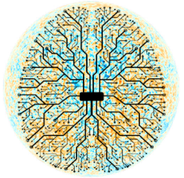
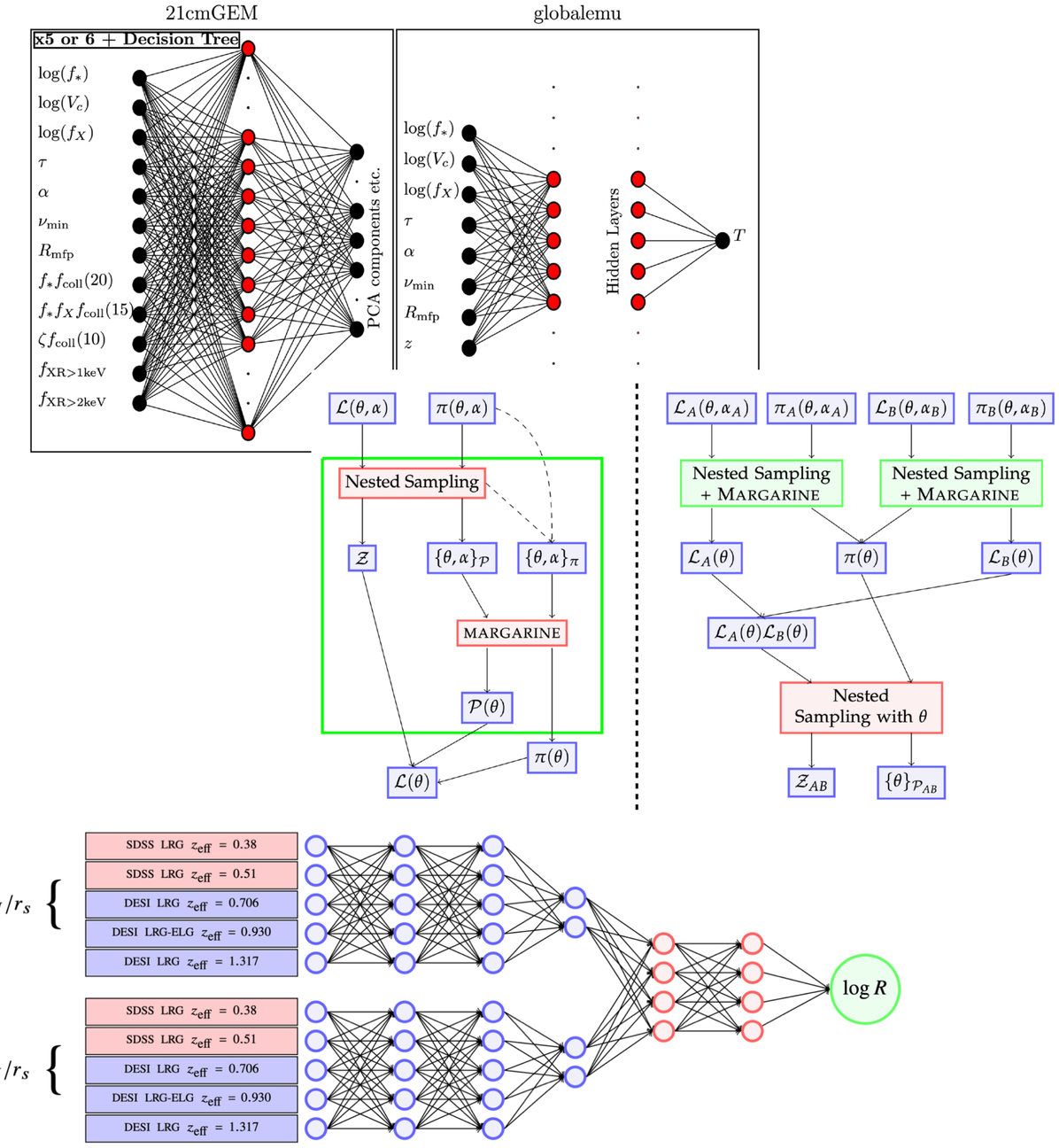
Machine Learning enhanced Bayesian Inference for Cosmology

Harry Bevins
With Will Handley (and others)


Contents

- I will discuss
  - Signal emulators [2104.04336]
  - Marginal Bayesian inference [2207.11457, 2205.12841]
  - Tension Statistics [2407.15478]



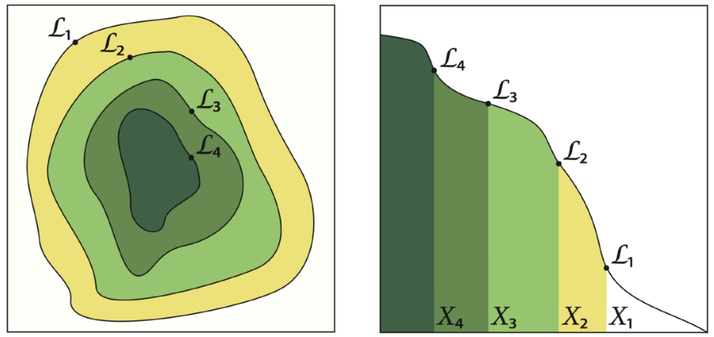
Signal Emulation

Bayesian Inference

$$P(\Theta \mid D, M) = \frac{P(D \mid \Theta, M) P(\Theta \mid M)}{P(D \mid M)} = \frac{L(\Theta)\pi(\Theta)}{Z}$$

- Posterior $P(\Theta \mid D, M)$
- Likelihood $L(\Theta) = P(D \mid \Theta, M)$
- Prior $\pi(\Theta) = P(\Theta \mid M)$
- Evidence $Z = P(D \mid M)$
- Lots of different ways to access the posterior and evidence (e.g. Metropolis Hastings, HMC, SMC, Nested Sampling, ML-enhanced Harmonic Means, *floZ,* Simulation Based Inference)
- Focusing on Nested Sampling here

emcee
The MCMC Hammer

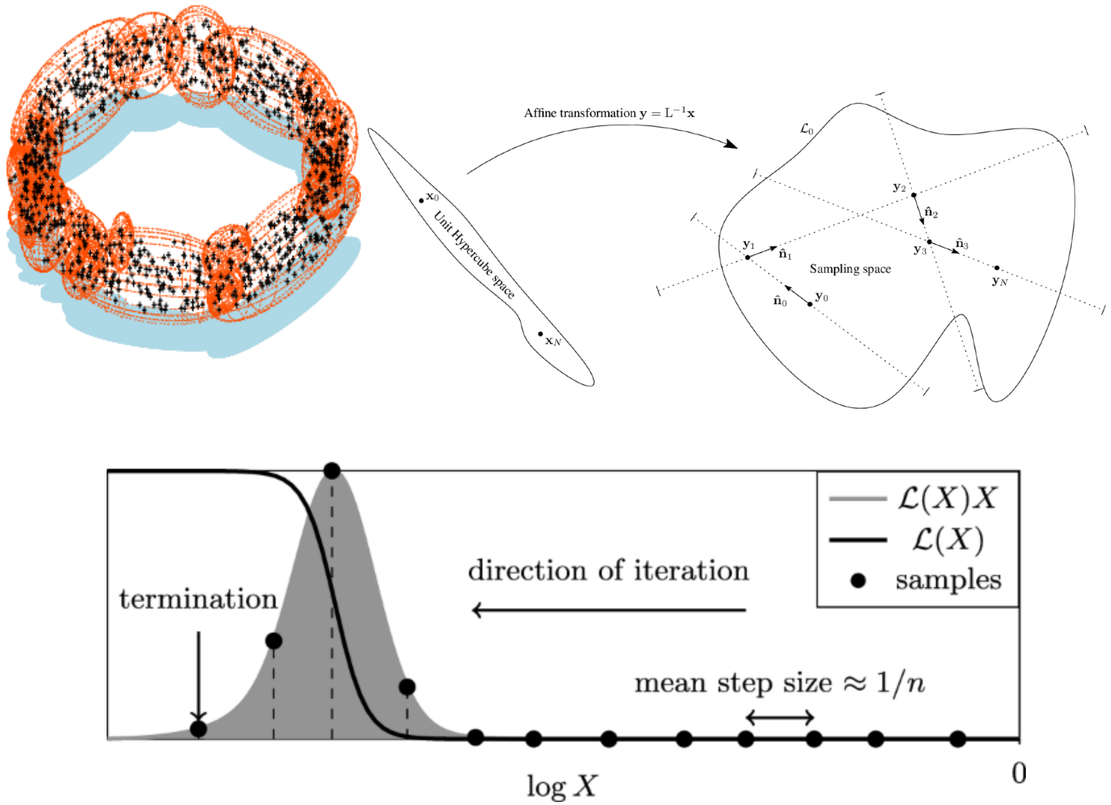


sbi

Scaling of Nested Sampling

$$T \propto n_{\text{live}} \times \langle T\{L(\Theta)\}\rangle \times \langle T\{\text{Impl.}\}\rangle \times D_{KL}(P\,||\,\pi)$$

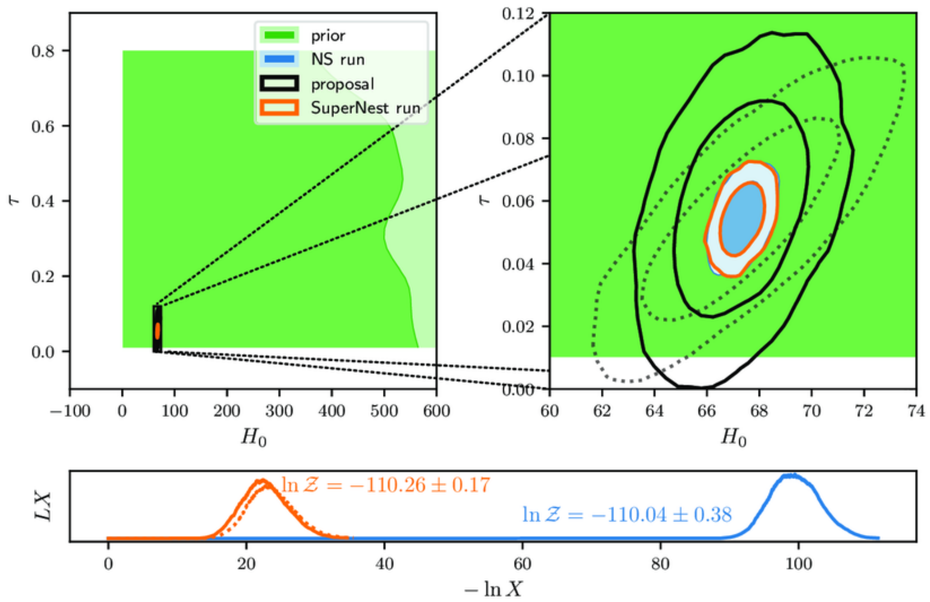
- n_{live} is number of explorers
- $\langle T\{L(\Theta)\}\rangle$ is the time complexity of the likelihood
- $\langle T\{Impl.\}\rangle$ is time complexity of sampling method e.g. slice sampling, region sampling etc
- $D_{KL}(P\,||\,\pi)$ is the KL divergence between posterior and prior



Scaling of Nested Sampling

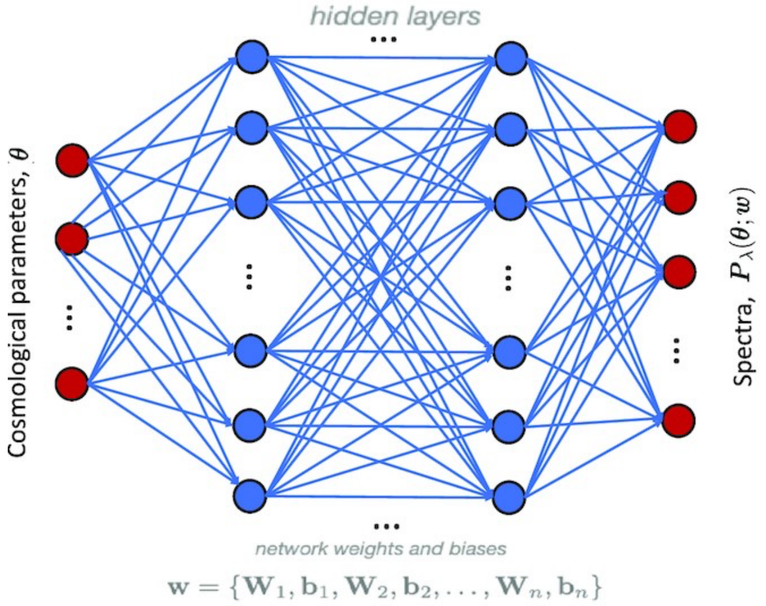
$$T \propto n_{\text{live}} \times \langle T\{L(\Theta)\}\rangle \times \langle T\{\text{Impl.}\}\rangle \times D_{KL}(P\,||\,\pi)$$

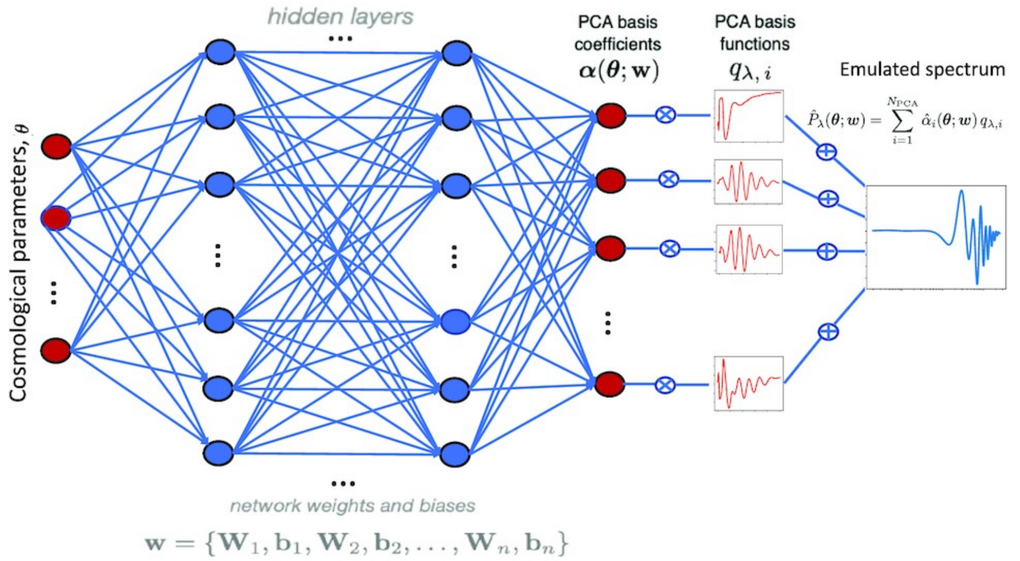
- Have to be careful fiddling with n_{live} (dynamic nested sampling)
- Improving $\langle T\{\text{Impl.}\}\rangle$ is hard!
- Reducing the KL-divergence is eminently doable (see 2212.01760)
- Speeding up our likelihood is also usually doable
- One way to do this is with signal emulation!



What is emulation?

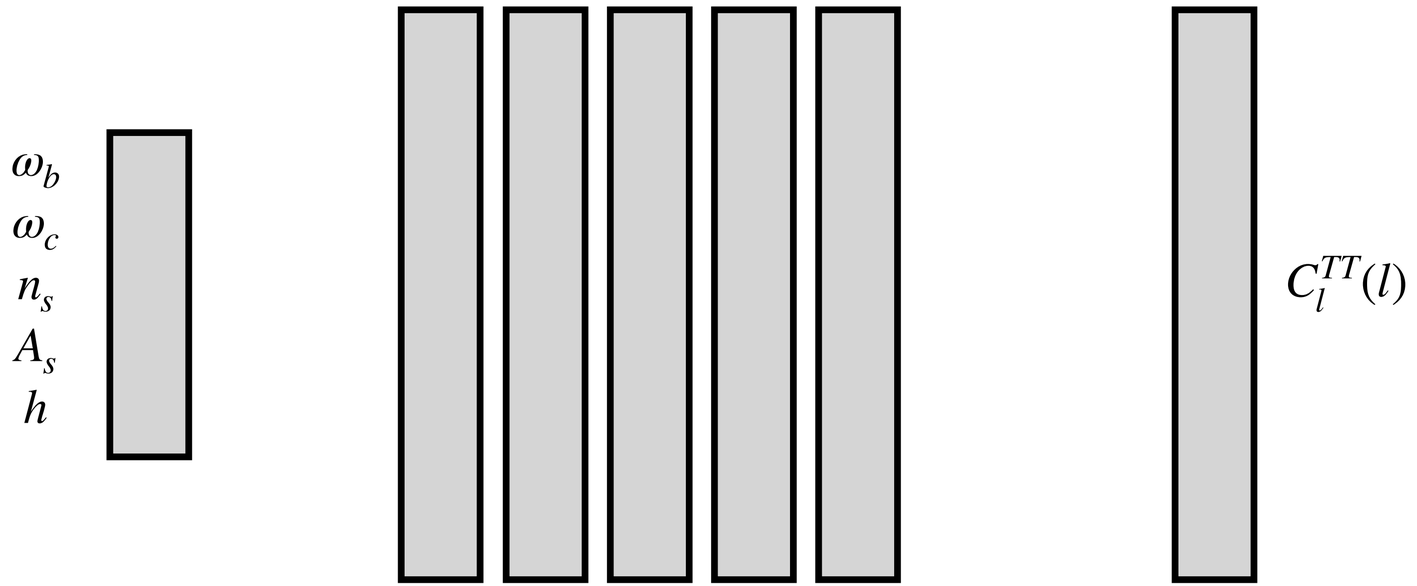
- Emulators are machine learning tools that learn to mimic the outputs of complex scientific simulations
- Probably all familiar with cosmopower which emulates CAMB
- Lots of work on emulators in the field of 21cm Cosmology
- MLPs, transformers, decision trees, variational autoencoders, CNNs etc





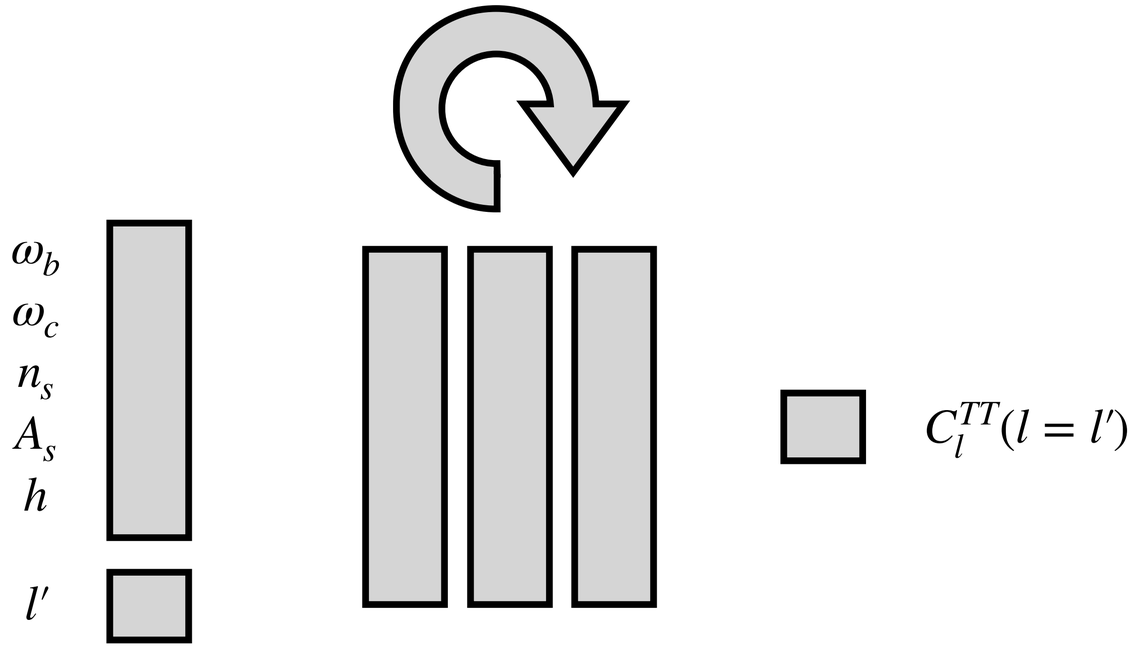
A novel approach to emulation

- Typically emulators take in some model parameters θ and predict a function $y(x)$
- For cosmopower this is $\theta = \{\omega_b, \omega_c, n_s, A_s, h\}$ and $y(x) = C_l^{TT}(l)$
- Cosmopower gets around this by doing PCA



A novel approach to emulation

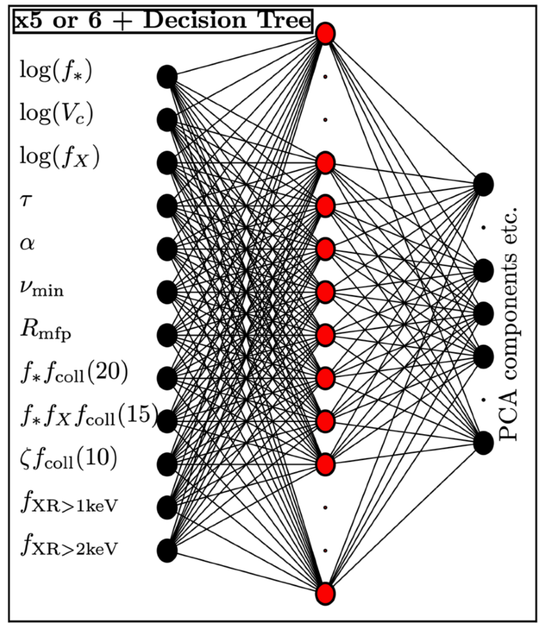
- We suggest having the independent variable as an input to the network
- Looping over the network to get a spectrum



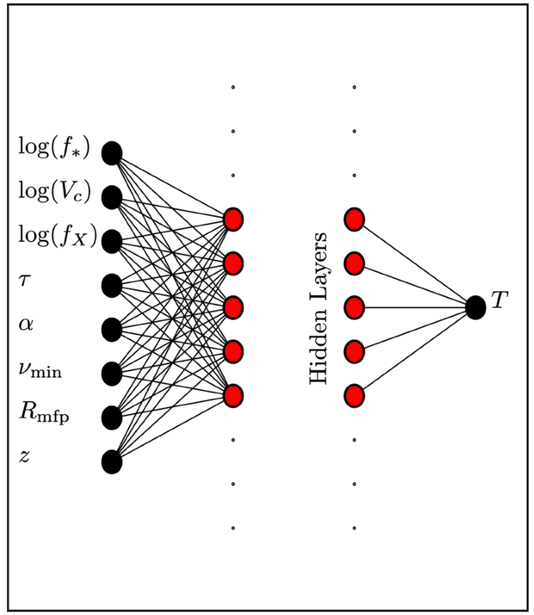
Emulating the 21cm signal

- We demonstrated this approach for the sky-averaged 21cm signal which is a function of redshift z
- Strong dependence on astrophysics
- Loop over the network to get the full signal
- Go from ~hours per signal to ~ms per signal
- Factor 100 improvement in runtime and 2 in accuracy
- See 2104.04336 and for applications 2212.00464, 2312.08828, 2312.08095, 2309.06942 among others!

21cmGEM

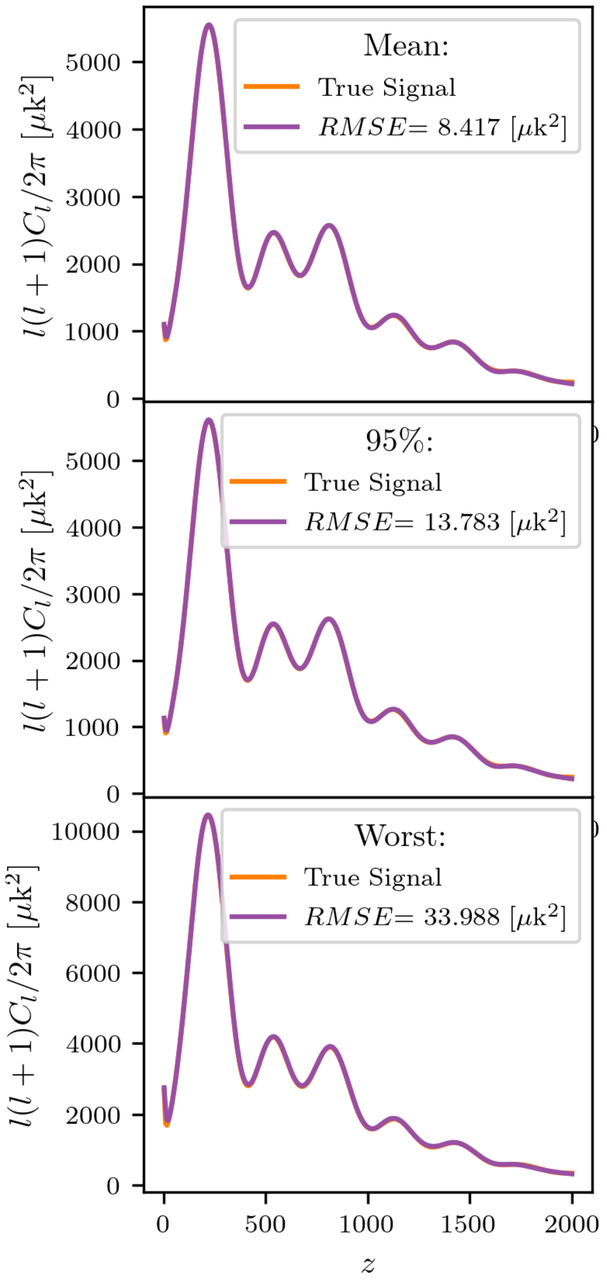


globalemu



Applications to the CMB?

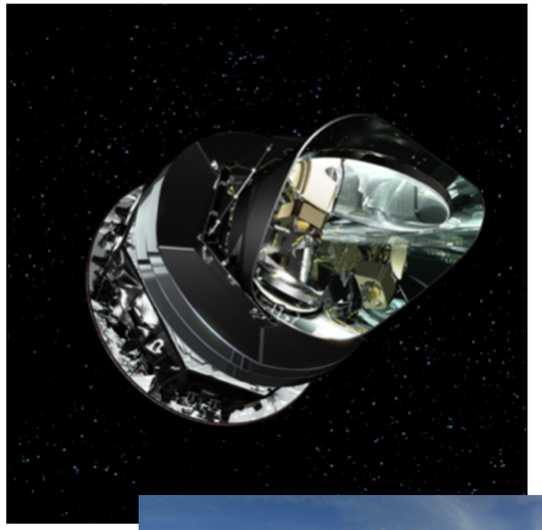
- In practice we can do this for the CMB as well
- Inputs $\theta = \{\omega_b, \omega_c, n_s, A_s, h, l'\}$ and output $C_l^{TT}(l = l')$
- Do not have to do PCA and we can use a smaller network
- Train on high resolution simulations
- Currently discussing with colleagues in Cambridge

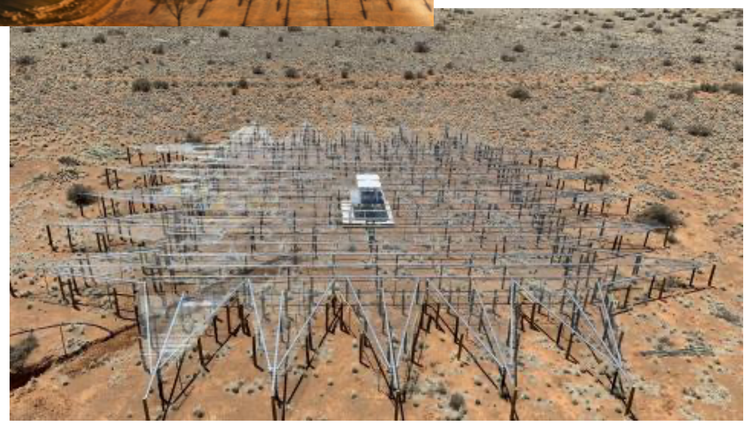


Marginal Bayesian Inference

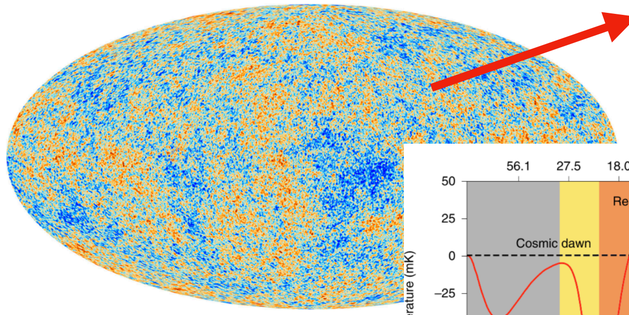
Why are we interested in the marginal space?

- Often we have nuisance parameters α in our modelling that describe instrumental effects or contaminating signals
- While they are interesting we usually are only *really* interested in a few cosmological parameters θ
- Nuisance parameters make joint analysis hard and make comparing experimental results difficult

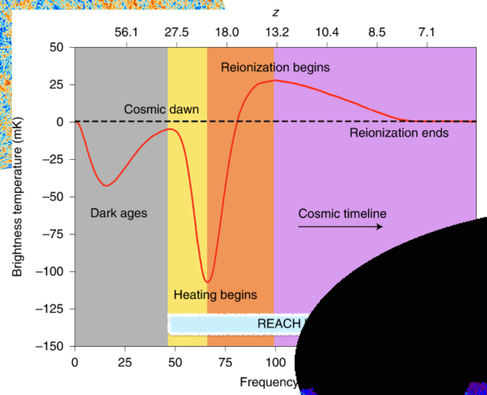




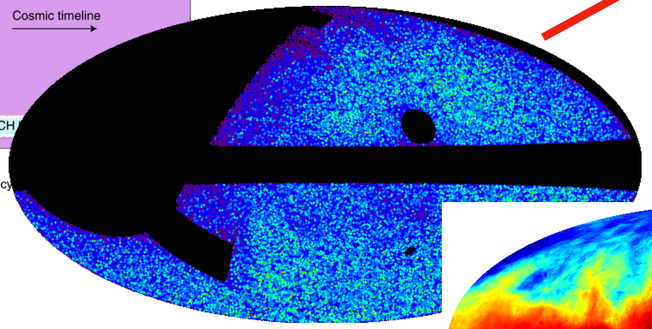
The 21-cm Line



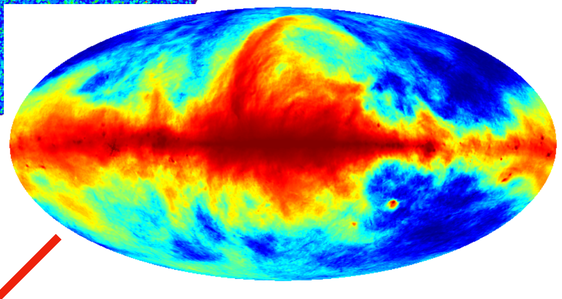
The Cosmic Microwave Background



The 21-cm signal
(What we are interested in!)



Foreground galaxies



Our own Galaxy

The atmosphere

The environment around the telescope

The marginal likelihood

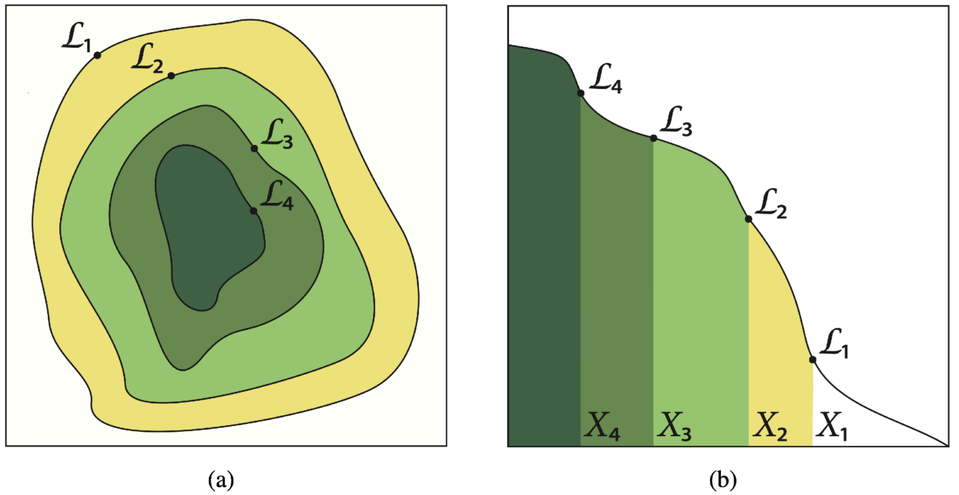
- If we define the marginal posterior and prior as

$$P(\theta) = \int P(\theta, \alpha) d\alpha \quad \text{and} \quad \pi(\theta) = \int \pi(\theta, \alpha) d\alpha$$

- Then we can define the nuisance marginalised likelihood as

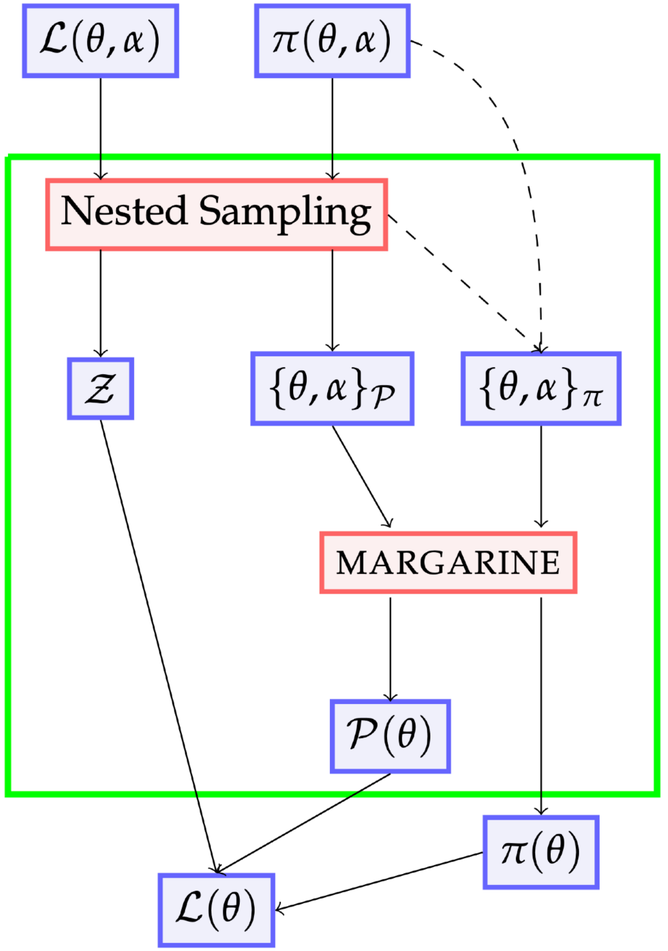
$$L(\theta) = \frac{\int L(\theta, \alpha)\pi(\theta, \alpha) d\alpha}{\int \pi(\theta, \alpha) d\alpha} = \frac{P(\theta)Z}{\pi(\theta)}$$

- Allows us to do efficient joint inference



Normalising Flows

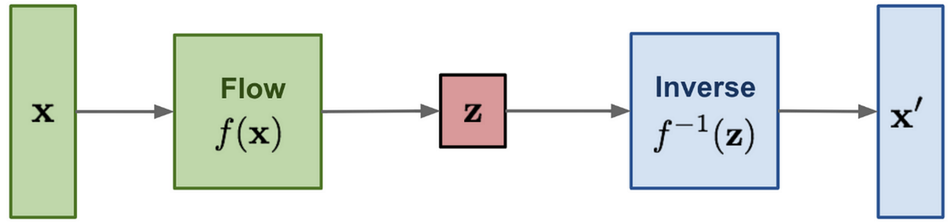
- The issue is that we have samples on $\Theta_P = \{\theta, \alpha\}_P \sim P(\theta, \alpha)$ and $\Theta_\pi = \{\theta, \alpha\}_\pi \sim \pi(\theta, \alpha)$ but not the marginal probabilities $P(\theta)$ and $\pi(\theta)$
- But we can access these with density estimation tools like Normalising Flows
- We implement Masked Autoregressive Flows and package our code up in to a python package called ***margarine***

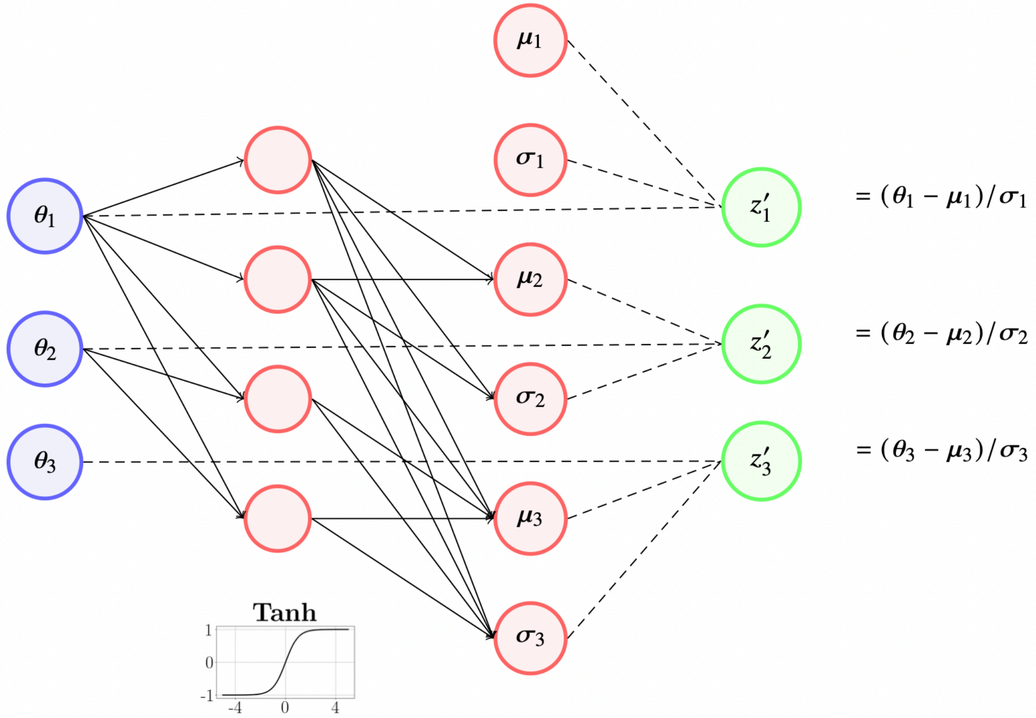


Normalising Flows

- Normalising flows are a class of machine learning model
- Learn a transformation from samples on a tractable distribution to samples on a more complex target like a posterior
- Invertible transformations

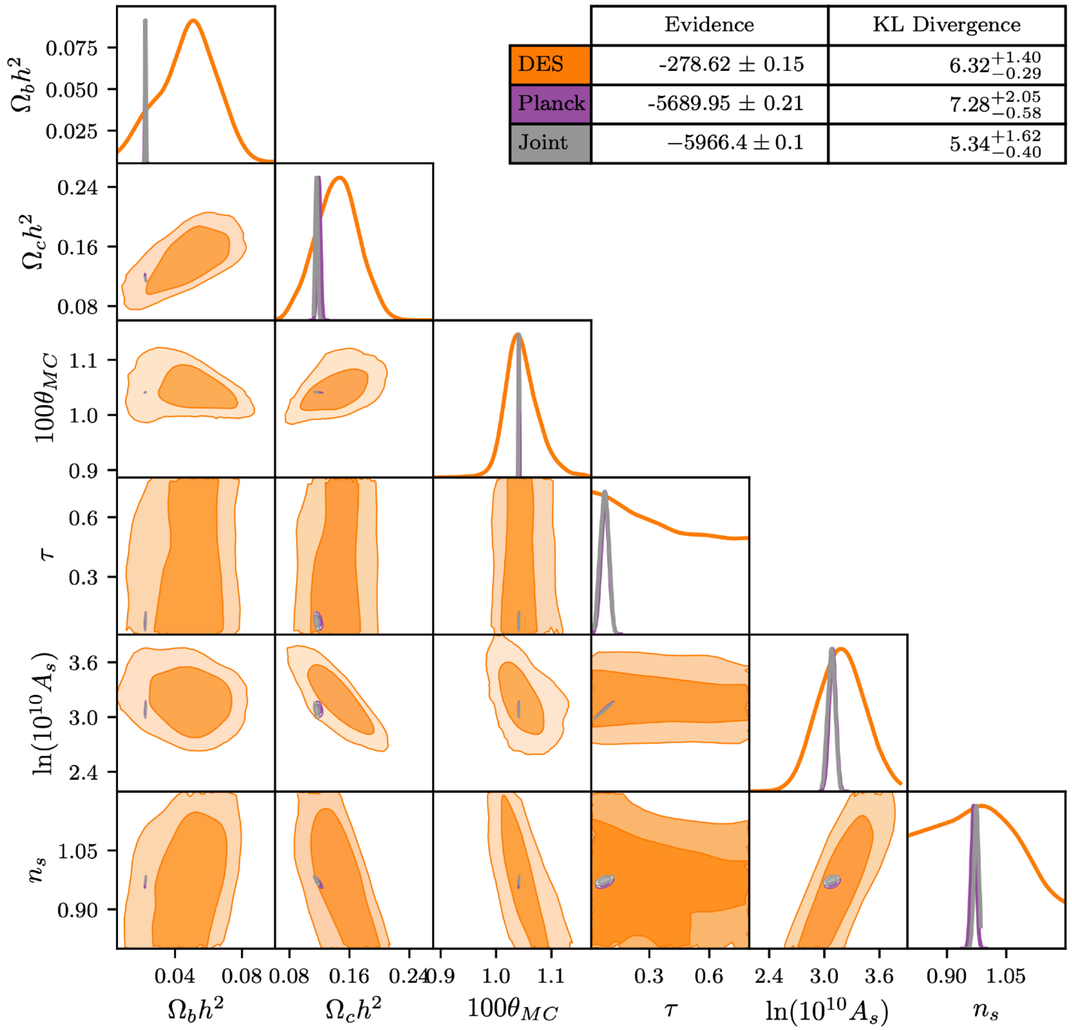
$$P(x) = P(f^{-1}(x)) \left| \det \left(\frac{df^{-1}(x)}{dx} \right) \right|$$





Efficiently combining Planck and DES

- So using normalising flows we can emulate $P(\theta)$, $\pi(\theta)$ and $L(\theta)$ for a given θ
- $L_{Pla.+DES}(\theta) = L_{Pla.}(\theta)L_{DES}(\theta)$
- Reduce dimensionality significantly
- For Nested Sampling the runtime scales as $T \propto d^3$
- From $T \propto 20^3 + 26^3 + 46^3$ to $T \propto 20^3 + 26^3 + 6^3$
- Correct evidence and no double counting of priors

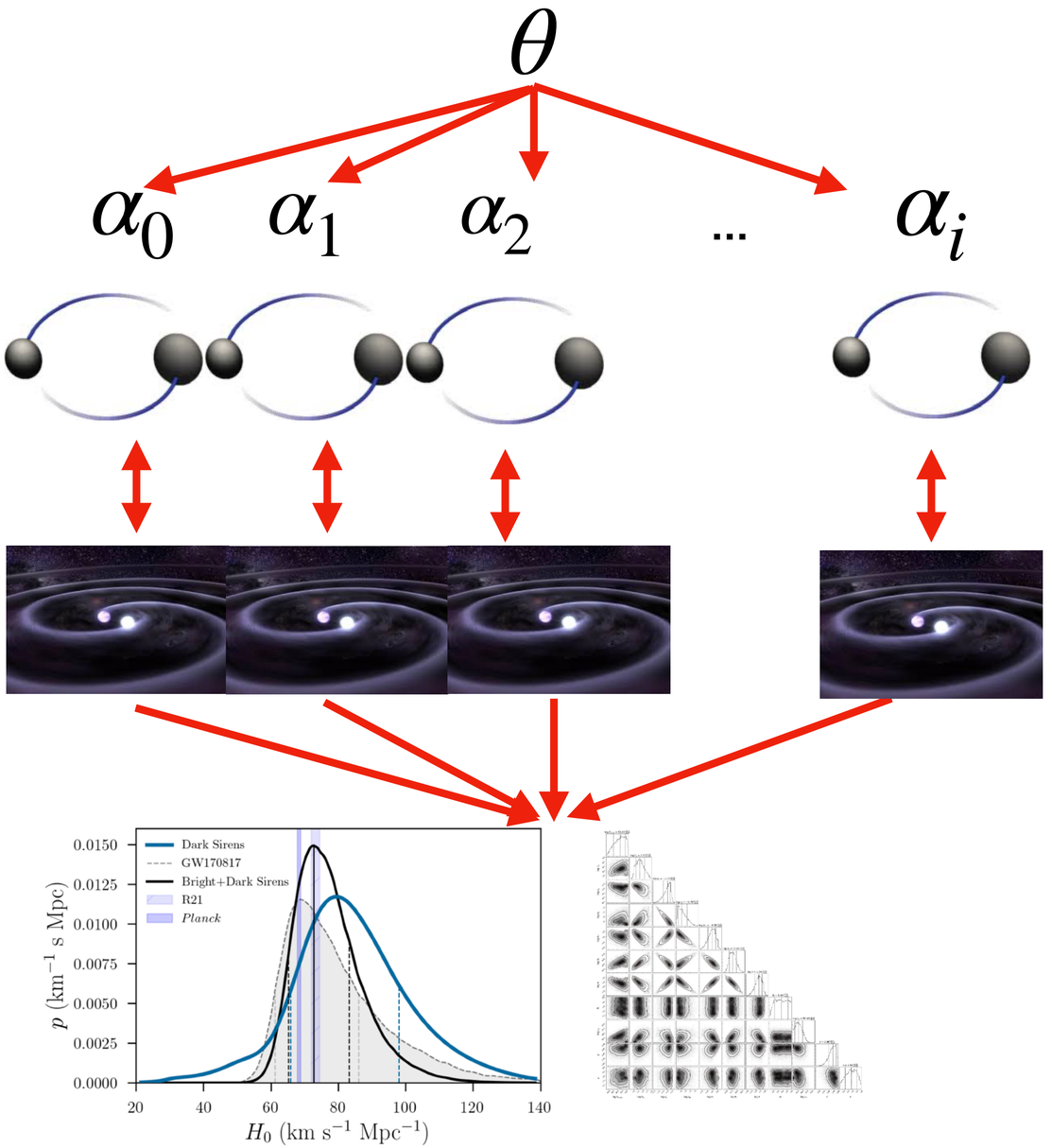


Marginal likelihood for population level inference

- Hierarchical modelling we have population level parameters θ and individual object parameters α
- Propose fitting each object (usually do this anyway) to get $P_i(\theta, \alpha \,|\, D_i, M)$
- Using margarine to evaluate $P_i(\theta \,|\, D_i, M) \rightarrow L_i(\theta)$
- Then assuming the objects are independent sample the likelihood

$$L(\theta) = \prod_i^{N_{gal}} L_i(\theta)$$

- If $N_\theta = 6$ and $N_\alpha = 5$ then for 10 waves $N_{dim} = 56$ and $T \propto 56^3$

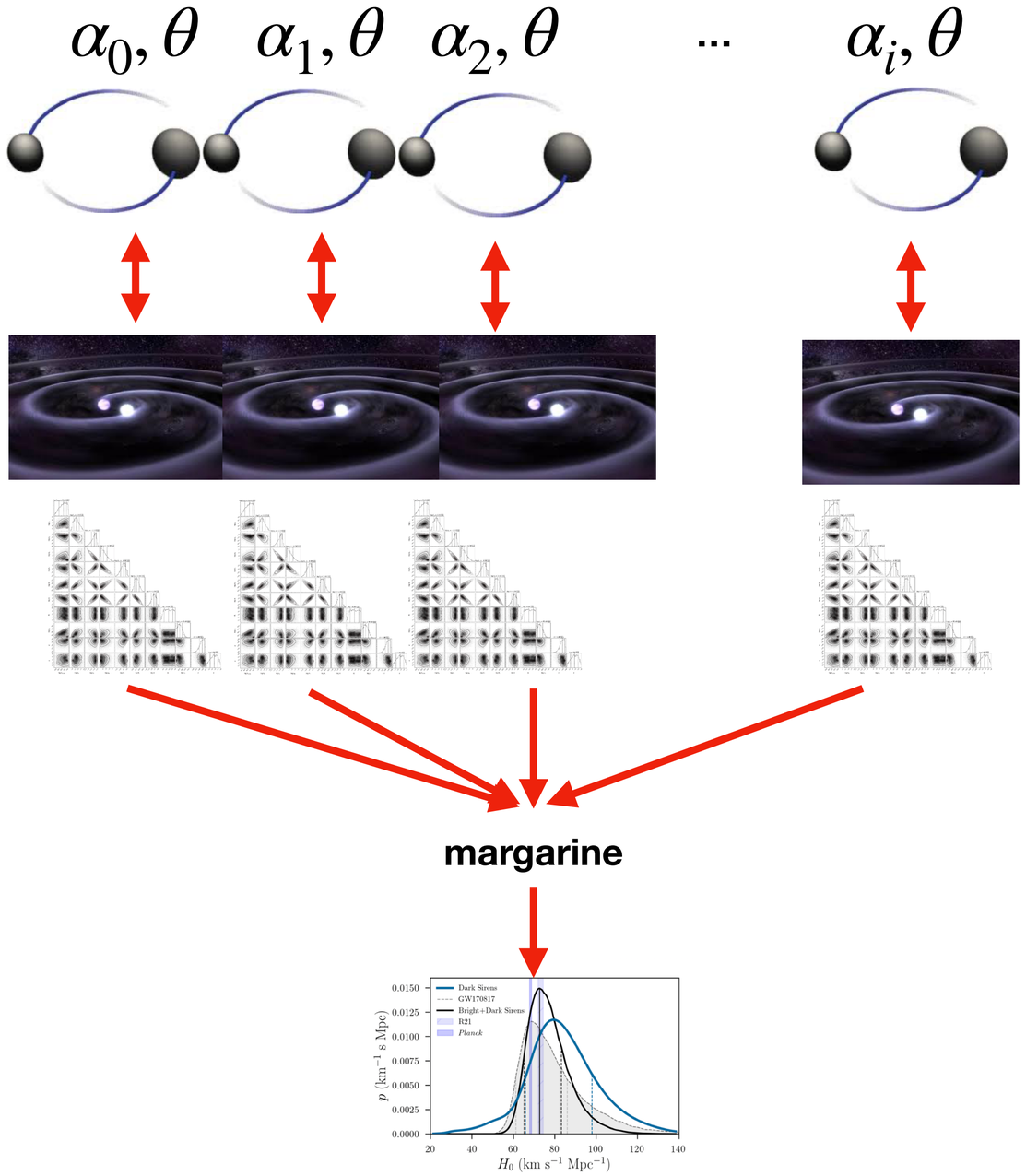


Marginal likelihood for population level inference

- As an example we can think about gravitational wave constraints on H_0 from dark and bright sirens
- For each object we can derive $P_i(H_0, M_1, M_2, S_1, S_2, \ldots \mid D_i, M)$
- We can then train flows on $P_i(H_0 \mid D_i, M)$ and $\pi(H_0)$ to evaluate $L_i(H_0)$
- And combine many observations to get

$$L(H_0) = \prod_i^{N_{GW}} L_i(H_0) \rightarrow P(H_0)$$

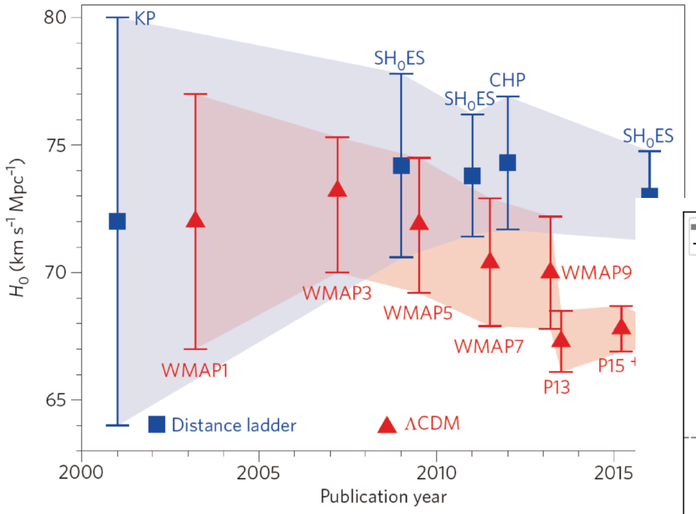
- $T \propto 10 \times 11^3 + 6^3 < 56^3$

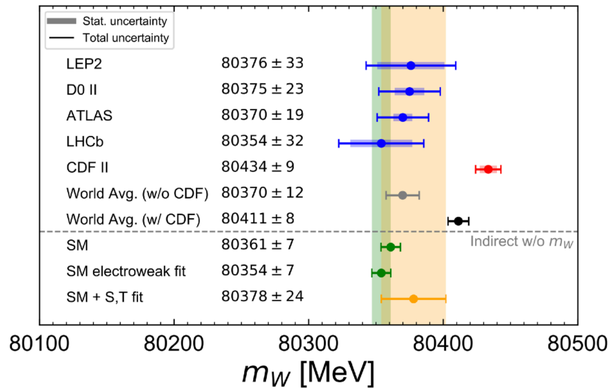


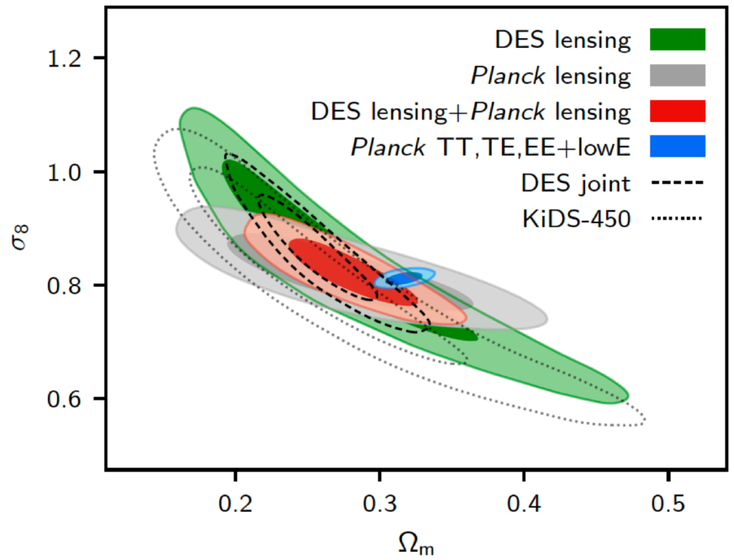
Tension Statistics

Why are we interested in tension?

- Important to be able to independently observe and confirm experimental results
- When two experiments give different results we call this a tension
- Understanding where tension comes from can lead us to new physics and a better understanding of our instruments







[2204.0526]
[1905.11122]
[1807.06209]

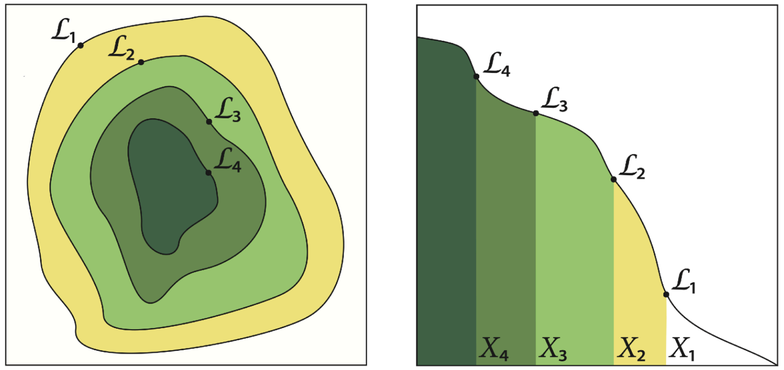
Quantifying tension

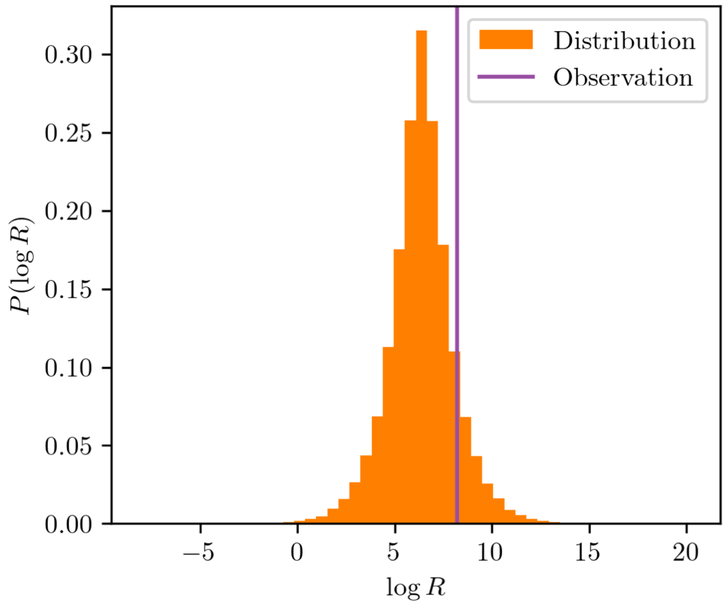
- Parameter differences, goodness of fit degradation, suspiciousness (see 2012.09554 for a review)

- Here, interested in evidence ratio

$$R = \frac{P(D_A, D_B)}{P(D_A)P(D_B)} = \frac{Z_{AB}}{Z_A Z_B}$$

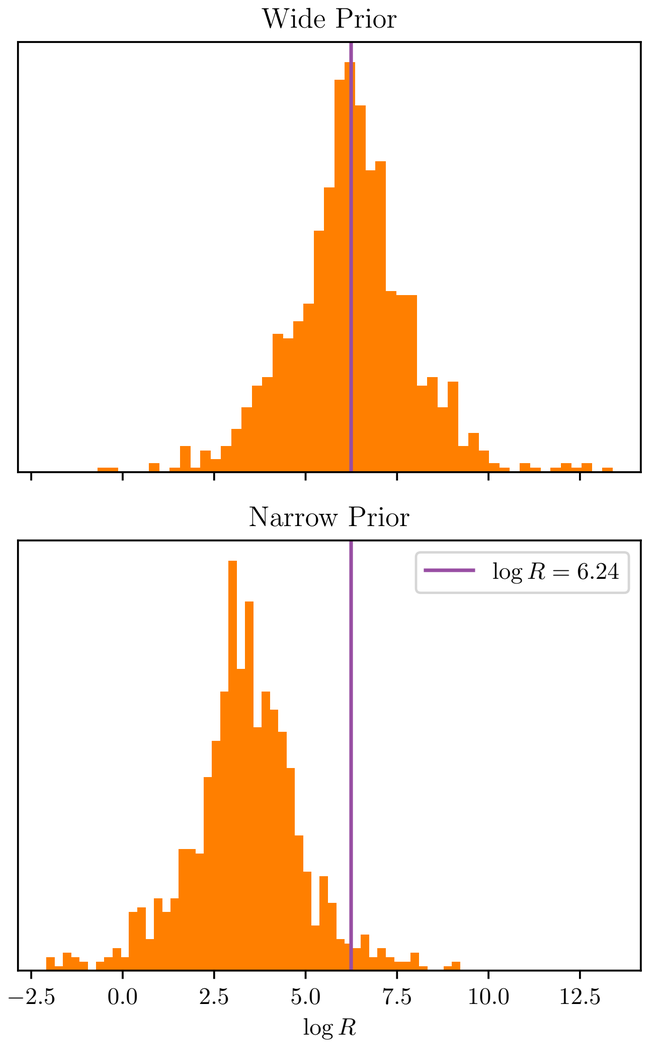
- For any pair of experiments, model and prior there is a distribution of in concordance R values





The issues with R

- R is dimensionally consistent, parameterisation invariant and symmetric
- However it has a strong prior dependence and its hard to interpret
- Usually say
  - $R \gg 1 \rightarrow$ in concordance
  - $R \ll 1 \rightarrow$ tension



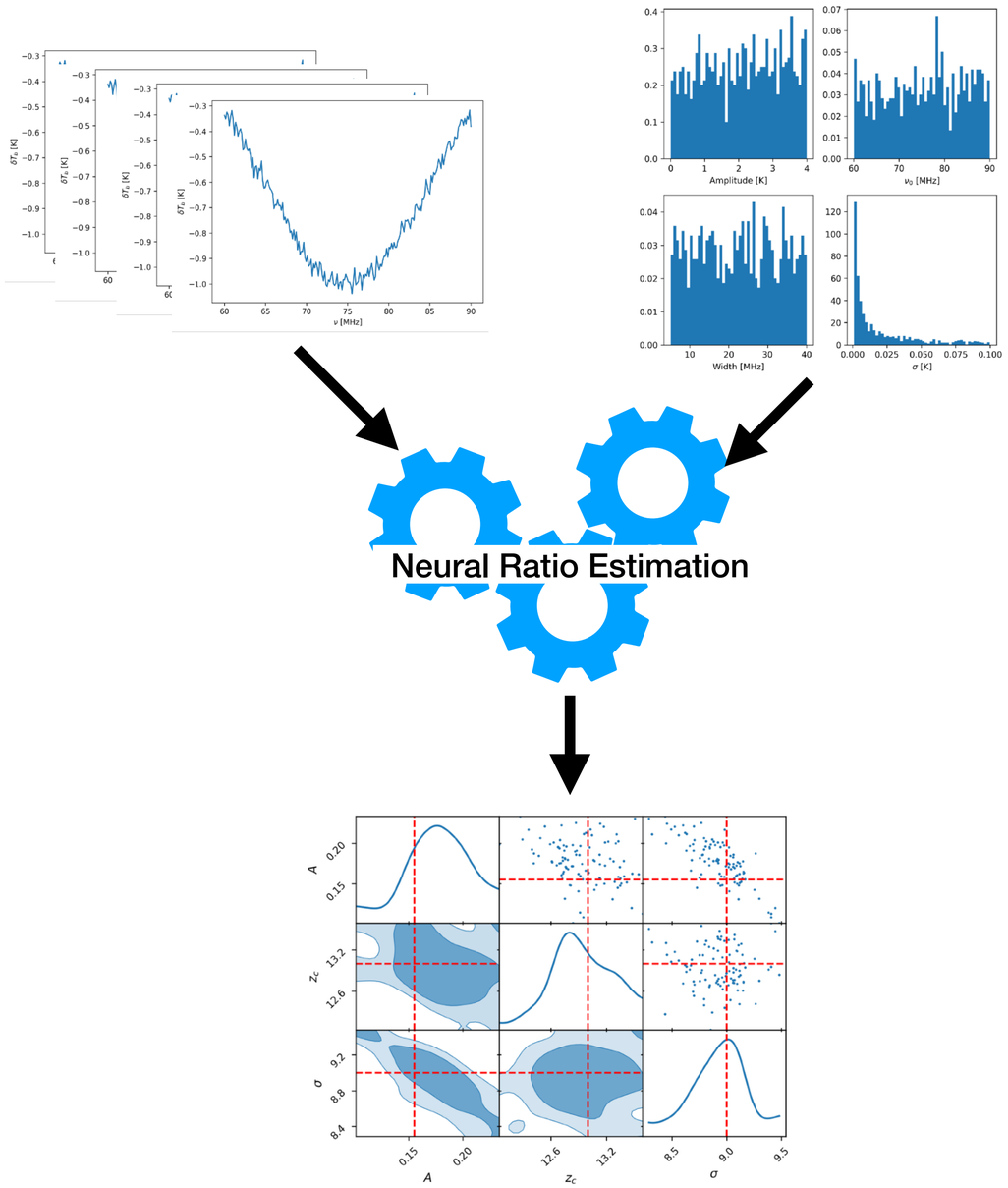
Neural Ratio Estimation

- Essentially just classifiers
- Take in two inputs A and B and estimate the probability that they are drawn from joint distribution vs independent

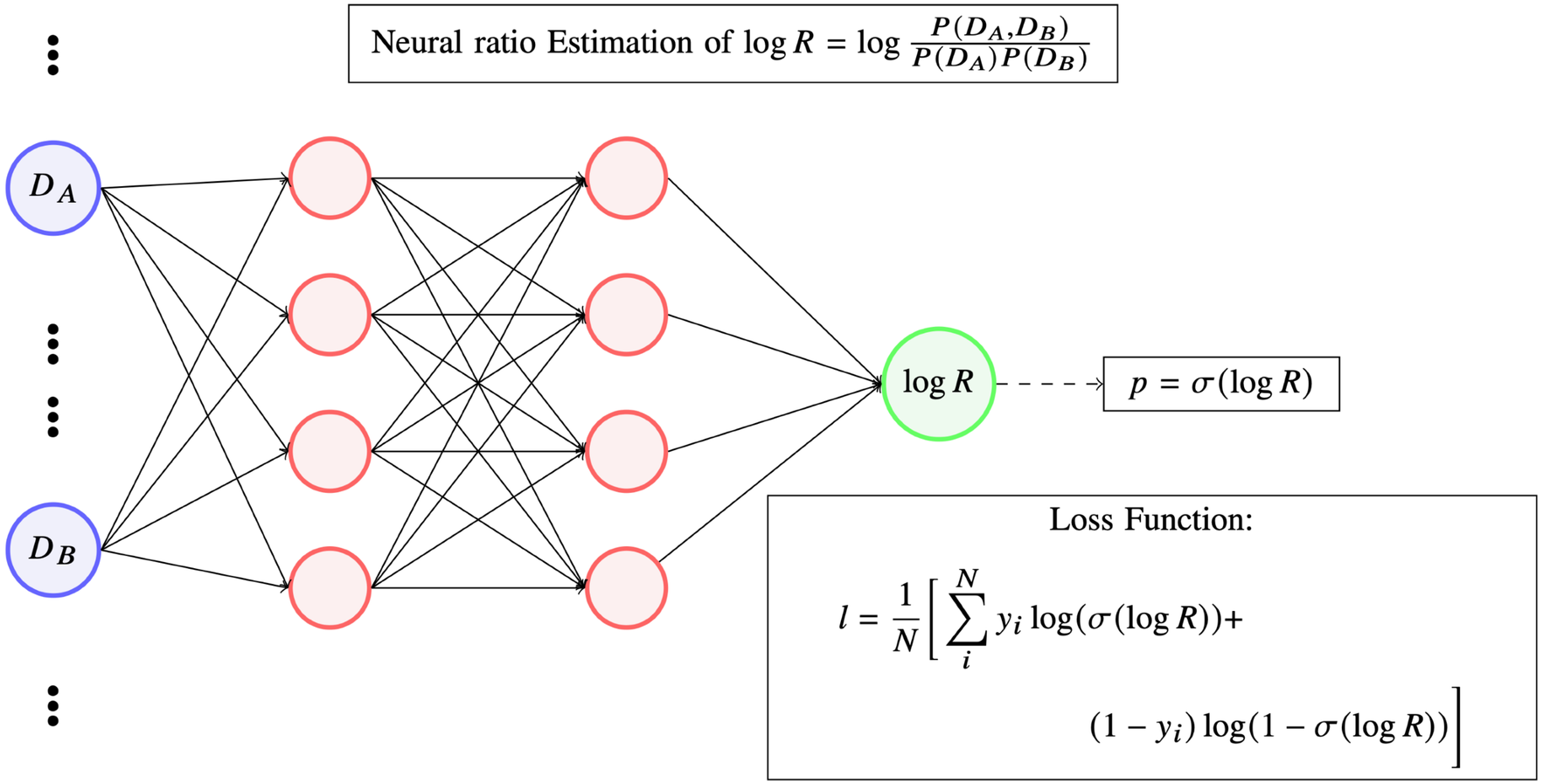
$$r = \frac{P(A, B)}{P(A)P(B)}$$

- Used for parameter inference

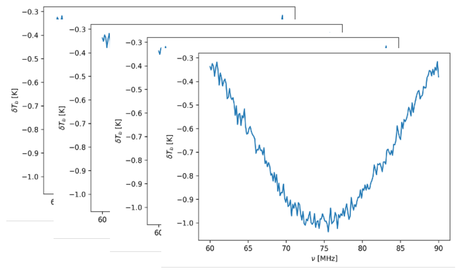
$$r = \frac{P(D, \theta)}{P(D)P(\theta)} = \frac{P(D \mid \theta)}{P(D)} = \frac{L(\theta)}{Z}$$

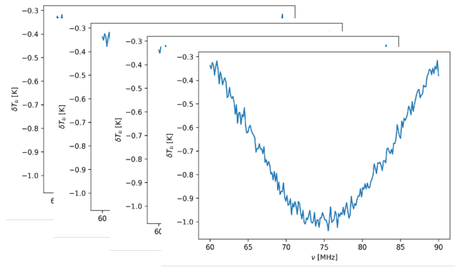


R with NREs

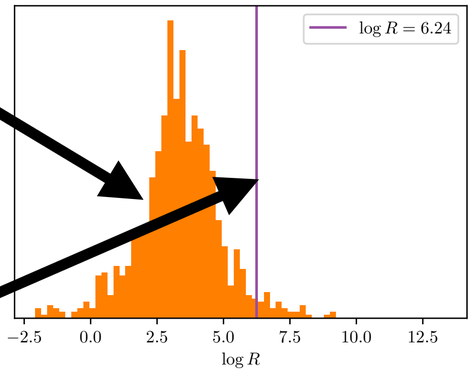

Neural ratio Estimation of $\log R = \log \frac{P(D_A, D_B)}{P(D_A)P(D_B)}$
D_A
D_B
$\log R$
$p = \sigma(\log R)$
Loss Function:
$l = \frac{1}{N}\left[\sum_i^N y_i \log(\sigma(\log R)) + (1 - y_i)\log(1 - \sigma(\log R))\right]$

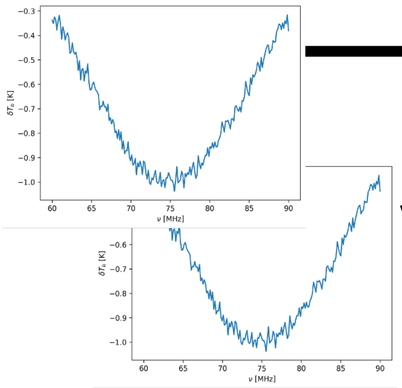
Direct predictions or calibration?

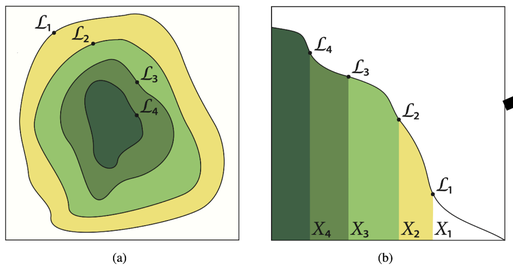




Neural Ratio Estimation

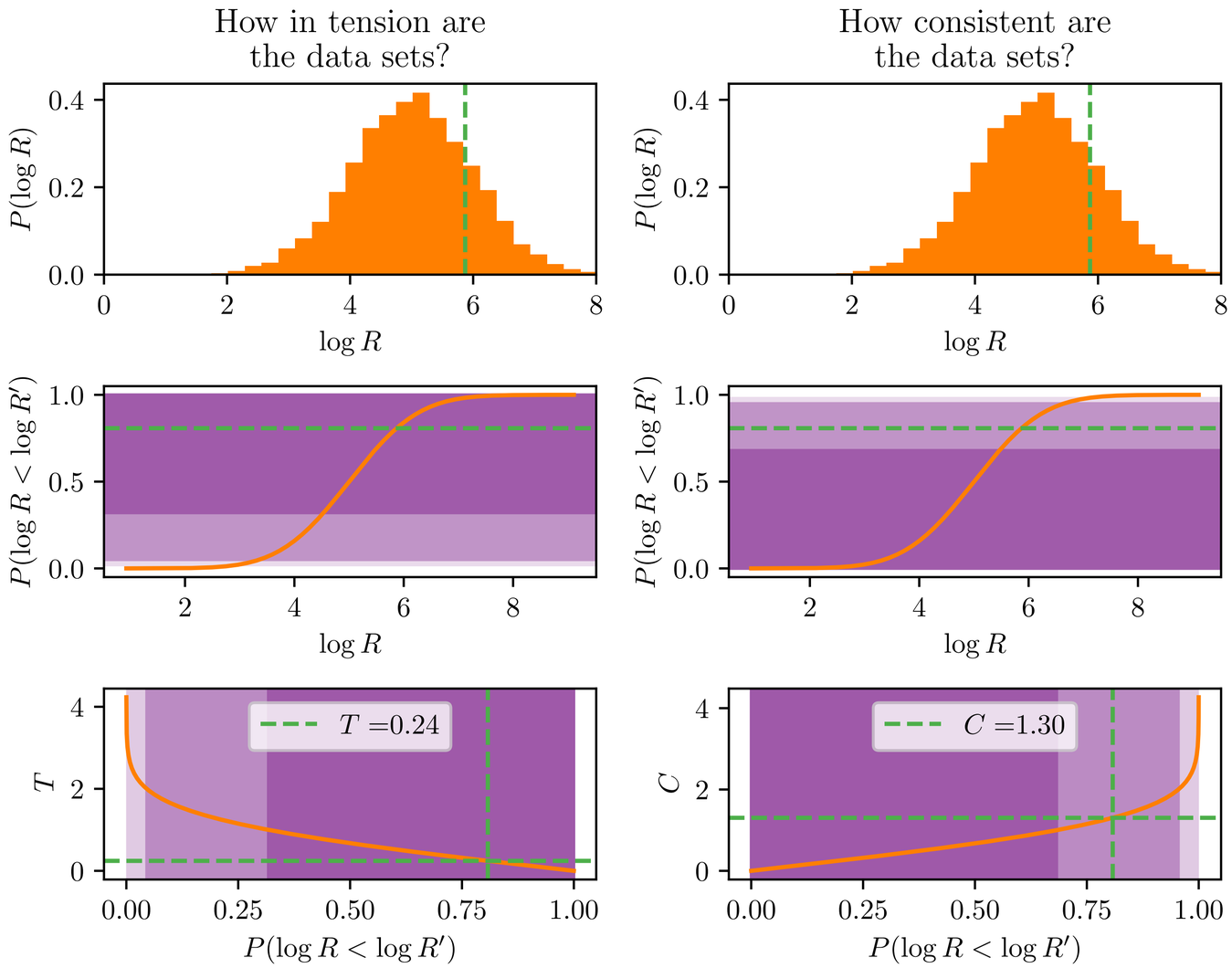






Nested Sampling

Calibration of R



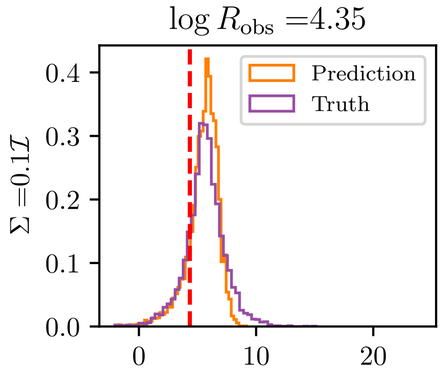
Analytic Example: Prior Dependence

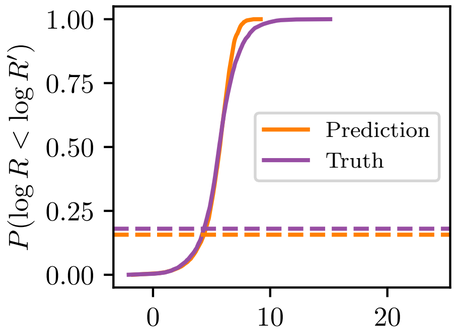
- Define a linear model

$$D_A = M_A\theta + m_A \pm \sqrt{C_A}$$

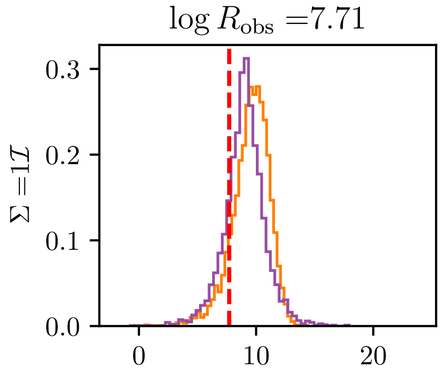
$$D_B = M_B\theta + m_B \pm \sqrt{C_B}$$

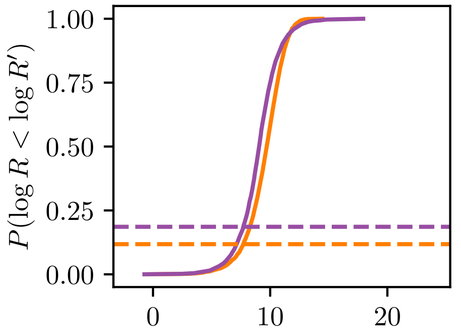
- $n_{dims} = 3, n_{data} = 50$
- Gaussian prior and likelihood



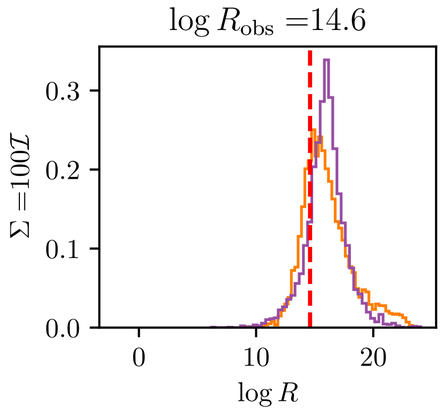


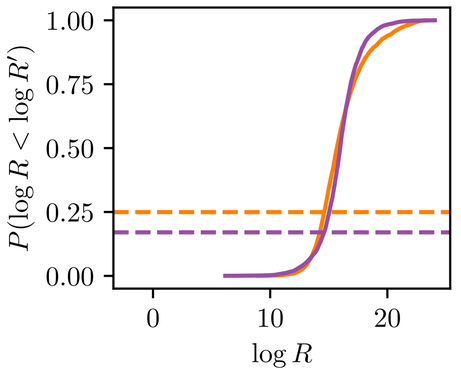
	T	C
Truth	1.340	0.228
TENSIONNET	1.416	0.198





	T	C
Truth	1.323	0.235
TENSIONNET	1.563	0.148

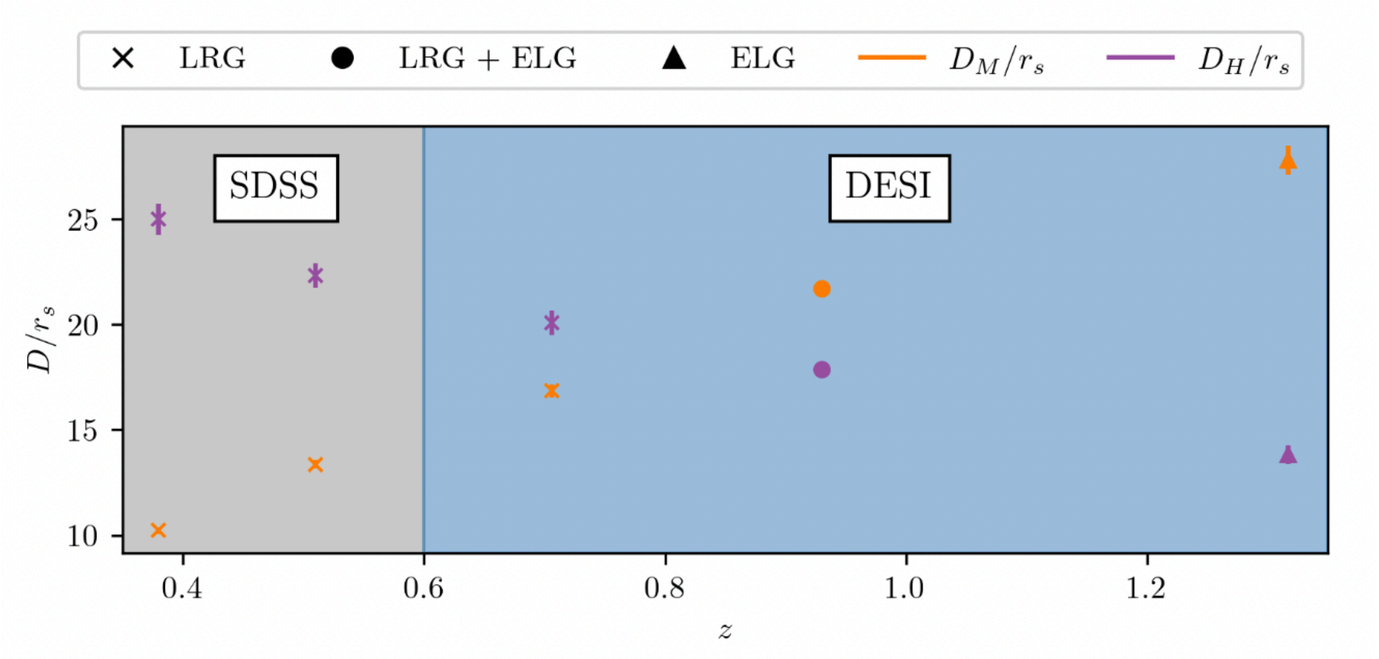




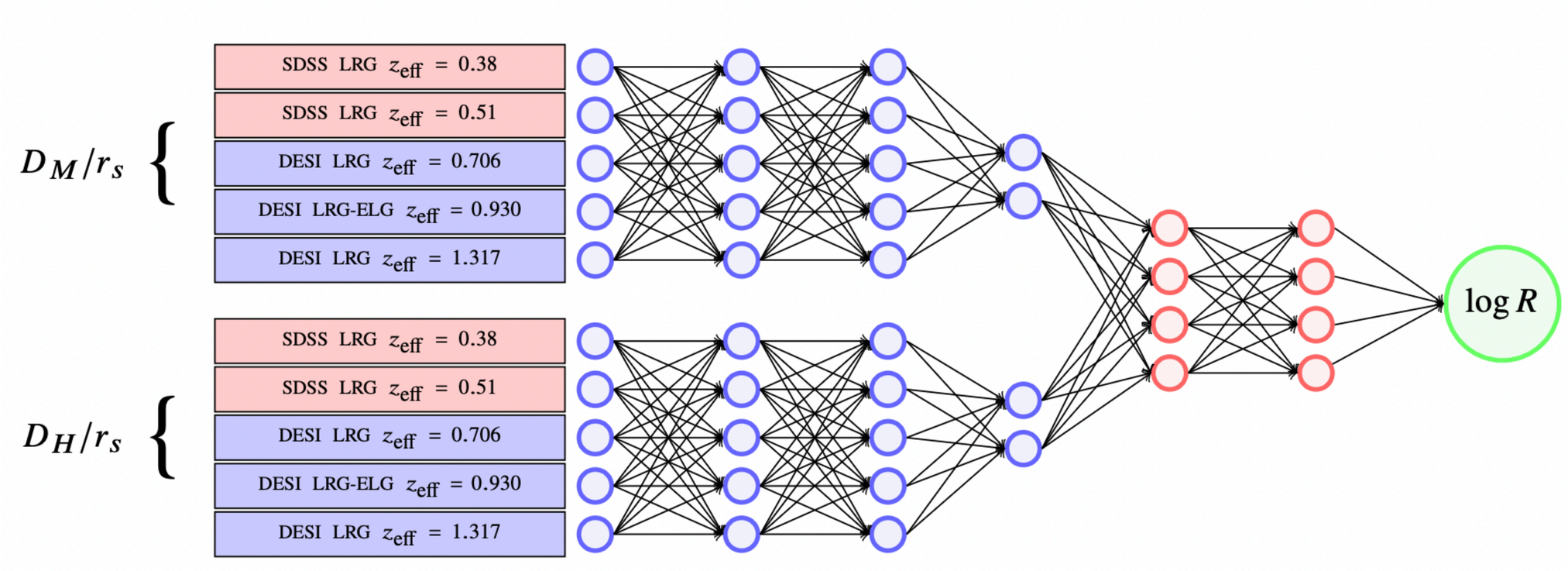
	T	C
Truth	1.371	0.215
TENSIONNET	1.151	0.318

DESI + SDSS: Joint Data Set

- No existing correlated likelihood to evaluate a true $R_{\rm obs}$ with Nested Sampling
- Select different measurements from each survey to maximise the effective volume [e.g. 2404.03002]
- Focusing on LRG and ELG
- Add Quasars and Lyα in the future

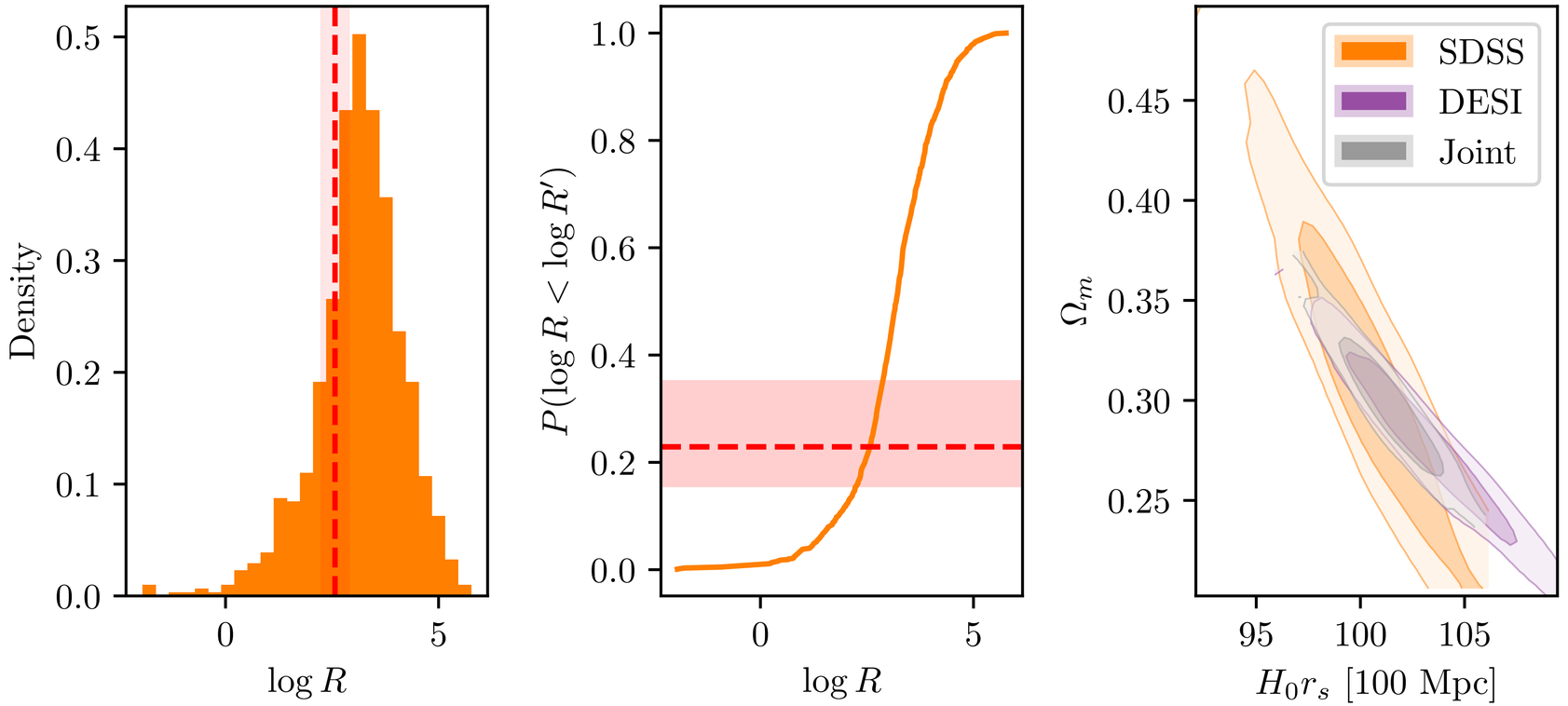


DESI + SDSS: NRE Set Up



DESI + SDSS: Results

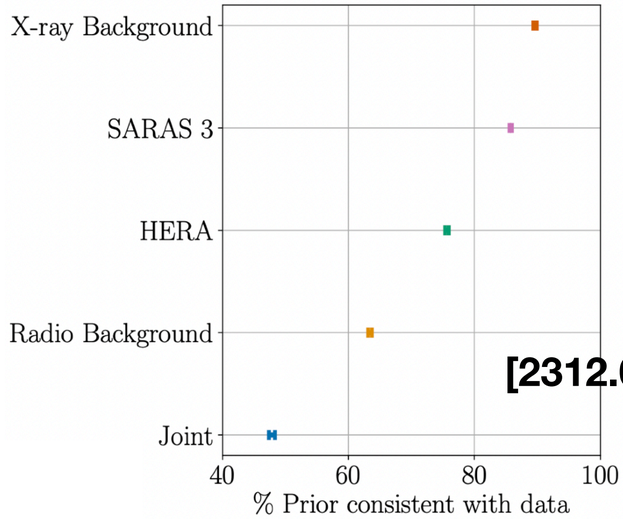
- We find $T = 1.22 \pm 0.20$



Conclusions

What I did not cover

- Posterior repartitioning with normalizing flows and temperature dependent normalizing flows
- Using normalizing flows to calculate marginal bayesian statistics
- Mutual information with neural ratio estimation
- Population level analysis of galaxy surveys with SBI
- Fully Bayesian forecasts [2309.06942]
- Radio cosmology!



[2312.08095]

[2210.07409]

What I did talk about

- Machine learning can offer us a lot when we are trying to interrogate our data
- We can use it to speed up inference and interrogate our data
  - With emulators
  - By marginalising away nuisance parameters
  - Or through simulation based inference
- Papers: arxiv.org/a/bevins_h_1
- Code and Talks: https://github.com/htjb