Optimal prior for Bayesian inference in a constrained parameter space

Steen Hannestad\textsuperscript{a} and Thomas Tram\textsuperscript{a,b}

\textsuperscript{a}Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK–8000 Aarhus C, Denmark
\textsuperscript{b}Aarhus Institute of Advanced Studies (AIAS), Aarhus University, DK–8000 Aarhus C, Denmark

E-mail: sth@phys.au.dk, thomas.tram@aias.au.dk

Abstract. Bayesian parameter inference depends on a choice of prior probability distribution for the parameters in question. The prior which makes the posterior distribution maximally sensitive to data is called the Jeffreys prior, and it is completely determined by the response of the likelihood to changes in parameters. Under the assumption that the likelihood is a Gaussian distribution, the Jeffreys prior is a constant, i.e. flat. However, if one parameter is constrained by physical considerations, the Gaussian approximation fails and the flat prior is no longer the Jeffreys prior.

In this paper we compute the correct Jeffreys prior for a multivariate normal distribution constrained in one dimension, and we apply it to the sum of neutrino masses $\Sigma m_\nu$ and the tensor-to-scalar ratio $r$. We find that one-dimensional marginalised posteriors for these two parameters change considerably and that the 68\% and 95\% Bayesian upper limits increase by 9\% and 4\% respectively for $\Sigma m_\nu$ and 22\% and 3\% for $r$. Adding the prior to an existing chain can be done as a trivial importance sampling in the final step of the analysis proces.
1 Introduction

The physical properties of our Universe are well described by the flat ΛCDM model with six free parameters. These parameters are the baryon and cold dark matter densities, the current expansion rate of the Universe, the redshift at which the Universe was reionised, and two parameters describing the primordial curvature powerspectrum in the form of an amplitude and a spectral tilt. Other parameters could play a role, but they are currently not required by data. Those include the spatial curvature, the sum of neutrino masses, and the amplitude of primordial tensor fluctuations. All these parameters (including a growing number of experimental nuisance parameters) must be determined by measurements of cosmological observables such as the CMB and the Large-Scale Structure (LSS) of the late time Universe.

The number of parameters in a typical analysis are in the range 30-50, so one must rely on stochastic methods for sampling the parameter space, see e.g. [1] for a recent review. This is called Bayesian inference, and several tools have been developed for exactly this purpose and made available to the cosmology community, for instance CosmoMC [2] and MULTINEST [5].

2 Priors in Bayesian inference and cosmology

2.1 Bayes theorem

The essence of Bayesian inference is encapsulated by Bayes’ theorem of conditional probabilities

\[ Pr(\mu) \times Pr(D|\mu) = Pr(D) \times Pr(\mu|d), \]

where \( D \) is the data and \( \mu \) is the vector of model-parameters. It is customary to rename the probability distributions in equation (2.1) as follows. \( Pr(\mu) \equiv \pi(\mu) \) is called the prior,
Pr(D|µ) ≡ L(µ) is called the likelihood\(^1\) and \(Pr(D) = \int L(µ)π(µ)dµ \equiv E\) is the evidence. \(Pr(µ|d) \equiv P(µ)\) is called the posterior and can be expressed as

\[
P(µ) = \frac{π(µ)L(µ)}{E},
\]

using Bayes theorem.

In some contexts the prior is used to describe prior knowledge of the posterior distribution, e.g. from other experiments. However, in cosmology we are always using multiple datasets simultaneously, so it is much more natural to include all available data in \(D\) and to let \(L(µ)\) denote the (properly) combined likelihood. The prior \(π(µ)\) then describes the theoretical prejudice one may have regarding the parameters \(µ\), independent of any experimental data \(D\). What could constitute such a theoretical prejudice? One example is a variation of Weinbergs argument [6] for the size of the cosmological constant \(Ω_Λ\): using only one bit of data, the existence of the Universe, we may infer the posterior distribution of \(Ω_Λ\) to be strongly peaked around the maximum possible value that would allow observers. The theoretical prejudice invoked in this argument (apart from the existence of a multiverse) is roughly that the generation mechanism would naturally produce a very large value, and could only be reduced through accidental cancelations of very large terms, i.e. finetuning. In such a setup, large values would be exponentially more likely than small values.

### 2.2 The Jeffreys prior

The danger of this kind of reasoning is obvious, since in the absence of data we are likely to be vulnerable to confirmation bias. To avoid this, one may instead choose to rely only on the data \(D\). In that case, we should choose the prior that maximises the effect of the data \(D\) on the posterior distribution \(P(µ)\). This prior is called the Jeffreys prior [7] (see e.g. [8] for a review), and is given by

\[
π_J(µ) = \sqrt{|F(µ)|},
\]

where \(|F(µ)|\) is the determinant of the Fisher information matrix

\[
F_{ij}(µ) = -E\left[\frac{∂^2}{∂µ_i∂µ_j}\log L(µ)\right].
\]

\(E[\cdot]\) denotes the expectation value over all realisations of the data. The appearance of the Fisher information matrix is not particularly surprising, since \(F\) precisely encodes how sensitive the likelihood function is to variations in the model parameter vector \(µ\). A very large value of \(F\) indicates that data is very sensitive and thus the prior should assign correspondingly large weight to this region.

Computing the Fisher matrix in all points in parameter space is of course completely impractical. However, if the likelihood behaves as a Gaussian distribution in the model parameters \(µ\), one can compute the Jeffreys prior explicitly. In this approximation, let us now compute the Jeffreys prior explicitly in one dimension. Assume that the likelihood for a parameter \(µ\) is given by the Gaussian distribution

\[
L(µ) = \frac{1}{\sqrt{2πσ^2}}e^{-\frac{(q-µ)^2}{2σ^2}},
\]

\(^1\)\(L(µ)\) is not the raw likelihood of the experiment, but can be thought of as composed of a function or code that maps the model-parameters \(µ\) to an observable \(O\) in the context of a given model, and \(O\) is then passed to the raw likelihood.
where $\sigma$ is a constant. $q$ represents some estimate of $\mu$, and the data $D$ provides a sample of such estimates. Note that technically $\mu$ denotes the true value, and not just the mean value inferred from the particular sample in $D$. The Jeffreys prior now becomes
\[
\pi_J(\mu) \propto \sqrt{\int_{-\infty}^{\infty} dq L(-) \frac{d^2}{d\mu^2} \log L} = \sqrt{\frac{1}{\sigma^2} \int_{-\infty}^{\infty} dq L} = \frac{1}{\sigma},
\] (2.6)
i.e. the Jeffreys prior is flat. This is reassuring, since this is the prior normally used in cosmology and for the parameters usually used, the combined likelihood is typically not far from Gaussian.

2.3 The logarithmic prior

The Jeffreys prior is invariant under re-parametrisations $\phi = f(\mu)$ which is another desirable quality. Using the transformation theorem, it is easy to show that
\[
\pi_J(\phi) = \sqrt{|F(\phi)|}. \quad (2.7)
\]
In one dimension, the re-parametrisation $\phi = \log \mu$ then gives
\[
\pi_J(\log \mu) = \sqrt{|F(\log \mu)|} = \sqrt{|\mu^2 F(\mu)|} = |\mu| \pi_J(\mu). \quad (2.8)
\]
This means that we can write the posterior likelihood as
\[
P(\log(\mu)) \propto \mu \pi_J(\mu) P(\mu), \quad (2.9)
\]
which is very different from the choice of using a uniform prior on $\log \mu$.

Assuming a uniform prior on $\log \mu$ assigns a very large weight to regions of very negative $\log \mu$ where the likelihood $L$ is large, but where data cannot separate $\mu$ from zero. This leads to a prior which is very far from Jeffreys in the sense that it is very far from allowing the data to influence the posterior distribution. This spurious effect of assigning large weight to the region of very negative $\log \mu$ is taken care of by the additional $\mu$ appearing in the Jeffreys prior on $\log \mu$. This assigns progressively smaller weight to regions of increasingly negative $\log \mu$ where the data has no discriminating power. For parameters where $\mu \gg \sigma$ (i.e. most cosmological parameters) there is little difference between using a uniform prior on $\mu$ and on $\log \mu$ because the factor $\mu$ in the prior changes little over the interval $\mu - \sigma$ to $\mu + \sigma$.

A flat prior on the logarithm of a parameter, a logarithmic prior, is sometimes also erroneously called a Jeffreys prior. This misunderstanding presumably originates outside of cosmology, where sometimes the mean value $\mu$ is a known quantity while its standard deviation $\sigma$ is not. In this case a calculation similar to equation (2.6) yields $\pi_J(\sigma) \propto \frac{1}{\sigma}$ which, by equation (2.8), is equivalent to a flat prior in $\log \sigma$.

2.4 The Jeffreys prior for a constrained parameter

In the preceding subsections, we have assumed that the estimates $q$ could be any real number. However, many parameters in cosmology are required to be positive, and this means that all estimates $q$ of $\mu$ will be positive; the data could never support a negative $\mu$. Let us now redo the calculation in equation (2.6) for the truncated Gaussian distribution
\[
L(\mu) = \begin{cases} 
\frac{2}{1 + \text{erf}(\frac{\mu}{\sqrt{2} \sigma})} \frac{1}{\sqrt{2\pi} \sigma^2} e^{-\frac{(q-\mu)^2}{2\sigma^2}} & q \geq 0, \\
0 & q < 0, 
\end{cases} \quad (2.10)
\]
where the normalisation factor now depends on $\mu$. The Jeffreys prior is now much more complicated and we find

$$
\pi_J = \frac{1}{\sigma} \sqrt{1 - \frac{\mu}{\sigma}} \left( \frac{e^{-\frac{\mu^2}{2\sigma^2}}}{1 + \text{erf} \left( \frac{\mu}{\sqrt{2\sigma}} \right)} \right) - \frac{2}{\pi} \left( \frac{e^{-\frac{\mu^2}{2\sigma^2}}}{1 + \text{erf} \left( \frac{\mu}{\sqrt{2\sigma}} \right)} \right)^2.
$$

(2.11)

In the limit $\frac{\mu}{\sigma} \to \infty$, the two brackets in equation (2.11) goes to zero and we recover the flat prior of the unconstrained Gaussian, equation (2.6). Using equation (2.11) we can compute the effect on the (Bayesian) 95% upper bound of using the Jeffreys prior compared to the flat prior. For $\mu = 0$, the effect is 5.1% with the Jeffreys prior yielding the looser bound because of the lower weight at small values of $q$. For $\mu = 1\sigma$, the effect has decreased to 3.2%, and for $\mu = 2\sigma$ to 1.3%.

It is clear that if even if a given parameter is constrained to be positive the flat prior is also the Jeffreys prior as long as the mean is well separated from 0 (i.e. at high significance). In this case the allowed range could bin principle be extended to the entire real axis without affecting any results (even if, from a physical point of view, negative values are unphysical). This means that even if most cosmological parameter are constrained to be positive the flat prior is close to Jeffreys as long as the parameters are known to be different from zero at high significance. This is true for parameters like $\Omega_m h^2$, $\Omega_b h^2$, etc., but untrue for the neutrino mass, $m_\nu$, and the tensor-to-scalar ratio, $r$, both of which are restricted to be positive, but not yet distinguishable from zero by cosmological data.

### 2.5 Multidimensional Gaussian with a single constrained parameter

We will now consider the more realistic case where one parameter is constrained to be positive while all others are allowed to be on the entire real axis. In this case, the Jeffreys prior can still be computed analytically, while in the more general case of multiple constrained parameters, it must be computed numerically. We start from the probability distribution

$$
\mathcal{L} = A(\mu_1)e^{-\frac{1}{2}(q-\mu)^T M(q-\mu)},
$$

(2.12)

where we have assumed $q_1$, the estimates of $\mu_1$, to be constrained to the positive real axis. All other parameters span the entire real axis, so that the norm, $A(\mu_1)$, depends only on $\mu_1$. The Hessian matrix of $\log \mathcal{L}$ is simply

$$
\partial_i \partial_j \log \mathcal{L} = \frac{d^2 \log A(\mu_1)}{d\mu_1^2} \delta_{i1} \delta_{j1} - M_{ij}.
$$

(2.13)

The expectation value of the Hessian matrix can now be computed as follows

$$
-E \left[ \partial_i \partial_j \log \mathcal{L} \right] = -\int dq_1 \partial_i \partial_j \left( \log \mathcal{L} \right) A(\mu_1)e^{-\frac{1}{2}(q-\mu)^T M(q-\mu)},
$$

(2.14)

$$
= \left[ M_{ij} - \frac{d^2 \log A(\mu_1)}{d\mu_1^2} \delta_{i1} \delta_{j1} \right] \int dq A(\mu_1)e^{-\frac{1}{2}(q-\mu)^T M(q-\mu)},
$$

(2.15)

$$
= M_{ij} - \frac{d^2 \log A(\mu_1)}{d\mu_1^2} \delta_{i1} \delta_{j1},
$$

(2.16)

where we have used the fact that $\mathcal{L}$ is normalised to 1.
It is now advantageous to write the matrix $M$ as
\begin{equation}
M = \begin{pmatrix} U_0 & V \\ V^T & W_0 \end{pmatrix}, \tag{2.17}
\end{equation}
where $U_0 \equiv M_{11}$ and $V$ is a vector of length $n - 1$. The determinant of equation (2.16) is
\begin{equation}
| - E \left[ \partial_i \partial_j \log \mathcal{L} \right] | = | M | - \frac{d^2 \log A(\mu_1)}{d\mu_1^2} | W_0 |. \tag{2.18}
\end{equation}

When $A$ is independent of $\mu_1$, we are left with the determinant of the inverse covariance matrix $M$ as expected.

The only remaining task is to compute the normalisation factor $A(\mu_1)$ and its second derivative. Using the general formula for the partial Gaussian integral, the normalisation condition becomes
\begin{equation}
1 = \int dq A(\mu_1) e^{-\frac{1}{2} (q - \mu)^T M (q - \mu)}, \tag{2.19}
\end{equation}
\begin{equation}
= A(\mu_1) \frac{(2\pi)^{(n-1)/2}}{\sqrt{|W_0|}} \int_0^\infty dq e^{-\frac{1}{2} q^2} \Rightarrow \tag{2.20}
\end{equation}
\begin{equation}
A(\mu_1) = \frac{2 |W_0|^2}{\sqrt{(2\pi)^n}} \left[ 1 + \text{erf} \left( \sqrt{\frac{U}{2}} \mu_1 \right) \right], \tag{2.21}
\end{equation}
where we have defined
\begin{equation}
U \equiv U_0 - VW_0^{-1} V^T. \tag{2.22}
\end{equation}
The second derivative is now
\begin{equation}
\frac{d^2 \log A(\mu_1)}{d\mu_1^2} = \frac{2}{\pi} \left[ \frac{e^{-U \mu_1^2}}{\left( 1 + \text{erf} \left( \sqrt{\frac{U}{2}} \mu_1 \right) \right)^2} + \frac{\sqrt{\pi} \frac{U}{2} e^{-\frac{1}{2} U \mu_1^2}}{1 + \text{erf} \left( \sqrt{\frac{U}{2}} \mu_1 \right)} \right], \tag{2.23}
\end{equation}
\begin{equation}
\equiv U \frac{2}{\pi} \left[ Z(x)^2 + \sqrt{\pi} x Z(x) \right]. \tag{2.24}
\end{equation}
In the last line we defined the convenient quantities
\begin{equation}
Z(x) \equiv \frac{e^{-x^2}}{1 + \text{erf}(x)}, \quad x \equiv \sqrt{\frac{U}{2}} \mu_1. \tag{2.25}
\end{equation}

It may sometimes be useful to approximate $Z(x)$ by expanding the error function in terms of a Padé approximant
\begin{equation}
Z(x) \simeq \frac{e^{-x^2}}{6x} \frac{1}{\sqrt{\pi} (x^2 + 3)} + 1. \tag{2.26}
\end{equation}
Figure 1. Jeffreys prior in the rescaled parameter $x = \sqrt{\frac{U}{2}} \mu_1$.

which is accurate at the $10^{-3}$ level for all $x$. Using $|M| = U |W_0|$ we can finally write the Jeffreys prior as

$$
\pi_J(\mu_1) = \sqrt{|M| - \frac{d^2 \log A(\mu_1)}{d \mu_1^2}|W_0|}, \\
\propto \sqrt{1 - \frac{1}{U} \frac{d^2 \log A(\mu_1)}{d \mu_1^2}}, \\
= \sqrt{1 - \frac{2}{\pi} \left[Z(x)^2 + \sqrt{\pi} x Z(x) \right]}.
$$

We have shown this prior in figure 1. For $x_1 = \mu_1 = 0$ we have $Z(0) = 1$, so the value of the prior at the origin is suppressed w.r.t. its asymptotic value by a factor $\sqrt{1 - 2/\pi} \approx 0.60$. For a one-dimensional truncated Gaussian, we have $U = U_0 = \frac{1}{\sigma^2}$ and equation (2.27) reduces to the one-dimensional result in equation (2.11).

3 Application to cosmology

3.1 Implementation

We have implemented the Jeffreys prior in equation (2.27) inside MONTEPYTHON in the form of a likelihood. It requires knowledge of the matrix $M$ in equation (2.12) which we compute as the inverse of the covariance matrix. The covariance matrix itself was computed from a separate run with the same datasets but without the prior, and the same covariance matrix was then used in the second Markov-Chain Monte Carlo (MCMC) run. This approach is very conservative since importance sampling of the Jeffreys prior should work very well. And since the importance sampling in this case is independent of the cosmology, a very simple procedure is in fact possible:

1. Do a standard sampling (e.g. MCMC) using a flat prior.
2. Compute the covariance matrix, and invert it to estimate $M$ in equation (2.12).
3. Compute $U$ for the constrained parameter from $M$ using equation (2.17) and (2.22).
4. Importance sample the chain with the prior by multiplying the multiplicity of each point by equation (2.27).

These steps can be done on-the-fly when analysing a chain and could be implemented in the analysis tools, i.e. GetDist [9] or in MontePython’s analyze.py. We will recommend this method for future use.

3.2 Models and data

We consider two one-parameter extensions of the $ΛCDM$ base model which we compute using class [10, 11]. The first one is $ΛCDM + Σ m_ν$ where the sum of neutrino masses $Σ m_ν$ is allowed to vary, $Σ m_ν ∈ [0, 10]$ assuming a degenerate hierarchy. The second model is $ΛCDM + r$ where the tensor-to-scalar ratio $r$ at the pivot-scale $k_{pivot} = 0.05\text{Mpc}^{-1}$ is allowed to vary in the region $r ∈ [0, 0.5]$. We used the following datasets:

CMB Temperature and polarisation likelihood from Planck 2015 [12] (TTTEE), as well as the lensing reconstruction from the CMB.

BAO Baryon Acoustic Oscillations (BAO) measurements from 6dF [13], BOSS LOWZ and CMASS [14] and the Main Galaxy Sample (MGS) of SDSS DR7 [15].

BK14 Polarisation likelihood from the Bicep-Keck collaboration [16] including the Gaussian priors on the spectral index of dust and synchroton radiation.

πJ Jeffreys prior on either $Σ m_ν$ or $r$.

For the $ΛCDM + Σ m_ν$ model we used the combination CMB+BAO with and without πJ, and the the $ΛCDM + r$ model we used CMB+BK14 with and without πJ.

3.3 Results

After imposing the prior, none of the posteriors change significantly except for the one in the constrained parameter. In figure 2 we have shown the effect of imposing the new prior on the

---

**Figure 2.** Marginalised posterior distributions for $Σ m_ν$ and $r$ with the same data but with or without the Jeffreys prior.
peaks sharply at the origin, but it reaches a plateau for small values of $\Sigma m$. We are showing the 95% upper bound.

Table 1. 68% and 95% bounds on the constrained parameters. Imposing the prior degrades the bounds somewhat, with the 68% bound being the most affected.

| $\Sigma m_\nu$ [eV] | Flat prior | $\pi J$ | Rel. dif. | Flat prior | $\pi J$ | Rel. dif. |
|---------------------|------------|---------|-----------|------------|---------|-----------|
| $\Sigma m_\nu$      | $< 0.081$  | $< 0.090$ | 9.9%      | $< 0.166$  | $< 0.173$ | 4.5%      |
| $r$                 | $< 0.039$  | $< 0.048$ | 22%       | $< 0.072$  | $< 0.075$ | 3.5%      |

Table 2. Mean values and 68% confidence limits for the cosmological parameters. For $\Sigma m_\nu$ and $r$ we are showing the 95% upper bound.

| $\Sigma m_\nu$ | CMB+BAO | CMB+BAO+$\pi J$ | CMB+BAO | CMB+BAO+$\pi J$ | CMB+BAO | CMB+BAO+$\pi J$ |
|-----------------|---------|----------------|---------|----------------|---------|----------------|
| $\Lambda$CDM + $\Sigma m_\nu$ | $1.0419^{+0.0003}_{-0.0003}$ | $1.0419^{+0.0003}_{-0.0003}$ | $1.0419^{+0.0003}_{-0.0003}$ | $1.0419^{+0.0003}_{-0.0003}$ | $1.0419^{+0.0003}_{-0.0003}$ | $1.0419^{+0.0003}_{-0.0003}$ |
| $10^{10}A_s$    | $2.16^{+0.05}_{-0.06}$ | $2.17^{+0.05}_{-0.06}$ | $2.14^{+0.05}_{-0.06}$ | $2.14^{+0.05}_{-0.06}$ | $2.14^{+0.05}_{-0.06}$ | $2.14^{+0.05}_{-0.06}$ |
| $\nu$           | $0.97^{+0.04}_{-0.04}$ | $0.967^{+0.004}_{-0.004}$ | $0.966^{+0.004}_{-0.004}$ | $0.966^{+0.004}_{-0.004}$ | $0.966^{+0.004}_{-0.004}$ | $0.966^{+0.004}_{-0.004}$ |
| $H_0$           | $67.9_{-0.5}^{+0.6}$ | $67.9_{-0.6}^{+0.6}$ | $67.6_{-0.6}^{+0.6}$ | $67.6_{-0.6}^{+0.6}$ | $67.6_{-0.6}^{+0.6}$ | $67.6_{-0.6}^{+0.6}$ |
| $\Omega_B$      | $0.049_{-0.007}^{+0.008}$ | $0.049_{-0.007}^{+0.008}$ | $0.049_{-0.007}^{+0.008}$ | $0.049_{-0.007}^{+0.008}$ | $0.049_{-0.007}^{+0.008}$ | $0.049_{-0.007}^{+0.008}$ |
| $\sigma_8$      | $0.818_{-0.009}^{+0.012}$ | $0.817_{-0.010}^{+0.013}$ | $0.816_{-0.009}^{+0.008}$ | $0.816_{-0.009}^{+0.008}$ | $0.816_{-0.009}^{+0.008}$ | $0.816_{-0.009}^{+0.008}$ |

one-dimensional posteriors for the constrained parameters. The posterior for $\Sigma m_\nu$ no longer peaks sharply at the origin, but it reaches a plateau for small values of $\Sigma m_\nu$. This makes sense, since the data cannot differentiate between 0eV and a very small but non-zero value like 0.02eV. For the posterior on $r$ we see how the small bump is somewhat enhanced by the prior. We must stress that this bump is not statistically significant before or after the prior is imposed. Using the prior slightly degrades the 95% upper bounds, but at the same time it improves the significance of peaks at small, non-zero values. Not using this prior therefore makes it harder to claim a 5$\sigma$-detection of a signal than it should be.

The 68% and 95% upper bounds for $\Sigma m_\nu$ and $r$ are given in Table 1. The 68% bound increases significantly when the Jeffreys prior is imposed, while the increase in the 95% bound if of the order 4%. In Table 2 we have given the full list of mean values and 68% confidence limits for all our cosmological parameters, although there are no significant changes when imposing the new prior. For completeness we have also included in Figures 3 and 4 the triangle plots for both models, which again shows that the new prior has no significant effect on the correlations between parameters.

4 Conclusion

The Jeffreys prior is the prior that optimises the information provided by the data, making it the preferred prior in many respects. The Jeffreys prior is flat if the response of the
likelihood is approximately Gaussian in the model parameters. However, this is no longer the case if a parameter is constrained to some region of parameters space, for instance by the requirement that it must be positive. In this paper we have computed the Jeffreys prior under the assumption of a multivariate Gaussian constrained in a single parameter. It is given by equation (2.27) which is our main result. It can be easily generalised to more than one constrained parameter, but in that case the integrals cannot be done analytically. However, the one-dimensional case should be very accurate as long as only one parameter is close to zero which is frequently the case in cosmology.

As an application we then considered the sum of neutrino masses $\Sigma m_\nu$ and the tensor-to-scalar ratio $r$ and found that their posteriors changed significantly. The change resulted in a 10-20% increase in the 68% upper limit and a 4% increase in the often quoted 95% upper limit. We have provided a simple recipe for implementing the prior directly in the analysis part of current codes such as \texttt{GetDist} or \texttt{MontePython}, and we propose this new prior to be used in future analyses with constrained parameters.

References

[1] R. Trotta, “Bayesian Methods in Cosmology,” 2017. \texttt{arXiv:1701.01467 [astro-ph.CO]}.
[2] A. Lewis and S. Bridle, “Cosmological parameters from CMB and other data: A Monte Carlo approach,” \textit{Phys. Rev.} \textbf{D66} (2002) 103511, \texttt{arXiv:astro-ph/0205436 [astro-ph]}.
[3] B. Audren, J. Lesgourgues, K. Benabed, and S. Prunet, “Conservative Constraints on Early Cosmology: an illustration of the Monte Python cosmological parameter inference code,” \textit{JCAP} \textbf{1302} (2013) 001, \texttt{arXiv:1210.7183 [astro-ph.CO]}.
[4] W. K. Hastings, “Monte Carlo Sampling Methods Using Markov Chains and Their Applications,” \textit{Biometrika} \textbf{57} (1970) 97–109.
[5] F. Feroz, M. P. Hobson, and M. Bridges, “MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics,” \textit{Mon. Not. Roy. Astron. Soc.} \textbf{398} (2009) 1601–1614, \texttt{arXiv:0809.3437 [astro-ph]}.
[6] S. Weinberg, “Anthropic Bound on the Cosmological Constant,” \textit{Phys. Rev. Lett.} \textbf{59} (1987) 2607.
[7] H. Jeffreys, “An Invariant Form for the Prior Probability in Estimation Problems,” \textit{Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences} \textbf{186} (1946) no. 1007, 453–461. http://www.jstor.org/stable/97883.
[8] R. E. Kass and L. Wasserman, “The Selection of Prior Distributions by Formal Rules,” \textit{Journal of the American Statistical Association} \textbf{91} (1996) no. 435, 1343–1370. http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1996.10477003.
[9] A. Lewis. \texttt{https://github.com/cmbant/getdist}.
[10] D. Blas, J. Lesgourgues, and T. Tram, “The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes,” \textit{JCAP} \textbf{1107} (2011) 034, \texttt{arXiv:1104.2933 [astro-ph.CO]}.
[11] J. Lesgourgues and T. Tram, “The Cosmic Linear Anisotropy Solving System (CLASS) IV: efficient implementation of non-cold relics,” \textit{JCAP} \textbf{1109} (2011) 032, \texttt{arXiv:1104.2935 [astro-ph.CO]}.
[12] \textbf{Planck} Collaboration, P. A. R. Ade \textit{et al.}, “Planck 2015 results. XIII. Cosmological parameters,” \textit{Astron. Astrophys.} \textbf{594} (2016) A13, \texttt{arXiv:1502.01589 [astro-ph.CO]}.
[13] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson, “The 6dF Galaxy Survey: Baryon Acoustic Oscillations and the Local Hubble Constant,” Mon. Not. Roy. Astron. Soc. 416 (2011) 3017–3032, arXiv:1106.3366 [astro-ph.CO].

[14] BOSS Collaboration, L. Anderson et al., “The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples,” Mon. Not. Roy. Astron. Soc. 441 (2014) no. 1, 24–62, arXiv:1312.4877 [astro-ph.CO].

[15] A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden, and M. Manera, “The clustering of the SDSS DR7 main Galaxy sample ? I. A 4 per cent distance measure at z = 0.15,” Mon. Not. Roy. Astron. Soc. 449 (2015) no. 1, 835–847, arXiv:1409.3242 [astro-ph.CO].

[16] BICEP2, Keck Array Collaboration, P. A. R. Ade et al., “Improved Constraints on Cosmology and Foregrounds from BICEP2 and Keck Array Cosmic Microwave Background Data with Inclusion of 95 GHz Band,” Phys. Rev. Lett. 116 (2016) 031302, arXiv:1510.09217 [astro-ph.CO].
Figure 3. Triangle plot of the cosmology parameters in the $\Lambda$CDM + $\Sigma m_\nu$ model. The correlations between $\Sigma m_\nu$ and the other parameters are mostly unaffected by changing the flat prior to the Jeffreys prior.
Figure 4. Triangle plot of the parameters in the $\Lambda$CDM + $r$ model. Like we saw in the $\Lambda$CDM + $\Sigma m_\nu$ model, the correlations between $r$ and the other parameters do not change significantly when imposing the Jeffreys prior.