Designing and testing inflationary models with Bayesian networks

Layne C. Price,[1,2] **Hiranya V. Peiris,**[3] **Jonathan Frazer,**[4,5,6] **and Richard Easther**[2]

[1]McWilliams Center for Cosmology, Department of Physics
   Carnegie Mellon University, Pittsburgh, PA 15213, USA
[2]Department of Physics, University of Auckland
   Private Bag 92019, Auckland, New Zealand
[3]Department of Physics and Astronomy, University College London
   London WC1E 6BT, UK
[4]Deutsches Elektronen-Synchrotron DESY, Theory Group, 22603, Hamburg, Germany
[5]Department of Theoretical Physics, University of the Basque Country, UPV/EHU
   48040 Bilbao, Spain
[6]IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain

E-mail: laynep@andrew.cmu.edu, h.peiris@ucl.ac.uk, jonathan.frazer@desy.de, r.easther@auckland.ac.nz

Abstract. Even simple inflationary scenarios have many free parameters. Beyond the variables appearing in the inflationary action, these include dynamical initial conditions, the number of fields, and couplings to other sectors. These quantities are often ignored but cosmological observables can depend on the unknown parameters. We use Bayesian networks to account for a large set of inflationary parameters, deriving *generative models* for the primordial spectra that are conditioned on a hierarchical set of prior probabilities describing the initial conditions, reheating physics, and other free parameters. We use N_f–quadratic inflation as an illustrative example, finding that the number of e-folds N_* between horizon exit for the pivot scale and the end of inflation is typically the most important parameter, even when the number of fields, their masses and initial conditions are unknown, along with possible conditional dependencies between these parameters.

Contents

1 Introduction

Precise measurements of the cosmic microwave background from *Planck* [1, 2] and WMAP [3] have put tight constraints on estimated values of the free parameters of simple inflationary models. However these analyses have only been performed for the simplest possible models. For instance, additional scalar fields, nontrivial couplings to gauge fields or the gravitational sector, or varying the initial conditions all introduce new variables into the model.

Parameters associated with these variables are implicitly set to zero in these treatments, and including them explicitly would make extracting constraints on the models considerably more challenging. These parameters can alter the predicted primordial density perturbations, and realistic treatments of these scenarios must account for this uncertainty in the *a priori* structure of the underlying model. This uncertainty should then be consistently propagated into the predictions for the early universe and enfolded into the comparisons with data. In this paper we address this challenge by introducing a Bayesian *hierarchical modelling* scheme for inflationary physics, which incorporates theoretical and statistical uncertainties within a unified framework.

Hierarchical models have been used in other cosmological contexts, including inference problems with Type Ia supernovae data [4–8], detection algorithms for spatially-localized features in the cosmic microwave background (CMB) [9], and the analysis of cosmic shear data [10]. Our approach treats each inflationary parameter as a random variable with an associated prior probability distribution. These priors may subsequently depend on other model parameters or *hyperparameters*, which can also be randomly distributed. For example, the mean and variance are hyperparameters for normally-distributed model parameters.[1] The data or predicted observables are generated stochastically by fixing the outermost hyperparameters in the hierarchy and sampling the chain of dependent prior probabilities. These complicated dependency chains are treated as *Bayesian networks* in Sect. 2, which give an easily visualized graphical depiction of the nested parameters. A toy Bayesian network for a simple inflation model is shown in Fig. 1.

Our methodology can be summarized as follows:

---

[1]In Sect. 3 the distribution of scalar masses $P(m^2 \mid \beta)$ has a hyperparameter β that is the ratio of the number of axions to the total number of moduli [11].
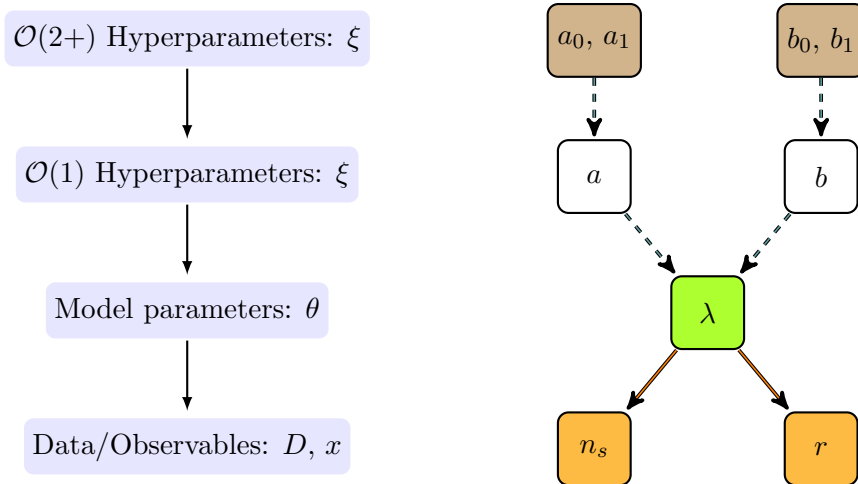
Figure 1: (*Left*) A schematic relationship between the parameters and hyperparameters in a Bayesian network. The arrows indicate that the target node depends on the source node. (*Right*) A Bayesian network for a single-field, slow-roll inflation model with inflaton self-couplings λ. The nodes correspond to observables (orange), model parameters (green), hyperparameters (white), and higher order hyperparameters (brown). The dashed, blue arrows show a dependency through conditional probabilities; the solid, orange arrows indicate a deterministic dependency. For this toy model the quasi-observables n_s and r are uniquely determined given the self-couplings λ, which are drawn uniformly in the range $a < b$ and the limits of this range are drawn uniformly from $a_0 < a_1$ and $b_0 < b_1$.

1. *Outline hierarchical dependencies.*— Carefully describe the nested set of conditional dependencies between the parameters and hyperparameters in the inflation model and write down a graphical Bayesian network. Subtly different choices can give substantially different predictions so this is a crucial step.

2. *Define prior probabilities & hyperprior probabilities.*— Treat every parameter and hyperparameter as a random variable with an associated prior probability distribution; these distributions are related by the details of the Bayesian network.

3. *Sample the Bayesian network.*— Fix the value of the outermost hyperparameters. Sample the prior probabilities for each subsequent layer of (hyper)parameters in an order set by their conditional dependencies. Calculate the prediction using the final set of model parameters. Repeat to build a large sample from which a probability distribution can be estimated.

These hierarchical methods are applicable to any inflation model, but are especially useful with scenarios with multiple active fields or couplings between the inflaton and other sectors. We illustrate this approach with three variants of N_f–quadratic inflation, where the Bayesian networks exhibit varying degrees of complexity. We present results for the distribution of the tensor-to-scalar ratio r, the spectral tilt n_s, and scalar running α_s in Sect. 3, with the standard definitions for the primordial spectra which are summarized in Sect. 3.1.

This model predicts a tensor-to-scalar ratio of $r \sim \mathcal{O}(10^{-1})$,[2] which is in tension with the joint *Planck* and BICEP2/*Keck Array* results for the ΛCDM+r model [1, 12–16], although it is more consistent with an extended ΛCDM scenario [17]. However, N_f–quadratic inflation is a useful example in that it is sufficiently complex to have a non-trivial Bayesian network, while still allowing an intuitive understanding of the inflationary dynamics. Importantly, the techniques we use here apply to any inflation model, including those that provide a better fit to the CMB data, *e.g.*, single-field models or models with non-canonical kinetic energies or non-minimal couplings to the gravity sector.

We confirm previous indications [18–20] that N_f–quadratic models are largely insensitive to the exact particle masses, initial conditions, and the total number of fields and conclude that the unknown reheating mechanism is the dominant source of uncertainty in these models, evidenced by a strong dependency on the number of e-folds N_* occurring between the moment at which the pivot scale k_* leaves the horizon and the end of inflation.

This paper is organized as follows: in Section 2 we introduce generative hierarchical models and Bayesian networks. In Section 3 we explore three Bayesian networks for N_f–quadratic inflation and give the slow-roll predictions for the spectral index n_s, the scalar running α_s, and the tensor-to-scalar ratio r at the end of inflation. In Section 4 we summarize our results.

2   Generative hierarchical models & Bayesian networks

To compare an inflation scenario's prediction to experimental data we need to know (A) how to make a prediction for the data with fixed model parameters and (B) the distribution from which these parameters are to be drawn. We divide a model's parameters and their prior probabilities into three qualitative classes:

1. *Model parameters & model priors*— Parameters θ for an inflation model that directly affect the prediction, and prior probabilities $P(\theta)$ that encode how we expect θ to be distributed based on theoretical or empirical knowledge. For example, self couplings λ, masses m^2, the number of fields N_f, or the number of e-folds N_* after the pivot scale leaves the horizon could all be model parameters:

$$\theta \sim \left\{ \lambda, m^2, N_f, N_*, \dots \right\}. \tag{2.1}$$

The model parameters traditionally define the physics.

2. *Hyperparameters & hyperpriors*— Variables ξ that define the model priors $P(\theta \,|\, \xi)$. For instance, if a model parameter θ is expected to be normally distributed as $\theta \sim \mathcal{N}(\mu, \sigma)$, then the mean μ and standard deviation σ are hyperparameters:

$$\xi \sim \{\mu, \sigma, \dots \}. \tag{2.2}$$

Hyperparameters are distributed according to hyperpriors $P(\xi)$. The hyperparameters/hyperpriors statistically generate realizations of the inflationary scenarios. A nested set of hyperparameters may be required if the lower order hyperparameters are inherently stochastic, if their true value is unknown, or if we wish to marginalize over prior probability distributions in order to assess the robustness of the Bayesian network.

---

[2]This estimate assumes that observables do not change during the post-inflationary epochs, although the evolution may not be adiabatic at the end of inflation for this model. Throughout this paper we ignore this potentially important dependence on post-inflationary physics, although in principle it could be included as an additional layer in the Bayesian network.

Model	Description	Reference
BayesNet A	A simple network with minimal dependencies	Fig. 3 & Sect. 3.2
BayesNet B	The prior for N_f maximizes the Shannon entropy	Fig. 3 & Sect. 3.3
BayesNet C	Incorporates a dependence on the initial energy and requires sub-Planckian field displacements	Fig. 7 & Sect. 3.4

Table 1: A legend for the three example N_f–quadratic models.

3. *Data & likelihoods*— The predicted data D are treated as random variables in the model with some likelihood $P(D\,|\,\theta,\xi)$. In place of simulating experimental data we can replace D with proxy quasi-observables x, such as parametrizations of the primordial spectra

$$x \sim \{A_s, n_s, r, f_{\mathrm{NL}}, \dots\}. \tag{2.3}$$

Throughout, we refer to data D or quasi-observables x interchangeably.

By sampling the prior and hyperprior distributions we obtain stochastic predictions for the data and/or quasi-observables in the model, which are conditioned on the prior probabilities. This defines a *generative model* for the primordial parameters:[3]

$$\{m^2, N_*, \dots\} \xrightarrow{\text{generative}} \{A_s, n_s, r, f_{\mathrm{NL}}, \dots\}. \tag{2.4}$$

We specify generative inflation models using Bayesian networks, which outline conditional dependencies between parameters, hyperparameters, priors, and the predicted data by a directed acyclic graph. Figure 1 shows a schematic depiction of a Bayesian network along with a toy network for an inflationary model where the spectral index and the tensor-to-scalar ratio are simple functions of the inflaton self-couplings λ. The arrows indicate conditional dependence between the target and source node. The network is directed so that the probability distribution for an interior variable in the model is determined exclusively from the higher levels in the graph. The interior parameters and the data are generated according to this chain of probabilistic dependencies. Since the graph is acyclic, there will be a set of outermost parameters ξ_{out}, which we set to fixed values. Repeated sampling of the generating distribution results in a sample of predicted data that follow the marginalized *a priori* distribution

$$P(D\,|\,\xi_{\mathrm{out}}) = \int P(D\,|\,\theta,\xi,\xi_{\mathrm{out}})P(\theta,\xi\,|\,\xi_{\mathrm{out}})\,d\theta\,d\xi. \tag{2.5}$$

The prior $P(\theta,\xi\,|\,\xi_{\mathrm{out}})$ can be factorized into conditional probabilities that define the structure of the Bayesian network. Finally, varying the ξ_{out} gives a qualitative measure of the sensitivity of the model's predictions to the values of the hyperparameters.

3 Example: N_f–quadratic inflation

3.1 The model parameters and observables

The potential.— In Sects. 3.2–3.4 we consider several specific scenarios that illustrate the utility of the hierarchical networks. We summarize the models in Table 1. This methodology

---

[3]Although the generative model could also directly predict the CMB data, in this paper we restrict attention to proxy quasi-observables.
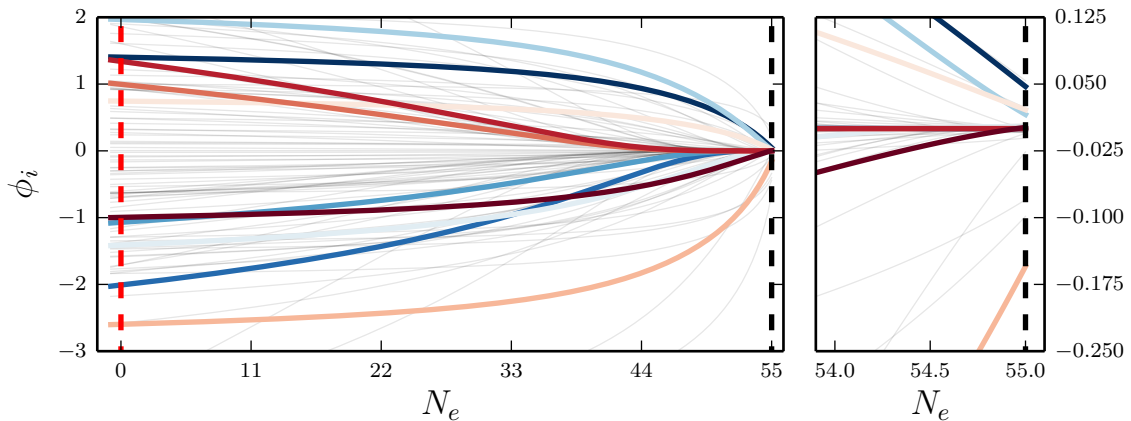
Figure 2: Evolution of the fields ϕ_i as a function of e-folds N_e after the pivot scale leaves the horizon at $N_e = 0$, with the N_f–quadratic potential (3.1) and $N_* = 55$. Masses are set by the Marčenko-Pastur distribution (3.18) and 100 fields are plotted, with 10 randomly chosen light fields emphasized in color. (*Left*) Initial conditions are set near the horizon crossing surface (*dashed red line*). (*Right*) Zoom-in near the end of inflation surface (*dashed black line*), demonstrating genuinely multifield behavior for the entirety of the inflationary period.

does not impose any constraints on the nature of the inflationary Lagrangian and applies equally well to complicated and simple scenarios. However, in most single-field models the relevance of various parameters tends to be obvious, so we focus on more complex models.

Consider the N_f–quadratic inflation model, which has N_f massive fields, a canonical kinetic energy term, and the sum-separable potential

$$V(\phi_1, \ldots, \phi_{N_f}) \equiv \sum_i V_i(\phi_i) = \frac{1}{2} \sum_i^{N_f} m_i^2 \phi_i^2. \tag{3.1}$$

If the number of fields N_f is large this potential can realize assisted inflation [21, 22], where each field undergoes a sub-Planckian field displacement while generating observable gravitational wave backgrounds, with $r \gtrsim 0.1$. If we assume each field has an initial value $|\phi_i| < 1$ and all $m_i^2 > 0$ this potential is the lowest-order description for a set of canonical fields rolling into a minimum. The curvature spectra derived from the potential (3.1) have been well studied [11, 21–27] and marginalized predictions for the power spectrum were obtained in Refs. [18–20, 28], but no complete generative models have been developed and analyzed for this scenario.

Even if all of the model parameters are fixed, observables can still depend on the initial states $\phi_{i,0}$ of the homogeneous background fields, particularly in multifield models. For generic initial conditions and masses the potential (3.1) has genuinely multifield behavior throughout the entirety of inflation, as the field evolution does not generally reach a single-field attractor. The trajectory of the field values as a function of the number of e-folds of expansion N_e during inflation is shown in Fig. 2, for a model with fields of different masses. At the end of inflation there are typically two or more active fields.

By contrast, a single-field model can use the monotonically decreasing field value ϕ as a clock during inflation, and within the slow-roll approximation a given field value can be

uniquely associated with the number of e-folds that will occur before the end of inflation. Multifield models only have a monotonically decreasing energy density ρ and the values of individual fields ϕ_i depend on the initial values. This initial state is unknown, so the *a priori* probability distribution of initial field values induces a probability distribution for the model's predictions. The inherently stochastic predictions of multifield models are well described by the generative modelling methodology of Sect. 2.

Methodology.— We will focus on the predictions of N_f–quadratic inflation for the primordial power spectrum of curvature perturbations $\mathcal{P}_\zeta(k)$ at the end of inflation. While higher order correlators could also be computed either analytically or numerically, the two-point function is most easily constrained by data. Furthermore, the N_f–quadratic model generally predicts a low level of non-Gaussianity well within current bounds from *Planck* [29, 30], *e.g.*, $|f_{\mathrm{NL}}| \sim 1/N_* \sim \mathcal{O}(10^{-2})$.

We will evaluate the spectral index n_s of the scalar spectrum, the running of the spectral index α_s, and the tensor-to-scalar ratio r using the δN formalism, which relates curvature perturbations ζ on constant density hypersurfaces to field perturbations $\delta\phi^i$ on flat hypersurfaces as

$$\zeta = \frac{\partial N_e}{\partial \phi_i}\delta\phi^i + \mathcal{O}(\delta\phi_j^2), \tag{3.2}$$

where N_e is the number of e-folds between the two hypersurfaces, given by

$$N_e = -\sum_i^{N_f} \int_{\phi_{i,0}}^{\phi_{i,c}} \left(\frac{V_i}{V_i'}\right) d\phi_i \tag{3.3}$$

assuming the slow-roll conditions. If the sub-horizon evolution of the field perturbations gives no cross-correlations $\left\langle \delta\phi_i \delta\phi_j^* \right\rangle$ at horizon exit, then any sum-separable inflation model like Eq. (3.1) predicts a spectral index of [28, 31, 32]

$$n_s - 1 = -2\epsilon_* - 4\left[1 - \sum_i^{N_f} \frac{\eta_i^* u_i^2}{2\epsilon_i^*}\right]\left[\sum_i^{N_f} \frac{u_i^2}{\epsilon_i^*}\right]^{-1}, \tag{3.4}$$

a tensor-to-scalar ratio of

$$r = 16\left[\sum_i^{N_f} \frac{u_i^2}{\epsilon_i^*}\right]^{-1}, \tag{3.5}$$

and a running of the spectral index of

$$\alpha_s = -8\epsilon_*^2 + 4\sum_i^{N_f} \epsilon_i^* \eta_i^* - 16\left[\sum_i^{N_f} \frac{u_i^2}{\epsilon_i^*}\right]^{-2}\left[1 - \sum_i^{N_f} \frac{\eta_i^* u_i^2}{2\epsilon_i^*}\right]^2$$

$$- 8\left[\sum_i^{N_f} \frac{u_i^2}{\epsilon_i^*}\right]^{-1}\sum_i^{N_f} \eta_i^* u_i\left[1 - \frac{\eta_i^* u_i^2}{2\epsilon_i^*}\right]$$

$$+ \frac{4}{\epsilon_*}\left[\sum_i^{N_f} \frac{u_i^2}{\epsilon_i^*}\right]^{-1}\sum_i^{N_f} \frac{\eta_i^* u_i^2}{2\epsilon_i^*} - 2\left[\sum_i^{N_f} \frac{u_i^2}{\epsilon_i^*}\right]^{-1}\sum_i^{N_f} \frac{\xi_i^* u_i^2}{\epsilon_i^*}. \tag{3.6}$$

We define the slow-roll parameters as

$$\epsilon \equiv \sum_i^{N_f} \epsilon_i \equiv \sum_i^{N_f} \frac{1}{2} \left(\frac{V_i'}{V} \right)^2, \qquad \eta_i \equiv \frac{V_i''}{V}, \qquad \text{and} \qquad \xi_i^2 \equiv \frac{V_i' V_i'''}{V^2}, \qquad (3.7)$$

where $V_i' \equiv dV_i/d\phi_i$ and similarly for higher derivatives. We have also used the functions

$$u_i \equiv \frac{V_i^* + Z_i^c}{V} \qquad (3.8)$$

and

$$Z_i^c \equiv \frac{1}{\epsilon_c} \sum_j^{N_f} V_j^c \left[\epsilon_i^c - \left(\sum_k^{N_f} \epsilon_k^c \right) \delta_{ij} \right], \qquad (3.9)$$

where δ_{ij} is the Kronecker delta. Quantities labeled by $(*)$ are evaluated as modes leave the horizon with $k = aH$ and quantities labelled by (c) in Eq. (3.9) are evaluated at the end of inflation. Ignoring possible cross-correlations at horizon crossing is a very good approximation for this model, as shown in comparison to exact numerical results in Refs. [18, 20]. In this paper we will describe n_s, α_s, and r only near the pivot scale k_*.

We evaluate Eqs. (3.4)– (3.6) numerically with the publicly available package MULTI-MODECODE [33–36], which numerically evolves the Klein-Gordon equations

$$\frac{d^2\phi_i}{dN_e^2} + (3 - \epsilon) \frac{d\phi_i}{dN_e} + \frac{1}{H^2} \frac{\partial V}{\partial \phi_i} = 0 \qquad (3.10)$$

from some initial state $\{\phi_{i,0}, \phi_{i,0}'\}$ until the end of inflation when $\epsilon \equiv \sum_i \epsilon_i = 1$. Given masses m_i and the number of e-folds N_* between when the pivot scale k_* leaves the horizon and the end of inflation, the field values at horizon crossing $\phi_{i,*}$ and the functions u_i and Z_i^c can be calculated numerically. The details of this procedure can be found in Ref. [36].

The horizon crossing approximation (HCA).— We can simplify the expressions for u_i and Z_i^c if we ignore any contributions to the observables from the end-of-inflation surface, which sets $Z_i^c \to 0$. This is the *horizon crossing approximation* (HCA) [25, 31], which has been shown to be a reliable estimate for n_s and r, even in the many-field limit $N_f \gg 1$ of N_f-quadratic inflation [18, 20, 28]. We discuss this further in Sect. 3.2. For chaotic potentials like Eq. (3.1), the number of e-folds between horizon exit and the end of inflation from Eq. (3.3) is

$$N_* = \frac{1}{4} \sum_i^{N_f} \left[\phi_{i,0}^2 - \phi_{i,c}^2 \right] \approx \frac{1}{4} \sum_i^{N_f} \phi_{i,0}^2, \qquad (3.11)$$

where we have ignored the end-of-inflation contribution. Consequently, an arbitrarily large amount of inflation can be generated by setting the initial conditions $\phi_{i,0}$ to larger values. This sets a close relationship between N_e and its derivatives:

$$\sum_i \left(\frac{\partial N_*}{\partial \phi_i} \right)^2 = N_*. \qquad (3.12)$$

If the two-point correlation functions of the field perturbations at horizon crossing are diagonal $\left\langle \delta\phi_i \delta\phi_j^* \right\rangle \sim \delta_{ij}$, then the field-field power spectrum is related directly to the curvature power spectrum in Eq. (3.2) by N_*.

Substituting V from Eq. (3.1) into Eq. (3.5), we get

$$\mathcal{P}_\zeta = \frac{H^2}{4\pi^2} N_*, \qquad\qquad n_s - 1 = -2\epsilon - \frac{1}{N_*},$$

$$\alpha_s = -8\epsilon^2 - \frac{1}{N_*^2} + 4\epsilon_i\eta_i, \qquad\qquad r = \frac{8}{N_*}. \qquad\qquad (3.13)$$

Importantly, n_s and α_s depend directly on the background field values at horizon crossing ϕ_i^*, the masses of each field m_i^2, and the number of fields N_f through the slow-roll parameters ϵ_i and η_i. However, we will find that the relationship Eq. (3.12) ensures that N_* remains the most important parameter in this model for predicting n_s and α_s.

We will not use the HCA when computing observables for the models in Sects. 3.2–3.4. However, we assume that the chain of dependencies revealed by the HCA are sufficiently complete that we can use it to deduce the structure of the model's Bayesian networks, *i.e.*, any conditional dependencies induced by the end-of-inflation hypersurface are negligible in comparison to other dependencies at higher energies. This assumption will not be valid in general, but we leave the detailed analysis of more complicated potentials for future work.

Isocurvature, reheating, and N_*.— We compute the power spectrum at the end of inflation, when the evolution of the universe may be non-adiabatic. The field perturbations at the end of inflation may still have relative entropy perturbations that yield a non-adiabatic pressure perturbation $\delta P_{\rm nad}$, which sources a change in the super-horizon curvature perturbation [37–39]

$$\dot\zeta = -\frac{H}{\sum_i \dot\phi_i^2} \delta P_{\rm nad}, \qquad\qquad (3.14)$$

where an overdot indicates a derivative with respect to cosmic time.[4] Consequently, the primordial spectrum $P_\zeta(k)$ can continue to evolve until it reaches the adiabatic limit, where $\delta P_{\rm nad} \to 0$. We showed in Ref. [18] that the non-adiabatic power spectrum increases with the number of fields N_f for the potential (3.1), although what this means for reheating dependence is still unknown. However, Refs. [40–43] demonstrate that for two fields perturbative reheating may yield a significant change in the predicted value of n_s, r, and $f_{\rm NL}$, if there are significant isocurvature perturbations at the end of inflation. Similarly, for some masses this model can realize a curvaton-like mechanism that may suppress $r \sim 0$ after inflation ends [44, 45]. Consequently, the predicted distributions for n_s, α_s, and r found in Sects. 3.2–3.4 do not include super-horizon effects from reheating and may thus be altered by the largely uncertain physics of the post-inflationary universe.

In principle, N_* also depends on the post-inflationary expansion history by [34, 46]

$$N_* \approx 56.12 + \frac{1}{3(1+w_r)} \log\left(\frac{2}{3}\right) + \frac{1}{4}\log\left(\frac{V_*}{V_c}\right) \qquad\qquad (3.15)$$
$$+ \frac{1-3w_r}{12(1+w_r)} \log\left(\frac{\rho_r}{V_c}\right) + \log\left(\frac{V_*^{\frac{1}{4}}}{10^{16}\,\text{GeV}}\right),$$

which is a generalization of the standard matching equation [47, 48] to include an integrated equation of state w_r for the reheating epoch. The primary uncertainty in N_* comes from the

---

[4] We have also assumed that $\zeta \approx \mathcal{R}$ when $k \gg aH$, where $\mathcal{R}$ is the curvature perturbation on comoving hypersurfaces.
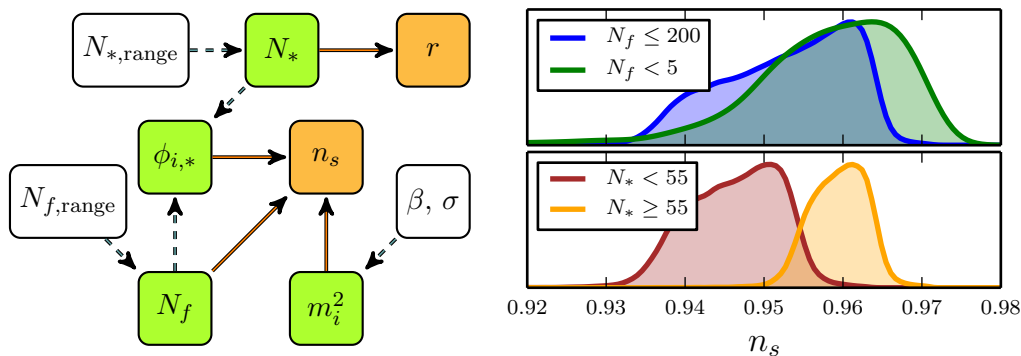
Figure 3: (*Left*) Bayesian network (BayesNet A) for N_f–quadratic inflation with masses m_i^2 distributed by Eq. (3.18). We fix $\beta = 0.5$ and $\sigma = 5 \times 10^{-6}$ and the number of fields is distributed as $N_f \sim \mathcal{U}[1, 200]$. Any background field values $\phi_{i,*}$ at horizon-crossing ($k_* = a_* H_*$) are equally likely according to the iso-N_* prior. We draw N_* uniformly from $N_* \sim \mathcal{U}[40, 70]$. (*Right*) Marginalized probability distributions $P(n_s)$ with varying ranges for N_f and N_*. The marginalized prediction is sensitively dependent only on the upper limit allowed for N_*.

unknown physics of reheating and the primordial dark ages [49] prior to neutrino decoupling, which is parametrized by w_r and the energy density ρ_r when the universe is guaranteed to have been thermalized. Consequently, a prior probability for N_* encodes an expectation on the dynamics of homogeneous background quantities during reheating. If we assume that $\rho_r^{1/4} \gtrsim \mathcal{O}(100 \, \text{MeV})$, then a matter-dominated equation of state $w_r = 0$ gives $N_* \gtrsim 40$, while a more exotic equation of state $w_r \approx 1$ gives an upper bound of $N_* \lesssim 70$ for N_f–quadratic inflation. Although the lower end of this range is generally favored *a priori* [33, 34, 48, 50], we set these values as our upper and lower limits for N_*.

3.2 BayesNet A: a simple model

Model.— Figure 3 displays the Bayesian network for our first version of N_f–quadratic inflation, a phenomenological description of assisted inflation around a minimum in a high dimensional field space. We use an uninformative, uniform prior probability for the number of e-folds of expansion $N_* \sim \mathcal{U}[40, 70]$. We set $N_f < 200$, which does not strongly affect the model's predictions, as we will discuss below, and choose a uniform prior probability over this range, $N_f \sim \mathcal{U}[1, 200]$. In Sect. 3.3 we update this prior to a maximum entropy prior.

We assume that the fields are slowly rolling, and the initial velocities $\dot{\phi}_{i,0}$ are those predicted by the slow-roll equations of motion (3.10). Following Refs. [18, 20, 28], we set the initial field values $\phi_{i,0}$ with the *iso-N_* prior*, a uniform probability distribution $f_{\text{iso}N_*}$ pulled back onto an N_f–dimensional hypersphere in field-space,

$$f_{\text{iso}N_*}(\phi_{i,0} \,|\, N_f) \propto \delta \left(\sum_i^{N_f} \phi_{i,0}^2 - r_{\text{iso}N_*} \right), \tag{3.16}$$

where δ is the Dirac delta function and $r_{\text{iso}N_*}$ is the radius of the hypersphere. We set $r_{\text{iso}N_*} = 2\sqrt{N_*}$ since, in the slow-roll limit, the number of e-folds between the initial condition and the end of inflation is approximately $N_{\text{tot}} \approx \frac{1}{4} \sum_i \phi_{i,0}^2$ if we ignore the contribution to N_{tot}

from $\phi_{i,c}$. Then $f_{\mathrm{iso}N_*}$ is defined on surfaces of constant e-foldings and by fixing $N_{\mathrm{tot}} = N_*$ we can identify the initial field values with those at horizon crossing $\phi_{i,0} = \phi_{i,*}$. This reduces the number of parameters in the model and allows efficient sampling of initial conditions, since sampling Eq. (3.16) amounts to picking uniformly distributed points on an $(N_f - 1)$-sphere. The specific choice of prior for the initial conditions has a very weak impact on the predictions for the primordial power spectra in this model [18].

Following Ref. [11] we assume that the masses m_i^2 are the eigenvalues of a random matrix

$$\mathcal{M}_{ij}^2 \sim A_{ik} A^k_{\ j}, \tag{3.17}$$

where A_{ij} is a real matrix with independent and identically distributed (iid) random entries that have a distribution with mean μ and standard deviation σ. We assume $\mu = 0$ and choose $\sigma = 5 \times 10^{-6}$ to obtain a primordial scalar spectrum that has amplitude $A_s \sim 10^{-9}$–10^{-10}. For convenience we fix σ, since it primarily affects only A_s, as demonstrated in Ref. [11]. If the number of fields is large, the distribution for the eigenvalues of $\mathcal{M}_{ij}^2$ limits toward the Marčenko-Pastur distribution with hyperparameters β and σ, defined by

$$P_{\mathrm{MP}}(m^2 \,|\, \sigma, \beta) = \frac{1}{2\pi m^2 \beta \sigma^2} \sqrt{(\beta_+ - m^2)(m^2 - \beta_-)}, \tag{3.18}$$

where $\beta_\pm \equiv \sigma^2(1 \pm \sqrt{\beta})^2$ and $\beta = \#^A_{\mathrm{row}}/\#^A_{\mathrm{colm}}$ is the ratio of the number of rows to the number of columns in the matrix A_{ij} in Eq. (3.17). We set $\beta = 0.5$, which is motivated by the expected ratio of axions to the total number of moduli in the stringy construction of Ref. [11]. This choice sets the typical mass range so that all of the masses are within an order of magnitude of each other.

Predictions.— The marginalized probability distribution for the tensor:scalar ratio r in the HCA can be obtained analytically by integrating Eq. (3.13) over N_* with the prior $N \sim \mathcal{U}[a, b]$, giving

$$P(r \,|\, a, b) = \begin{cases} [8/(b-a)] \ r^{-2} & 8/b \leq r \leq 8/a \\ 0 & \text{else} \end{cases}. \tag{3.19}$$

This distribution maximizes at $r = 8/b$, which is $r = 0.114$ for the upper limit on $N_* \leq 70$ for this model. Consequently, when using this generative model to predict the tensor-to-scalar ratio there will be a strong dependence on the upper limit that we allow for N_*. In practice we do not use the HCA, but solve Eq. (3.5) directly for r to obtain correct results to first order in the slow-roll parameters, but the two approaches are in good agreement, consistently with Ref. [18]. The generic prediction of $r \sim 0.1$ holds for any reasonable range of allowed values for N_*. Unless r is modified by reheating effects not incorporated in this analysis, this value is in tension with the latest results from the joint *Planck* and BICEP2/*Keck Array* [1, 12–16]. Here, N_f–quadratic inflation was chosen only as an example model to display the Bayesian hierarchical modelling techniques, but in a different inflation scenario with $r \lesssim 0.1$ the exact prediction for r and its relative dependencies on the hyperparameters would be highly important.

Figures 3 and 4 show the more complicated predictions for the spectral tilt n_s and scalar running α_s. Figure 3 plots $P(n_s)$ as a function of the hyperparameters that control the ranges of N_f and N_*. When the upper limit on the number of fields is between 20 and
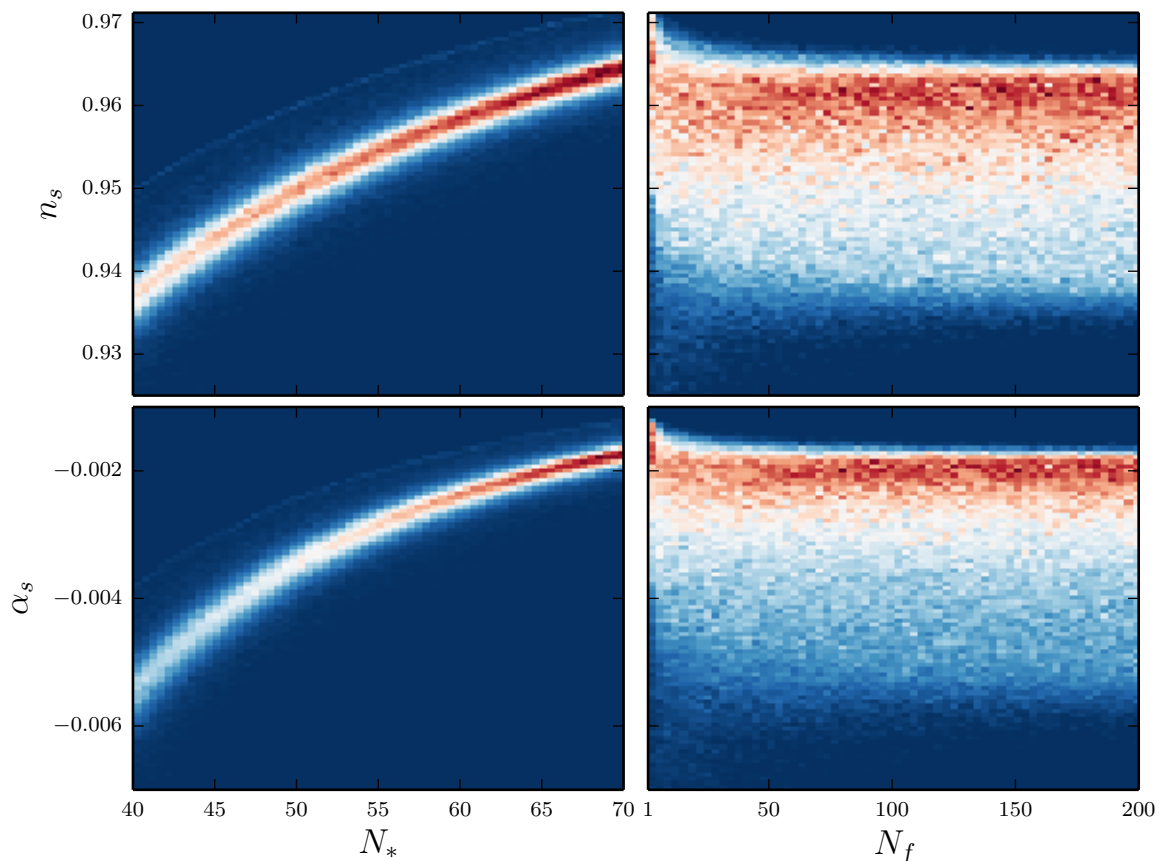
Figure 4: Histogram estimation for the marginalized distributions (*upper left*) $P(n_s \,|\, N_*)$, (*upper right*) $P(n_s \,|\, N_f)$, (*lower left*) $P(\alpha_s \,|\, N_*)$, and (*lower right*) $P(\alpha_s \,|\, N_f)$ of BayesNet A. Each vertical column of bins is independently normalized. The prediction for r is given by Eq. (3.19).

The marginalized distributions for n_s and α_s change only minimally for $N_f \gtrsim 10$ but depends strongly on the choice of N_*.

200 the predicted values for n_s or α_s do not change. For example, the distribution

$$P(n_s \,|\, N_f < 20) \equiv \sum_{i=1}^{20} P(n_s \,|\, N_f = i) \, P(N_f = i) \tag{3.20}$$

closely matches the plotted curve $P(n_s \,|\, N_f < 200)$ in Fig. 3. However, if the upper limit on the number of fields is less than ten, the marginal distribution $P(n_s)$ can depart substantially from the distribution $P(n_s \,|\, N_f < 200)$. Figure 4 plots the conditional probabilities $P(n_s \,|\, N_f)$ and $P(\alpha_s \,|\, N_f)$, demonstrating that the distributions quickly approach a limiting distribution

$$P_\infty(n_s) \equiv P(n_s \,|\, N_f \to \infty) \qquad \text{and} \qquad P_\infty(\alpha_s) \equiv P(\alpha_s \,|\, N_f \to \infty) \tag{3.21}$$

for $N_f \gtrsim 20$. Consequently, the marginalized distributions $P(n_s)$ and $P(\alpha_s)$ will be dominated by the contributions from the large-N_f region of parameter space:

$$P(n_s) \sim \int P(n_s \,|\, N_f) \, dN_f \sim P_\infty(n_s) \tag{3.22}$$
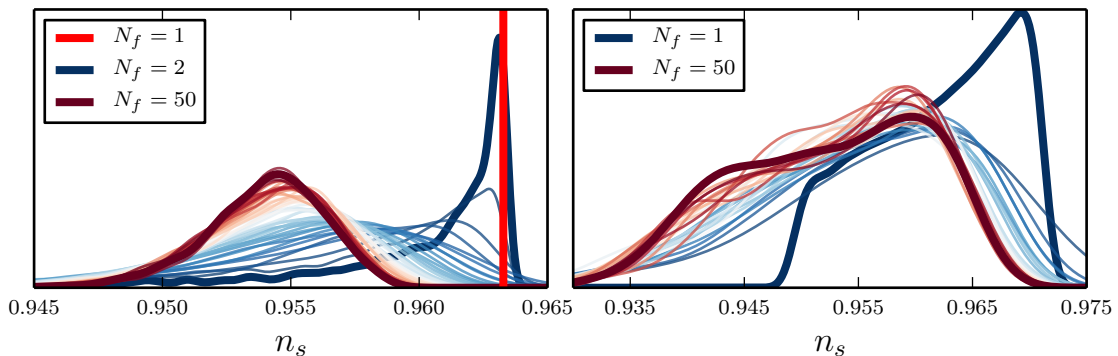
Figure 5: (*Left*) The probability distribution $P(n_s \,|\, N_f, N_* = 55)$ with $1 \leq N_f \leq 50$ and (*Right*) the probability distribution $P(n_s \,|\, N_f)$ with N_* marginalized over a uniform distribution $\mathcal{U}[40, 70]$ as in Fig. 4 for BayesNet A.

and

$$P(\alpha_s) \sim P_\infty(\alpha_s), \tag{3.23}$$

assuming that the upper limit for N_f is substantially larger than $N_f \gtrsim 20$.

Figure 5 illustrates the transition to a limiting distribution as N_f is increased, with N_* both fixed and as a free parameter. The distribution $P(n_s \,|\, N_f = 1, N_* = 55)$ is single-valued at $n_s = 0.963$, matching $m^2\phi^2$ inflation. As N_f is increased the peak of the distribution remains near $n_s = 0.963$ with an increasing variance for $N_f \lesssim 5$. For larger values of N_f the distribution approaches a normal distribution with mean of $n_s = 0.954$, as detailed in Ref. [18]. The central limit theorem suggests that the variance of the distribution for n_s will be inversely proportional to N_f. However, when N_* is marginalized, the transition to the limiting behavior $P_\infty(n_s)$ is very rapid. For single-field inflation, the predicted spectral tilt is $n_s \sim N_*^{-1}$ which can be analytically marginalized, as shown in Eq. (3.19). However, for more fields the distribution $P(n_s \,|\, N_f)$ quickly approaches a limiting case, which does not substantially differ over the range $2 < N_f < 50$ that is plotted in Fig. 5. We do not expect the variance of this distribution to vanish in the $N_f \rightarrow \infty$ limit due to the inherent uncertainty in N_* that is being integrated over in $P(n_s)$. Therefore, we can estimate the shape of $P_\infty(n_s)$ by the limiting curve in Fig. 5.

While they are largely independent of the total number of fields, $P(n_s)$ and $P(\alpha_s)$ depend strongly on N_*. Figure 3 gives the predicted spectral tilt when the upper and lower limits for the allowed value of N_* are changed. If N_* is uniformly distributed in the range $55 \leq N_* \leq 70$, the mean prediction is $\langle n_s \rangle = 0.959$ with a standard deviation of $\sigma = 3.89 \times 10^{-3}$. However, if N_* is uniformly distributed in the range $40 \leq N_* < 55$, then $\langle n_s \rangle = 0.946$ and the standard deviation grows to $\sigma = 6.03 \times 10^{-3}$. As in Eq. (3.19) for the distribution of r, $P(n_s)$ is most strongly dependent on the upper limit allowed for N_*. The left and right columns of Fig. 4 show that for all values of N_* the variance of the distributions $P(n_s \,|\, N_*)$ and $P(\alpha_s \,|\, N_*)$, which marginalizes over N_f, is significantly less than $P(n_s \,|\, N_f)$ and $P(\alpha_s \,|\, N_f)$, which marginalizes over N_*. While this demonstrates qualitatively that N_* is the most important parameter for the prediction of n_s and α_s, not all observables will be sensitively dependent on N_* in this model, *e.g.*, the prediction for the tensor consistency relation $P(n_t/r)$ has a variance that scales like $\sigma^2 \sim 1/N_f$ even after marginalizing over N_* [20].

The effectiveness of the HCA.— In the horizon crossing approximation (HCA) we ignore the contribution of the field values at the end of inflation to the total number of e-folds of expansion during inflation, as in Eq. (3.11). While this approximation holds for $\phi_{i,*} \gg \phi_{i,c}$, Fig. 2 shows that N_f–quadratic inflation has many light fields ϕ_i with small masses, which evolve only minimally during inflation. However, even though this model does not always have a large hierarchy between the horizon crossing and end-of-inflation field values, the marginalized predictions obtained in this Section are still closely approximated by the HCA. We interpret this as a consequence of the relative unimportance of the parameter N_f in determining the primordial power spectrum in this model.

The model's predictions depend on the rate of change of the number of e-folds as a function of the horizon crossing field values [32], which is given by

$$dN = \left(\frac{\partial N}{\partial \phi_{i,*}} + \frac{\partial N}{\partial \phi_{j,c}} \frac{\partial \phi_{j,c}}{\partial \phi_{i,*}} \right) d\phi_{i,*} \tag{3.24}$$
$$= \sum_{j=1}^{N_f} \left[\left(\frac{V_j}{V_j'} \right)_* - \sum_i \frac{\partial \phi_{i,c}}{\partial \phi_{j,*}} \left(\frac{V_j}{V_j'} \right)_c \right] d\phi_{j,*}.$$

The HCA is equivalent to ignoring the second term in the brackets. To zeroth order the field values for the lightest fields at horizon crossing are approximately the same as at the end of inflation, which yields

$$\frac{\partial \phi_{i,c}}{\partial \phi_{j,*}} \approx \delta_{ij}. \tag{3.25}$$

Substituting Eq. (3.25) into Eq. (3.24) reduces the range of the summation so that only the fields that evolve substantially during inflation contribute to the derivatives of N, which reduces the effective N_f for this model. Since Fig. 4 shows that the marginalized predictions for the spectral tilt and scalar running reach a limiting distribution for $N_f \gtrsim 20$, this moderate reduction of N_f will not typically affect the prediction at the end of inflation. Since the HCA is a good estimate for those fields with large field displacements during inflation and since removing the light fields from the summation does not change the prediction, the HCA should therefore closely resemble the exact solution for this model even when $\phi_{i,*} \sim \phi_{i,c}$.

3.3 BayesNet B: maximum entropy prior for N_f

Model.— In BayesNet A of Sect. 3.2 the predictions for n_s and α_s do not depend sensitively on the number of fields and quickly approaches a limiting value for $N_f \gtrsim 20$. However, if the prior probability for the number of fields were weighted toward $N_f \sim 1$ this parameter may have more impact on the model's predictions. Since we had no fundamental reason for choosing the uniform distribution $N_f \sim \mathcal{U}[1, 200]$ in Sect. 3.2, we now assess how the results of the previous model depend on the chosen prior distribution for N_f.

The resulting Bayesian network is similar to that of Fig. 3, but uses a maximum entropy prior for N_f, *i.e.*, a prior probability that contains the minimal amount of Shannon information [51]. We allow $N_f = \{1, 2, \dots\}$ and fix the expected size of N_f *a priori*, which is necessary if we are to define a maximum entropy distribution over the positive integers. The geometric prior probability $\mathcal{G}_p(N_f)$ is the probability mass function that maximizes the Shannon entropy for this range of allowed values,

$$P\left(N_f \mid \langle N_f \rangle\right) \equiv \mathcal{G}_p(N_f) = p\,(1-p)^{N_f - 1} \tag{3.26}$$
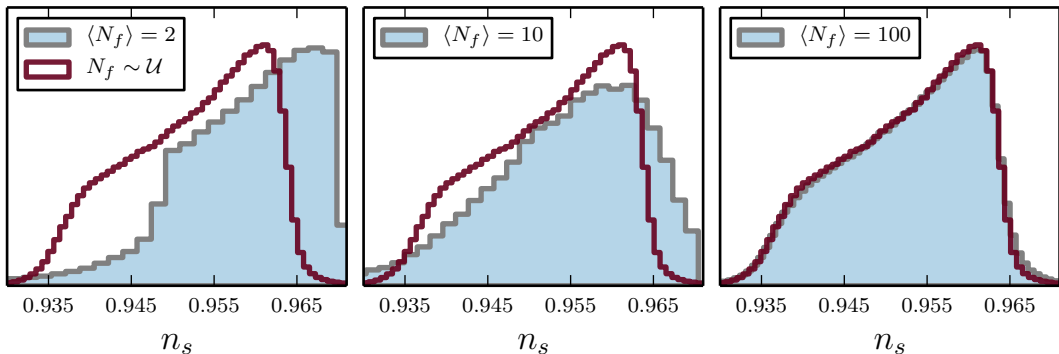
Figure 6: Histogram estimation of $P(n_s \,|\, \langle N_f \rangle)$ for BayesNet B with $N_f \sim \mathcal{G}_p$ in Eq. (3.26) as compared to the prediction from the uniform prior $N_f \sim \mathcal{U}[1, 200]$ of BayesNet A.

for $p \equiv 1/\langle N_f \rangle$. The mean and variance of $\mathcal{G}_p$ is p and $(1 - p)/p^2$, respectively. The distribution peaks at $N_f = 1$ for all p, giving a larger weighting toward lower numbers of fields relative to a uniform distribution. We then add the *a priori* expected number of fields $\langle N_f \rangle$ to the Bayesian network of Fig. 3 as a hyperparameter for N_f. Although this seems like additional structure that we have added to BayesNet A, the principle of maximum entropy ensures that our arbitrary choice of boundary for the uniform prior $N_f \sim \mathcal{U}[1, 200]$ contains more assumptions than Eq. (3.26) when measured according to the Shannon information.

Predictions.— Figure 6 compares the marginalized prediction $P(n_s \,|\, \langle N_f \rangle)$ for three different choices of $\langle N_f \rangle$ to the uniformly distributed case of Sect. 3.2. We plot distributions for a model with few expected fields $\langle N_f \rangle = 2$, an intermediate number of fields $\langle N_f \rangle = 10$, and many fields $\langle N_f \rangle = 100$, where the specific numbers were chosen based on the behavior of $P(n_s \,|\, N_f)$ in Fig. 4. If we pick $\langle N_f \rangle = 1$, then there is zero probability for $N_f > 1$ and we recover the single-field inflation results. If we expect many fields $\langle N_f \rangle = 100$, then we obtain the same limiting distribution $P_\infty(n_s)$ that we saw in Fig. 5. Out of a sample of $975,000$ points the largest simulated value was $N_f = 983$ and 50% of the sample had $N_f < 70$. The mean predicted value of the spectral index is $\langle n_s \rangle = 0.953$, with a standard deviation of $\sigma = 8.73 \times 10^{-3}$. If we expect a large number of fields, then the marginalized prediction will be largely independent of the exact value of N_f.

As we expect from $P(n_s \,|\, N_f)$ in Fig. 4, for lower $\langle N_f \rangle$ the scalar spectrum is marginally more blue. However, it is not well approximated by the single-field prediction, which has a mean value of $\langle n_s \rangle = 0.962$, due to the sharp transition in the functional form of $P(n_s \,|\, N_f)$ when the number of fields changes from $N_f = 1$ to $N_f = 2$, as displayed in Fig. 5. For $\langle N_f \rangle = 2$ the vast majority of samples have $N_f \leq 2$ and the largest simulated value was $N_f = 20$ in a $199,000$ element sample. Approximately 50% of this sample has $N_f > 1$, which results in a redder spectrum than if N_f were fixed to unity, with a mean prediction of $\langle n_s \rangle = 0.958$ and a standard deviation of $\sigma = 1.47 \times 10^{-2}$. For $\langle N_f \rangle = 10$ the distribution for n_s is intermediate between these two cases, giving $\langle n_s \rangle = 0.955$ and $\sigma = 1.29 \times 10^{-2}$. Consequently, if we expect $\langle N_f \rangle \gtrsim 10$, then the predictions for n_s from the maximum entropy prior for N_f should be well approximated by the results of Sect. 3.2.
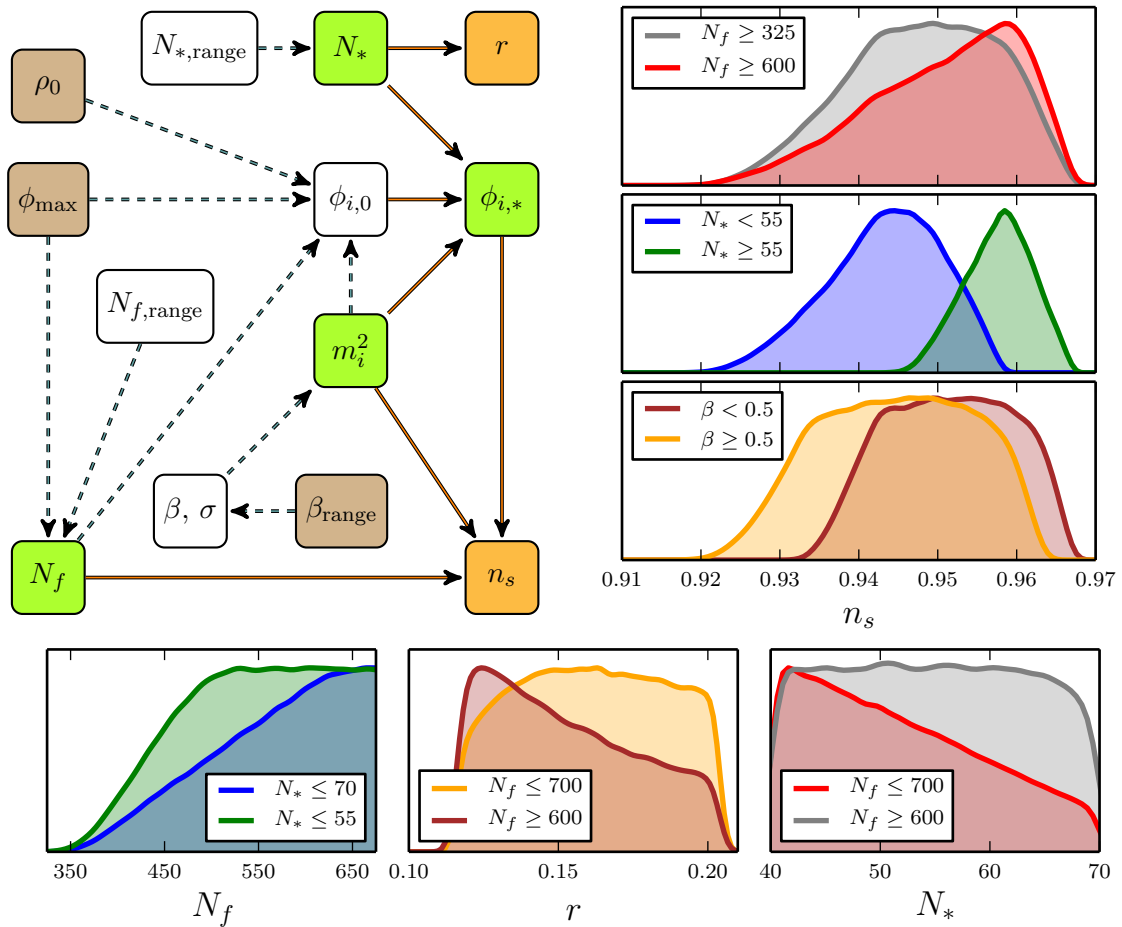
Figure 7: (*Top-left*) BayesNet C with initial energy $\rho_0 = M_{\rm Pl}$ and restrictions $|\phi_{i,0}| < M_{\rm Pl} \equiv \phi_{\max}$ and $N_{\rm tot} > N_*$. The priors are $m_i^2 \sim$ Eq. (3.18), $\sigma = 5 \times 10^6$, and $\beta \sim \mathcal{U}[0.25, 0.75]$; we also sample $N_f \sim \mathcal{U}[200, 700]$, $\phi_{i,0}(\rho_0) \sim$ iso$-E$, and $N_* \sim \mathcal{U}[40, 70]$, but enforce the restriction $N_{\rm tot} > N_*$ (Eq. (3.29)) after N_f, $\phi_{i,0}$, and N_* are sampled, which results in non-standard priors for these parameters, which we plot in Fig. 8. (*Top-right*) Marginalized distributions for n_s with varying hyperparameters. (*Bottom*) Marginalized prediction for the tensor-to-scalar ratio r and the non-uniform priors $P(N_f)$ and $P(N_*)$, with varying hyperparameter ranges.

3.4 BayesNet C: high-energy dependence

Model.— In Sects. 3.2 and 3.3 we restricted the dimensionality of parameter space by equating the initial conditions $\phi_{i,0}$ and horizon crossing values $\phi_{i,*}$ by using the iso-N_* prior, as well as fixing the hyperparameter $\beta = 0.5$ in the mass distribution of Eq. (3.18). Here we relax some of these assumptions and demonstrate how the predictions of the inflation model change when a dependence on an initial energy density ρ_0 is incorporated.

Figure 3 is a Bayesian network for a model that involves a higher number of parameters and conditional dependencies for N_f–quadratic inflation. We use the same uniform prior $N_* \sim \mathcal{U}[40, 70]$ as in Sect. 3.2 and draw masses from the Marčenko-Pastur distribution (3.18), but allow the hyperparameter β to vary uniformly in the range $\beta \sim \mathcal{U}[0.25, 0.75]$. We impose

a uniform prior on N_f, with the range $N_f \sim \mathcal{U}[1, 700]$, where the upper bound is chosen to ensure that the numerical calculations remain tractable. We set the background initial conditions $\phi_{i,0}$ and $\dot{\phi}_{i,0}$ according to the *iso-E* prior from Refs. [18, 52, 53]. This is a uniform probability distribution defined on a surface of constant energy density ρ_0, which is defined as

$$f_{\mathrm{iso}E}(\phi_{i,0}, \dot{\phi}_{i,0}) \propto \delta\left(\frac{1}{2} \sum_i^{N_f} \left[\dot{\phi}_{i,0}^2 + m_i^2 \phi_{i,0}^2 \right] - \rho_0 \right), \tag{3.27}$$

where δ is the Dirac delta function. We fix the initial energy scale via the conservative choice $\rho_0 = M_{\mathrm{Pl}}$, since at this scale new physics is almost guaranteed to appear. Refs. [54, 55] have a similar prior for initial conditions defined on surfaces of constant ρ, finding that the predictions resulting from this prior are independent of the relatively arbitrary choice of ρ_0. However, Ref. [18] demonstrated that the exact choice of initial conditions prior has a relatively minor effect on the predicted power spectrum, so we will assume Eq. (3.27) sufficiently captures the initial conditions dependence.

To preserve the interpretation of the N_f–quadratic potential as a Taylor expansion around a local minimum we need to specify the domain of validity of the expansion, requiring

$$|\phi_{i,0}| \leq \phi_{\max} \equiv M_{\mathrm{Pl}}. \tag{3.28}$$

The basis in which the domain of validity is given by Eq. (3.28) is not in general the same as the basis which diagonalises the mass matrix [56, 57], but we will assume here that the two bases are the same.

Since $\phi_{i,0} \neq \phi_{i,*}$ we also require that the total number of e-folds N_{tot} between the initial state and the end of inflation exceeds N_*. In the slow-roll limit and ignoring contributions from the end of inflation, Eq. (3.11) gives the joint constraint

$$\frac{1}{4} \sum_i^{N_f} \phi_{i,0}^2 \gtrsim N_*, \tag{3.29}$$

which implies that $N_f \gg 4N_*$ given the maximum value of $\phi_{i,0} < 1$. These conditions were not needed in Sects. 3.2 and 3.3 as the models allowed super-Planckian field displacements.

We sample $f_{\mathrm{iso}E}$ subject to the additional constraint (3.28) using the techniques described in Ref. [52] and enforce $N_{\mathrm{tot}} > N_*$ with rejection sampling, *i.e.*, discarding those samples that do not realize enough e-folds of inflation. Rejection sampling alters the naïve priors $N_f \sim \mathcal{U}$, $N_* \sim \mathcal{U}$, $m_i^2 \sim P_{\mathrm{MP}}$, or $\phi_{i,0} \sim f_{\mathrm{iso}E}$. The deviation from the naïve priors depends on the order in which the conditional dependencies in the Bayesian network in Fig. 7 are sampled. Here, we first select N_f, then the initial field values, then N_* from their naïve priors and throw away all combinations of parameters $\{N_f, N_*, \phi_{i,0}, \dots\}$ that would violate Eq. (3.29). Conversely, after generating N_f and the initial conditions we could have sampled an *a priori* conditional distribution $P(N_* \,|\, N_* \lesssim N_{\mathrm{tot}})$, where we calculate N_{tot} from the equations of motion. If N_{tot} for a given configuration is larger than the minimum allowed value for N_*, then N_f preserves a uniform prior, with only $P(N_*)$ changed. These two priors should give different predictions, since they will place different weightings on N_*. Furthermore, they describe different expectations on the physics of reheating and post-inflationary evolution, since N_* depends on the entire expansion history of the universe. We will show that the constraints in Eqs (3.28)–(3.29) introduce a change in the predictions from this
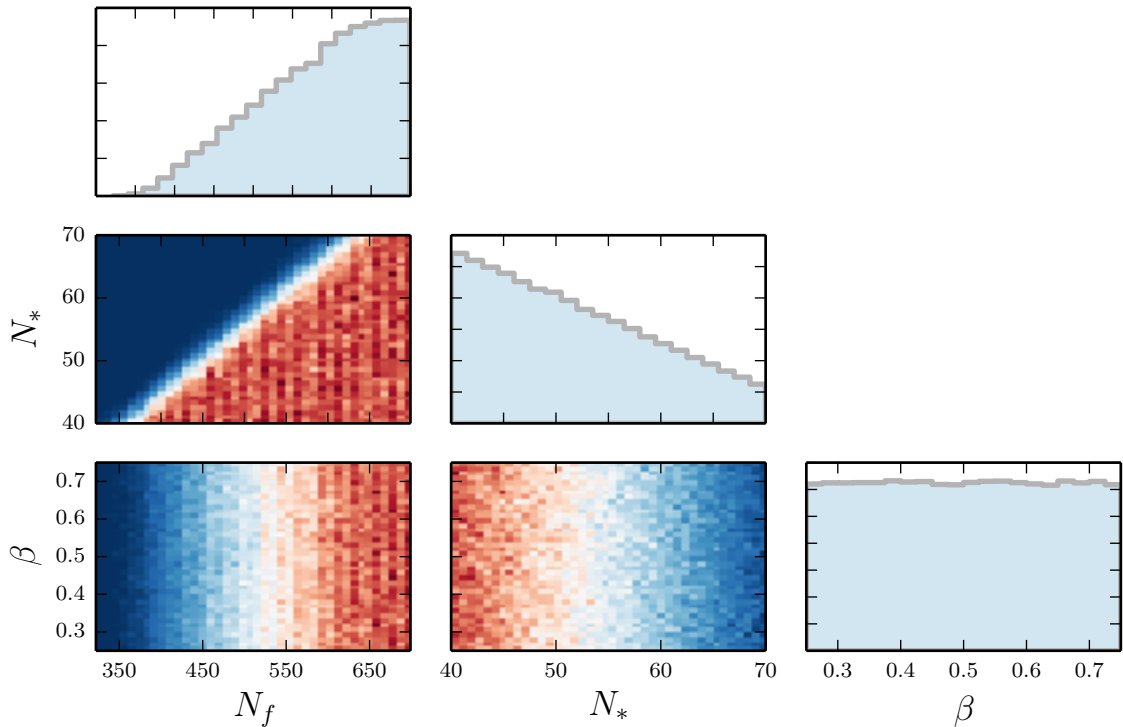
Figure 8: Histogram-estimated prior probability distributions for BayesNet C, after enforcing the constraint (3.29) by rejection sampling. These are not generally uniform, since we require sub-Planckian field displacements (3.28).

Bayesian network that dominates any differences resulting from the other novel parameters in the model.

The non-uniform prior probability on N_f and N_* that results from our rejection sampling technique can be seen in Figs. 7 and 8. We set the minimum value of N_* at 40 e-folds, which requires a minimum of $N_f \gtrsim 160$ to get a sufficient number of e-folds of inflation, but more typically needs $N_f \gtrsim 350$. If N_f is relatively small, we need a larger field displacement to ensure $N_{\rm tot} > N_*$, which conflicts with the maximum allowed range for the initial field values. Similarly, if N_* is relatively large, then more fields are required to get enough inflation. This results in an *a priori* preference for smaller values of N_* and larger values of N_f as compared to the uniform priors. However, the parameter β, which sets the variance of the prior for the masses (3.18), is not affected by these considerations, as shown in Fig. 8. In the slow-roll calculation, the total number of e-folds does not depend on the masses (3.11), which gives β a negligible dependence on the non-trivial prior space for N_*.

Predictions.— The prior probability distribution $P(N_*)$ approaches the uniform distribution $\mathcal{U}[40, 70]$ as the number of fields grows large. Figure 7 gives a nearly flat distribution for N_* if we restrict the minimum number of fields to $N_f \geq 600$, which is modified so that it is approximately inversely proportional to N_* when the full range $N_f \leq 700$ are included. The non-uniform prior on N_* gives a marginalized distribution for r which differs from Eq. (3.19), which is also plotted in Fig. 7. Since $r \propto N_*^{-1}$ the approximate inverse power law for $P(N_*)$ yields an almost uniform distribution for the tensor-to-scalar ratio in
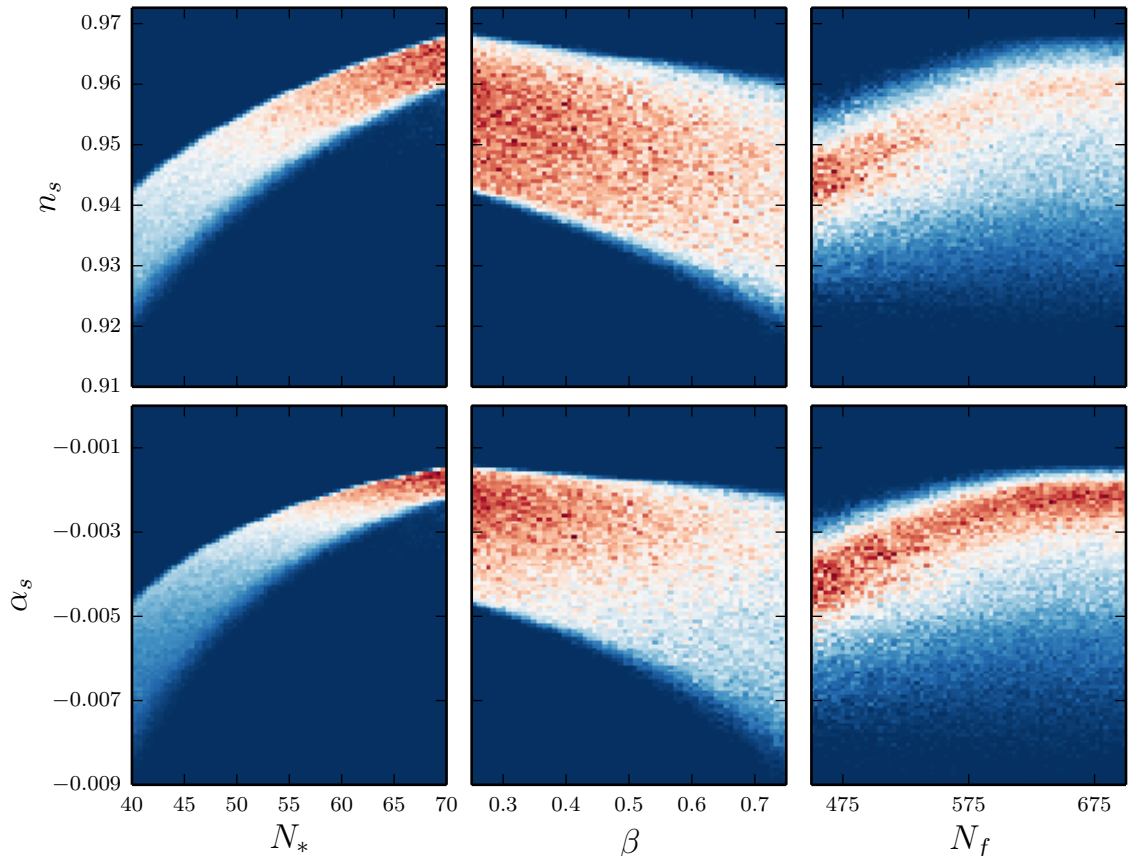
Figure 9: Histogram estimates for the distributions (*left*) $P(n_s \,|\, N_*)$ and $P(\alpha_s \,|\, N_*)$, (*middle*) $P(n_s \,|\, \beta)$ and $P(\alpha_s \,|\, \beta)$, and (*right*) $P(n_s \,|\, N_f)$ and $P(\alpha_s \,|\, N_f)$ for BayesNet C. Each vertical column of bins is independently normalized as a PDF. The distributions depend strongly on N_* and $N_f \lesssim 600$ and moderately on β. The non-uniform prior probability $P(N_f)$ in Fig. 7 makes the predictions moderately dependent on N_f. The prediction for r does not change radically from Fig. 7.

the range $0.11 < r < 0.21$. This deemphasizes the effect of the upper limit of N_* on the marginalized distribution $P(r)$, which was seen in Eq. (3.19). The prior probability for N_f is also given in Fig. 7 and begins to approach a flat distribution for smaller values of N_*. For $N_* \leq 55$ almost all $N_f \gtrsim 500$ are almost equally likely, with a decreasing probability for smaller numbers of fields. In our sample, which is composed of $250,000$ elements, the smallest value was $N_f = 322$, which sets an effective lower bound on the expected number of fields in this model.

Figures 7 and 9 display probability distributions for n_s and α_s. The marginalized distributions $P(n_s)$ in Fig. 7 are plotted with different allowed ranges of N_f, N_*, and β. The distributions for both n_s and α_s are sensitively dependent on N_* and moderately dependent on β and $N_f \lesssim 500$, which can be seen by the relative thickness of the distributions in Fig. 9. Keeping β in the range $0.25 \leq \beta \leq 0.50$ gives a mean prediction of $\langle n_s \rangle = 0.951$ with standard deviation $\sigma = 7.80 \times 10^{-3}$, while the range $0.50 \leq \beta \leq 0.75$ gives $\langle n_s \rangle = 0.945$ and $\sigma = 9.50 \times 10^{-3}$. In comparison, if we restrict $55 \leq N_* \leq 70$, then the mean increases to

$\langle n_s \rangle = 0.958$ with standard deviation $\sigma = 4.26 \times 10^{-3}$, while the range $40 \leq N_* \leq 55$ gives the redder value of $\langle n_s \rangle = 0.943$ and $\sigma = 7.25 \times 10^{-3}$. The dependence on small N_f in the marginalized prediction results from the covariance between N_f and N_*, since smaller values of N_f need smaller N_*, which gives a redder spectrum as seen in Fig. 9. The plot for N_f in Fig. 9 is truncated at 450 because the lower ranges of N_f are not well sampled and the histogram estimator is noisy.

We see from this analysis that imposing a simple bound on the field values specified by Eq. (3.28) has dramatic implications for the model's prediction. Interestingly, although the priors for N_* and N_f incorporate information about different sectors of fundamental particle theory, this requirement creates a joint *a priori* dependency between these two hyperparameters.[5] Consequently, the parameter N_* affects the allowed number of fields, causing N_f to have more impact on the distributions $P(n_s \,|\, N_f)$ and $P(\alpha_s \,|\, N_f)$ than it does for BayesNet A or BayesNet B. However, at large N_f we see behaviour similar to BayesNet A and BayesNet B, since the *a priori* covariance between N_f and N_* vanishes in this limit according to Fig. 8. We conclude that the unknown reheating physics, as encoded in the joint prior $P(N_*, N_f)$, remains vitally important for the predictions of BayesNet C, despite the more complicated dependence on N_f and β in Figs. 7 and 9.

4   Summary & conclusions

In Sect. 2 we introduced generative hierarchical models and Bayesian networks. We emphasized the role played by hyperparameters and hyperpriors in fixing the predictions of specific inflationary models. While more layers of hyperparameters increases the variance in the predicted distributions of observables, the results are more robust in that they more fully reflect the unknown physics of inflation.

In Sect. 3.2 we looked at a phenomenologically motivated Bayesian network with minimal dependencies between the fields' initial conditions $\phi_{i,0}$, masses m_i^2, the total number of fields, and N_*. We required that the model gives exactly N_* e-folds of expansion and stipulated that the prior probability on the initial conditions is uniformly weighted on the field space regions satisfying this relationship. We chose the total number of e-folds uniformly from the range $40 \leq N_* \leq 70$, the number of fields uniformly from the range $1 \leq N_f \leq 200$, and set the prior probability for the fields' masses to the Marčenko-Pastur distribution. Examining the marginalized distributions $P(n_s)$ and $P(\alpha_s)$ we found that the most important parameter is N_*, indicating that the inherent uncertainty in the predictions of N_f–quadratic inflation are far more sensitively dependent on the unknown physics of reheating than on the fields' masses, initial conditions, or the total number of active fields. The dominant source of uncertainty in this model is therefore the same as that which plagues even the simplest single field models.

Section 3.3 used the same Bayesian network as Sect. 3.2, but updated the prior probability for the number of fields to the geometric distribution, which maximizes the Shannon entropy, with the *a priori* expected number of fields $\langle N_f \rangle$ entering as new hyperparameter. Although this prior substantially reweights the number of active fields in a given inflationary realization, the model's predictions are indistinguishable from BayesNet A of Sect. 3.2 for $\langle N_f \rangle \gtrsim 20$, but are noticeably different for $\langle N_f \rangle \ll 20$.

---

[5]Note that N_f is fixed within the inflationary sector, while N_* depends on the integrated post-inflationary expansion of the universe, and is thus sensitive to unknown physics down to LHC scales, and possibly below [34, 46].

Section 3.4 explored the dependency on the initial energy scale ρ_0 in a model where the fields are required to have sub-Planckian displacements $|\phi_{i,0}| \ll M_{\mathrm{Pl}}$, while varying the expected ratio of axions to the total number moduli fields in the string theory motivated version of N_f–quadratic inflation of Ref. [11]. The reduced range for $\phi_{i,0}$ requires more fields to get at least N_* e-folds of inflation [21, 22]. Since lower values of N_* need fewer fields to achieve enough inflation, there is a significant coupling between the prior probabilities for N_* and N_f, resulting in predictions for n_s and α_s differing from BayesNet A or BayesNet B. Examining the marginalized distributions for n_s and α_s, we still qualitatively identify N_* as the most important parameter for determining the model's prediction, although there are now more complicated interdependencies between N_*, β, and N_f. With a non-trivial prior for N_f and allowing the ratio of axions to the total moduli to vary in the mass distribution, this model has a greater dependence on the total number of fields and masses than BayesNet A or BayesNet B, especially for $N_f \lesssim 600$. Despite this, we can identify a limiting behavior for large N_f where N_* is again clearly the most important parameter.

While our results are restricted to the case of N_f–quadratic inflation, the hierarchical modelling methods that we have used can be applied to any inflation model. These techniques would be especially well suited to studying models that have non-standard features, such as couplings to other sectors and non-canonical kinetic terms. Most importantly, the methods developed here will be indispensable for obtaining posterior probabilities on the parameters of complex inflation models with CMB or large scale structure data.

Acknowledgments

References

[1] **Planck** Collaboration, P. Ade et al., *Planck 2013 results. XXII. Constraints on inflation*, *Astron.Astrophys.* **571** (2014) A22, [arXiv:1303.5082].

[2] **Planck** Collaboration, P. A. R. Ade et al., *Planck 2015 results. XX. Constraints on inflation*, arXiv:1502.02114.

[3] **WMAP** Collaboration, G. Hinshaw et al., *Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results*, *Astrophys. J. Suppl.* **208** (2013) 19, [arXiv:1212.5226].

[4] K. S. Mandel, W. M. Wood-Vasey, A. S. Friedman, and R. P. Kirshner, *Type Ia Supernova Light Curve Inference: Hierarchical Bayesian Analysis in the Near Infrared*, Astrophys. J. **704** (2009) 629–651, [arXiv:0908.0536].

[5] M. C. March, R. Trotta, P. Berkes, G. D. Starkman, and P. M. Vaudrevange, *Improved constraints on cosmological parameters from SNIa data*, Mon. Not. Roy. Astron. Soc. **418** (2011) 2308–2329, [arXiv:1102.3237].

[6] K. S. Mandel, R. J. Foley, and R. P. Kirshner, *Type Ia Supernova Colors and Ejecta Velocities: Hierarchical Bayesian Regression with Non-Gaussian Distributions*, Astrophys. J. **797** (2014), no. 2 75, [arXiv:1402.7079].

[7] D. Rubin, G. Aldering, K. Barbary, K. Boone, G. Chappell, et al., *UNITY: Confronting Supernova Cosmology's Statistical and Systematic Uncertainties in a Unified Bayesian Framework*, arXiv:1507.01602.

[8] H. Shariff, X. Jiao, R. Trotta, and D. A. van Dyk, *BAHAMAS: new SNIa analysis reveals inconsistencies with standard cosmology*, arXiv:1510.05954.

[9] S. M. Feeney, M. C. Johnson, J. D. McEwen, D. J. Mortlock, and H. V. Peiris, *Hierarchical Bayesian Detection Algorithm for Early-Universe Relics in the Cosmic Microwave Background*, Phys. Rev. **D88** (2013) 043012, [arXiv:1210.2725].

[10] J. Alsing, A. Heavens, A. H. Jaffe, A. Kiessling, B. Wandelt, and T. Hoffmann, *Hierarchical Cosmic Shear Power Spectrum Inference*, arXiv:1505.07840.

[11] R. Easther and L. McAllister, *Random matrices and the spectrum of N-flation*, JCAP **0605** (2006) 018, [hep-th/0512102].

[12] **Planck** Collaboration, P. Ade et al., *Planck 2013 results. XVI. Cosmological parameters*, Astron.Astrophys. **571** (2014) A16, [arXiv:1303.5076].

[13] M. J. Mortonson and U. Seljak, *A joint analysis of Planck and BICEP2 B modes including dust polarization uncertainty*, JCAP **1410** (2014), no. 10 035, [arXiv:1405.5857].

[14] R. Flauger, J. C. Hill, and D. N. Spergel, *Toward an Understanding of Foreground Emission in the BICEP2 Region*, JCAP **1408** (2014) 039, [arXiv:1405.7351].

[15] **BICEP2, Planck** Collaboration, P. Ade et al., *Joint Analysis of BICEP2/Keck Array and Planck Data*, Phys. Rev. Lett. **114** (2015) 101301, [arXiv:1502.00612].

[16] **Planck** Collaboration, P. A. R. Ade et al., *Planck 2015 results. XIII. Cosmological parameters*, arXiv:1502.01589.

[17] E. Di Valentino, A. Melchiorri, and J. Silk, *Beyond six parameters: extending ΛCDM*, arXiv:1507.06646.

[18] R. Easther, J. Frazer, H. V. Peiris, and L. C. Price, *Simple predictions from multifield inflationary models*, Phys.Rev.Lett. **112** (2014) 161302, [arXiv:1312.4035].

[19] D. Wenren, *Tilt and Tensor-to-Scalar Ratio in Multifield Monodromy Inflation*, arXiv:1405.1411.

[20] L. C. Price, H. V. Peiris, J. Frazer, and R. Easther, *Gravitational wave consistency relations for multifield inflation*, Phys.Rev.Lett. **114** (2015), no. 3 031301, [arXiv:1409.2498].

[21] A. R. Liddle, A. Mazumdar, and F. E. Schunck, *Assisted inflation*, Phys.Rev. **D58** (1998) 061301, [astro-ph/9804177].

[22] S. Dimopoulos, S. Kachru, J. McGreevy, and J. G. Wacker, *N-flation*, JCAP **0808** (2008) 003, [hep-th/0507205].

[23] P. Kanti and K. A. Olive, *Assisted chaotic inflation in higher dimensional theories*, Phys.Lett. **B464** (1999) 192–198, [hep-ph/9906331].

[24] S. A. Kim and A. R. Liddle, *Nflation: multi-field inflationary dynamics and perturbations*, *Phys.Rev.* **D74** (2006) 023513, [`astro-ph/0605604`].

[25] S. A. Kim and A. R. Liddle, *Nflation: Non-Gaussianity in the horizon-crossing approximation*, *Phys.Rev.* **D74** (2006) 063522, [`astro-ph/0608186`].

[26] S. A. Kim and A. R. Liddle, *Nflation: observable predictions from the random matrix mass spectrum*, *Phys.Rev.* **D76** (2007) 063515, [`arXiv:0707.1982`].

[27] S. A. Kim, A. R. Liddle, and D. Seery, *Non-gaussianity in axion N-flation models: detailed predictions and mass spectra*, *Phys.Rev.* **D85** (2012) 023532, [`arXiv:1108.2944`].

[28] J. Frazer, *Predictions in multifield models of inflation*, *Journal of Cosmology and Astroparticle Physics* **2014** (2013), no. 01 028, [`arXiv:1303.3611`].

[29] **Planck** Collaboration, P. A. R. Ade et al., *Planck 2013 Results. XXIV. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **571** (2014) A24, [`arXiv:1303.5084`].

[30] **Planck** Collaboration, P. A. R. Ade et al., *Planck 2015 results. XVII. Constraints on primordial non-Gaussianity*, [`arXiv:1502.01592`].

[31] F. Vernizzi and D. Wands, *Non-gaussianities in two-field inflation*, *JCAP* **0605** (2006) 019, [`astro-ph/0603799`].

[32] T. Battefeld and R. Easther, *Non-Gaussianities in Multi-field Inflation*, *JCAP* **0703** (2007) 020, [`astro-ph/0610296`].

[33] M. J. Mortonson, H. V. Peiris, and R. Easther, *Bayesian Analysis of Inflation: Parameter Estimation for Single Field Models*, *Phys.Rev.* **D83** (2011) 043505, [`arXiv:1007.4205`].

[34] R. Easther and H. V. Peiris, *Bayesian Analysis of Inflation II: Model Selection and Constraints on Reheating*, *Phys.Rev.* **D85** (2012) 103533, [`arXiv:1112.0326`].

[35] J. Norena, C. Wagner, L. Verde, H. V. Peiris, and R. Easther, *Bayesian Analysis of Inflation III: Slow Roll Reconstruction Using Model Selection*, *Phys.Rev.* **D86** (2012) 023505, [`arXiv:1202.0304`].

[36] L. C. Price, J. Frazer, J. Xu, H. V. Peiris, and R. Easther, *MultiModeCode: An efficient numerical solver for multifield inflation*, *JCAP* **1503** (2015), no. 03 005, [`arXiv:1410.0685`].

[37] J. García-Bellido and D. Wands, *Metric perturbations in two field inflation*, *Phys.Rev.* **D53** (1996) 5437–5445, [`astro-ph/9511029`].

[38] D. Wands, K. A. Malik, D. H. Lyth, and A. R. Liddle, *A new approach to the evolution of cosmological perturbations on large scales*, *Phys.Rev.* **D62** (2000) 043527, [`astro-ph/0003278`].

[39] K. A. Malik, D. Wands, and C. Ungarelli, *Large scale curvature and entropy perturbations for multiple interacting fluids*, *Phys.Rev.* **D67** (2003) 063516, [`astro-ph/0211602`].

[40] G. Leung, E. R. Tarrant, C. T. Byrnes, and E. J. Copeland, *Reheating, Multifield Inflation and the Fate of the Primordial Observables*, *JCAP* **1209** (2012) 008, [`arXiv:1206.5196`].

[41] I. Huston and A. J. Christopherson, *Isocurvature Perturbations and Reheating in Multi-Field Inflation*, [`arXiv:1302.4298`].

[42] G. Leung, E. R. Tarrant, C. T. Byrnes, and E. J. Copeland, *Influence of Reheating on the Trispectrum and its Scale Dependence*, *JCAP* **1308** (2013) 006, [`arXiv:1303.4678`].

[43] J. Meyers and E. R. M. Tarrant, *Perturbative Reheating After Multiple-Field Inflation: The Impact on Primordial Observables*, *Phys. Rev.* **D89** (2014), no. 6 063535, [`arXiv:1311.3972`].

[44] R. J. Hardwick and C. T. Byrnes, *Bayesian evidence of the post-Planck curvaton*, *JCAP* **1508** (2015), no. 08 010, [`arXiv:1502.06951`].

[45] V. Vennin, K. Koyama, and D. Wands, *Encyclopdia curvatonis*, *JCAP* **1511** (2015), no. 11

008, [arXiv:1507.07575].

[46] P. Adshead, R. Easther, J. Pritchard, and A. Loeb, *Inflation and the Scale Dependent Spectral Index: Prospects and Strategies*, *JCAP* **1102** (2011) 021, [arXiv:1007.3748].

[47] A. R. Liddle and D. H. Lyth, *The Cold dark matter density perturbation*, *Phys. Rept.* **231** (1993) 1–105, [astro-ph/9303019].

[48] A. R. Liddle and S. M. Leach, *How long before the end of inflation were observable perturbations produced?*, *Phys. Rev.* **D68** (2003) 103503, [astro-ph/0305263].

[49] L. A. Boyle and P. J. Steinhardt, *Probing the early universe with inflationary gravitational waves*, *Phys. Rev.* **D77** (2008) 063504, [astro-ph/0512014].

[50] J. Martin and C. Ringeval, *First CMB Constraints on the Inflationary Reheating Temperature*, *Phys. Rev.* **D82** (2010) 023511, [arXiv:1004.5525].

[51] C. E. Shannon, *A mathematical theory of communication*, *ACM SIGMOBILE Mobile Computing and Communications Review* **5** (2001), no. 1 3–55.

[52] R. Easther and L. C. Price, *Initial conditions and sampling for multifield inflation*, *JCAP* **1307** (2013) 027, [arXiv:1304.4244].

[53] R. Easther, L. C. Price, and J. Rasero, *Inflating an Inhomogeneous Universe*, *JCAP* **1408** (2014) 041, [arXiv:1406.2869].

[54] A. Corichi and D. Sloan, *Inflationary Attractors and their Measures*, *Class.Quant.Grav.* **31** (2014) 062001, [arXiv:1310.6399].

[55] D. Sloan, *Minimal Coupling and Attractors*, *Class.Quant.Grav.* **31** (2014) 245015, [arXiv:1407.3977].

[56] T. C. Bachlechner, M. Dias, J. Frazer, and L. McAllister, *Chaotic inflation with kinetic alignment of axion fields*, *Phys. Rev.* **D91** (2015), no. 2 023520, [arXiv:1404.7496].

[57] T. C. Bachlechner, C. Long, and L. McAllister, *Planckian Axions in String Theory*, arXiv:1412.1093.