sigma}}^{-1}\underline{\mu} + \underline{\underline{\Lambda}}^{-1}\underline{\xi}$$

UNIVERSITY OF CAMBRIDGE

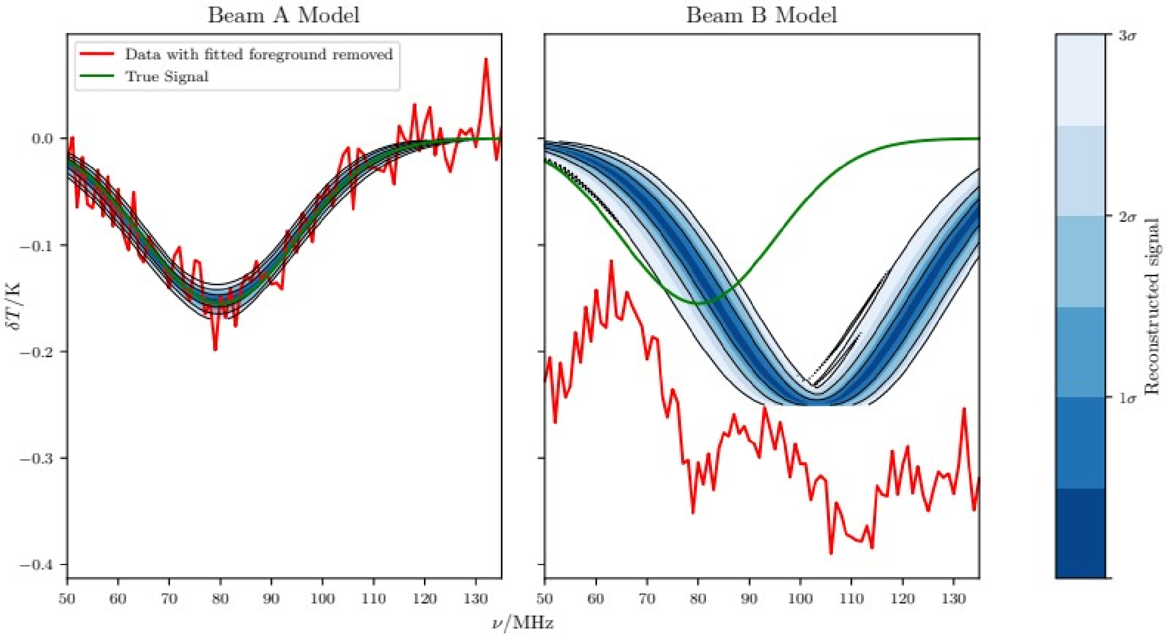
Use Case – 21cm cosmology

Recall from the previous lecture 21cm example, the model of the antenna's directivity was assumed to be exactly known

$$T_{\mathrm{F}}\left(\nu, \theta_{\mathrm{F}}\right) = \frac{1}{4\pi} \int D\left(\Omega, \nu\right) \left[T_{\mathrm{base}}\left(\Omega\right) - T_{\mathrm{CMB}}\right] \left(\frac{\nu}{\nu_{\mathrm{base}}}\right)^{-\beta} \mathrm{d}\Omega + T_{\mathrm{CMB}}$$

Foreground known exactly

Beam error of <0.5%



UNIVERSITY OF CAMBRIDGE

Fitting a beam

Analytical models

$$\mathcal{M} = a\sin\left(\frac{x}{b} - c\right)$$

None exist

Linear Models

$$\mathcal{M} = \sum_i a_i X_i$$

Forward Models

$$\mathcal{M} = F\left(\theta\right)$$

Requires many large-scale EM simulations. Very slow

$$D\left(\Omega,\nu\right) = \sum_k^{N_{\mathrm{basis}}} \Gamma_k\left(\nu,\theta\right) Y_k\left(\Omega\right)$$

Parametrised coefficient functions

Basis functions

$$T_{\mathrm{F}}\left(\nu,\theta_{\mathrm{F}}\right) = \frac{1}{4\pi}\int \sum_k^{N_{\mathrm{basis}}} \Gamma_k\left(\nu,\theta_{\mathrm{beam}}\right) Y_k\left(\Omega\right)\left[T_{\mathrm{base}}\left(\Omega\right) - T_{\mathrm{CMB}}\right]\left(\frac{\nu}{\nu_{\mathrm{base}}}\right)^{-\beta} \mathrm{d}\Omega + T_{\mathrm{CMB}}$$

Fitting a beam

- Spherical Harmonics:
  > 1000s of basis functions

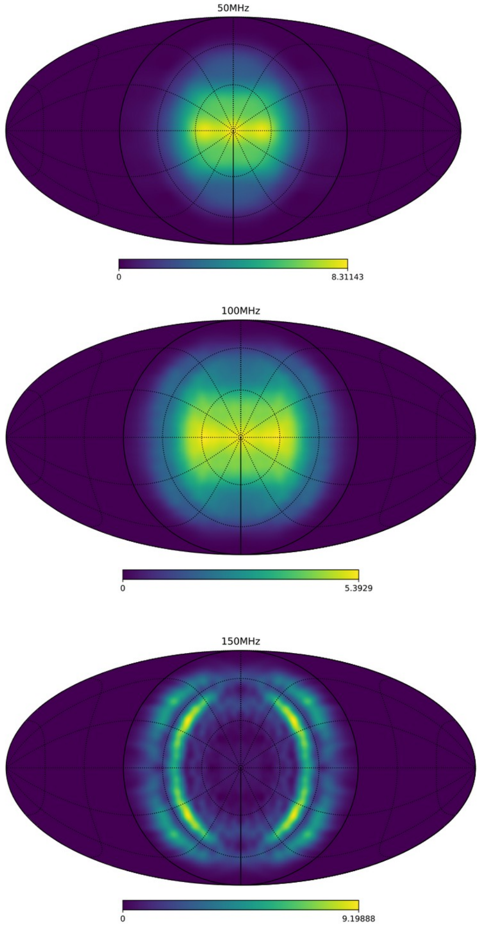
- More sophisticated basis functions
  ~ 10-30

Each basis functions requires a parametrised coefficient function.
Order 5-10 parameters each.

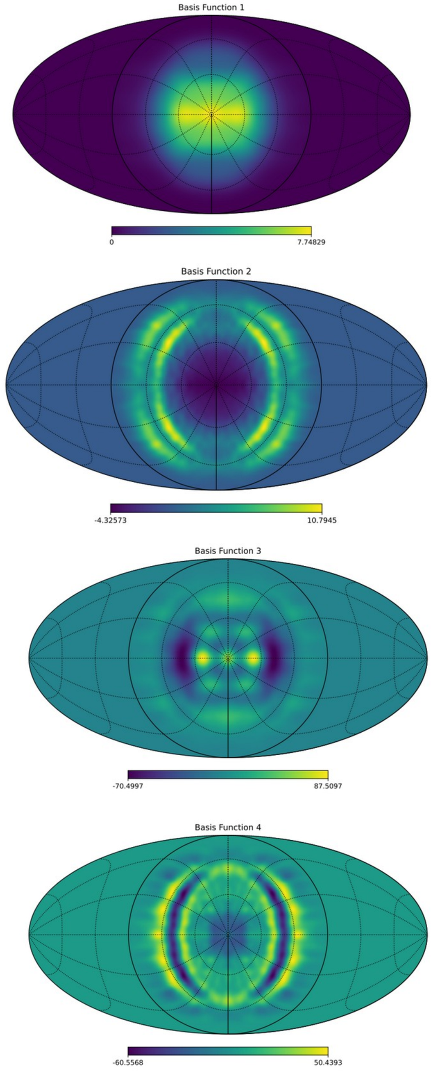
Current best case requires of order ~50-100 parameters.

Parameters are linear by construction

Beam



Basis functions



UNIVERSITY OF CAMBRIDGE

Fitting a beam

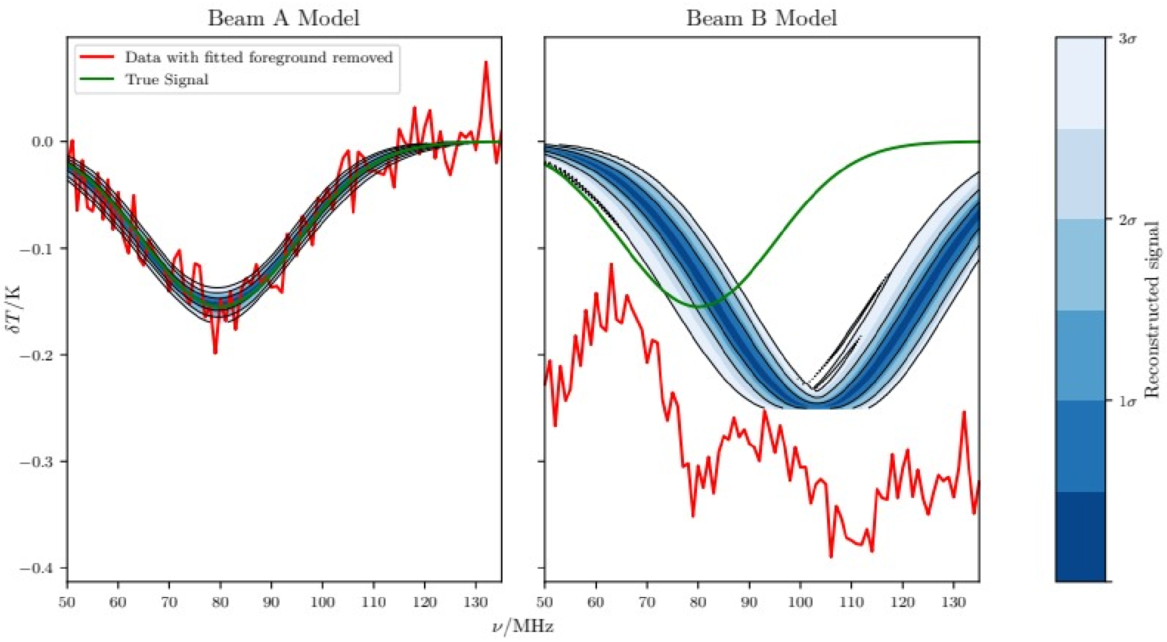
$$\mathcal{M}\left(\theta_{\mathrm{S}},\theta_{\mathcal{N}}\right)=\underline{\underline{A}}\left(\theta_{\mathcal{S}}\right)\underline{\theta_{\mathcal{N}}}=\underline{\underline{A}}\,\underline{\theta}$$

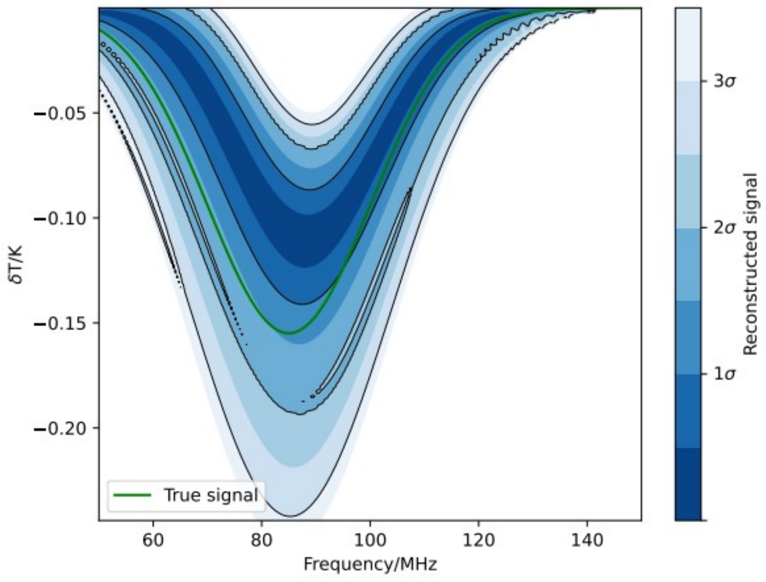
$$\underline{\mu}=\underline{\underline{\Sigma}}\,\underline{\underline{A}}^{T}\underline{\underline{C}}^{-1}\underline{\mathcal{D}}$$

Requires inverting

$$\underline{\underline{\Sigma}}^{-1}=\underline{\underline{A}}^{\mathrm{T}}\underline{\underline{C}}^{-1}\underline{\underline{A}}$$

$$\log\mathcal{L}_{\mathrm{eff}}=-\frac{1}{2}\log\left|\left(2\pi\right)^{n}\underline{\underline{C}}\right|-\frac{1}{2}\underline{\mathcal{D}}^{\mathrm{T}}\underline{\underline{C}}^{-1}\underline{\mathcal{D}}+\frac{1}{2}\underline{\mu}^{\mathrm{T}}\underline{\underline{\Sigma}}^{-1}\underline{\mu}+\frac{1}{2}\log\left|\left(2\pi\right)^{k}\underline{\underline{\Sigma}}\right|$$





UNIVERSITY OF CAMBRIDGE

Marginal Statistics

What if the marginal likelihood is needed but isn't analytically tractable?

$$\mathcal{P}\left(\theta_\mathcal{S}, \theta_\mathcal{N}\right) = \frac{\mathcal{L}\left(\theta_\mathcal{S}, \theta_\mathcal{N}\right)\Pi\left(\theta_\mathcal{S}, \theta_\mathcal{N}\right)}{\mathcal{Z}}$$

$$\Pi_\mathrm{eff}\left(\theta_\mathcal{S}\right) = \int \Pi\left(\theta_\mathcal{S}, \theta_\mathcal{N}\right)\mathrm{d}\theta_\mathcal{N}$$

$$\mathcal{P}_\mathrm{eff}\left(\theta_\mathcal{S}\right) = \int \mathcal{P}\left(\theta_\mathcal{S}, \theta_\mathcal{N}\right)\mathrm{d}\theta_\mathcal{N}$$

$$\mathcal{P}_\mathrm{eff}\left(\theta_\mathcal{S}\right)\mathcal{Z} = \int \mathcal{L}\left(\theta_\mathcal{S}, \theta_\mathcal{N}\right)\Pi\left(\theta_\mathcal{S}, \theta_\mathcal{N}\right)\mathrm{d}\theta_\mathcal{N}$$

$$\mathcal{P}_\mathrm{eff}\left(\theta_\mathcal{S}\right)\mathcal{Z} = \mathcal{L}_\mathrm{eff}\left(\theta_\mathcal{S}\right)\Pi_\mathrm{eff}\left(\theta_\mathcal{S}\right) \longrightarrow \mathcal{L}_\mathrm{eff}\left(\theta_\mathcal{S}\right) = \frac{\mathcal{P}_\mathrm{eff}\left(\theta_\mathcal{S}\right)\mathcal{Z}}{\Pi_\mathrm{eff}\left(\theta_\mathcal{S}\right)}$$

$$\mathcal{L}_\mathrm{eff}\left(\theta_\mathcal{S}\right) = \frac{\int \mathcal{L}\left(\theta_\mathcal{S}, \theta_\mathcal{N}\right)\Pi\left(\theta_\mathcal{S}, \theta_\mathcal{N}\right)\mathrm{d}\theta_\mathcal{N}}{\int \Pi\left(\theta_\mathcal{S}, \theta_\mathcal{N}\right)\mathrm{d}\theta_\mathcal{N}}$$

Slow to evaluate

The marginal likelihood can be inferred from the marginal posterior

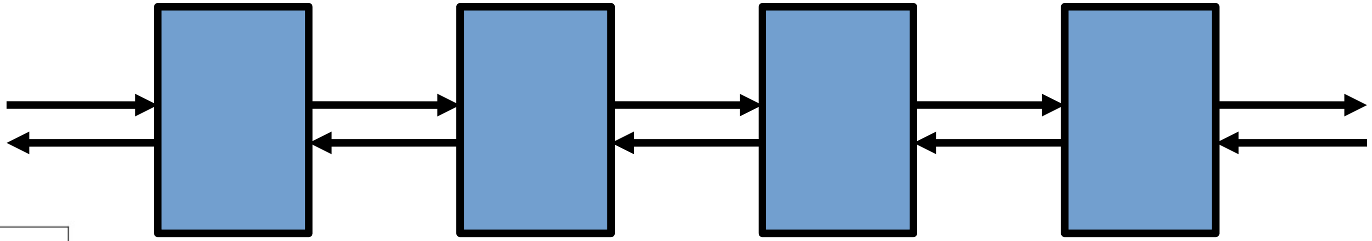
Evaluating Marginal Posteriors

Only have samples from the posterior

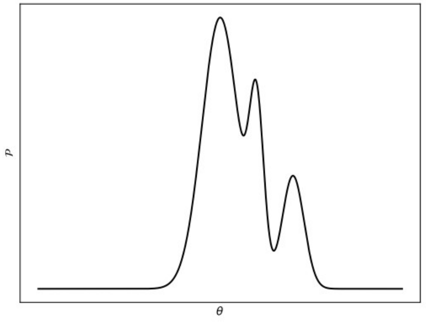
Learn the underlying distribution using normalising flows

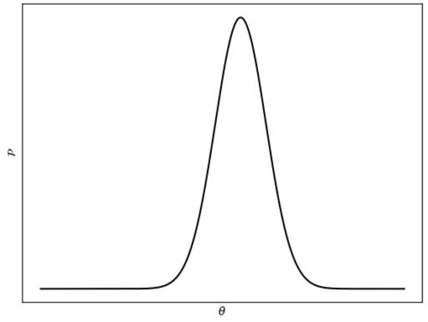
$$\mathcal{L}_{\text{eff}}\left(\theta_{\mathcal{S}}\right) = \frac{\mathcal{P}_{\text{eff}}\left(\theta_{\mathcal{S}}\right)\mathcal{Z}}{\Pi_{\text{eff}}\left(\theta_{\mathcal{S}}\right)}$$

Sampled distribution



Known target distribution (usually Gaussian)





Sequence of bijector neural nets

Feed known samples through and set loss function based on how well the resulting samples match the target

Can convert samples or the pdf of the target into samples or the pdf of the initial

UNIVERSITY OF CAMBRIDGE

Joint Constraints

Suppose we have two data sets with different models and nuisance parameters but with a few shared parameters of interest. Can both be used in a joint constraint?

Model 1

$$\underbrace{\{\overset{3}{\theta_{\mathcal{S}}}, \overset{14}{\theta_{\mathcal{N}1}}\}}_{17}$$

Model 2

$$\underbrace{\{\overset{3}{\theta_{\mathcal{S}}}, \overset{20}{\theta_{\mathcal{N}2}}\}}_{23}$$

$$\underbrace{}_{37}$$

$$\log \mathcal{L}\left(\theta_{\mathcal{S}}\right) = \log \mathcal{L}_{\mathrm{eff}\,1}\left(\theta_{\mathcal{S}}\right) + \log \mathcal{L}_{\mathrm{eff}\,2}\left(\theta_{\mathcal{S}}\right)$$

More efficient to fit nuisance models independently, then only jointly fit the marginals

PolyChord (Handley et al.) has a runtime that scales with $\mathcal{O}\left(n^3\right)$

$$37^3 > 17^3 + 23^3 + 3^3$$

UNIVERSITY OF CAMBRIDGE

Use Case – Multi-Instrument Early Universe Constraints

Recent measurements of the upper limits on the 21cm signal:

SARAS3:

- wideband monopole radiometer
- observes in the band 55-85MHz
- based in India

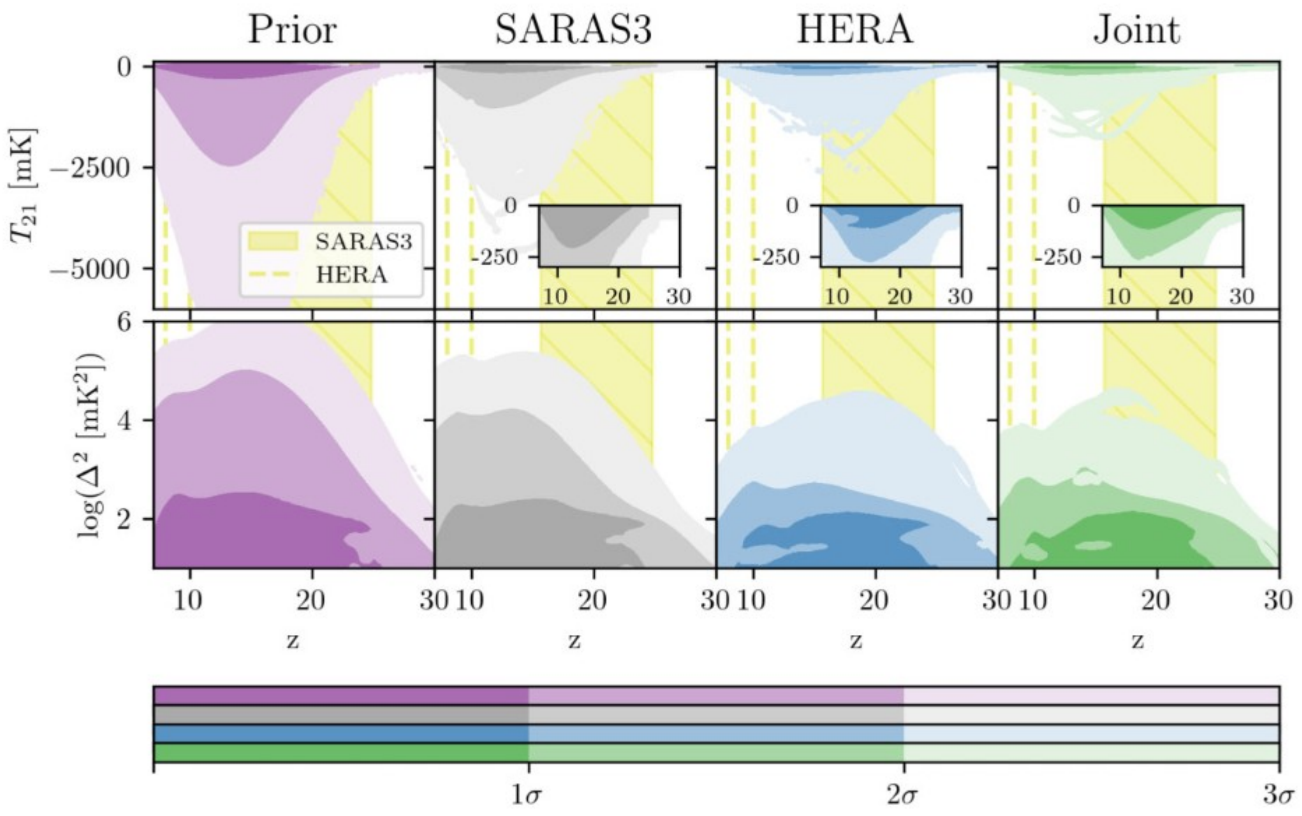
Singh et al. 2022

HERA:

- dense packed interferometer
- constraints in the band ~120-190MHz
- based in South Africa

Abdurashidova et al. 2022

Use Case – Multi-Instrument Early Universe Constraints



From Bevins et al. (2024)

UNIVERSITY OF CAMBRIDGE

Summary

Learned about 4 more advanced Bayesian statistical techniques and examples of appropriate circumstances for their application:

- Conjugate priors – For certain analytical likelihood, there exists a specific analytical conjugate prior, which gives a posterior in of the same family as the prior that can be evaluated quickly and analytically – e.g. in quick on-site radiometer calibration

- Likelihood reweighting – If a likelihood can be defined that takes the same parameters and has a similar posterior to another, slower likelihood, the slow posterior can be found by fitting the fast and reweighting the resultant samples – e.g. flagging transient RFI

- Analytical marginalisation – For certain likelihoods and prior pairs, linear nuisance parameters can be marginalised analytically in the likelihood, allowing the fit to avoid needing to sample those parameters – e.g. high dimensional antenna beam modelling

- Normalising flow marginal statistics – Marginal likelihoods can be derived from marginal posteriors and learnt with normalising flows, enabling efficient joint marginal fitting of multiple data sets – e.g. joint constraints on 21cm limits