Simulation Based Inference

theory, sampling & model comparison

Will Handley
<wh260@cam.ac.uk>

Royal Society University Research Fellow
Astrophysics Group, Cavendish Laboratory, University of Cambridge
Kavli Institute for Cosmology, Cambridge
Gonville & Caius College
`willhandley.co.uk/talks`

12th January 2024

How all SBI talks finish

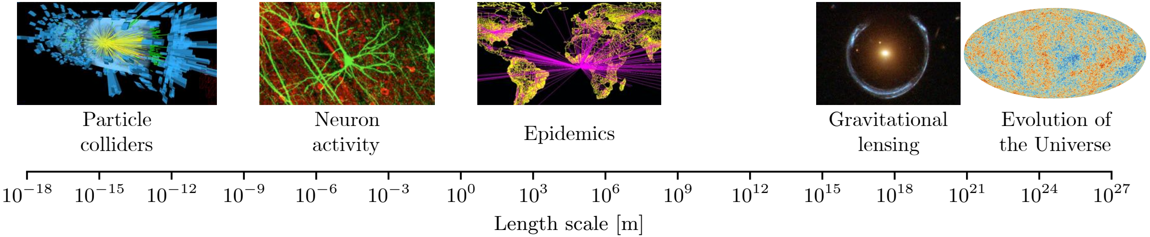
- There is a standard exchange that tends to happen after giving an SBI talk:

  audience: Surely you're only as good as your simulations —
  What if your forward model is missing physics X?

  speaker: The exact same thing affects likelihood-based analysis —
  All SBI does is make these assumptions explicit.

- I will try to unpack why I think both sides have a point.

Frontiers of simulation based inference [1911.01429]

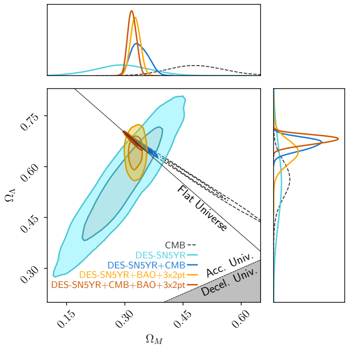


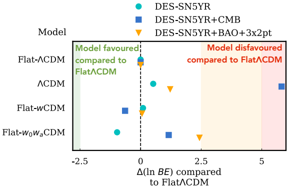
LBI: Likelihood-based inference

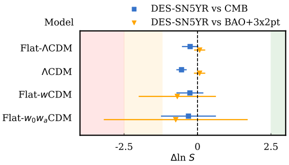
The standard approach if you are fortunate enough to have a likelihood function $P(D|\theta)$:

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

1. Define prior $\pi(\theta)$
   - spend some time being philosophical
2. Sample posterior $\mathcal{P}(\theta|D)$
   - use out-of-the-box MCMC tools such as `emcee` or `MultiNest`
   - make some triangle plots
3. Optionally compute evidence $\mathcal{Z}(D)$
   - e.g. nested sampling or parallel tempering
   - do some model comparison (i.e. science)
   - talk about tensions e.g. [2401.02929]





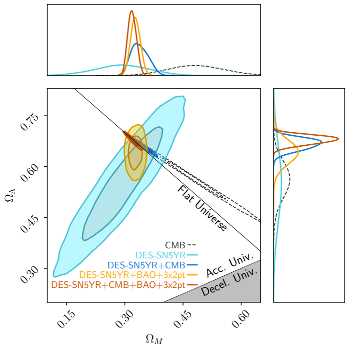


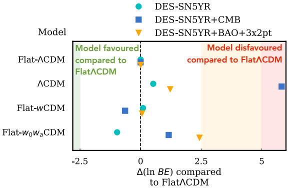
LBI: Likelihood-based inference

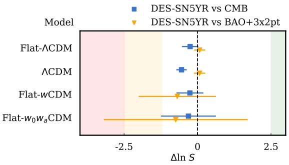
The standard approach if you are fortunate enough to have a likelihood function $P(D|\theta)$:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \qquad \text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

1. Define prior $\pi(\theta)$
   - spend some time being philosophical
2. Sample posterior $\mathcal{P}(\theta|D)$
   - use out-of-the-box MCMC tools such as `emcee` or `MultiNest`
   - make some triangle plots
3. Optionally compute evidence $\mathcal{Z}(D)$
   - e.g. nested sampling or parallel tempering
   - do some model comparison (i.e. science)
   - talk about tensions e.g. [2401.02929]



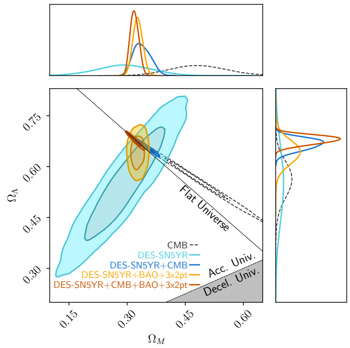




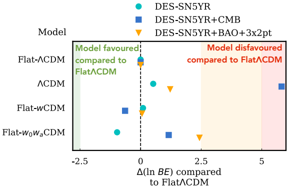
LBI: Likelihood-based inference

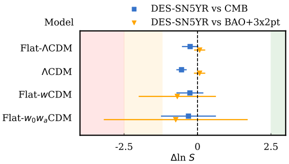
The standard approach if you are fortunate enough to have a likelihood function $\mathcal{L}(D|\theta)$:

$$\mathcal{P}(\theta|D) = \frac{\mathcal{L}(D|\theta)\pi(\theta)}{\mathcal{Z}(D)} \qquad \text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$



1. Define prior $\pi(\theta)$
   - spend some time being philosophical
2. Sample posterior $\mathcal{P}(\theta|D)$
   - use out-of-the-box MCMC tools such as `emcee` or `MultiNest`
   - make some triangle plots
3. Optionally compute evidence $\mathcal{Z}(D)$
   - e.g. nested sampling or parallel tempering
   - do some model comparison (i.e. science)
   - talk about tensions e.g. [2401.02929]

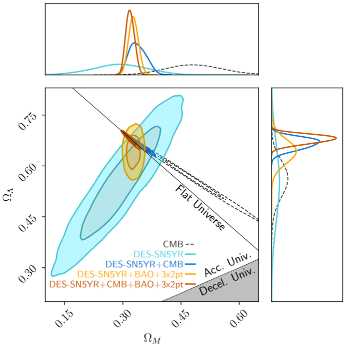


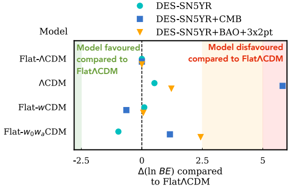


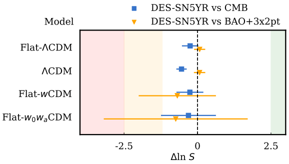
The standard approach if you are fortunate enough to have a likelihood function $\mathcal{L}(D|\theta)$:

$$\mathcal{P} \times \mathcal{Z} = \mathcal{L} \times \pi$$

1. Define prior $\pi(\theta)$
   - spend some time being philosophical
2. Sample posterior $\mathcal{P}(\theta|D)$
   - use out-of-the-box MCMC tools such as `emcee` or `MultiNest`
   - make some triangle plots
3. Optionally compute evidence $\mathcal{Z}(D)$
   - e.g. nested sampling or parallel tempering
   - do some model comparison (i.e. science)
   - talk about tensions e.g. [2401.02929]





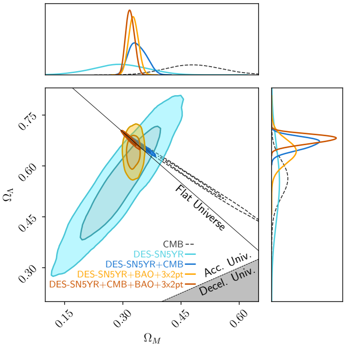


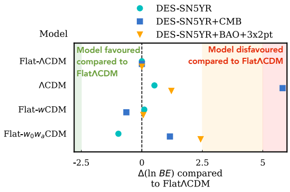
LBI: Likelihood-based inference

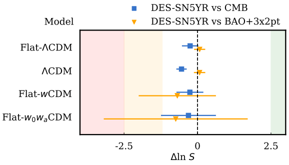
The standard approach if you are fortunate enough to have a likelihood function $\mathcal{L}(D|\theta)$:

$$\mathcal{P} \times \mathcal{Z} = \mathcal{J} = \mathcal{L} \times \pi, \qquad \text{Joint} = \mathcal{J} = P(D,\theta)$$

1. Define prior $\pi(\theta)$
   - spend some time being philosophical
2. Sample posterior $\mathcal{P}(\theta|D)$
   - use out-of-the-box MCMC tools such as `emcee` or `MultiNest`
   - make some triangle plots
3. Optionally compute evidence $\mathcal{Z}(D)$
   - e.g. nested sampling or parallel tempering
   - do some model comparison (i.e. science)
   - talk about tensions e.g. [2401.02929]

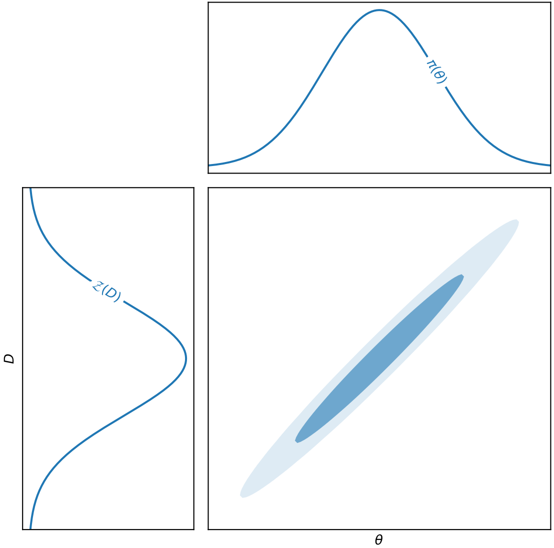






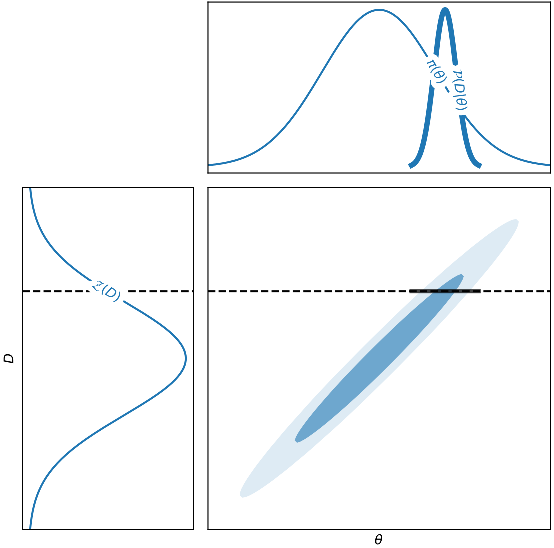
SBI: Simulation-based inference

- Only have access to a forward model $\theta \to D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$
  the "probability of everything".
- Task of SBI is then to go from joint $\mathcal{J}$ to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- SBI & forward modelling force us to think about data space D & parameter space θ.



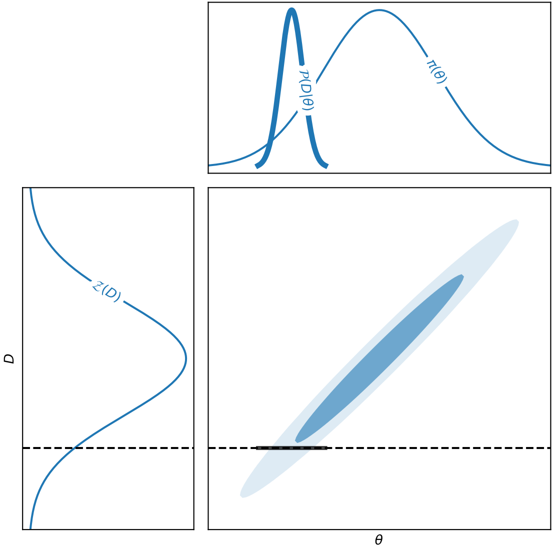
SBI: Simulation-based inference

- Only have access to a forward model $\theta \to D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$
  the "probability of everything".
- Task of SBI is then to go from joint $\mathcal{J}$ to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- SBI & forward modelling force us to think about data space D & parameter space θ.



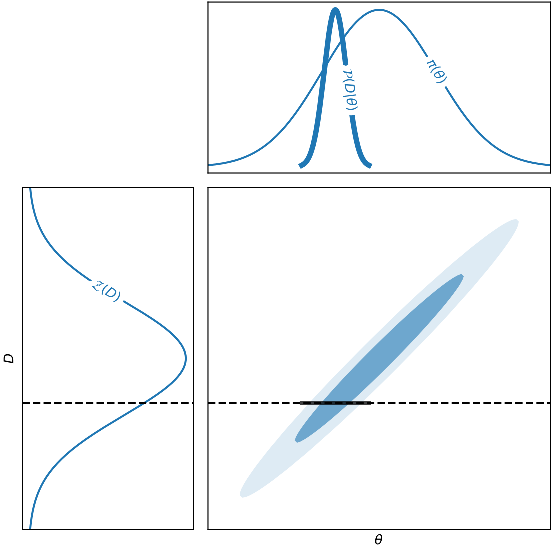
SBI: Simulation-based inference

- Only have access to a forward model $\theta \to D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- SBI & forward modelling force us to think about data space D & parameter space θ.



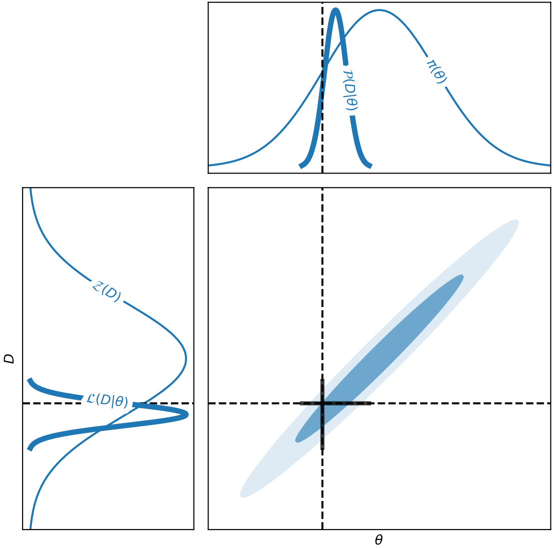
SBI: Simulation-based inference

- Only have access to a forward model $\theta \to D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- SBI & forward modelling force us to think about data space D & parameter space θ.



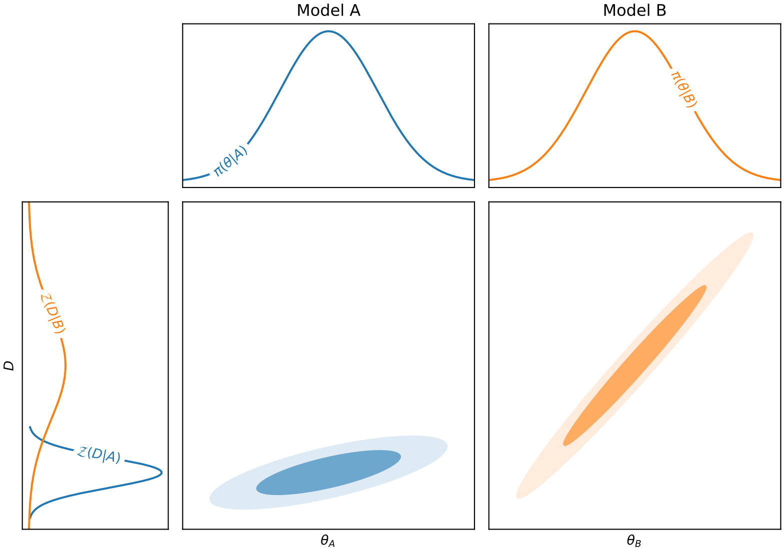
SBI: Simulation-based inference

- Only have access to a forward model $\theta \to D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$
  the "probability of everything".
- Task of SBI is then to go from joint $\mathcal{J}$ to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- SBI & forward modelling force us to think about data space D & parameter space θ.



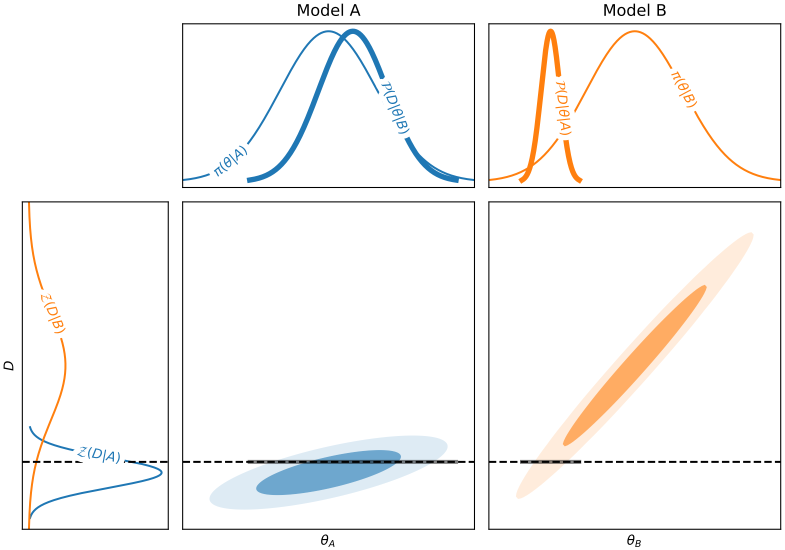
Simulation-based inference & model comparison

- Extend: models A and B.
- Each with own separate parameters θ_A and θ_B (can be same).
- The evidence $\mathcal{Z}(D|M)$ compares models
- Occams razor: more predictive $\equiv$ more probable (due to normalisation).



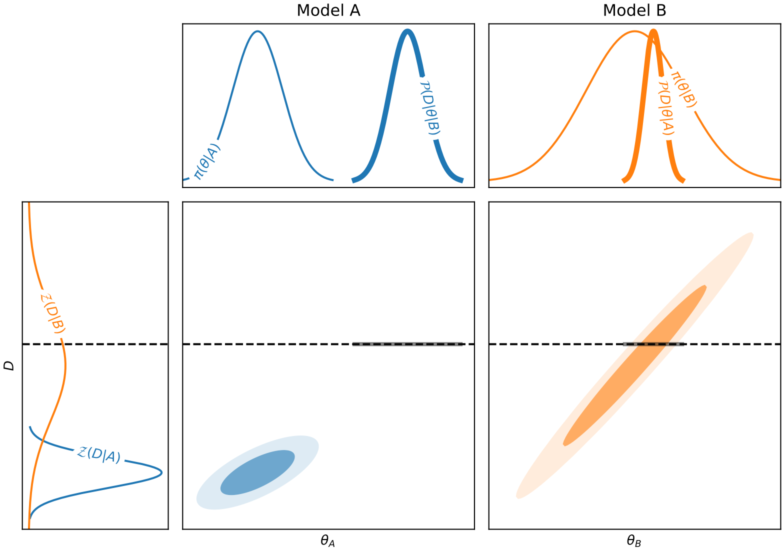
Simulation-based inference & model comparison

- Extend: models A and B.
- Each with own separate parameters θ_A and θ_B (can be same).
- The evidence $\mathcal{Z}(D|M)$ compares models
- Occams razor: more predictive $\equiv$ more probable (due to normalisation).

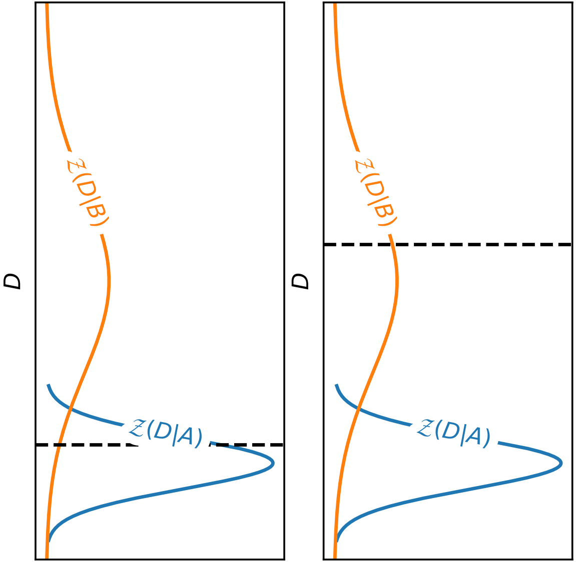


Simulation-based inference & model comparison

- Extend: models A and B.
- Each with own separate parameters θ_A and θ_B (can be same).
- The evidence $\mathcal{Z}(D|M)$ compares models
- Occams razor: more predictive $\equiv$ more probable (due to normalisation).

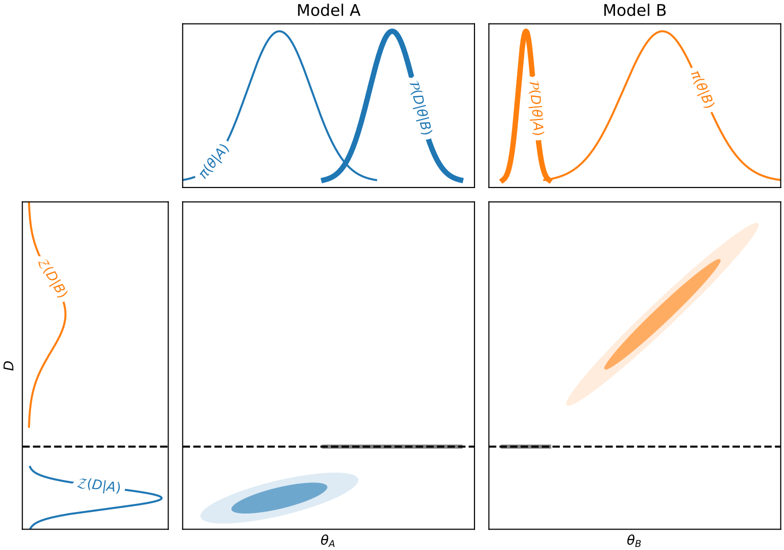


- Procedure proposed by Jeffrey & Wandelt:
  1. Generate labelled data from model A and model B.
  2. Train a probabilistic classifier to distinguish between the two.
  3. Use neural ratio trick to extract Bayes Factor $B = P(D|A)/P(D|B)$.
- NRE for data
- Fully marginalises out parameters
- Only works in the data space
- Model comparison without nested sampling!
- Can be extremely effective



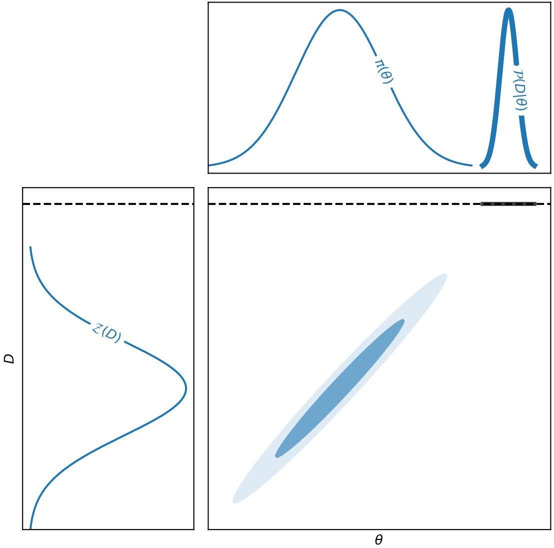
A word of caution on data-space modelling

- In practice the situation is more like this $\Rightarrow$
  - *"No models are true, (but some are useful)"*
- Curse of dimensionality means real data may not lie in either/any evidence distribution $\mathcal{Z}(D)$.
- e.g. if you are training an ML method, it will have never seen simulated data like the real data.



A word of caution on data-space modelling

- This concern affects any amortised method
  - means trainining method on simulations...
  - ...and then pass in the real data
  - They are amortised (over the data) because they can be re-used for any new data.
- Observed data is only thing we surely know.
- As scientists we should be suspicious of a method that leaves D_{obs} until the end.
- This can be amelioriated by fitting θ.
  - Fitting concentrates parameters & simulations around the posterior/real data
  - See this in truncated approaches & ABC

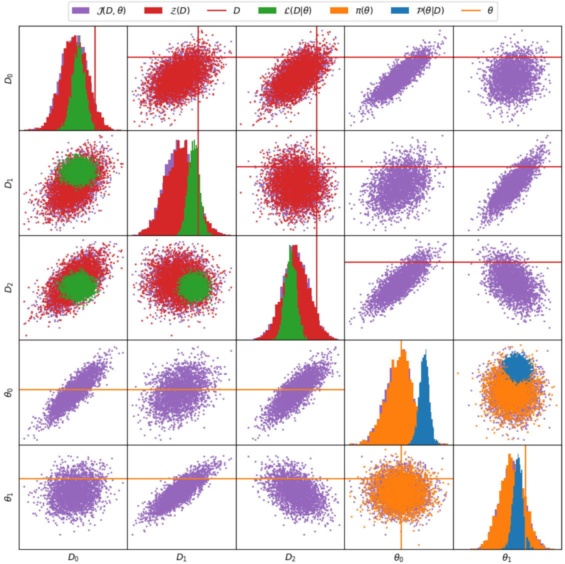


Why do amortised methods often work so well?

Whilst these concerns sound worrying, many succesful amortised methods exist. Why?

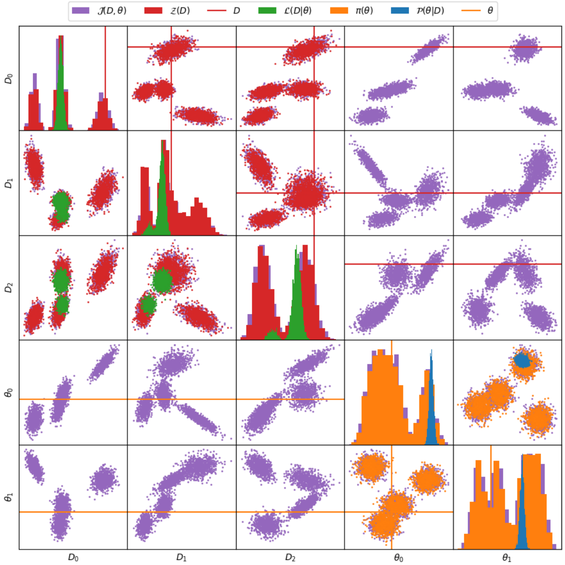
1. Some methods are only validated on data generated from the same simulator as the one used for inference.
2. Some methods are only validated on simple Gaussian examples, since it's possible to compute the ground truth in these cases.
   - Recommendation: also test on Gaussian mixture models, for which full analytics are also known
3. Real data may not actually be that challenging!

- If the final point holds, then in many cases we may not need expressive ML/AI methods
- Often it is the data-intensive “plug-and-play” power of ML packages that is most useful.
- If your ML is just learning a simple decision boundary, why not just use a linear model?
- `lsbi` is a python package that implements plug-and-play the fiddly linear mathematics.
- Also pedagogically useful for persuading people that SBI $\neq$ ML.
- Beta-testers wanted:
- `lsbi`: github.com/handley-lab/lsbi (PyPI & conda)



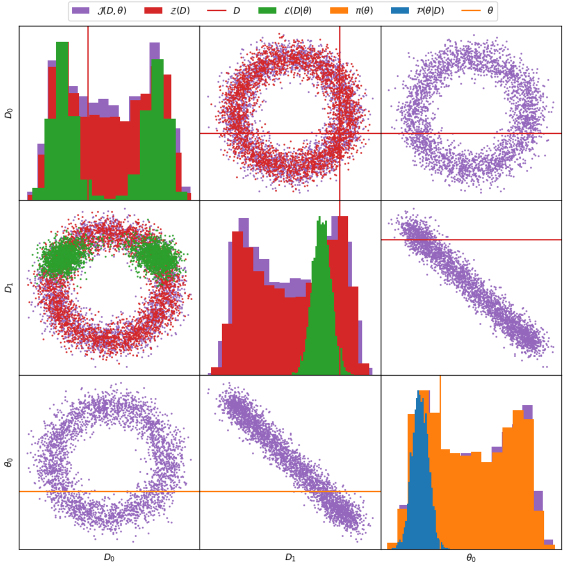
`lsbi`: Linear Simulation Based Inference

- If the final point holds, then in many cases we may not need expressive ML/AI methods
- Often it is the data-intensive "plug-and-play" power of ML packages that is most useful.
- If your ML is just learning a simple decision boundary, why not just use a linear model?
- `lsbi` is a python package that implements plug-and-play the fiddly linear mathematics.
- Also pedagogically useful for persuading people that SBI $\neq$ ML.
- Beta-testers wanted:
- `lsbi`: github.com/handley-lab/lsbi (PyPI & conda)



`lsbi`: Linear Simulation Based Inference

- If the final point holds, then in many cases we may not need expressive ML/AI methods
- Often it is the data-intensive "plug-and-play" power of ML packages that is most useful.
- If your ML is just learning a simple decision boundary, why not just use a linear model?
- `lsbi` is a python package that implements plug-and-play the fiddly linear mathematics.
- Also pedagogically useful for persuading people that SBI $\neq$ ML.
- Beta-testers wanted:
- `lsbi`: github.com/handley-lab/lsbi (PyPI & conda)



How this SBI talk finishes

- There is a standard exchange that tends to happen after giving an SBI talk:

  audience Surely you're only as good as your simulations —
  What if your forward model is missing physics X?

  speaker The exact same thing affects likelihood-based analysis —
  All SBI does is make these assumptions explicit.

- The audience is implicitly making a query about the danger of working in data space D, whilst the speaker's comment only applies to parameter space θ.
- Discussion point: We should therefore focus on SBI approaches which have tunable parameter spaces (i.e. interpretable posteriors).

Conclusions

github.com/handley-lab

- These musings emerged from conversations with:
  - David Yallup
  - Mike Hobson
  - Ben Wandelt
  - Justin Alsing
  - Niall Jeffrey
- As scientists, we should be cautious of amortised approaches
- `lsbi` preview: a package-driven attempt to free SBI from ML

Cosmological forecasting

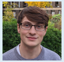
Have you ever done a Fisher forecast, and then felt Bayesian guilt?

- Cosmologists are interested in forecasting what a Bayesian analysis of future data might produce.
- Useful for:
  - white papers/grants,
  - optimising existing instruments/strategies,
  - picking theory/observation to explore next.
- To do this properly:
  1. start from current knowledge $\pi(\theta)$, derived from current data
  2. Pick potential dataset $D \sim \mathcal{Z}(D)$ that might be collected from $P(D)$ $(= \mathcal{Z})$
  3. Derive posterior $\mathcal{P}(\theta|D)$
  4. Summarise science (e.g. constraint on θ, ability to perform model comparison)
- This procedure should be marginalised over:
  1. All possible parameters θ (consistent with prior knowledge)
  2. All possible data D
- i.e. marginalised over the joint $P(\theta, D) = P(D|\theta)P(\theta)$.
- Historically this has proven very challenging.
- Most analyses assume a fiducial cosmology θ_*, and/or a Gaussian likelihood/posterior (c.f. Fisher forecasting).
- This runs the risk of biasing forecasts by baking in a given theory/data realisation.

Fully Bayesian Forecasting [2309.06942]

Thomas Gessey-Jones

PhD



- Simulation based inference gives us the language to marginalise over parameters θ and possible future data D.
- Evidence networks give us the ability to do this at scale for forecasting.
- Demonstrated in 21cm global experiments, marginalising over:
  - theoretical uncertainty
  - foreground uncertainty
  - systematic uncertainty
- Able to say "at 67mK radiometer noise", have a 50% chance of 5σ Bayes factor detection.
- Can use to optimise instrument design
- Re-usable package: `prescience`