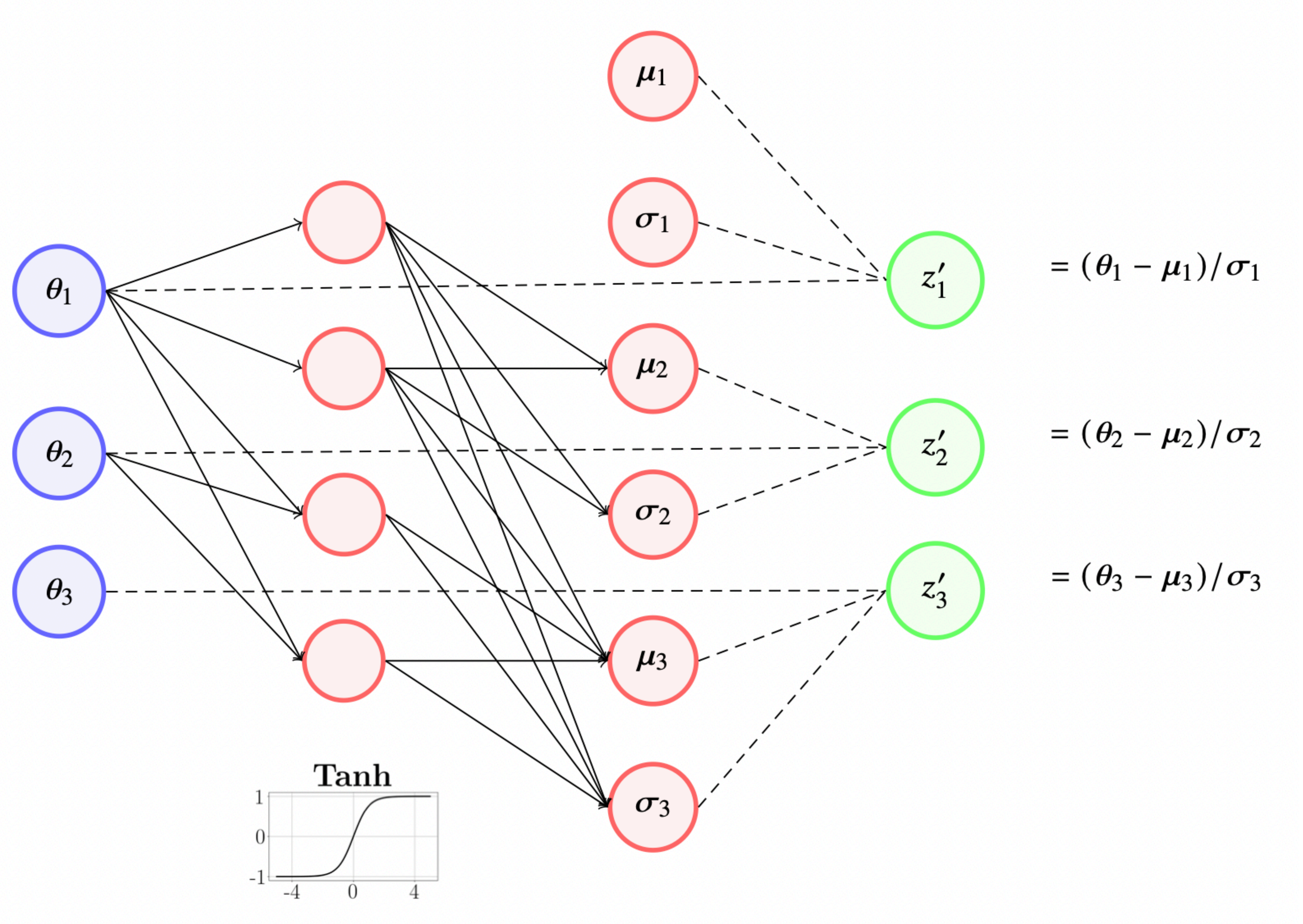
Marginal Bayesian Statistics and Post Processing of Bayesian Constraints

Harry Bevins
With Will Handley + REACH collaboration

Contents

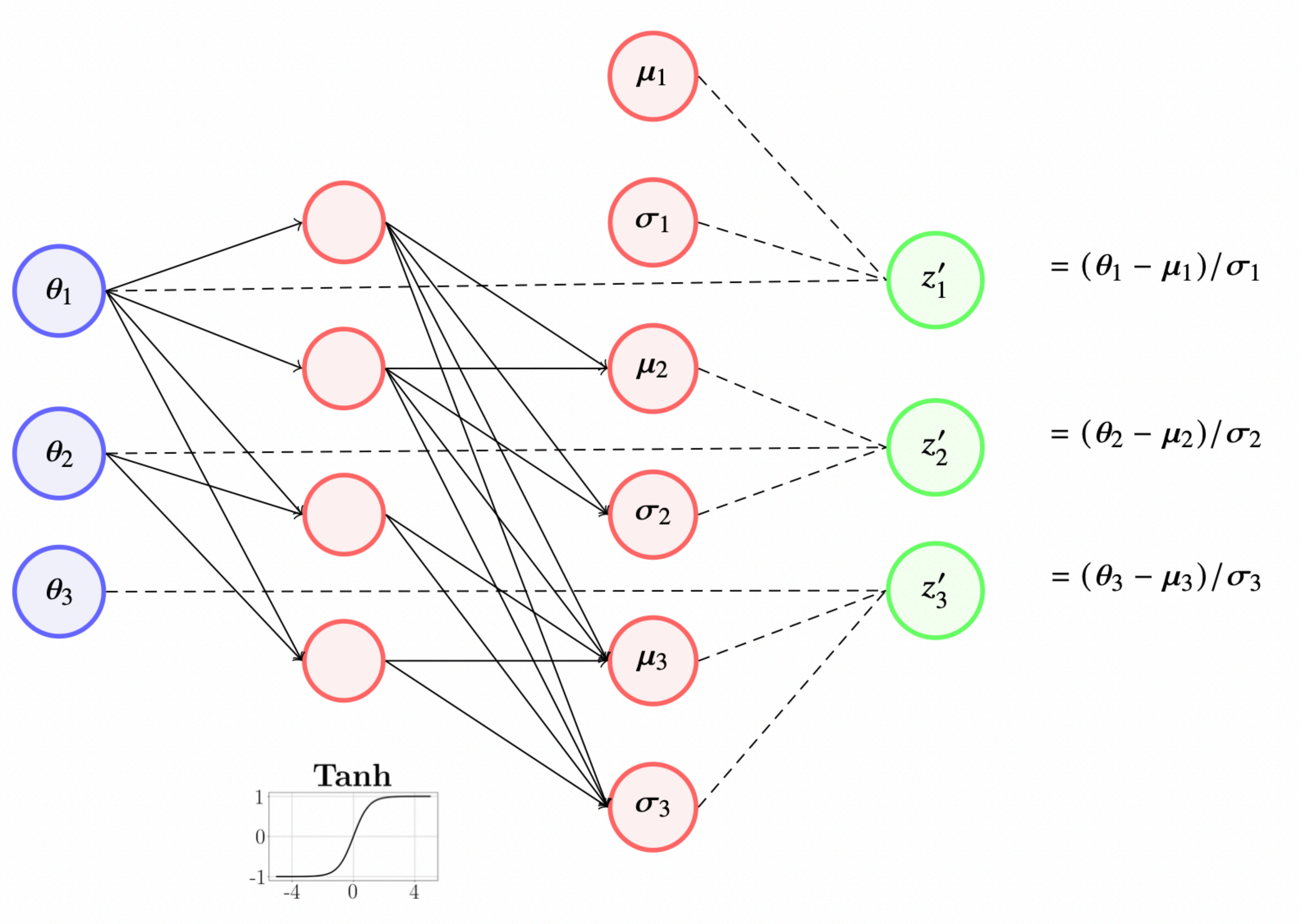
- Brief introduction to normalising flows
- Marginal statistics
- Joint analysis
- Mutual information



Normalising Flows

Normalising Flows (NF)

- Generative density estimation tool
- Learn an invertible transformation f between a known base distribution (e.g. $z \sim \mathcal{N}(0,1)$) and samples on a target distribution
- For us the target distribution corresponds to a posterior or prior
- Many different types of NF

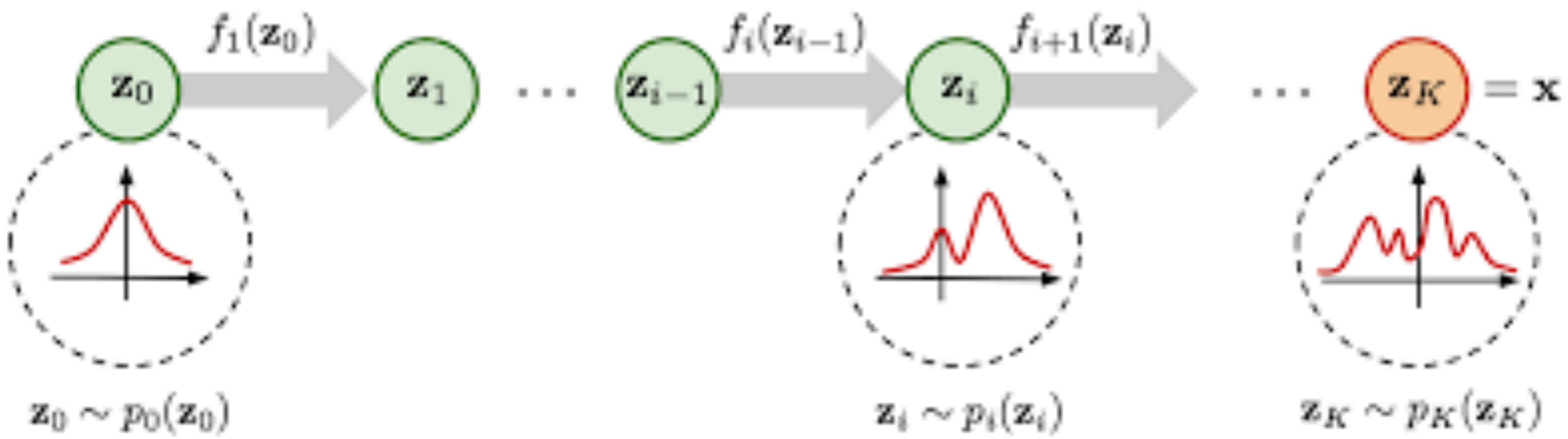


Normalising Flows (NF)

- Once trained we can generate samples from the flows
- And calculate log probabilities by taking gradients across the network and using

$$P(\theta) = \mathcal{N}(f^{-1}(\theta)) \left| det\left(\frac{df^{-1}(\theta)}{d\theta} \right) \right|$$

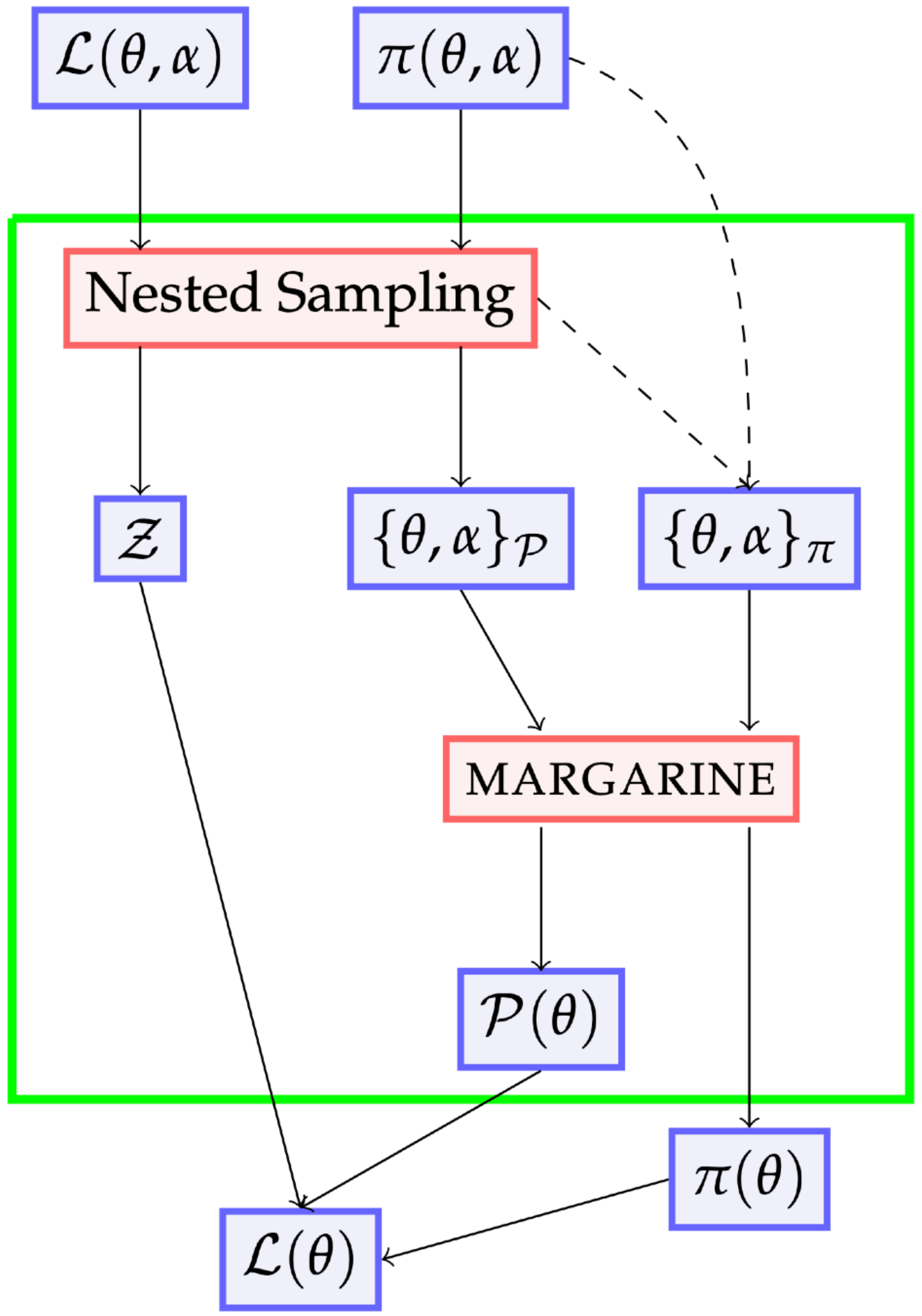
- Where $f^{-1}(\theta_i) \approx z_i$



Marginal Probabilities

$$P(\theta) = \mathcal{N}(f^{-1}(\theta)) \left| det\left(\frac{df^{-1}(\theta)}{d\theta} \right) \right|$$

- Models typically have signal parameters θ and nuisance parameters α
- Not really interested in α
- But given samples on $\Theta = \{\theta, \alpha\}$ we can use NFs to get access to $P(\theta)$

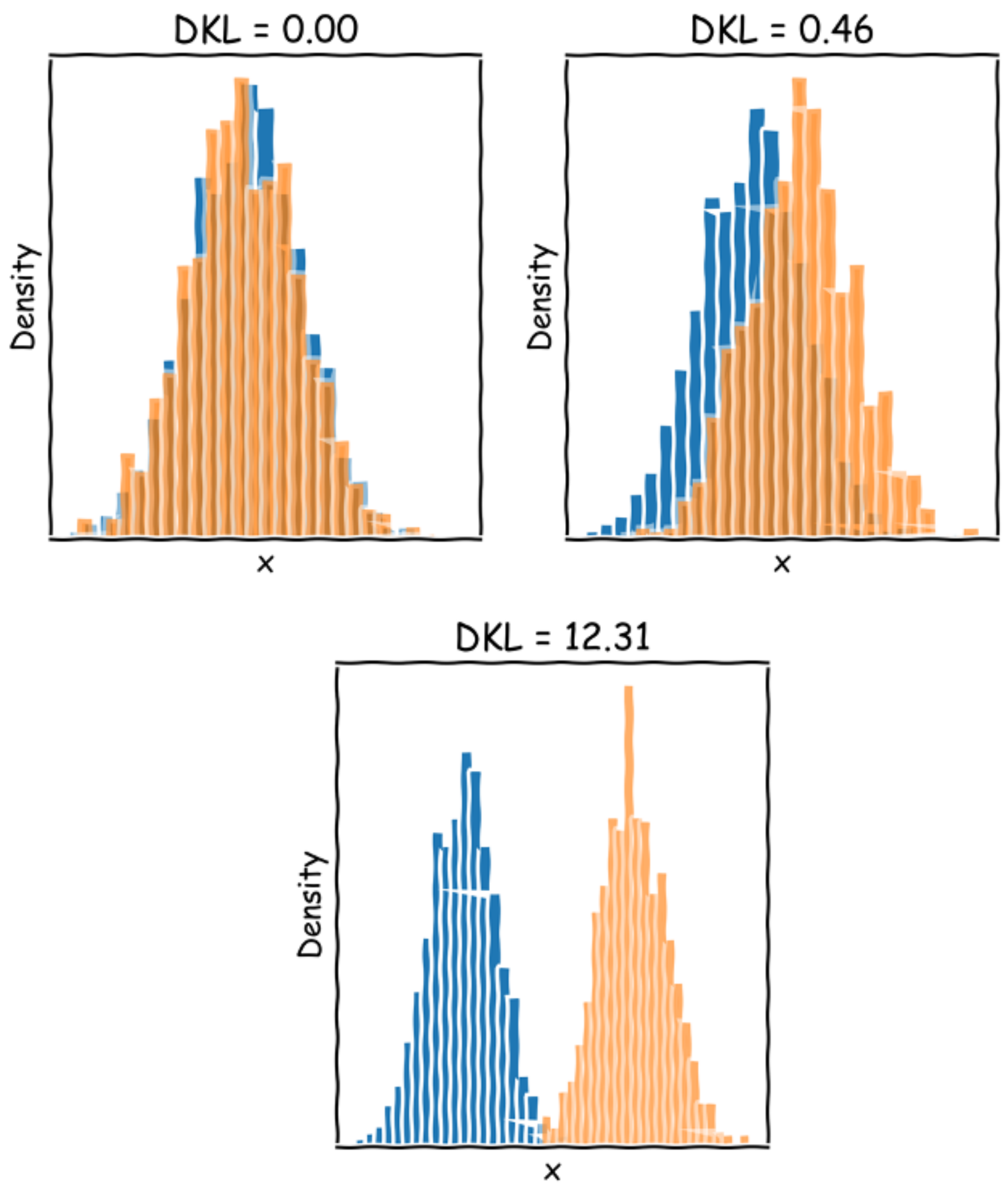


Marginal Statistics

How do we estimate information gain?

- Typically we are interested in how informative an experiment is
- In a Bayesian sense this corresponds to the information gain between the prior and the posterior
- We can measure this with the Kullback-Leibler Divergence

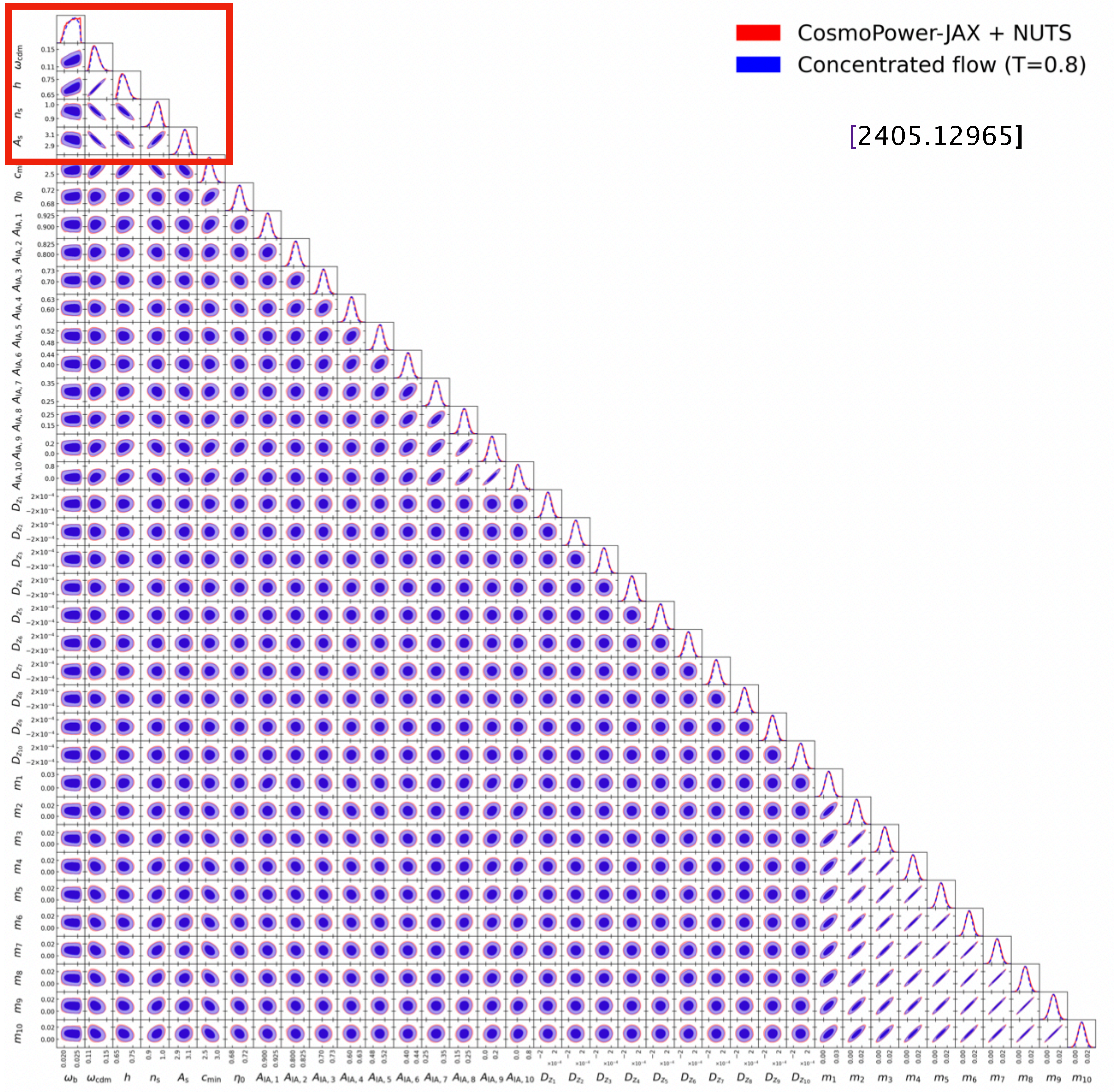
$$D_{KL}(P\,||\,\pi) = \int P(\Theta)\log\frac{P(\Theta)}{\pi(\Theta)}d\Theta$$



Comparing KL divergences

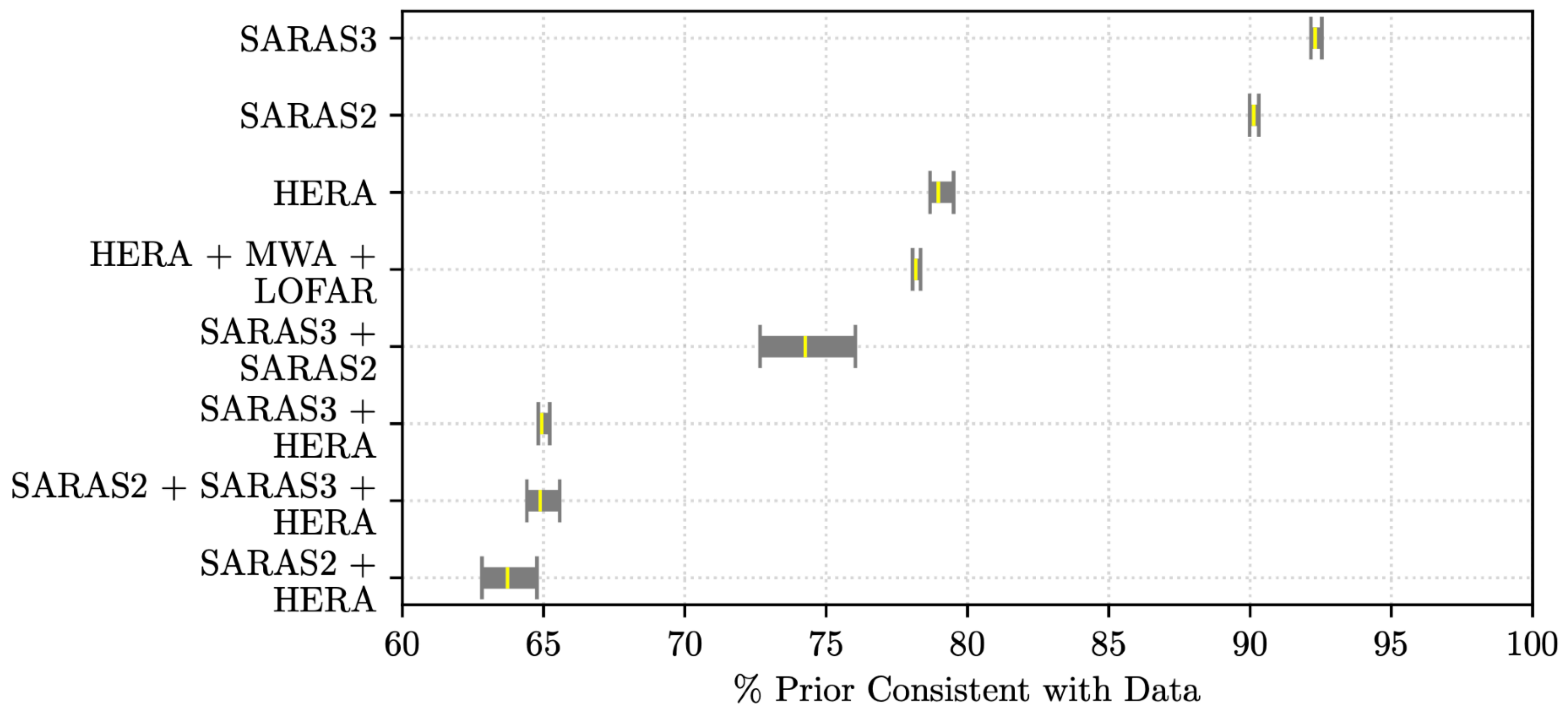
$$D_{KL}(P||\pi) = \int P(\Theta)\log\frac{P(\Theta)}{\pi(\Theta)}d\Theta$$

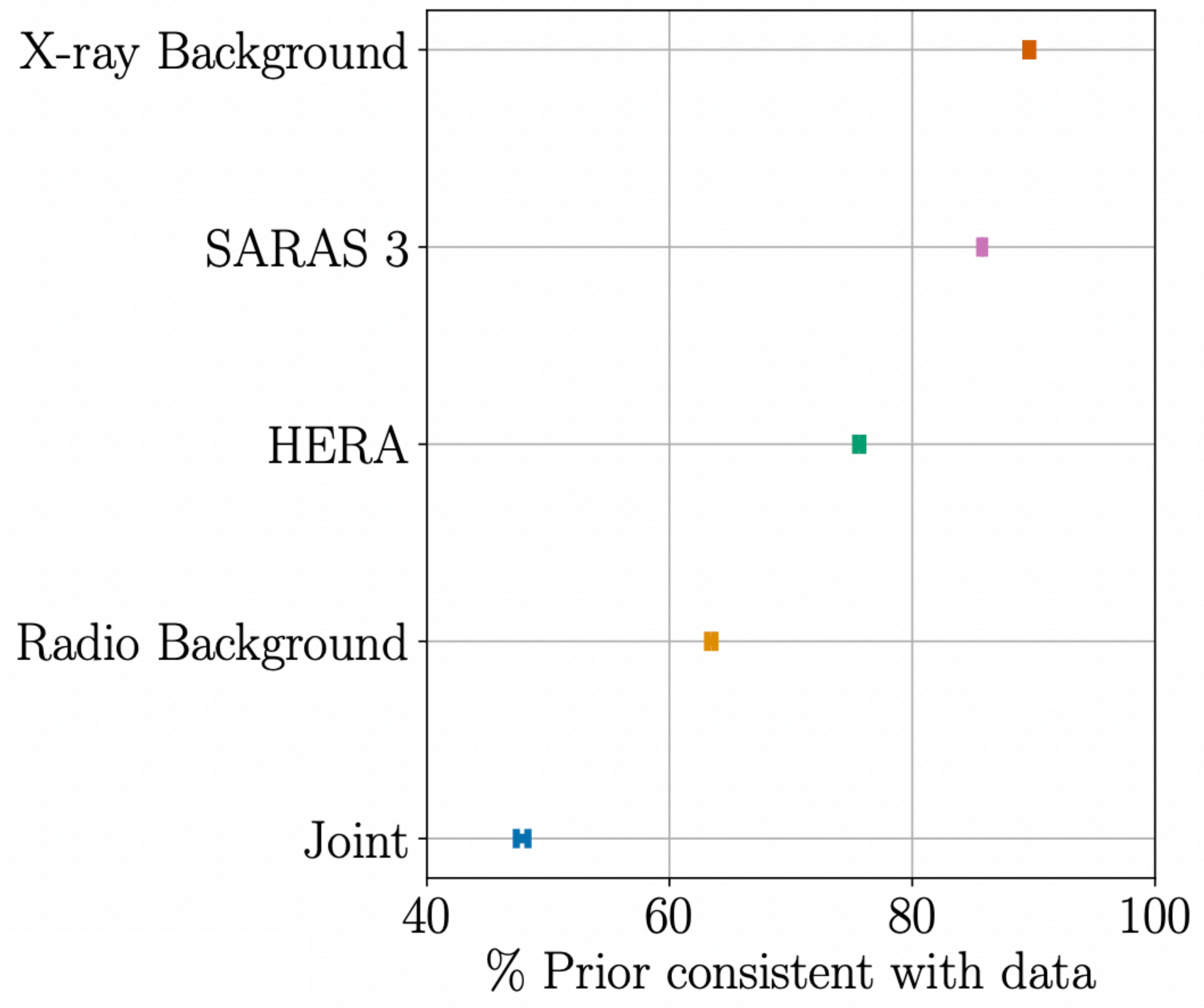
- What if we want to compare the information gain from different experiments?
- Well these experiments often have different sets of nuisance parameters so $\Theta = \{\theta, \alpha\}$
- Only really interested in the constraining power on the common parameters so we need $P(\theta)$



Why is this useful?

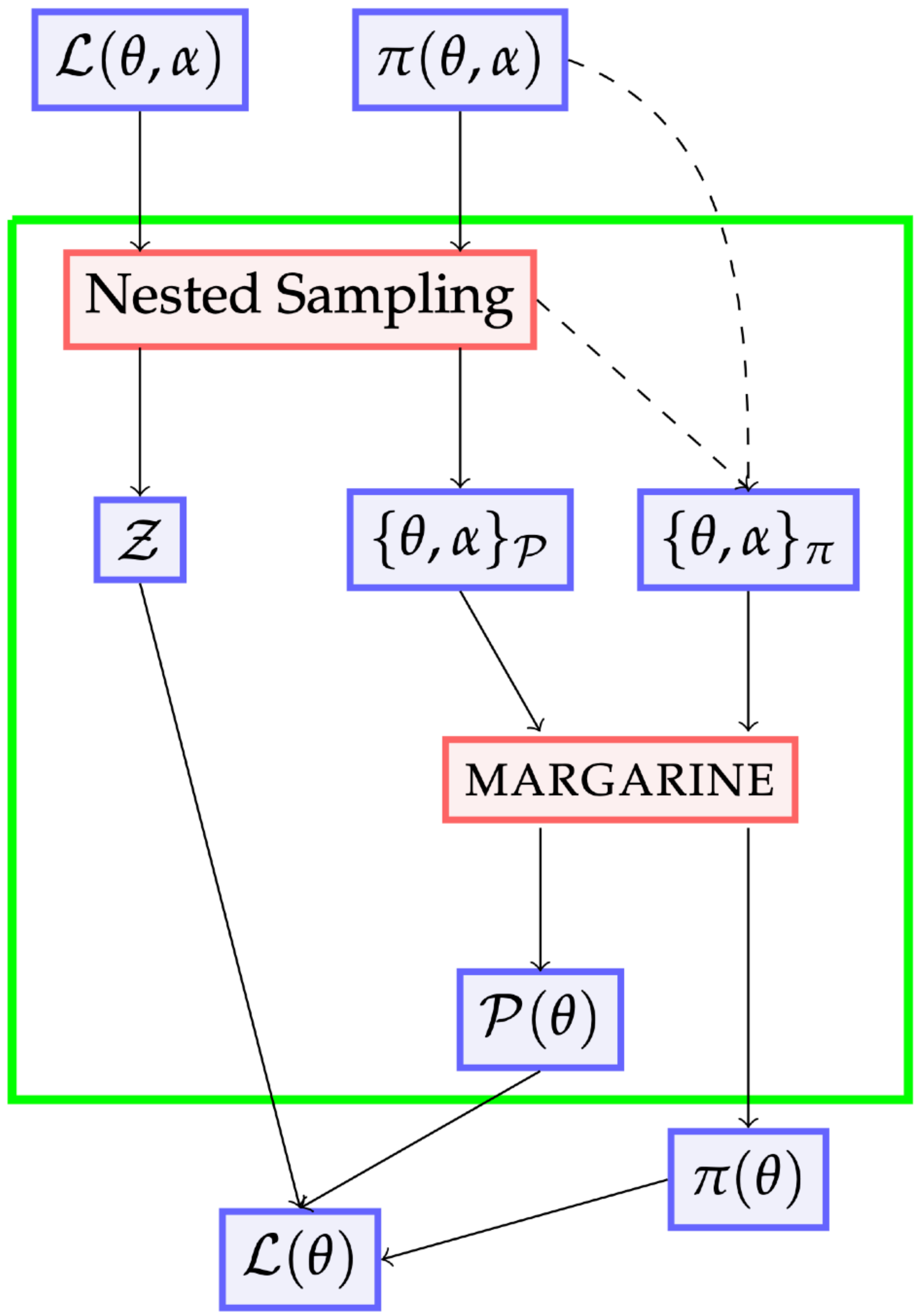
- Use Normalising Flows to marginalise out the nuisance parameters and estimate $P(\theta)$
- Allows us to compare different experimental approaches
- Use this to inform experimental design



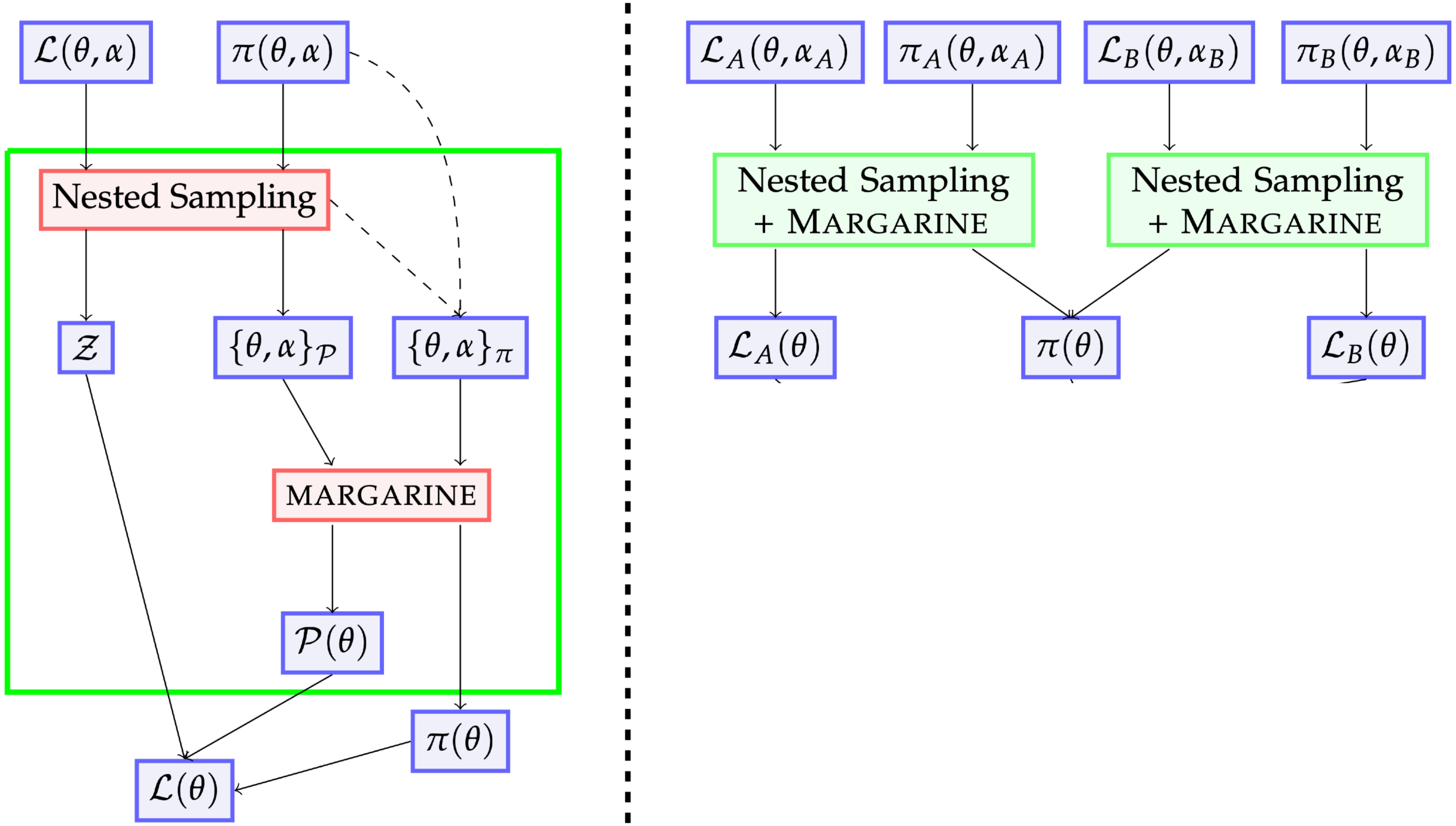


Joint Analysis with NFs

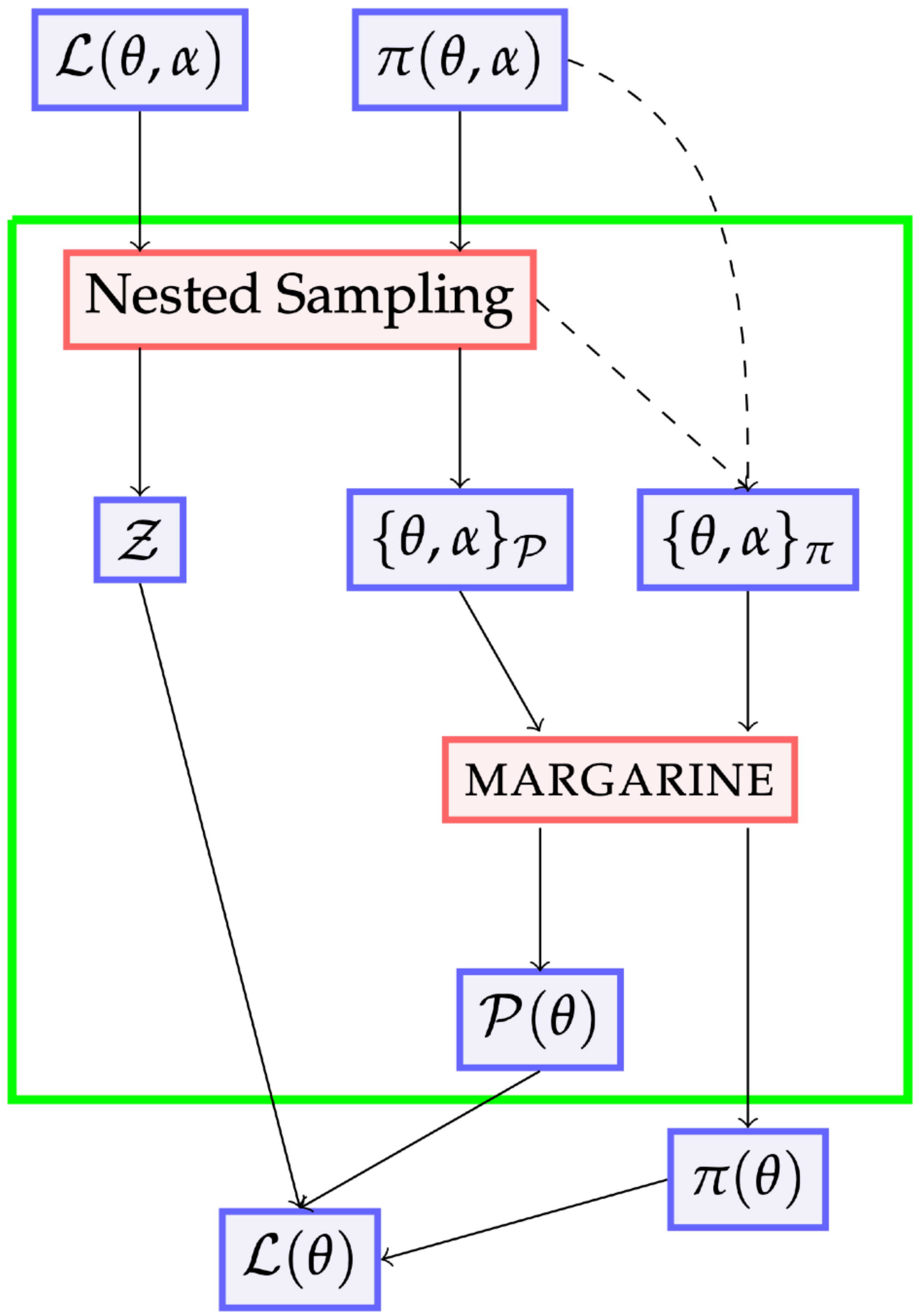
Efficiently joint analysis

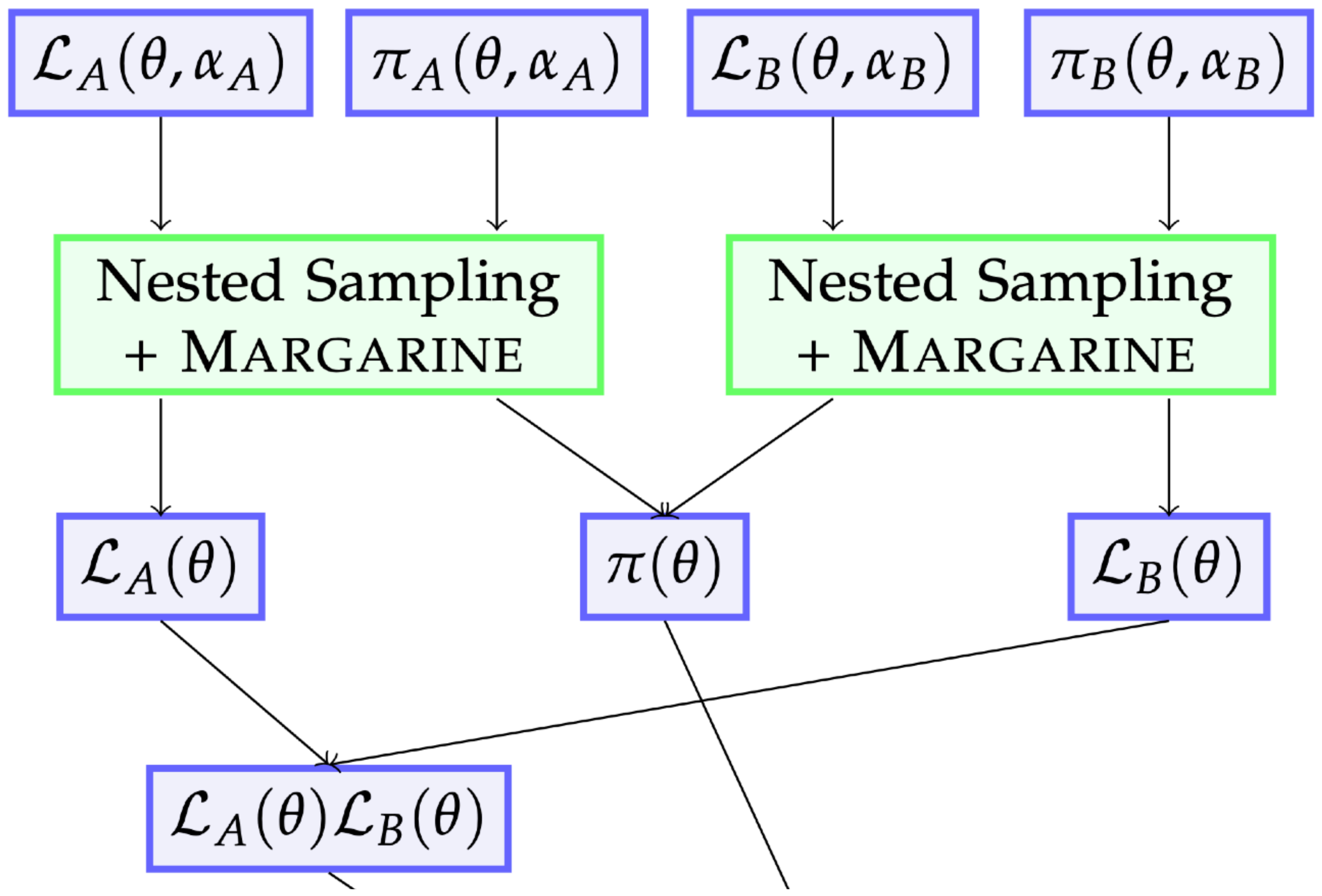


Efficiently joint analysis

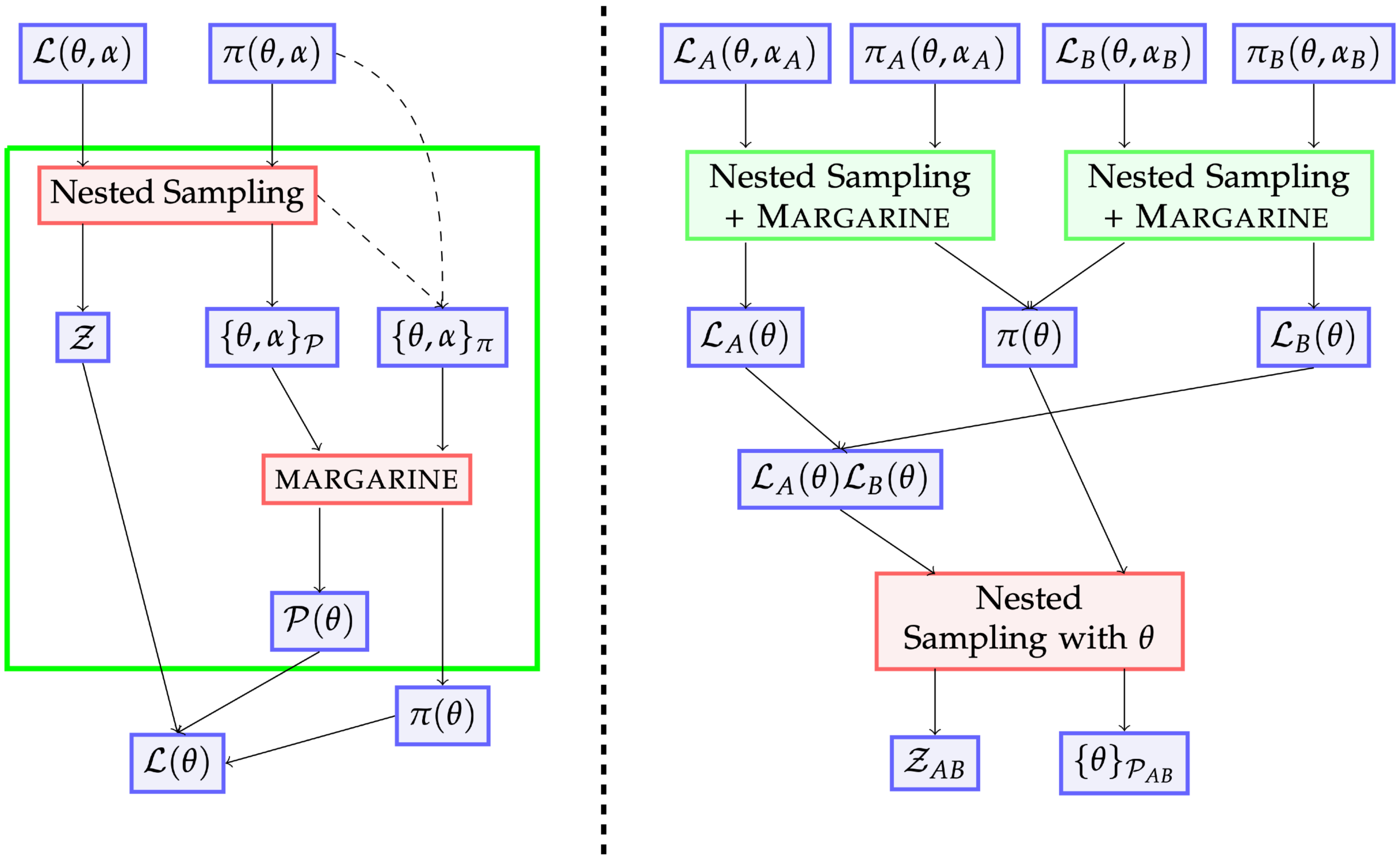


Efficiently joint analysis



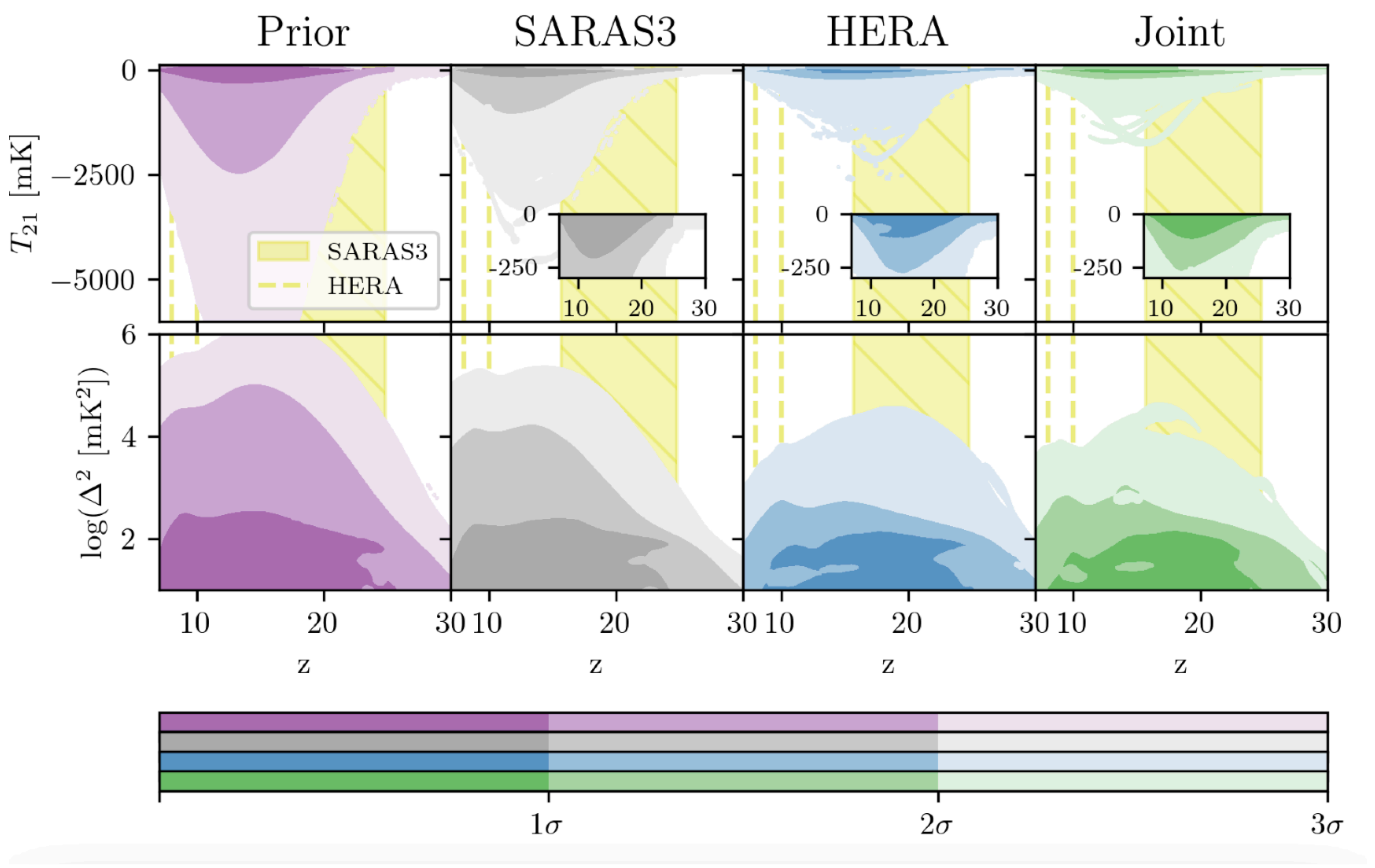


Efficiently joint analysis



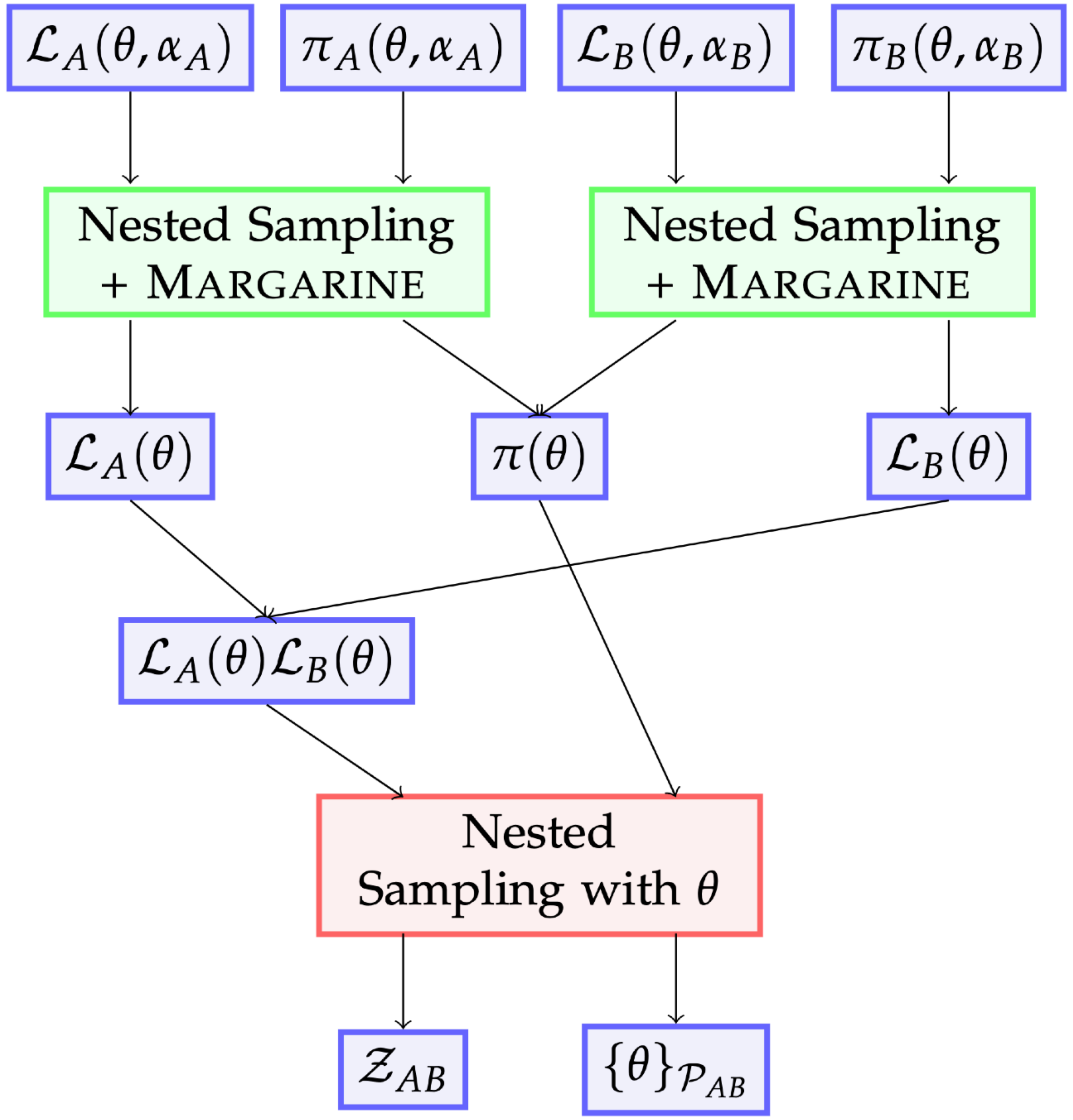
Example 1: Efficiently combining SARAS and HERA

- So using normalising flows we can emulate $P(\theta)$, $\pi(\theta)$ and $L(\theta)$ for a given θ marginalising over α
- $L_{SARAS+HERA}(\theta) = L_{SARAS}(\theta)L_{HERA}(\theta)$
- Reduce dimensionality significantly
- For Nested Sampling the runtime scales as $T \propto d^3$
- Correct evidence and no double counting of priors
- Work continued by Simon Pochinda [2312.08095] and Thomas Gessy-Jones [2312.08828]



Example 2: Time dependent effects in 21cm

- Work led by Joe Pattison [2408.06012]
- Showed that we can combine observations from different time bins with different soil properties using ***margarine***
- Basic idea is to fit each time bin with a different beam based on our understanding of the soil at that time
- Then combine the constraints from the different fits



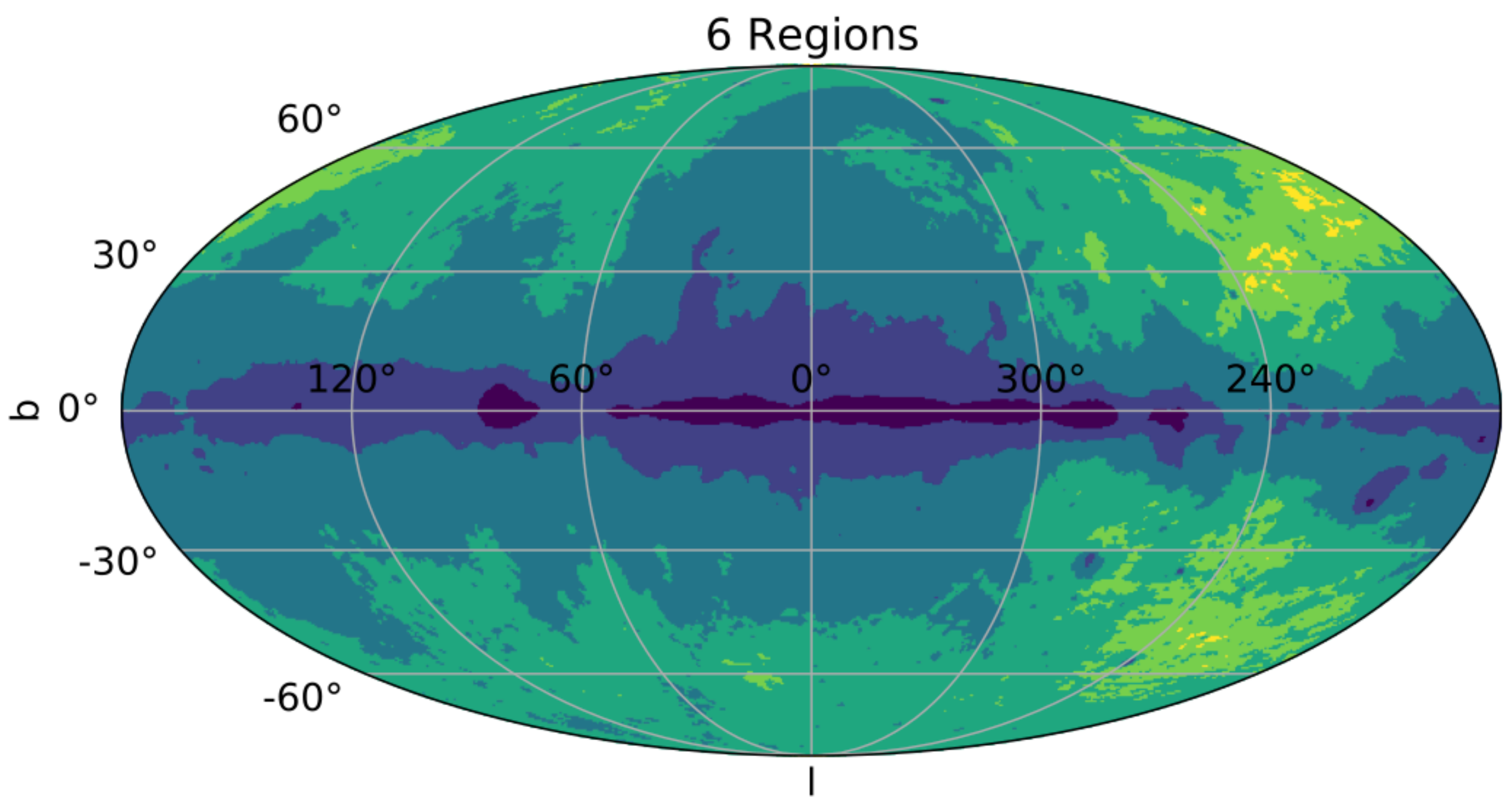
Mutual Information

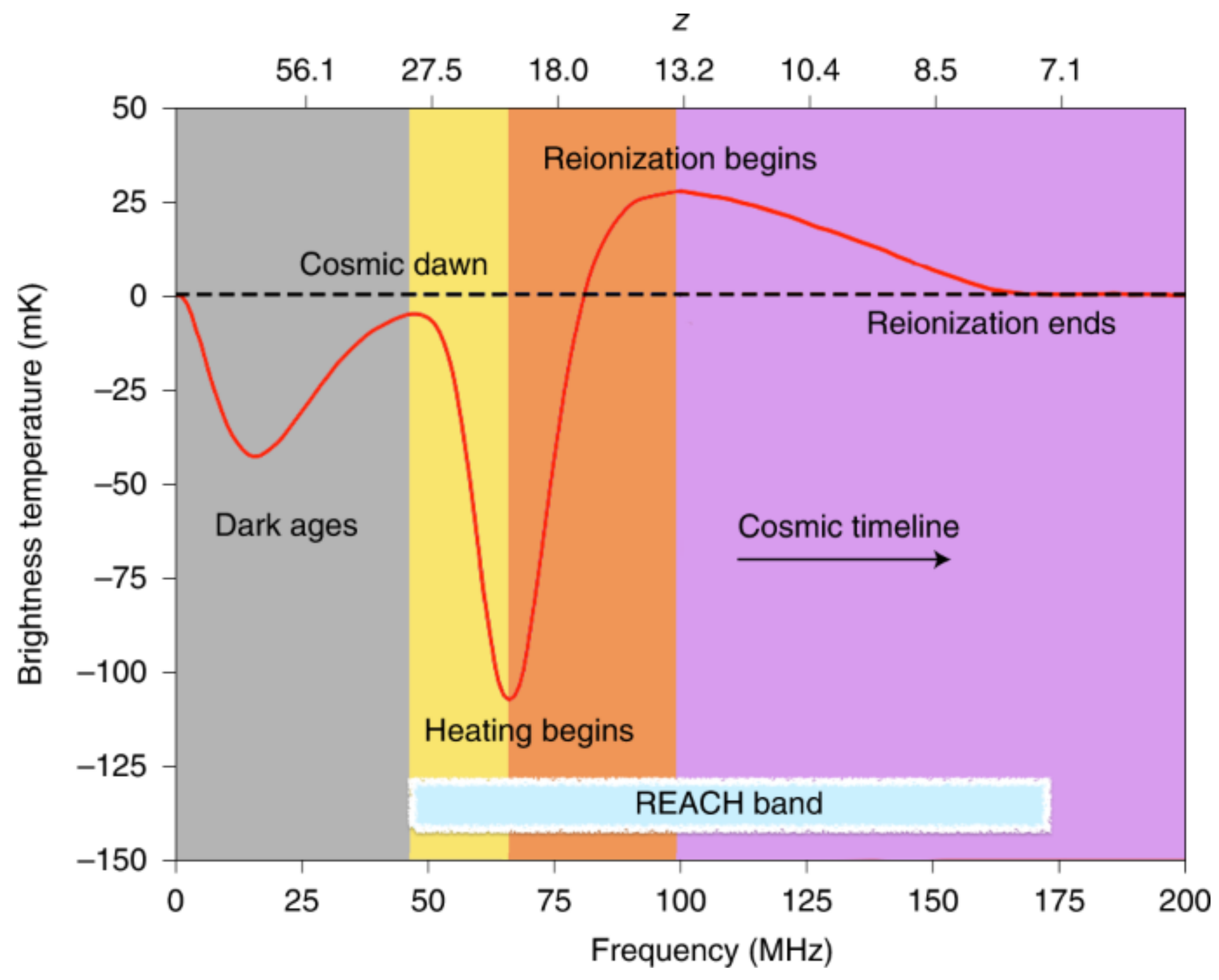
Mutual Information

- Measure of the dependence between two different sets of parameters

$$MI = \left\langle \log \frac{P(\theta, \phi)}{P(\theta)P(\phi)} \right\rangle_{P(\theta,\phi)}$$

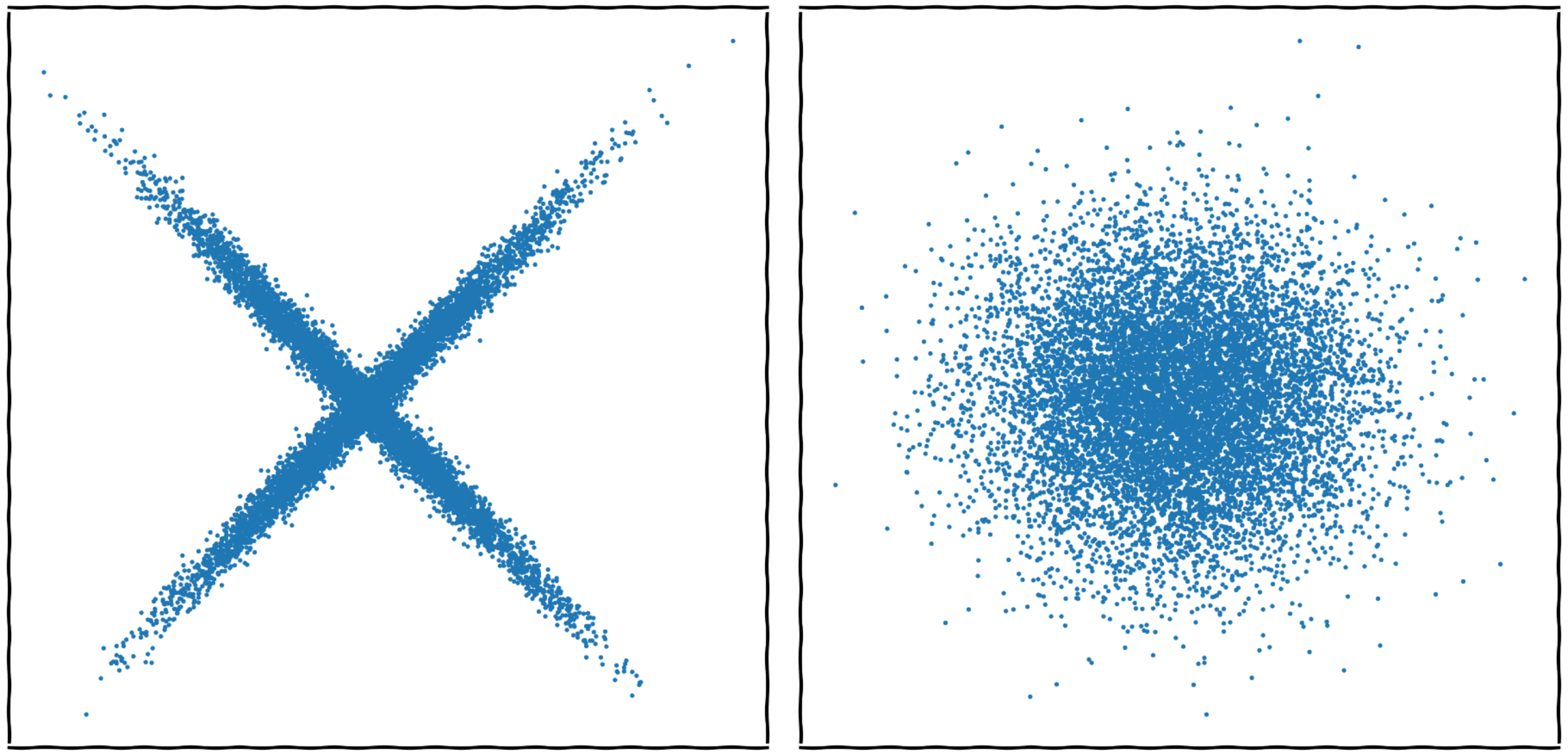
- Need to estimate the marginals $P(\theta)$ and $P(\phi)$ which we can do with NFs





Why is this interesting?

- Mutual information captures relationships that covariance cannot
- These two distributions have the same covariance
- However the mutual information for the X is approximately 1.3 compared to 0 for the Gaussian
- For 21cm we can use MI to quantify the dependence between foreground and signal parameters for example



Conclusions

Conclusions

- We can use NFs (or other density estimation tools) in many different ways
- We can learn about the constraining power of different experiments
- Perform efficient joint analysis
- Estimate correlations between sub spaces
- Papers: 2205.12841, 2207.11457, 2301.03298, 2408.06012, 2305.02930
- Code: https://github.com/htjb/margarine