Journal of Cosmology and
Astroparticle Physics

A new measure of tension between experiments

To cite this article: Saroj Adhikari and Dragan Huterer JCAP01(2019)036

Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

A new measure of tension between experiments

Saroj Adhikari and Dragan Huterer

Department of Physics, University of Michigan,
450 Church St, Ann Arbor, MI 48109-1040, U.S.A.
Leinweber Center for Theoretical Physics, University of Michigan,
450 Church St, Ann Arbor, MI 48109-1040, U.S.A.

E-mail: saroj@umich.edu, huterer@umich.edu

Received September 25, 2018
Accepted January 9, 2019
Published January 18, 2019

Abstract. Tensions between cosmological measurements by different surveys or probes have always been important — and are presently much discussed — as they may lead to evidence of new physics. Several tests have been devised to probe the consistency of datasets given a cosmological model, but they often have undesired features such as dependence on the prior volume, or burdensome requirements such as that of near-Gaussian posterior distributions. We propose a new quantity, defined in a similar way as the Bayesian evidence ratio, in which these undesired properties are absent. We test the quantity on simple models with Gaussian and non-Gaussian likelihoods. We then apply it to data from the *Planck* satellite: we investigate the consistency of ΛCDM model parameters obtained from TT and EE angular power spectrum measurements, as well as the mutual consistency of cosmological parameters obtained from large scale (multipoles, $\ell < 1000$) and small scale ($\ell \geq 1000$) portions of each measurement and find no significant discrepancy in the six-dimensional ΛCDM parameter space.

Keywords: cosmological parameters from CMBR, cosmological parameters from LSS

ArXiv ePrint: 1806.04292

 https://doi.org/10.1088/1475-7516/2019/01/036

Contents

1 Introduction

The use of Bayesian statistics in cosmology is now commonplace: most of the results on cosmological parameters from cosmic microwave background (CMB) experiments [1] and large-scale structure (LSS) surveys [2] are reported as posterior distributions. In addition, various Bayesian methods are used for model comparison [3].

Along with the increase in the number of cosmological surveys and the improvement in their precision, a number of tensions between parameters derived from different experiments have been observed. For example, the Hubble constant H_0 measured using the distance ladder in the local universe disagrees with that derived from *Planck* CMB observations; in the standard six parameter ΛCDM model, the disagreement is about 3.4σ [4, 5]. There is also some tension between the measurements of the amplitude of fluctuations σ_8 and the matter density Ω_m from weak lensing to that of the measurement from *Planck* CMB data [6–8]. As a result, a number of statistics have been developed to compare datasets in cosmology. The primary goal of these statistics is to determine if two datasets are consistent realizations of the same model, that is, with a single set of cosmological parameters (see [9–11] for discussions and comparisons of some of the popular methods). For an alternative approach using hyperparameters, see [12, 13].

The Bayesian evidence-based metric of [14] has been widely used [15–19], but is known to strongly depend on the priors given to parameters. This has led to the use of other measures [20–23] that do not have the prior-volume dependence, but at the expense of losing the simplicity of an evidence ratio. In this work, we define an evidence-based quantity which fixes the problem of prior volume dependence and which can be evaluated on a easy-to-interpret scale.

Consider two datasets $\mathbf{d}_1$ and $\mathbf{d}_2$, and let $\boldsymbol{\theta}$ be the parameters of a model. Let us assume that both the datasets and the combination of them can be modeled by a particular ΛCDM realization, and the priors are wide enough to include the parameter posteriors preferred by both the datasets individually. Most commonly, a Bayesian analysis is used to determine posterior probability distributions for the model parameters $\boldsymbol{\theta}$. Suppose the two datasets separately give two (normalized) posterior distributions

$$\begin{aligned} p_1(\boldsymbol{\theta}|\mathbf{d}_1) &= \frac{\mathcal{L}(\mathbf{d}_1|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{E(\mathbf{d}_1)}, \\ p_2(\boldsymbol{\theta}|\mathbf{d}_2) &= \frac{\mathcal{L}(\mathbf{d}_2|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{E(\mathbf{d}_2)}, \end{aligned} \tag{1.1}$$

where $\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})$ denotes the likelihood of the data $\mathbf{d}$ given the model defined by a set of parameters $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta})$ is the prior probability of the model parameters, and E is called the marginal likelihood or the evidence,

$$E(\mathbf{d}) = \int d\boldsymbol{\theta} \mathcal{L}(\mathbf{d}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}). \tag{1.2}$$

We will always use normalized probability density functions for the likelihood ($\int \mathcal{L}(\mathbf{d}|\theta) \cdot d\mathbf{d} = 1$) and the prior ($\int \pi(\boldsymbol{\theta})d\boldsymbol{\theta} = 1$). The posterior for the combination of datasets $\mathbf{d}_1, \mathbf{d}_2$ is:

$$\begin{aligned} p_{12}(\boldsymbol{\theta}|\mathbf{d}_1, \mathbf{d}_2) &= \frac{\mathcal{L}(\mathbf{d}_1, \mathbf{d}_2, \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{E(\mathbf{d}_1, \mathbf{d}_2)} \\ &= \frac{\mathcal{L}(\mathbf{d}_1, \boldsymbol{\theta})\mathcal{L}(\mathbf{d}_2, \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{E(\mathbf{d}_1, \mathbf{d}_2)}, \end{aligned}$$

where the second equality assumes that the combined likelihood is approximated well by the product of the two likelihoods.

The ratio of the evidences obtained using two different models, called the Bayes Factor, is a widely used measure for model comparison. In this work, we will define a similar ratio to compare two sets of parameter constraints of a model obtained using different datasets or experiments.

2 Evidence for model parameters

We first define the marginal likelihood (or the evidence) for the maximum likelihood model parameters, $\boldsymbol{\theta}^{\mathrm{ML}}$, instead of the usual definition of the evidence for the data, $\mathbf{d}$. We do so as our primary goal is to quantify the level of consistency between model parameters obtained from different datasets or experiments. Analogous to eq. (1.2), we define the evidence for the maximum likelihood model parameters

$$E(g(\boldsymbol{\theta}^{\mathrm{ML}})) = \int d\boldsymbol{\theta} \mathcal{L}(g(\boldsymbol{\theta}^{\mathrm{ML}})|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{2.1}$$

where, instead of the measured data, we have used the maximum likelihood values of the data realization given the model, $g(\boldsymbol{\theta}^{\mathrm{ML}})$; see figure 1 for an illustration. Here $g(\boldsymbol{\theta})$ is the function that computes the model prediction for the data given the parameters $\boldsymbol{\theta}$; for example, in the case of the CMB temperature fluctuation data, the model prediction is represented by the theory angular power spectra $g^{TT}_{\mathrm{cmb}}(\boldsymbol{\theta}) = \{C^{TT}_\ell\}$. If the likelihood in the above equation is a combination of two experiments, then we can define evidences for the maximum likelihood parameters obtained through the combination of the two datasets (denoted by i, j), $E(g(\boldsymbol{\theta}^{\mathrm{ML}}_{ij}))$. Alternatively, we can define an evidence so that each part of the data vector in the evidence integral uses its own maximum likelihood parameter values, obtained by analyzing each experiment separately, $E(\{g_i(\boldsymbol{\theta}^{\mathrm{ML}}_i), g_j(\boldsymbol{\theta}^{\mathrm{ML}}_j)\})$. As we will show, the ratio of the two evidences can quantify the tension between the parameter constraints obtained from two different datasets.

3 Evidence-based dataset comparison

For simplicity, consider two datasets that have independent likelihoods, $\mathcal{L}_i(\mathbf{d}|\boldsymbol{\theta}), \mathcal{L}_j(\mathbf{d}|\boldsymbol{\theta})$, and let the measured data vector for each experiment be denoted by $\mathbf{d}_i, \mathbf{d}_j$. Let us further

JCAP01(2019)036

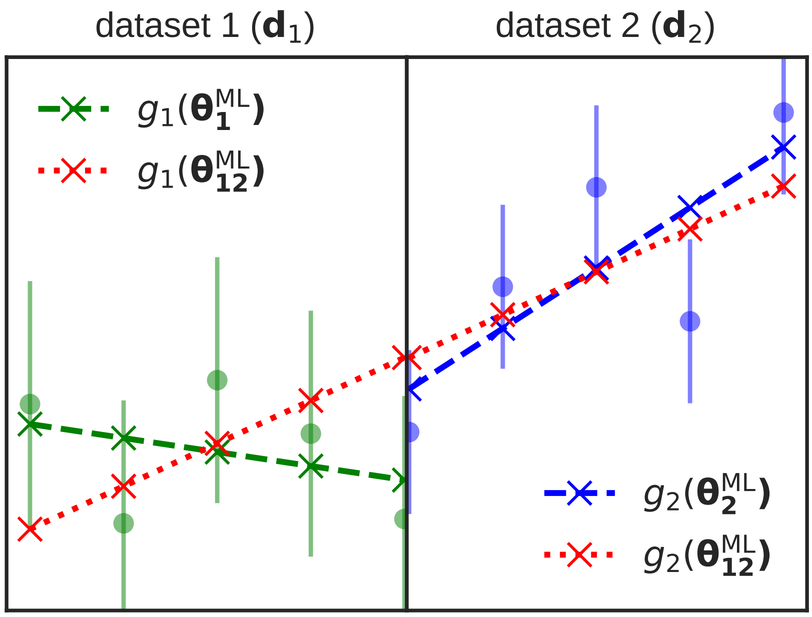

Figure 1. Illustration of how we make use of the maximum-likelihood data-realizations (crosses) as opposed to the actual measurements (dots) in our evidence integrals. Doing this gets rid of the data scatter and therefore makes our statistic only sensitive to the model parameters and not the spread in the data realizations.

assume that the maximum likelihood parameters of the model, ΛCDM for example, are known for three different cases: two datasets analyzed separately, $\boldsymbol{\theta}_i^{\mathrm{ML}}, \boldsymbol{\theta}_j^{\mathrm{ML}}$, and their combined analysis, $\boldsymbol{\theta}_{ij}^{\mathrm{ML}}$.

Our null hypothesis $\mathcal{H}_0$ is that both the datasets are realizations of a single set of parameters, $\boldsymbol{\theta}_{ij}^{\mathrm{ML}}$, from the combined fit. The alternative, more complicated, hypothesis $\mathcal{H}_1$ is that each of the datasets are realizations of their own set of parameters, $\boldsymbol{\theta}_i^{\mathrm{ML}}, \boldsymbol{\theta}_j^{\mathrm{ML}}$. Then, using the Bayes theorem similarly to the derivation of the Bayes factor, we get

$$
\begin{aligned}
\frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)} &= \frac{\int d\boldsymbol{\theta}\pi(\boldsymbol{\theta})\mathcal{L}_i(g_i(\boldsymbol{\theta}_i^{\mathrm{ML}})|\boldsymbol{\theta})\mathcal{L}_j(g_j(\boldsymbol{\theta}_j^{\mathrm{ML}})|\boldsymbol{\theta})}{\int d\boldsymbol{\theta}\pi(\boldsymbol{\theta})\mathcal{L}_i(g_i(\boldsymbol{\theta}_{ij}^{\mathrm{ML}})|\boldsymbol{\theta})\mathcal{L}_j(g_j(\boldsymbol{\theta}_{ij}^{\mathrm{ML}})|\boldsymbol{\theta})} \\
&= \frac{E_{ij}^{\mathrm{sep}}}{E_{ij}^{\mathrm{com}}} \equiv R_{ij},
\end{aligned}
\tag{3.1}
$$

where the superscripts `sep` and `com` in the formula above stand for separate and combined maximum likelihood parameters, respectively. We will first consider a case where the above expression can be evaluated analytically.

Consider two likelihoods given by two N_{params}-dimensional multivariate Gaussian distributions with arbitrary covariance matrices Σ_1 and Σ_2,

$$
\begin{aligned}
\mathcal{L}_1(\boldsymbol{\theta}) &= \frac{1}{\sqrt{\det(2\pi\Sigma_1)}} \exp\left[-\frac{1}{2}(\mathbf{d}_1 - \boldsymbol{\theta})^T\Sigma_1^{-1}(\mathbf{d}_1 - \boldsymbol{\theta})\right] \\
\mathcal{L}_2(\boldsymbol{\theta}) &= \frac{1}{\sqrt{\det(2\pi\Sigma_2)}} \exp\left[-\frac{1}{2}(\mathbf{d}_2 - \boldsymbol{\theta})^T\Sigma_2^{-1}(\mathbf{d}_2 - \boldsymbol{\theta})\right].
\end{aligned}
$$

In this simple example, we have taken $g_i(\boldsymbol{\theta}) = g_j(\boldsymbol{\theta}) = \boldsymbol{\theta}$ so that the expressions are easy to evaluate analytically. If we further assume that the prior on each of the parameters is uniform and wide (compared to the constraint on the parameter), we get [24],

$$\begin{aligned}E_{12}^{\rm sep} &\propto \int d\boldsymbol{\theta}\mathcal{L}_1(\mathbf{d}_1|\boldsymbol{\theta})\mathcal{L}_2(\mathbf{d}_2|\boldsymbol{\theta}) = \frac{1}{\sqrt{\det\left[2\pi(\Sigma_1+\Sigma_2)\right]}} \\ &\qquad \times \exp\left[-\frac{1}{2}(\mathbf{d}_1-\mathbf{d}_2)^T(\Sigma_1+\Sigma_2)^{-1}(\mathbf{d}_1-\mathbf{d}_2)\right] \\ E_{12}^{\rm com} &\propto \int d\boldsymbol{\theta}\mathcal{L}_1(\mathbf{d}_{12}|\boldsymbol{\theta})\mathcal{L}_2(\mathbf{d}_{12}|\boldsymbol{\theta}) = \frac{1}{\sqrt{\det\left[2\pi(\Sigma_1+\Sigma_2)\right]}}\end{aligned} \tag{3.2}$$

so that

$$R_{12} = \exp\left[-\frac{1}{2}(\mathbf{d}_1-\mathbf{d}_2)^T(\Sigma_1+\Sigma_2)^{-1}(\mathbf{d}_1-\mathbf{d}_2)\right], \tag{3.3}$$

the negative logarithm of which ($-\ln R_{12}$) is the two-experiment index of inconsistency (IOI) defined in [11]. Under these conditions assuming that the null hypothesis is true, $(-2\ln R_{ij})$ is χ^2 distributed with $N_{\rm params}$ degrees of freedom (dof) [25] (see their definition and discussion of $Q_{\rm DM}$). More generally, the ratio of probabilities of two hypotheses (evidence ratio) is similar to a likelihood-ratio test [26], and the distribution of $(-2\ln R_{ij})$ asymptotically approaches $\chi^2_{N_{\rm params}}$ by Wilks theorem [27]. Here, $N_{\rm params} = {\rm dof}(\mathcal{H}_0) - {\rm dof}(\mathcal{H}_1)$, when comparing two datasets. We will, therefore, evaluate the probability-to-exceed (PTE) value of observed $\ln R_{ij}$ values by taking $(-2\ln R_{ij})$ to be $\chi^2_{N_{\rm params}}$ distributed.

We can also obtain analytic expression for $\ln R_{ij}$ for a linear Gaussian model, where $g_i(\boldsymbol{\theta}) = F_i + M_i\boldsymbol{\theta}$, rather than the simple $g_i(\boldsymbol{\theta}) = \boldsymbol{\theta}$ model used above. A linear Gaussian model is often employed to analytically calculate the results of various tension measures [9]. Under the assumption of the linear model, we can calculate the separate and combined evidences for model parameters. When the likelihood is Gaussian in data and the parameters have uniform priors, the posterior distribution of parameters in the Gaussian linear model is also Gaussian, with mean $\boldsymbol{\theta}_i^{\rm ML}$ and covariance $\Sigma_i^p = (M_i^T\Sigma_i^{-1}M_i)^{-1}$ [20]. Then, the individual likelihoods used in the computations of the evidence integrals for model parameters can be written in terms of the mean and covariance of the parameter posteriors:

$$\mathcal{L}_i(g(\boldsymbol{\theta}^{\rm ML})|\boldsymbol{\theta}) = \frac{1}{\sqrt{\det(2\pi\Sigma_i)}}\exp\left[-\frac{1}{2}(\boldsymbol{\theta}_i^{\rm ML}-\boldsymbol{\theta})(\Sigma_i^p)^{-1}(\boldsymbol{\theta}_i^{\rm ML}-\boldsymbol{\theta})\right] \tag{3.4}$$

from which, assuming that the two datasets are uncorrelated, it follows that

$$\ln R_{ij} = -\frac{1}{2}(\boldsymbol{\theta}_i^{\rm ML}-\boldsymbol{\theta}_j^{\rm ML})^T(\Sigma_i^p+\Sigma_j^p)^{-1}(\boldsymbol{\theta}_i^{\rm ML}-\boldsymbol{\theta}_j^{\rm ML}) \tag{3.5}$$

which only depends on the parameter posterior distributions and is independent of the prior volume.

A particularly instructive case is that of two one-dimensional Gaussian likelihoods. Then $\mathcal{L}_1 = \mathcal{N}(d_1,\sigma_1)$ and $\mathcal{L}_2 = \mathcal{N}(d_2,\sigma_2)$, and we get $(-2\ln R_{12}) = (d_1-d_2)^2/(\sigma_1^2+\sigma_2^2)$. The application of our new measure to the marginalized Hubble constant H_0 likelihoods from *Planck* $\mathcal{L}_1 \sim \mathcal{N}(66.93, 0.62)$ [1] and distance ladder $\mathcal{L}_2 \sim \mathcal{N}(73.24, 1.74)$ [4], therefore, trivially gives us the values expected from Gaussian statistics i.e. $\ln R_{12} = -5.83$ [28] with a p-value 6.4×10^{-4} or 3.4σ.

JCAP01(2019)036

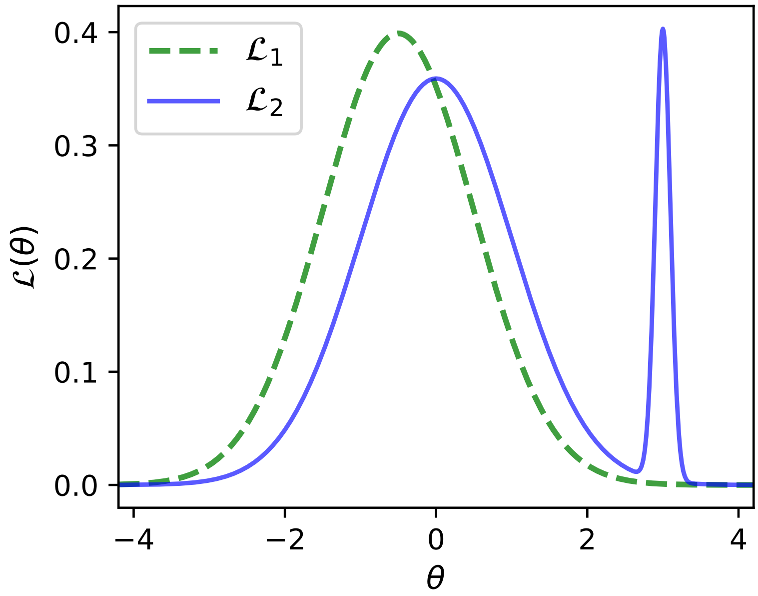

Figure 2. In this example, we verify that the two distributions shown in the figure: (i) Gaussian (dashed), and (ii) non-Gaussian (solid), are consistent with each other; the effect of the peak around $\theta = 3$ in the non-Gaussian distribution is small with $\ln R_{12} = -0.064$ (compared to $\ln R_{12} = -0.0625$ if $\mathcal{L}_2$ had no second peak at $\theta = 3$).

We note that our new measure is related to the *tension* measure defined in [29], because in some situations $E_{ij}^{\rm com}$ can be approximated by shifting one of the posterior probability density functions while preserving its shape. However, there can be ambiguity in the process of shifting one or both of the posterior distributions (for non-Gaussian and multimodal distributions), as discussed in section X.B. of [11]. That ambiguity is removed in our definition, as we reference the likelihood functions directly. We provide an example in figure 2, in which the Gaussian likelihood is simply, $\mathcal{L}_1(d|\theta) = \mathcal{N}(d, 1)$. The non-Gaussian likelihood is a (normalized) sum of two Gaussians, defined as $\mathcal{L}_2(d|\theta) = 0.9\mathcal{N}(d, 1) + 0.1\mathcal{N}(d + 3, 0.1)$. The distributions plotted in figure 2 are $\mathcal{L}_1(d = -0.5|\theta)$ (dashed) and $\mathcal{L}_2(d = 0|\theta)$ (solid). Because the combined fit is insensitive to the narrow peak near $\theta = 3$, we get $\ln R_{12} = -0.064$ without any ambiguity in how to shift the distributions, which shows that the two sets of parameters $\theta_1 = -0.5$ and $\theta_2 = 0$ from the two likelihoods are consistent, as expected. Without the additional peak at $\theta = 3$, the level of consistency is slightly better: $\ln R_{12} = -0.5 \times (0.5^2)/2 = -0.0625$, a simple verification that the new measure gets contribution from non-Gaussian features.

Next, we calculate $\ln R_{12}$ using different pairs of datasets (e.g. TT vs. EE) from the *Planck* satellite, in which case $g_i(\boldsymbol{\theta})$ is no more a simple linear function but has to evaluated numerically.

4 Application to Planck data

We use the binned and foreground-marginalized `plik_lite` likelihood from the Planck collaboration [30] which includes multipoles 30–2508 for TT power spectrum, and multipoles 30–1996 for EE power spectrum. We fix the *Planck* calibration factor y_p to 1; see section C.6.2 of [30],

Parameter	Range	Parameter	Range
$\ln(10^{10}A_s)$	[2.7, 3.4]	n_s	[0.8, 1.2]
Ω_c	[0.1, 0.45]	Ω_b	[0.044, 0.055]
H_0	[50, 95]	τ	[0.005, 0.2]

Table 1. Cosmological parameters and their prior ranges. A_s and n_s are the amplitude and spectral index of primordial scalar fluctuations, Ω_c and Ω_b are cold dark matter and baryonic matter densities, H_0 is the Hubble constant, and τ is the optical depth to reionization.

from which the CMB-only Gaussian `plik_lite` likelihood is:

$$\ln \mathcal{L}(\tilde{C}_b^{\rm CMB}|C_b^{\rm th}) = -\frac{1}{2}\mathbf{x}^T\tilde{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}\ln\left[\det(2\pi\tilde{\Sigma})\right], \tag{4.1}$$

where $\mathbf{x} = \tilde{C}_b^{\rm CMB}/y_p^2 - C_b^{\rm th}$. The binned and marginalized mean $\tilde{C}_b^{\rm CMB}$ and covariance matrix $\tilde{\Sigma}$ are provided by the *Planck* team. To evaluate the likelihood in eq. (4.1), we compute lensed $C_\ell^{\rm th}$ for a given set of parameters $\boldsymbol{\theta}$ using `camb` [31, 32] and bin the $C_\ell^{\rm th}$ using the appropriate weights to get $C_b^{\rm th}$.

Without low-multipole polarization data, the optical depth to reionization τ is only weakly constrained and is strongly degenerate with the amplitude of scalar fluctuations A_s. To break this degeneracy, we use a low-ℓ polarization prior $\tau = 0.07 \pm 0.02$. The evidences we compute are:

$$\begin{aligned} E^{\rm sep}_{\rm TT,EE} &= \int d\boldsymbol{\theta}\,\pi(\boldsymbol{\theta})\mathcal{L}(\{C_\ell^{\rm TT}(\boldsymbol{\theta}_T^{\rm ML}), C_\ell^{\rm EE}(\boldsymbol{\theta}_E^{\rm ML})\}|\boldsymbol{\theta}) \\ E^{\rm com}_{\rm TT,EE} &= \int d\boldsymbol{\theta}\,\pi(\boldsymbol{\theta})\mathcal{L}(\{C_\ell^{\rm TT}(\boldsymbol{\theta}_C^{\rm ML}), C_\ell^{\rm EE}(\boldsymbol{\theta}_C^{\rm ML})\}|\boldsymbol{\theta}) \end{aligned} \tag{4.2}$$

where $\boldsymbol{\theta}_T^{\rm ML}$ and $\boldsymbol{\theta}_E^{\rm ML}$ are obtained individually by using the respective TT and EE data, while $\boldsymbol{\theta}_C^{\rm ML}$ is the maximum likelihood model parameters from the combined fit. We obtain the maximum likelihood values $\boldsymbol{\theta}^{\rm ML}$ by using a global optimization algorithm `differential_evolution` [33] implemented in `scipy` [34]. We calculate the evidences using the `MultiNest` package [35, 36], and quote results and statistical errorbars produced by the importance nested sampling method [37]. For evidence calculations, we take uniform priors on six cosmological parameters listed in table 1.

The results are shown in table 2 where, in addition to $\ln R$, we also quote the corresponding probability-to-exceed (p-value) and Gaussian n-σ values. For the discrepancy between model parameters obtained from TT and EE spectra, we obtain $\ln R_{\rm TT,EE} = -1.93 \pm 0.03$ (approximately 0.4σ). Previous studies also find no indication of strong discrepancy between these datasets [38], albeit by using more complicated methods, or by directly using the posteriors [28].

We perform another test using the *Planck* power spectrum data, by splitting the temperature data into $\ell < 1000$ and $\ell \geq 1000$ samples and calculating $\ln R$ for these two datasets. We again find that the level of inconsistency is small with $\ln R_{\rm TTsplit} = -4.13 \pm 0.16$ or approximately 1.2σ, which agrees with the significance obtained using simulated data sets in [39]. Note that, to obtain the values in table 2, we are using the `plik_lite` likelihood in which low-ℓ ($\ell < 30$) multipoles are not included; inclusion of these large-scale multipoles would likely increase the discrepancy as their amplitude is known to be anomalously low.

datasets	TT,EE	TTlow, TThigh ($\ell_{\rm split} = 1000$)	EElow, EEhigh ($\ell_{\rm split} = 1000$)
$\ln R$	-1.93 ± 0.03	-4.13 ± 0.16	-0.83 ± 0.16
p-value	$0.7(0.4\sigma)$	$0.22(1.2\sigma)$	$0.95(0.1\sigma)$

Table 2. Summary of applications of our new statistic to *Planck* data, discussed in the text. The second column shows $\ln R$ for TT and EE as the two datasets using *Planck* `plik_lite` likelihood. The last two columns show $\ln R$ for splitting the data into two multipole ranges at $\ell_{\rm split} = 1000$ using only TT or only EE data. Each p-value is computed by taking $(-2\ln R)$ as $\chi^2_{N_{\rm params}}$ distributed, and the corresponding Gaussian n-σ value is also quoted in parentheses.

To estimate the effect of low-ℓ part of the TT likelihood, we implement an approximation to the low-ℓ likelihood following [39] (see their section 3.2 for details), which they have tested to find that the approximation gives similar cosmological parameters compared to the computationally more demanding pixel-space likelihood. To summarize: $f_\ell(2\ell+1)\hat{C}_\ell/C_\ell$ is drawn from a $\chi^2_{f_\ell(2\ell+1)}$ probability distribution function, where f_ℓ are mask-dependent fitting factors determined for the `commander` mask. Here $\hat{C}_\ell$ is the mask-deconvolved power spectrum, which we take to be the *Planck* `commander` quadratic maximum likelihood (QML) C_ℓs. Any correlation between different multipoles for $\ell < 30$ and with the `plik_lite` multipole bins is ignored. For $\ell_{\rm split} = 1000$, including the approximate low-ℓ likelihood, we now get $\ln R_{\rm TTsplit} = -5.32 \pm 0.05$ or approximately 1.6σ, which again agrees with the significance quoted in [39] obtained using simulations.

We finally carry out a similar analysis with the polarization data: we split the *Planck* EE data in multipole, using the `plik_lite` likelihood for each multipole range. The large and small scale multipole split for the EE spectrum results in consistent ΛCDM parameters: $\ln R_{\rm EEsplit} = -0.83 \pm 0.16$ (or approximately 0.1σ).

5 Summary and conclusion

We have introduced a new statistic to quantify tension between experiments. The statistic is based upon Bayesian evidence, and has advantages of not depending on the prior volumes of the parameters, and of being straightforward to apply to multiparameter, non-Gaussian likelihood distributions. We have shown that our new measure reduces to the expected discrepancy measure for Gaussian distributed posteriors, and gives sensible results in the non-Gaussian tests that we performed.

Applying the new statistic to the *Planck* power spectrum data, we find that the cosmological parameters obtained from TT and EE spectra are consistent, and that the level of discrepancy of the parameters obtained from the TT spectrum split into smaller and larger scales at $\ell_{\rm split} = 1000$ is slightly larger at about 1.6σ.

We have limited our application to just the *Planck* data in this work. It is worthwhile to apply the new measure to comparing the *Planck* constraints with weak-lensing constraints [2] and smaller-scale CMB constraints [40]. It will also be useful to consider using the statistic in the context of ΛCDM extensions. Further, we have only carefully investigated the ratio for comparing two datasets. A straightforward application of the ratio for more than two datasets might be possible by evaluating $(-2\ln R)$ as χ^2 distributed with $N_{\rm params} \times (N_{\rm sets} - 1)$ degrees of freedom, but detailed investigation of this possibility, application to likelihoods with very non-Gaussian posteriors, and application to other cosmological datasets is left for future study.

Acknowledgments

The authors are supported by NASA under contract 14-ATP14-0005. DH is also supported by DOE under Contract No. DE-FG02-95ER40899. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. We thank Marius Millea for providing the necessary f_ℓ coefficients and example code to implement the low-ℓ approximated likelihood. We are grateful to Wayne Hu, Marco Raveri, Vivian Miranda, Weikang Lin, Mustapha Ishak-Boushaki, Pavel Motloch and Michael Hobson for insightful comments.

References

[1] Planck collaboration, *Planck 2015 results. XIII. Cosmological parameters*, *Astron. Astrophys.* **594** (2016) A13 [arXiv:1502.01589] [INSPIRE].

[2] DES collaboration, *Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing*, *Phys. Rev.* **D 98** (2018) 043526 [arXiv:1708.01530] [INSPIRE].

[3] R. Trotta, *Bayes in the sky: Bayesian inference and model selection in cosmology*, *Contemp. Phys.* **49** (2008) 71 [arXiv:0803.4089] [INSPIRE].

[4] A.G. Riess et al., *A* 2.4% *Determination of the Local Value of the Hubble Constant*, *Astrophys. J.* **826** (2016) 56 [arXiv:1604.01424] [INSPIRE].

[5] J.L. Bernal, L. Verde and A.G. Riess, *The trouble with* H_0, *JCAP* **10** (2016) 019 [arXiv:1607.05617] [INSPIRE].

[6] H. Hildebrandt et al., *KiDS-450: Cosmological parameter constraints from tomographic weak gravitational lensing*, *Mon. Not. Roy. Astron. Soc.* **465** (2017) 1454 [arXiv:1606.05338] [INSPIRE].

[7] DES collaboration, *Dark Energy Survey Year 1 results: Cosmological constraints from cosmic shear*, *Phys. Rev.* **D 98** (2018) 043528 [arXiv:1708.01538] [INSPIRE].

[8] DES collaboration, *Survey geometry and the internal consistency of recent cosmic shear measurements*, *Mon. Not. Roy. Astron. Soc.* **479** (2018) 4998 [arXiv:1804.10663] [INSPIRE].

[9] S. Seehars, S. Grandis, A. Amara and A. Refregier, *Quantifying Concordance in Cosmology*, *Phys. Rev.* **D 93** (2016) 103507 [arXiv:1510.08483] [INSPIRE].

[10] T. Charnock, R.A. Battye and A. Moss, *Planck data versus large scale structure*, *Phys. Rev.* **D 95** (2017) 123535 [arXiv:1703.05959] [INSPIRE].

[11] W. Lin and M. Ishak, *Cosmological discordances: A new measure, marginalization effects and application to geometry versus growth current data sets*, *Phys. Rev.* **D 96** (2017) 023532 [arXiv:1705.05303] [INSPIRE].

[12] M.P. Hobson, S.L. Bridle and O. Lahav, *Combining cosmological datasets: hyperparameters and bayesian evidence*, *Mon. Not. Roy. Astron. Soc.* **335** (2002) 377 [astro-ph/0203259] [INSPIRE].

[13] J. Luis Bernal and J.A. Peacock, *Conservative cosmology: combining data with allowance for unknown systematics*, *JCAP* **07** (2018) 002 [arXiv:1803.04470] [INSPIRE].

[14] P. Marshall, N. Rajguru and A. Slosar, *Bayesian evidence as a tool for comparing datasets*, *Phys. Rev.* **D 73** (2006) 067302 [astro-ph/0412535] [INSPIRE].

[15] M.C. March, R. Trotta, L. Amendola and D. Huterer, *Robustness to systematics for future dark energy probes*, *Mon. Not. Roy. Astron. Soc.* **415** (2011) 143 [arXiv:1101.1521] [INSPIRE].

[16] L. Amendola, V. Marra and M. Quartin, *Internal Robustness: systematic search for systematic bias in SN Ia data*, *Mon. Not. Roy. Astron. Soc.* **430** (2013) 1867 [arXiv:1209.1897] [INSPIRE].

[17] J. Martin, C. Ringeval, R. Trotta and V. Vennin, *Compatibility of Planck and BICEP2 in the Light of Inflation*, *Phys. Rev.* **D 90** (2014) 063501 [arXiv:1405.7272] [INSPIRE].

[18] S. Joudaki et al., *CFHTLenS revisited: assessing concordance with Planck including astrophysical systematics*, *Mon. Not. Roy. Astron. Soc.* **465** (2017) 2033 [arXiv:1601.05786] [INSPIRE].

[19] M. Raveri, *Are cosmological data sets consistent with each other within the Λ cold dark matter model?*, *Phys. Rev.* **D 93** (2016) 043522 [arXiv:1510.00688] [INSPIRE].

[20] S. Seehars, A. Amara, A. Refregier, A. Paranjape and J. Akeret, *Information Gains from Cosmic Microwave Background Experiments*, *Phys. Rev.* **D 90** (2014) 023533 [arXiv:1402.3593] [INSPIRE].

[21] S. Grandis, D. Rapetti, A. Saro, J.J. Mohr and J.P. Dietrich, *Quantifying tensions between CMB and distance data sets in models with free curvature or lensing amplitude*, *Mon. Not. Roy. Astron. Soc.* **463** (2016) 1416 [arXiv:1604.06463] [INSPIRE].

[22] S.M. Feeney et al., *Prospects for resolving the Hubble constant tension with standard sirens*, arXiv:1802.03404 [INSPIRE].

[23] A. Nicola, A. Amara and A. Refregier, *Consistency tests in cosmology using relative entropy*, arXiv:1809.07333 [INSPIRE].

[24] K.B. Petersen and M.S. Pedersen, *The matrix cookbook*, Technical University of Denmark (2012).

[25] M. Raveri and W. Hu, *Concordance and Discordance in Cosmology*, arXiv:1806.04649 [INSPIRE].

[26] R.E. Kass and A.E. Raftery, *Bayes factors*, *J. Am. Statist. Assoc.* **90** (1995) 773 [INSPIRE].

[27] S.S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, *Annals Math. Statist.* **9** (1938) 60 [INSPIRE].

[28] W. Lin and M. Ishak, *Cosmological discordances II: Hubble constant, Planck and large-scale-structure data sets*, *Phys. Rev.* **D 96** (2017) 083532 [arXiv:1708.09813] [INSPIRE].

[29] L. Verde, P. Protopapas and R. Jimenez, *Planck and the local Universe: Quantifying the tension*, *Phys. Dark Univ.* **2** (2013) 166 [arXiv:1306.6766] [INSPIRE].

[30] Planck collaboration, *Planck 2015 results. XI. CMB power spectra, likelihoods and robustness of parameters*, *Astron. Astrophys.* **594** (2016) A11 [arXiv:1507.02704] [INSPIRE].

[31] A. Lewis, A. Challinor and A. Lasenby, *Efficient computation of CMB anisotropies in closed FRW models*, *Astrophys. J.* **538** (2000) 473 [astro-ph/9911177] [INSPIRE].

[32] C. Howlett, A. Lewis, A. Hall and A. Challinor, *CMB power spectrum parameter degeneracies in the era of precision cosmology*, *JCAP* **04** (2012) 027 [arXiv:1201.3654] [INSPIRE].

[33] R. Storn and K. Price, *Differential evolution — a simple and efficient heuristic for global optimization over continuous spaces*, *J. Global Optim.* **11** (1997) 341.

[34] E. Jones, T. Oliphant and P. Peterson, *SciPy: Open source scientific tools for Python*, (2001) http://www.scipy.org/.

[35] F. Feroz and M.P. Hobson, *Multimodal nested sampling: an efficient and robust alternative to MCMC methods for astronomical data analysis*, *Mon. Not. Roy. Astron. Soc.* **384** (2008) 449 [arXiv:0704.3704] [INSPIRE].

[36] F. Feroz, M.P. Hobson and M. Bridges, *MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics*, *Mon. Not. Roy. Astron. Soc.* **398** (2009) 1601 [arXiv:0809.3437] [INSPIRE].

JCAP01(2019)036

[37] F. Feroz, M.P. Hobson, E. Cameron and A.N. Pettitt, *Importance Nested Sampling and the MultiNest Algorithm*, arXiv:1306.2144 [INSPIRE].

[38] A. Shafieloo and D.K. Hazra, *Consistency of the Planck CMB data and ΛCDM cosmology*, *JCAP* **04** (2017) 012 [arXiv:1610.07402] [INSPIRE].

[39] Planck collaboration, *Planck intermediate results. LI. Features in the cosmic microwave background temperature power spectrum and shifts in cosmological parameters*, *Astron. Astrophys.* **607** (2017) A95 [arXiv:1608.02487] [INSPIRE].

[40] SPT collaboration, *A Comparison of Cosmological Parameters Determined from CMB Temperature Power Spectra from the South Pole Telescope and the Planck Satellite*, *Astrophys. J.* **850** (2017) 101 [arXiv:1706.10286] [INSPIRE].