An Informational Approach to Cosmological Parameter Estimation

Michelle Stephens\(^\star\) and Marcelo Gleiser\(^\dagger\)

Department of Physics and Astronomy
Dartmouth College, Hanover, NH 03755, USA

(Dated: June 25, 2019)

We introduce a new approach for cosmological parameter estimation based on the information-theoretical Jensen-Shannon Divergence (JSD), calculating it for models in the restricted parameter space \(\{H_0, w_0, w_a\}\), where \(H_0\) is the value of the Hubble constant today, and \(w_0\) and \(w_a\) are dark energy parameters, with the other parameters held fixed at their best-fit values from the Planck 2018 data. As an application, we investigate the \(H_0\) tension between the Planck data and the local astronomical data by comparing the \(\Lambda\)CDM model with the \(w\)CDM and the \(w_0w_a\)CDM dynamic dark energy models. We find agreement with other works using the standard Bayesian inference for parameter estimation, but, in addition, show that while the JSD is equally minimized for both values of \(H_0\) along the \((w_0, w_a)\) plane, the lines of degeneracy are different for each value of \(H_0\). This allows for distinguishing between the two, once the value of either \(w_0\) or \(w_a\) is known.

I. INTRODUCTION

The concordance cosmological model, \(\Lambda\)CDM, is in excellent agreement with data spanning a broad range of redshifts, including the temperature anisotropies of the cosmic microwave background (CMB) \(^1\), the large-scale galaxy clustering feature of baryon acoustic oscillations (BAO) \(^2\), and the luminosity-redshift relation of local sources, calibrated primarily with type Ia supernovae data (SNeIa) \(^3\). It has six independent parameters, assuming the dark energy (DE) equation of state is a constant \(w = -1\), and a flat universe: the amplitude and spectral index of the primordial density perturbations, \(A_s\) and \(n_s\) respectively, the reionization optical depth \(\tau\), the present-day Hubble parameter \(H_0\), and the present-day physical baryon and dark matter densities \(\Omega_b h^2\) and \(\Omega_c h^2\), respectively, where \(h = H_0/100\).

The Friedmann equation for a spatially flat universe with a cosmological constant \(\Lambda\), in the matter-dominated era \((z \ll 3200)\) can be written as \(\Omega_0 + \Omega_\Lambda = 1\), where \(H_0\) determines the critical density normalization on \(\Omega_X\). Thus, the cosmic expansion history and structure formation in the universe is sensitive to the relative contributions of \(\Omega_m = \Omega_b + \Omega_c\), and \(\Lambda\). Despite the overall success of \(\Lambda\)CDM, statistically significant tensions exist between early-universe parameter inference and their direct local measurement, most notably in the value of the Hubble parameter today, \(H_0\) \(^4\) \(^5\).

Measurements of the CMB anisotropies at \(z \approx 1100\) by the Planck \(^6\) and WMAP \(^7\) missions constrain the combinations \(\Omega_b h^2\) and \(\Omega_c h^2\), but degeneracies prevent constraints of \(H_0\) alone \(^8\) \(^10\). Local measurements can probe \(H(z)\) directly through the luminosity-redshift relation, but distances to sources must be carefully calibrated to avoid systematic error. Uncertainties have been reduced to the sub-percent level in the case of the Planck analysis, and to the 1\% level with recent advances in the local distance-ladder determinations \(^11\) \(^12\). Excluding an as-yet unknown source of error in either of these analyses, the discrepancy may point to new physics.

Several possible resolutions to the Hubble tension have been proposed, including evolving DE with a phantom-like equation of state \(^13\), additional neutrinos \(^14\) \(^15\), local voids \(^16\), and pre-recombination modifications to DE (early dark energy) \(^17\), among many others \(^18\) \(^19\) \(^20\). Given the many data sets, extending the cosmological parameter space and performing a Markov-Chain Monte Carlo analysis to determine the most likely parameters is a computationally expensive problem \(^21\), and involves many complications in constructing the likelihood function arising from particular instrumentation, data set considerations, and prior choices \(^22\).

Here, we propose an alternative approach to cosmological parameter estimation based on a measure from information theory known as the Jensen-Shannon Divergence (JSD). We apply it to a one-parameter extension of the \(\Lambda\)CDM model, the \((w_0, w_a)\) parametrization of an evolving dark energy component. In section II we review the standard maximum likelihood method before introducing the JSD and the information theory needed for its interpretation, with a toy example. In section III we provide motivation for using the JSD to examine linear evolving DE models and detail the numerical approach. Results are presented in section IV. In section V we highlight prospects for extending this study in future work.

II. PARAMETER ESTIMATION

II.1. Maximum Likelihood Estimation

The standard approach for cosmological parameter estimation is a problem of Bayesian inference: we begin with a dataset \(D\), which we wish to accurately represent with a model parameterized by \(\theta\). We assume a prior distribution over the parameters, \(p(\theta)\). The prior ideally represents our best knowledge of the parameters, but in practice is commonly taken to be uniform. The model is specified by the form of the likelihood function, \(L(\theta) \equiv p(D|\theta)\) — the probability that the data
II.2. Jensen-Shannon Divergence

Claude Shannon’s seminal 1948 paper [27] provides the foundation for the definition and interpretation of the JSD. We typically think of describing a message or event in terms of a distinct encoding scheme — a set of symbols $\mathcal{N} = \{n_1, n_2, \ldots, n_L\}$. For instance, in English we encode words with the 26 letters of the alphabet, and full messages with additional characters for punctuation. The information content of a particular symbol is $I(n) = -\log_2 p(n)$, where $p(n)$ is the probability distribution over symbols in $\mathcal{N}$ determined from some collection of events encoded by $\mathcal{N}$. The expected value of information in a particular event is then

$$
\langle I \rangle = -\sum_{n \in \mathcal{N}} p(n) \log_2 p(n). \quad (1)
$$

The Jensen-Shannon Divergence (JSD) is a measure of the difference between two probability distributions, based on the Kullback-Leibler divergence (DKL) between two distributions $p(n)$ and $q(n)$, defined as [28]

$$
D_{KL}(p \mid\mid q) = \sum_n p_n \log \left( \frac{p_n}{q_n} \right). \quad (2)
$$

If $q$ is treated as a model for some “true” distribution $p$, DKL is a measure of the information lost in using $q$ rather than $p$. DKL is positive-definite, and is zero only if the two distributions are the same (known as the identity of indiscernibles). However, it is not symmetric and does not satisfy the triangle inequality: while we picture it as a “distance” between two distributions, it is not a metric. The Jensen-Shannon Divergence (JSD) is a symmetrized extension of DKL that can be treated as a true metric on the space of probability distributions [29]. It is defined as

$$
D_{JS} = \frac{1}{2} D_{KL}(p \mid\mid q) + \frac{1}{2} D_{KL}(q \mid\mid p), \quad (3)
$$

where $r = \frac{1}{2}(p + q)$. $0 \leq D_{JS} \leq 1$ if the logarithm used in the DKL is base 2. In this case, information is measured in bits. With the exception of the next example, we use the natural logarithm, so that $0 \leq D_{JS} \leq \ln(2)$.

As a simple illustration of the JSD, consider a collection of short messages in English. Let’s suppose we eliminate the spaces, capitalization, punctuation, etc., so that the set of symbols from which each message is drawn is simply the 26-character English alphabet. Using the JSD, we can find out how well each of these messages models the phrase “Cosmology Rocks”.

![FIG. 1. The JSD between the model messages and the reference message is shown, where the messages are case-insensitive and drawn from the English alphabet of 26 letters. This illustrates that the JSD is sensitive only to the identity and relative frequencies of letters in a message. It is maximal for distributions with nothing in common, and minimizes (at zero) for identical distributions.](image-url)
with the multipole moment $l$, and the frequency with $C_l$, it is the location and relative scaling of the acoustic peaks that provides the bulk of the CMB’s sensitivity to cosmological parameters. Finally, we note that if $q_n = p_n + \delta p_n$, with $\delta p_n \ll p_n$, expanding the JSD to first order in $\delta p_n$ gives a measure proportional to the chi-square.

III. METHODS

III.1. Dark Energy and the $H_0$ Tension

The $H_0$ tension can be stated in this way: late-time scale factor expansion is occurring faster than we would expect from $\Lambda$CDM, with parameter constraints inferred from early universe data. Framed this way, it is easy to see why most of the proposed resolutions involve modifying DE in some way. Pre-recombination modifications to DE can alter the sound horizon $r_s$ and thus change the inferred $H_0$, while minute shifts in other parameters maintain the agreement with CMB anisotropies [17]. Late-time modifications are an obvious mechanism to alter the expansion history and galaxy clustering, but are constrained by other measurements, notably BAO [30–32]. Constraining the DE equation of state is challenging, because density parameters and $H(z)$ are sensitive to a function of its integral over redshift.

Future observational surveys, like the DESI probe planned for 2019 [33], will be able to provide direct constraints on $w(z)$. Until then, phenomenological models have been introduced to capture what the general behavior of $w \neq -1$ might look like, and its influence on cosmological observables. A common parametrization for evolving DE is the linear evolution model $w = w_0 + w_a(1 - a)$, where $w_0$ is the value today and $w_a = -dw/da$ [34]. In this framework, $\Lambda$CDM corresponds to $w_0 = -1, w_a = 0$, and other constant-$w$ models can be considered by setting $w_a = 0$.

Several studies have extended the $\Lambda$CDM basic six-parameter model to include these parameters, constraining them in the extended space via standard MCMC max-likelihood methods. References [13, 21, 35, 36] are an incomplete list.

III.2. Data and Numerical Approach

We compare a model’s prediction for the angular power spectrum to the Planck 2018 data by computing $\mathcal{D}_{\text{JS}}(F_l^{\text{mod}} || F_l^{\text{plk}})$ as in Equation 3 where $F_l^{\text{mod}}$ and $F_l^{\text{plk}}$ are determined from the model-predicted and Planck data-calculated angular power spectra, respectively. That is,

$$F_l = \frac{D_l}{\sum_i D_i},$$

where $D_l = l(l+1)C_l/2\pi$. We use $D_l$ to compute the JSD since it more clearly distinguishes the acoustic features.

The unbinned $C_l$ computed from the temperature fluctuations observed by Planck can be found on the Planck Legacy Archive [37]; we use public release 3’s baseline high-$l$ Planck TT power spectrum. The cosmological Boltzmann code CAMB (Code for Anisotropies in the Microwave Background) [38] is used to compute the angular power spectrum for a given model. The base set of cosmological parameters and their best-fit values as determined by the Planck 2018 analysis are summarized in Table I. All parameters except $H_0$ are left fixed at their best-fit values: $H_0$ is then set to be either 67.32 or 74.03 km/s/Mpc, the values reported by Planck 2018 (hereafter P18) and Riess, et al. 2019 (hereafter R19) [12], respectively.

![Table I. Planck 2018 best-fit parameter values](image)

Note that $H_0$ is an inferred value from the fitted $100\Theta_r$; they may be used interchangeably in the base parameter set.

For each $H_0$, we allow the DE equation of state to vary in the linear parametrization $w = w_0 + w_a(1 - a)$. The distance, in terms of the JSD, from each model to the Planck data can be summarized by two surfaces: $\mathcal{D}_{\text{JS}}(w_0, w_a \mid h = 0.6732)$ and $\mathcal{D}_{\text{JS}}(w_0, w_a \mid h = 0.7403)$. Our goal here is not to determine which value of $H_0$ is the “correct” one, but to investigate whether a model with a modified DE equation of state can shift the inferred $H_0$ closer to R19. Such a model would have a shorter distance (in the sense of the JSD) to the Planck data, and provide an alternative method of parameter estimation. A forthcoming full analysis will allow $H_0$ to vary along with the other parameters in Table I.

IV. RESULTS

Figure 2 shows the $\mathcal{D}_{\text{JS}}$ as a function of $w_0$ and $w_a$ for both the P18 and R19 values of $H_0$. Both surfaces display a valley running along a degenerate minimum curve where $\mathcal{D}_{\text{JS}} \simeq 1.82 \times 10^{-3}$. Given the interpretation of $\mathcal{D}_{\text{JS}}$ as a metric between the two distributions, its minimum represents the preferred parameter data set. On the $H_0 = 67.32$ surface, the $\Lambda$CDM model is identified with a larger red point. The bottom plot of Figure 2 shows the degenerate curves projected onto the $(w_0, w_a)$ plane and fitted with a second-order polynomial. Since the two curves are different, once one of the two parameters is known, this approach allows for the determination
of the other, thus breaking the degeneracy in a predictive way.

FIG. 2. Top: $D_{ls}$ surface for the $\Lambda$CDM and R19 values of $H_0$, $w_0$ and $w_a$ have been allowed to vary, but the other parameters were fixed at their best-fit values from Planck 2018 (see Table I). The red and green curves are the lines of degeneracy in the model space; their distance from the Planck data is the same. Bottom: The lines of degeneracy from above are projected into the $w_0-w_a$ plane and fit with a second order polynomial, $w_a = c_0 w_0^2 + c_1 w_0 + c_2$. $[c_0, c_1, c_2] = [-1.196, -6.084, -4.897]$ and $[-0.982, -6.440, -6.411]$ for P18 $H_0$ and R19 $H_0$, respectively.

We also considered that data for the Hubble parameter as a function of redshift, $H(z)$, along with the BAO volume-averaged effective distance ratio $D_V(z)/r_s(z_{dec})$ could break the degeneracy in this model space compared to using the CMB data only. While in future work we will include these data sets a priori in the minimization of the JSD for a model, we used them here to examine their effect on the degeneracy found using only the Planck data. Using models along the degenerate curves, we used $\chi^2$ to fit 38 measurements of $H(z)$, compiled from [39] and references therein, and 12 measurements of the BAO data, compiled from [31, 32]. The $\chi^2(w_0)$ for each $H_0$ and both data sets is shown in Figure 3.

The additional, late-time data sets do reduce the degeneracy. Table II reports the values of $w_0$ and $w_a$ that minimize the $\chi^2$ for the BAO and $H(z)$ data, for both the P18 and R19 $H_0$. The $H(z)$ data slightly favors the P18 $H_0$.

FIG. 3. For models along the curves of degeneracy found in Figure 2, the $\chi^2$ is plotted as a function of $w_0$ for (top) the model’s prediction of the BAO $D_V(z)/r_s(z_{dec})$ and the measured data, and (bottom) the model’s prediction for $H(z)$ and the measured data. The red and green curves come from using the P18 value and R19 value of $H_0$, respectively.

| Data $H_0$ (km/s/Mpc) | $w_0$  | $w_a$  | $\chi^2$ |
|----------------------|--------|--------|----------|
| BAO P18             | −0.89  | −0.43  | 0.137    |
| BAO R19             | −1.26  | 0.14   | 0.137    |
| $H(z)$ P18          | −1.06  | 0.21   | 40.1     |
| $H(z)$ R19          | −1.40  | 0.68   | 43.3     |

TABLE II. Values of $w_0$ and $w_a$ along the curves of degeneracy from Figure 2 that minimize the $\chi^2$ to the BAO and $H(z)$ data sets.

V. CONCLUDING REMARKS

In this work, we introduced a new method to estimate cosmological parameters, based on the Janssen-Shannon divergence $D_{JS}$ of information theory. As a first application, we examined here the current tension in the value of the expansion rate $H_0$, comparing the extended $\Lambda$CDM temperature anisotropy spectrum for models with dynamic DE parameterized in ($w_0, w_a$) space with the Planck 2018 temperature anisotropy data. For both values of $H_0$, we found that there are curves of degeneracy in the $(w_0, w_a)$ plane, characterized as nearly indistinguishable minima of the $D_{JS}(w_0, w_a)$ surface. The two curves are different, however, allowing for degeneracy breaking and for the resolution of the $H_0$ tension, once one of the two parameters is known. Extending our analysis to
include $H(z)$ and BAO at different redshifts, we found that the data upholds the tension by slightly favoring the lower value of $H_0$ inferred by Planck 2018.

In a forthcoming paper, we plan to extend this analysis by running an MCMC to minimize the $D_M$ in the full seven-parameter space, and to include data from BAO, $H(z)$, and the Planck polarization spectrum in the minimization. Obviously, this more complete approach will probably change the results plotted in Fig. 2, which should be considered our method's first illustrative example. With this more complete analysis, we will thus be able to directly compare parameter confidence intervals from our information-based analysis to others reported in the literature. Our current results warrant further investigation of $D_M$ as an alternative and transparent method of cosmological parameter estimation.

**ACKNOWLEDGMENTS**

MG and MS were supported in part by a Department of Energy grant [de-sc0010386]. MS is a Hull doctoral fellow at Dartmouth College.

[1] Y. Akrami et al. [Planck Collaboration], arXiv:1807.06205 [astro-ph.CO].
[2] A. G. Sanchez, et al., arXiv:1203.6616 [astro-ph].
[3] M. Lopez-Corredoira, F. Melia, E. Lusso and G. Risaliti, Int. J. Mod. Phys. D 25, no. 05, 1650060 (2016) doi:10.1142/S0218271816500607 [arXiv:1602.06743 [astro-ph.CO]].
[4] W. L. Freedman, Nat. Astron. 1, 0121 (2017) doi:10.1038/s41550-017-0121 [arXiv:1706.02739 [astro-ph.CO]].
[5] J. L. Bernal, L. Verde and A. G. Riess, JCAP 1610, no. 10, 019 (2016) doi:10.1088/1475-5164/2016/10/019 [arXiv:1607.05617 [astro-ph.CO]].
[6] N. Aghanim et al. [Planck Collaboration], arXiv:1807.06209 [astro-ph.CO].
[7] C. L. Bennett et al., ApJS 208, 20.
[8] M. Zaldarriaga, D. N. Spergel and U. Seljak, Astrophys. J. 488, 1 (1997) doi:10.1086/304692 [astro-ph/9702157].
[9] G. Efstathiou and J. R. Bond, Mon. Not. Roy. Astron. Soc. 304, 75 (1999) doi:10.1046/j.1365-8711.1999.02274.x [astro-ph/9807103].
[10] C. Howlett. A. Lewis, A. Hall and A. Challinor, arXiv:1201.3565 [astro-ph.CO].
[11] A. G. Riess et al., Astrophys. J. 826, no. 1, 56 (2016) doi:10.3847/0004-637X/826/1/56 [arXiv:1604.01421 [astro-ph.CO]].
[12] A. G. Riess, S. Casertano, W. Yuan, L. M. Macri and D. Scolnic, arXiv:1903.07603 [astro-ph.CO].
[13] E. Di Valentino, A. Melchiorri, E. V. Linder and J. Silk, Phys. Rev. D 96, no. 2, 023523 (2017) doi:10.1103/PhysRevD.96.023523 [arXiv:1704.00762 [astro-ph.CO]].
[14] S. Carneiro, P. C. de Holanda, C. Pigozzo and F. Sobreira, arXiv:1812.06004 [astro-ph.CO].
[15] V. Poulin, K. K. Boddy, S. Bird and M. Kamionkowski, Phys. Rev. D 97, no. 12, 123504 (2018) doi:10.1103/PhysRevD.97.123504 [arXiv:1803.02474 [astro-ph.CO]].
[16] V. Marra, L. Amendola, I. Sawicki and W. Valkenburg, Phys. Rev. Lett. 110, no. 24, 241305 (2013) doi:10.1103/PhysRevLett.110.241305 [arXiv:1303.3121 [astro-ph.CO]].
[17] V. Poulin, T. L. Smith, T. Karwal and M. Kamionkowski, arXiv:1811.04083 [astro-ph.CO].
[18] E. Di Valentino, E. V. Linder and A. Melchiorri, Phys. Rev. D 97, no. 4, 043528 (2018) doi:10.1103/PhysRevD.97.043528 [arXiv:1710.02153 [astro-ph.CO]].
[19] E. Di Valentino, A. Melchiorri and O. Mena, Phys. Rev. D 96, no. 4, 043503 (2017) doi:10.1103/PhysRevD.96.043503 [arXiv:1704.08342 [astro-ph.CO]].
[20] K. Bolejko, Phys. Rev. D 97, no. 10, 103529 (2018) doi:10.1103/PhysRevD.97.103529 [arXiv:1712.02667 [astro-ph.CO]].
[21] E. Di Valentino, A. Melchiorri and J. Silk, Phys. Rev. D 92, no. 12, 121302 (2015) doi:10.1103/PhysRevD.92.121302 [arXiv:1507.06646 [astro-ph.CO]].
[22] P. A. R. Ade et al. [Planck Collaboration], Astron. Astrophys. 571, A15 (2014) doi:10.1051/0004-6361/201321573 [arXiv:1303.5075 [astro-ph.CO]].
[23] R. Trotta, Contemp. Phys. 49, 71 (2008) doi:10.1080/00107510802066753 [arXiv:0803.4089 [astro-ph]].
[24] J. R. Bond, G. Efstathiou and M. Tegmark, Mon. Not. Roy. Astron. Soc. 291, L33 (1997) doi:10.1093/mnras/291.L33 [astro-ph/9702100].
[25] M. White, D. Scott and J. Silk, Ann. Rev. Astron. Astrophys. 32, 319-70 (1994).
[26] A. Lewis and S. Bridle, Phys. Rev. D 66, 103511 (2002) doi:10.1103/PhysRevD.66.103511 [astro-ph/0205436].
[27] C. E. Shannon, The Bell System Technical Journal, 27, 379 (1948).
[28] S. Kullback and R. A. Leibler, Ann. Math. Statist. 22, 79–86 (1951).
[29] J. Lin, IEEE Transactions on Information Theory, 37, no. 1 Jan 1991.
[30] A. Aubourg et al., Phys. Rev. D 92, no. 12, 123516 (2015) [arXiv:1411.1074 [astro-ph.CO]].
[31] B. S. Haridasu, V. V. Lukovi and N. Vittorio, JCAP 1805, no. 05, 033 (2018) doi:10.1088/1475-7516/2018/05/033 [arXiv:1711.03929 [astro-ph.CO]].
[32] C. Cheng and Q. G. Huang, Sci. China Phys. Mech. Astron. 58, no. 9, 59801 (2015) doi:10.1007/s11433-015-5684-5 [arXiv:1409.6119 [astro-ph.CO]].
[33] M. Vargas-Magana et al., arXiv:1901.01581.
[34] M. Chevallier and D. Polarski, Int. J. Mod. Phys. D 10, 213 (2001) doi:10.1142/S0218271801000822 [gr-qc/0009008].
[35] W. L. Freedman and B. F. Madore, Ann. Rev. Astron. Astrophys. 48, 673 (2010) doi:10.1146/annurev-astro-082708-101829 [arXiv:1004.1856 [astro-ph.CO]].
[36] Q. Xia, H. Li and X. Zhang, Phys. Rev. D 88, 063501 (2013) doi:10.1103/PhysRevD.88.063501
[37] Planck Legacy Archive, URL: [http://pla.esac.esa.int/]

[38] A. Lewis, A. Challinor and A. Lasenby, Astrophys. J. **538**, 473 (2000) [arXiv:astro-ph/9911177].

[39] F. K. Anagnostopoulos and S. Basilakos, Phys. Rev. D **97**, no. 6, 063503 (2018) doi:10.1103/PhysRevD.97.063503 [arXiv:1709.02356 [astro-ph.CO]].

http://camb.info/