Normalising Flow for Joint Cosmological Analysis

Arrykrishna Mootoovaloo *
Department of Astrophysics
University of Oxford
Keble Road, Oxford, OX1 3RH, UK

Carlos García-García
Department of Astrophysics
University of Oxford
Keble Road, Oxford, OX1 3RH, UK

David Alonso
Department of Astrophysics
University of Oxford
Keble Road, Oxford, OX1 3RH, UK

Jaime Ruiz-Zapatero
Department of Astrophysics
University of Oxford
Keble Road, Oxford, OX1 3RH, UK

Abstract

We develop a method to learn the joint posterior of cosmological parameters using MCMC chains from various experiments by leveraging normalising flow to learn the density of each chain. These models are quick to train and, once stored, allow efficient sampling of joint posteriors from any experiment combination. Applied to test cases in cosmology, the method reveals robust accuracy and precision, even when known tensions in parameters like σ_8 exist. The flow model can also be used as a prior in likelihood analyses, significantly speeding up inference by eliminating the need for repeated computationally intensive analyses. Sampling the joint posterior using pre-trained models takes about 15 minutes, making this method highly efficient for cosmological studies.

1 Introduction

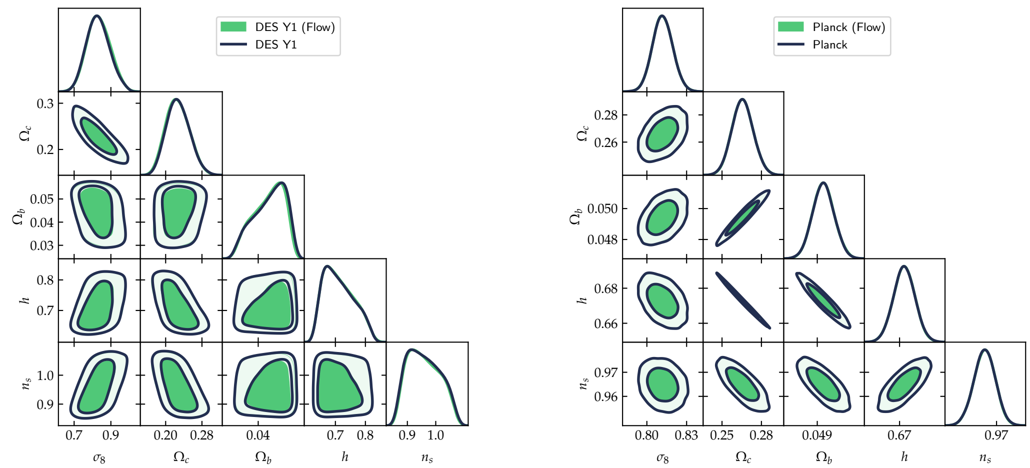
In Cosmology, numerous publicly available chains for cosmological and nuisance parameters have been obtained using MCMC-based approaches. The question is whether we can exploit these chains to perform joint analysis of different probes. A similar concept was investigated by Heavens et al. (2017a,b), who used publicly available MCMC chains to estimate the marginal likelihood. In this work, we demonstrate how normalising flows can be employed to learn the probability distribution of the cosmological parameters alone, effectively marginalising over the nuisance parameters, and subsequently performing a joint analysis. If we were to take a standard approach, gathering the right data and doing a principled analysis is not straightforward. Moreover, sampling the joint posterior would require the computations of many expensive and different power spectra, hence leading to an overhead of computations.

Normalising flows have been used in various applications in Cosmology. For example, Alsing & Handley (2021) combined normalising flow models with nested sampling. Recently, Srinivasan et al. (2024) developed a codebase, `flowZ` to estimate the Bayesian Evidence from posterior samples. Normalising flows have also been used in the estimation of Bayesian Evidence via the harmonic mean estimation (McEwen et al. 2021; Polanska et al. 2024). Our contributions in this work are as follows.

1. We train normalising flow models on separate MCMC chains from different experiments and use them for joint cosmological analysis.
2. We demonstrate that a pre-trained normalising flow model can serve as a prior in joint analysis scenarios. This approach is especially beneficial when combining an expensive

*arrykrishna.mootoovaloo@physics.ox.ac.uk

Machine Learning and the Physical Sciences Workshop, NeurIPS 2024.



(a) DES Y1 cosmological parameters

(b) Planck 2018 cosmological parameters

Figure 1: Panel (a) shows the posterior of the cosmological parameters, sampled from the pre-trained normalising flow (in green) for the DES Y1 analysis. The black contours show the original MCMC samples. Panel (b) shows the same but for the Planck 2018 MCMC samples. The flow models are able to capture the different geometry of the high dimensional posterior distribution.

likelihood with a fast normalising flow model, significantly enhancing efficiency for joint analysis tasks.

2 Bayesian Analysis with Normalising Flows

Normalising flows are a class of generative model which transform simple distributions into complex ones via invertible functions. They are known for density estimation and sampling and both are generally quite fast. In Cosmology, we have many MCMC chains which are publicly available and the question posed is whether we can take advantage of these to learn a density function.

Suppose we have N experiments, each having its own set of cosmological parameters, $\boldsymbol{\theta}_i$, nuisance parameters, $\boldsymbol{\beta}_i$ and data $\boldsymbol{x}_i$. For simplicity, we will assume that the we have a common set of cosmological parameters across all experiments. Let us also assume that we have samples of $\{\boldsymbol{\theta}_i,\, \boldsymbol{\beta}_i\}$, which are obtained by sampling the posterior of all parameters in each experiment. The marginalised posterior distribution of the cosmological parameters only correspond to the columns of the cosmological parameters in a typical MCMC file. In other words,

$$p(\boldsymbol{\theta}_i|\boldsymbol{x}_i) = \int p(\boldsymbol{\theta}_i,\, \boldsymbol{\beta}_i|\boldsymbol{x}_i)\, \mathrm{d}\boldsymbol{\beta}_i. \tag{1}$$

If we were to do a joint analysis among the different experiments, the total dimensionality of the problem can become large. For example, if we assume we have b cosmological parameters and each experiment E_i has c_i nuisance parameters, the total number of parameters is $b + \sum_i c_i$. Standard sampling schemes such as Metropolis-Hastings may struggle to learn the full posterior distribution of all parameters. Furthermore, as we incorporate more experiments into the analysis, it may become increasingly computationally intensive. Our proposal is to sample the parameters in each experiment (or use publicly available MCMC chains), followed by data fusion where the approximate joint posterior is:

$$p(\boldsymbol{\theta}|\boldsymbol{x}_1,\, \boldsymbol{x}_2,\, \ldots \boldsymbol{x}_N) = k \prod_{i=1}^{N} p_{\text{nf}}(\boldsymbol{\theta}_i|\boldsymbol{x}_i), \tag{2}$$

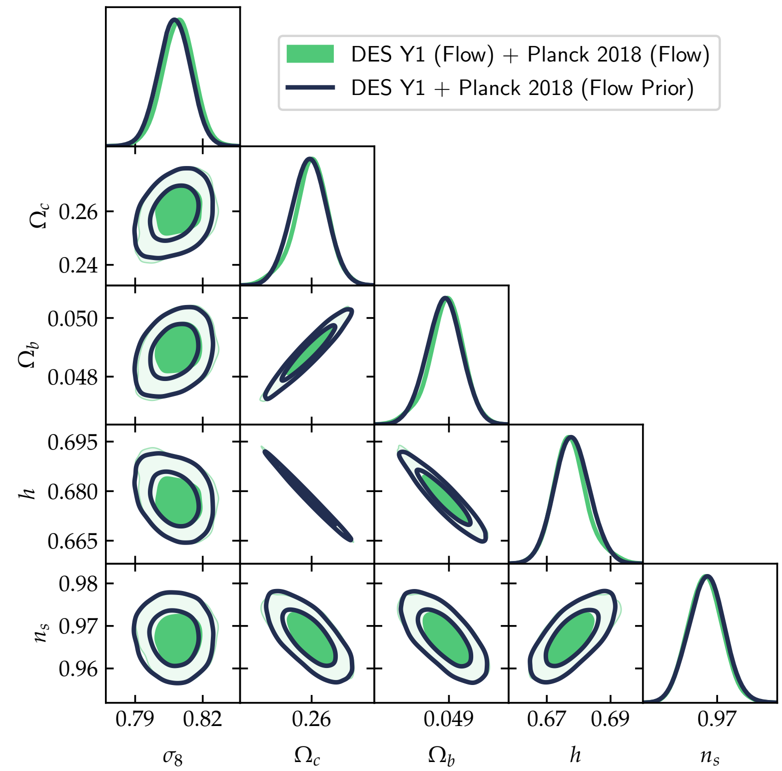


Figure 2: The figure shows the joint posterior of the cosmological parameters, marginalised over 20 nuisance parameters related to DES Y1. In this particular case, we have used the normalising flow built upon the public Planck 2018 chains (`base_plikHM_TTTEEE_lowl_lowE`). The green contours show the results when using normalising flow models for both experiments while the black contours show the joint posterior with the DES Y1 likelihood code, with the pre-trained normalising model for Planck 2018 being used as a prior. See Equation 5.

where $p_{\text{nf}}(\boldsymbol{\theta}_i|\boldsymbol{x}_i)$ is the learned normalizing flow model for each experiment and k is just a normalisation constant. The approximate joint log-posterior is:

$$\log \hat{p}(\boldsymbol{\theta}|\boldsymbol{x}_1,\, \boldsymbol{x}_2,\, \ldots \boldsymbol{x}_N) = \sum_{i=1}^{N} \log p_{\text{nf}}(\boldsymbol{\theta}_i|\boldsymbol{x}_i) + \log k. \tag{3}$$

If we have access to $\log p_{\text{nf}}(\boldsymbol{\theta}_i|\boldsymbol{x}_i)$, we can draw samples from this approximate distribution and we can also compute the log-density. This is crucial because we can then 1) use the log-density for joint analysis and 2) use the learned density as a prior in a completely new cosmological data analysis problem.

3 DES Y1 and Planck 2018

The DES Y1 data consists of 405 band powers data, $\boldsymbol{x}_1$ (García-García et al. 2021). The forward model consists of 5 cosmological parameters, θ:

$$\boldsymbol{\theta} = \{\sigma_8, \Omega_c,\, \Omega_b,\, h,\, n_s\} \tag{4}$$

and it also has 20 nuisance parameters, $\boldsymbol{\beta}$. We first sample the posterior of the cosmological and nuisance parameters, and we build a normalising flow model on top of the MCMC samples corresponding to $\boldsymbol{\theta}$. The MCMC samples for Planck 2018 (whose data we denote as $\boldsymbol{x}_2$) are publicly available and different analyses based on different cosmological models have been performed. We use the `base_plikHM_TTTEEE_lowl_lowE` MCMC samples to select the subset of cosmological parameters, $\boldsymbol{\theta}$. Both normalising flow models are trained using 15 000 training points and a learning rate of 10^{-3}. Training each flow takes around 2 minutes only. Using Equation 3, the joint posterior of $\boldsymbol{\theta}$ due to the two experiments are sampled using `EMCEE` (Foreman-Mackey et al. 2013).

Next, we also investigate the case where we use the pre-trained normalising flow model for Planck, $p_{\text{nf}}(\boldsymbol{\theta}|\boldsymbol{x}_2)$ as a prior. Using Bayes' theorem, the joint posterior distribution of the cosmological and nuisance parameters (from the DES Y1 likelihood) is:

$$p(\boldsymbol{\theta},\,\boldsymbol{\beta}|\boldsymbol{x}_1,\,\boldsymbol{x}_2) \propto p(\boldsymbol{x_1}|\boldsymbol{\theta},\,\boldsymbol{\beta})\,p(\boldsymbol{\beta})\;p(\boldsymbol{\theta})\left[\frac{p_{\text{nf}}(\boldsymbol{\theta}|\boldsymbol{x_2})}{p(\boldsymbol{\theta})}\right]. \tag{5}$$

It is important to weight the normalising flow model by the prior during analysis. We jointly sample the cosmological and nuisance parameters using the standard Metropolis-Hastings technique. On the other hand, it is known that obtaining MCMC samples for the Planck data alone, or in combination with other probes, typically requires several days.

4 Results

Panel (a) and Panel (b) of Figure 1 show the original MCMC samples in black and the normalising flow samples in green. Sampling from the normalising flow model is straightforward and very quick. This process involves drawing samples from the base distribution, $p(\boldsymbol{z})$, and mapping them to the cosmological parameters through a series of bijective transformations. Figure 1 also shows that the flow models can robustly learn the complex joint posterior of the cosmological parameters only, that is, $p(\boldsymbol{\theta}|\boldsymbol{x}_i)$.

Figure 2 shows the joint posterior of the cosmological for the joint analysis of DES Y1 and Planck. For any test point, $\boldsymbol{\theta}_*$, we can compute the log-density from each pre-trained normalising flow model. Hence, sampling from the joint distribution is possible via Equation 3. We generate 120 000 MCMC samples in around 15 minutes only and the samples are shown in green in Figure 2. On the other hand, if we couple the Planck normalising flow to the DES Y1 likelihood for a joint analysis of DES Y1 and Planck, the sampling procedure takes around 5 hours and the samples are shown by the black contours. The fact that both contours (green and black) lie almost on top of each other demonstrates the robustness of the method proposed in this work. In order to quantify the difference between the two distributions, we have also computed, $\delta_q = \frac{|\mu-\hat{\mu}|}{\sqrt{\sigma^2+\hat{\sigma}^2}}$, which quantifies the relative tension between the two distributions. The smaller the number the smaller the difference between the distribution in 1D. The minimum and maximum δ_q are 0.05 and 0.16, corresponding to n_s and σ_8 respectively.

5 Challenges

A particular challenge we encountered in this work is to re-run the Planck likelihood for a direct comparison of the method in different combination of experiments. A single MCMC run of Planck takes $\sim$ 25 days, which is also a motivation for developing the method described in this work.

6 Conclusion

In this study, we demonstrated the importance of normalising flow models, trained on existing MCMC samples, for deriving cosmological parameter constraints in a joint analysis of various probes. The method is not only robust to shifts and expansions of the posterior, ensuring precision and accuracy, but it is also very fast. We have also trained normalising flow models on other MCMC samples, demonstrating that the joint posterior obtained using two flow models is comparable to that from explicitly sampling the joint posterior with MCMC methods. Additionally, we have built a library of normalising flow models based on public MCMC chains ranging from large-scale structure, CMB, and BAO experiments such as DES Y3, KiDS-1000, ACT DR4, and SDSS. Users can easily load these pre-trained models for joint analysis by sampling multiple flows together or by integrating the flows with their own likelihood code. Furthermore, users can quickly train their own normalising flow models. We hope this work will facilitate various upcoming cosmological analyses and enable the generation of MCMC samples for joint analyses in under an hour.

References

Alsing, J. & Handley, W. 2021, , 505, L95

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, , 125, 306

García-García, C., Ruiz-Zapatero, J., Alonso, D., et al. 2021, , 2021, 030

Heavens, A., Fantaye, Y., Mootoovaloo, A., et al. 2017a, arXiv e-prints, arXiv:1704.03472

Heavens, A., Fantaye, Y., Sellentin, E., et al. 2017b, , 119, 101301

McEwen, J. D., Wallis, C. G. R., Price, M. A., & Spurio Mancini, A. 2021, arXiv e-prints, arXiv:2111.12720

Polanska, A., Price, M. A., Piras, D., Spurio Mancini, A., & McEwen, J. D. 2024, arXiv e-prints, arXiv:2405.05969

Srinivasan, R., Crisostomi, M., Trotta, R., Barausse, E., & Breschi, M. 2024, arXiv e-prints, arXiv:2404.12294