Marginal Bayesian Statistics with Masked Autoregressive Flows and Kernel Density Estimators

Harry T. J. Bevins

University of Cambridge

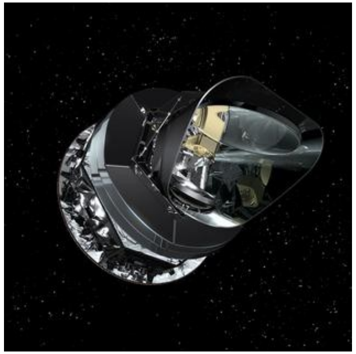
With Will Handley, Pablo Lemos, Peter Sims, Eloy de Lera Acedo and Anastasia Fialkov

The Problem...

- Combining constraints from experiments probing different aspects of the same physics is computationally expensive.

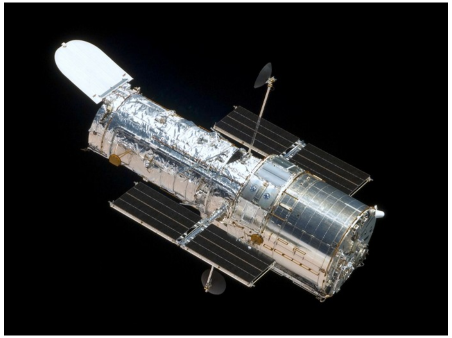
- Often only interested in ~6 cosmological parameters, θ, where each experiment can have an additional ~20 'nuisance' parameters, α, describing systematics and foregrounds.

Planck: $\theta=6$, $\alpha=15$

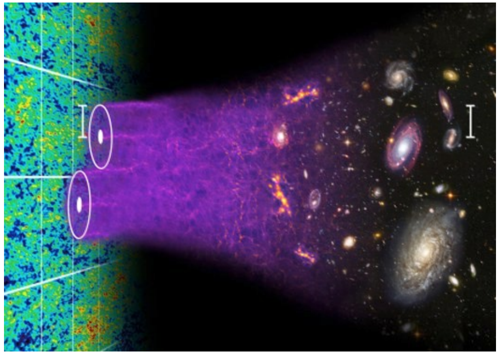


DES: $\theta=6$, $\alpha=20$

SH_0ES: $\theta=6$, $\alpha=0$

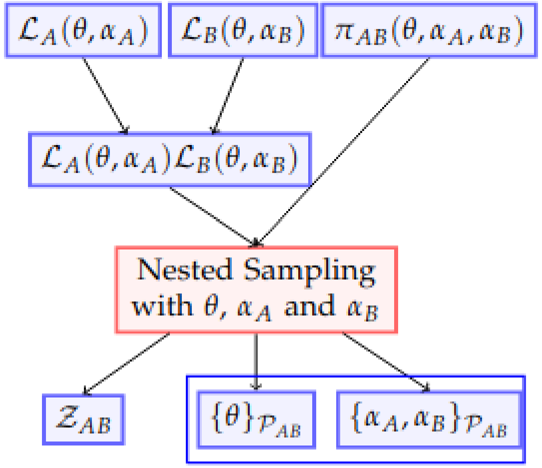


BOSS: $\theta=6$, $\alpha=0$



The Full Nested Sampling Approach

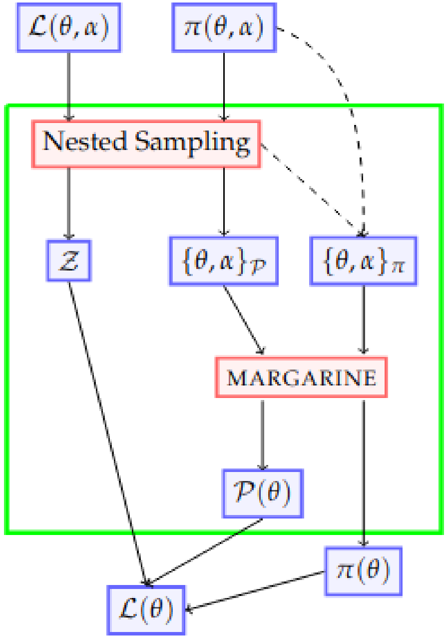
- Typically combination of constraints is performed by sampling over a full joint likelihood.
- Returns the Bayesian Evidence and posterior samples for both nuisance parameters and cosmological parameters.



$$\mathcal{L}_A(\theta, \alpha_A)\mathcal{L}_B(\theta, \alpha_B)\pi_{AB}(\theta, \alpha_A, \alpha_B) = \mathcal{P}_{AB}(\theta, \alpha_A, \alpha_B)\mathcal{Z}_{AB}$$

The *margarine* Approach

- *margarine* uses density estimators to replicate samples output from nested sampling.
- These density estimators are built from known base distributions meaning that we can use them to estimate *P(θ)* and *π(θ)*.
- This can then give us the nuisance-free likelihood along with the KL divergence.



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

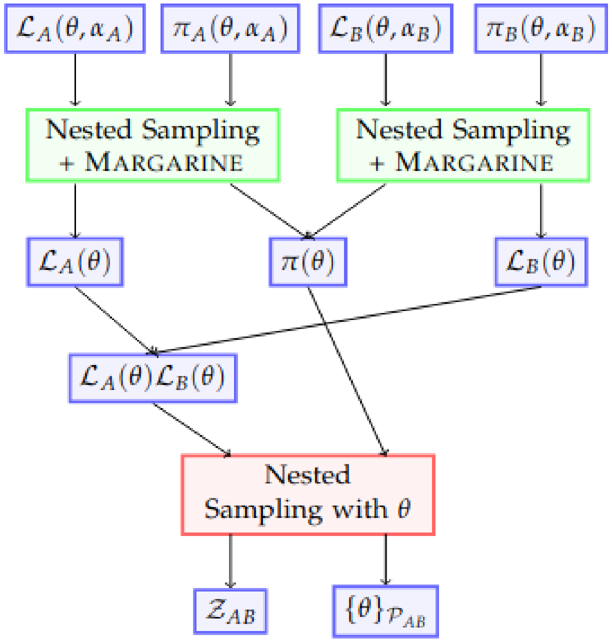
$$\mathcal{D}(\mathcal{P}||\pi) = \int \mathcal{P}(\theta)\log\frac{\mathcal{P}(\theta)}{\pi(\theta)}d\theta$$

The *margarine* Approach

- We can therefore combine nuisance-free likelihoods calculated from samples for experiment A and B.

- We show mathematically that the method is consistent with a full nested sampling run.

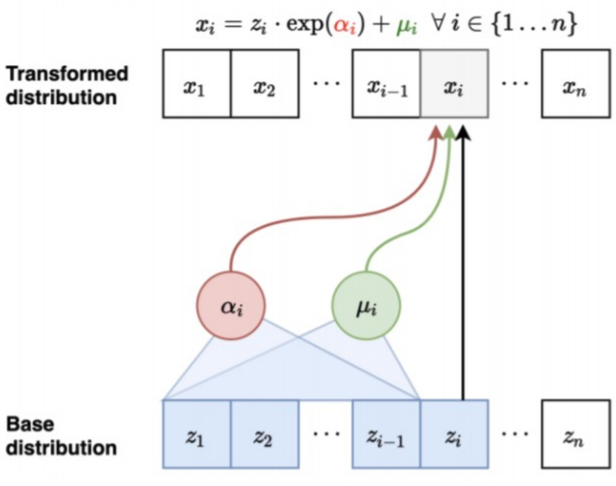
$$\mathcal{L}_A(\theta)\mathcal{L}_B(\theta)\pi(\theta) = \mathcal{P}_{AB}(\theta)\mathcal{Z}_{AB}$$

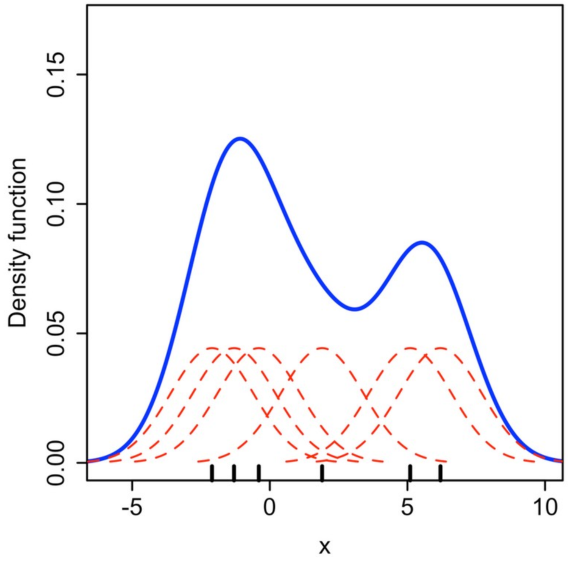


Types of Density Estimators

- **Masked Autoregressive Flows** which shift and scale a base distribution to look like the target posterior or prior.

- **Kernel Density Estimator** which models the posterior or prior distribution using Gaussian kernels.





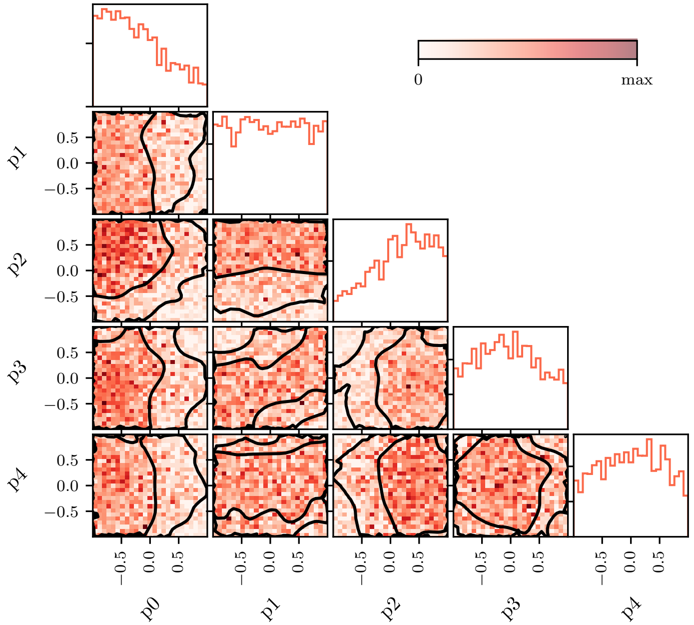
Toy Example

Experiment A:

$$\log \mathcal{L}_A = \sum_i \frac{1}{2}(x_i - \theta_i)^2,$$

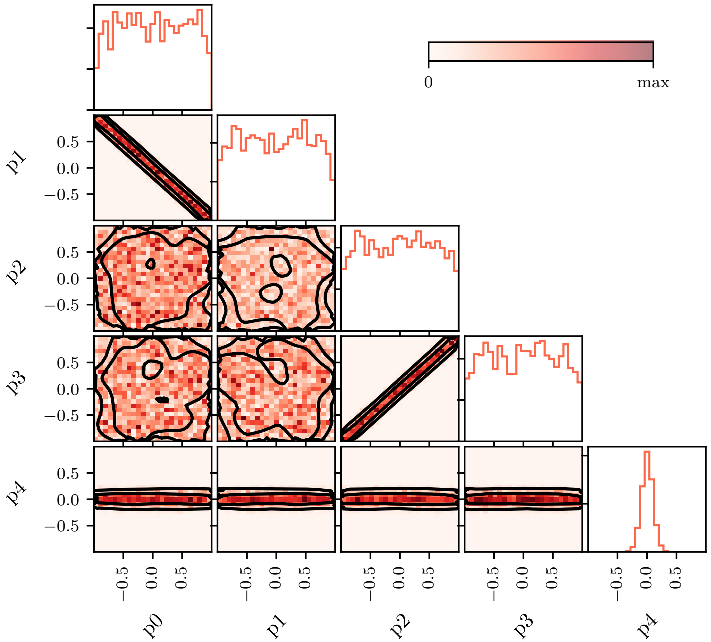
$$x = [-0.8, 0.5, 0.6, -0.2, 0]$$

$$\theta = [p_0,\ p_1 \times p_3,\ p_2,\ p_3,\ p_4]$$



Experiment B:

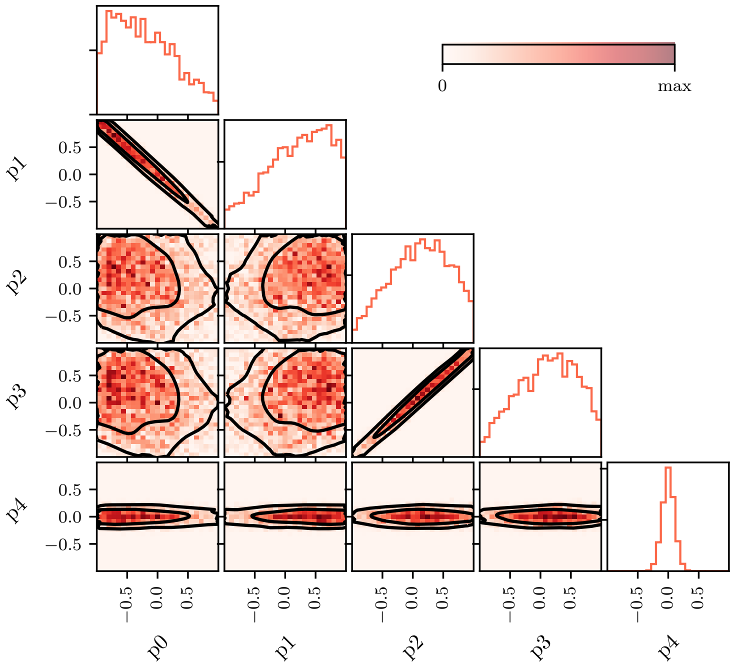
$$\log \mathcal{L}_B = \frac{1}{2}((p_0 + p_1)^2 + (p_2 - p_3)^2 + p_4^2)$$



Toy Example

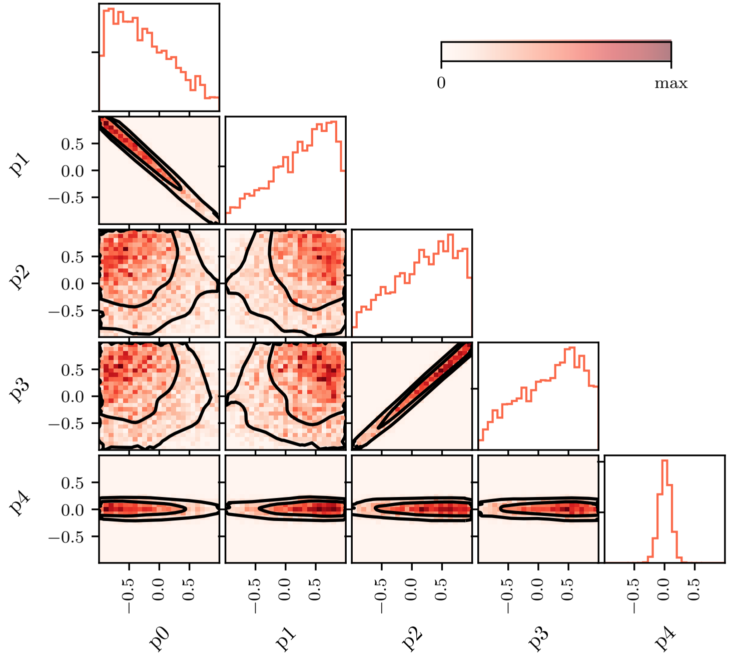
Combined w/ Nested Sampling
log(Z) = -11.7 ± 0.1

$$\log \mathcal{L}_{AB} = \sum \frac{1}{2}(x_i - \theta_i)^2 + \frac{1}{2}((p_0 + p_1)^2 + (p_2 - p_3)^2 + p_4^2)$$



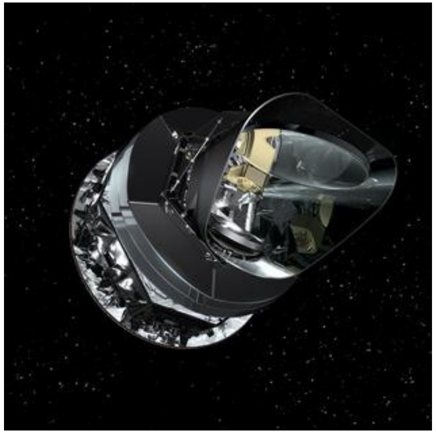
Combined w/ Nested Sampling + *margarine*
log(Z) = -11.8 ± 0.1

$$\log \mathcal{L}_{AB} = \log \mathcal{L}_A^{margarine} + \log \mathcal{L}_B^{margarine}$$



Planck + DES

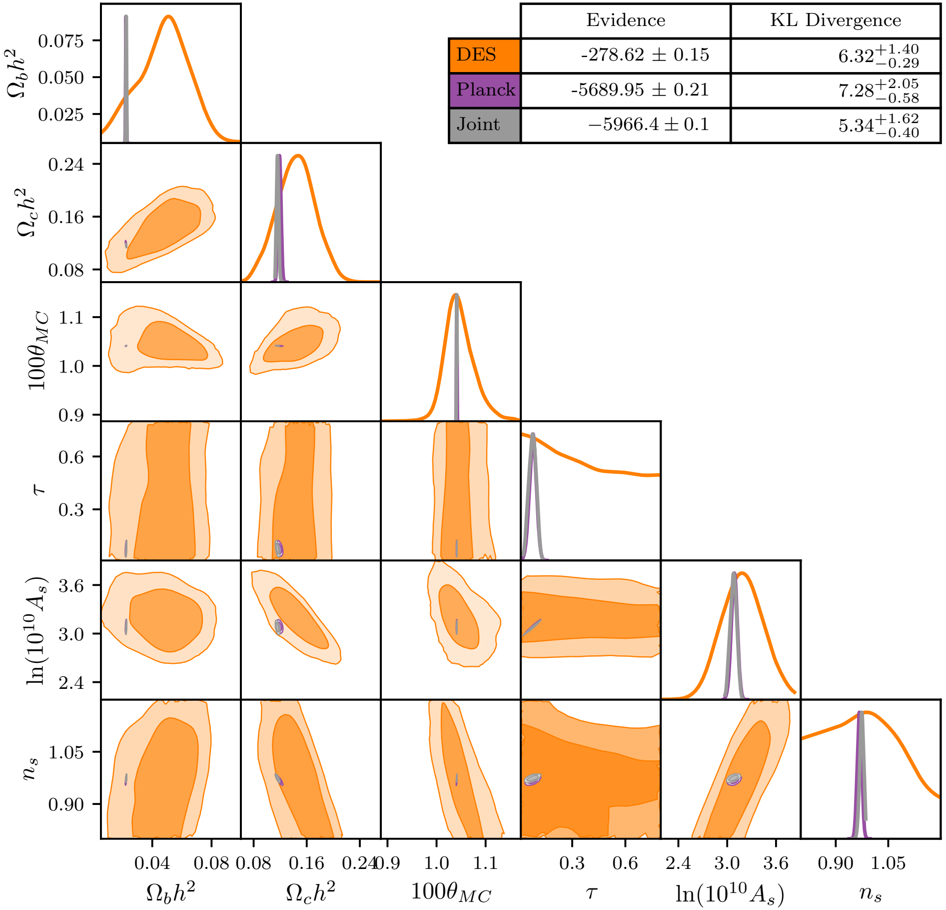
- To model both DES and Planck we have to fit 41 parameters of which 35 are nuisance parameters…
- Previously been done in Handley and Lemos 2019a and 2019b
- $log(Z) = -5965.7 \pm 0.3$
- $D = 6.17 \pm 0.36$



Planck + DES

- *margarine* is much more computationally efficient than fitting for 41 parameters.
- Using nuisance-free likelihoods we can sample over just the six cosmological parameters.
- *log(Z) = -5966.4 ± 0.1*
- *D = 5.34 +(-) 1.62(0.40)*
- We use *margarine* to perform importance sampling.

$$\mathcal{Z}_B = \mathcal{Z}_A \left\langle \frac{\pi_B(\theta)}{\pi_A(\theta)} \right\rangle_{\mathcal{P}_A} \qquad w_B^{(i)} = w_A^{(i)} \frac{\pi_B(\theta^{(i)})}{\pi_A(\theta^{(i)})}$$



	Evidence	KL Divergence
DES	-278.62 ± 0.15	$6.32^{+1.40}_{-0.29}$
Planck	-5689.95 ± 0.21	$7.28^{+2.05}_{-0.58}$
Joint	-5966.4 ± 0.1	$5.34^{+1.62}_{-0.40}$

margarine

https://github.com/htjb/margarine

https://arxiv.org/abs/2205.12841

setup.py — Log probabilities and likelihoods (#5) — 26 days ago

README.rst

margarine: Posterior Sampling and Marginal Bayesian Statistics

Introduction

margarine:	Marginal Bayesian Statistics
Authors:	Harry T.J. Bevins
Version:	0.2.0
Homepage:	https://github.com/htjb/margarine
Documentation:	https://margarine.readthedocs.io/

docs passing | launch binder | astro.IM arXiv:2205.12841

Removing the fat from your posterior samples with MARGARINE

Harry T. J. Bevins[*,1] William J. Handley[1,2] Pablo Lemos[3,4] Peter H. Sims[5]

Eloy de Lera Acedo[1,2] Anastasia Fialkov[2,3]

Justin Alsing[7]

[1] Astrophysics Group, Cavendish Laboratory, Cambridge, CB3 0HE, UK
[2] Kavli Institute for Cosmology, Cambridge, CB3 0HA, UK
[3] Department of Physics & Astronomy, University College London, London, WC1E 6BT, UK
[4] Department of Physics and Astronomy, University of Sussex, Brighton, BN1 9QH, UK
[5] Department of Physics and McGill Space Institute, McGill University, Montreal, QC H3A 2T8, Canada
[6] Institute of Astronomy, University of Cambridge, Cambridge CB3 0HA, UK
[7] Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, Stockholm SE-106 91, Sweden
*htjb2@cam.ac.uk

Conclusions

- *margarine* offers a computationally more efficient path to combined Bayesian analysis that is consistent with a full Nested Sampling run.
- Our approach is lossless in θ as it recovers the same marginal posterior and total evidence.
- The work paves the way for the development of a publicly available library of cosmological likelihood emulators.
- There are applications beyond cosmology.