*Statistical Challenges in 21st Century Cosmology*
*Proceedings IAU Symposium No. 306, 2014*
*A. F. Heavens, J.-L. Starck & A. Krone-Martins, eds.*

doi:10.1017/S1743921314013623

# Information Gains in Cosmological Parameter Estimation

**Sebastian Seehars[1], Adam Amara, Alexandre Refregier, Aseem Paranjape and Joël Akeret**

ETH Zurich, Department of Physics, Wolfgang-Pauli-Strasse 27, 8093 Zurich, Switzerland
[1]email: seehars@phys.ethz.ch

**Abstract.** Combining datasets from different experiments and probes to constrain cosmological models is an important challenge in observational cosmology. We summarize a framework for measuring the constraining power and the consistency of separately or jointly analyzed data within a given model that we proposed in earlier work (Seehars *et al.* 2014). Applying the Kullback-Leibler divergence to posterior distributions, we can quantify the difference between constraints and distinguish contributions from gains in precision and shifts in parameter space. We show results from applying this technique to a combination of datasets and probes such as the cosmic microwave background or baryon acoustic oscillations.

**Keywords.** methods: data analysis, cosmology

## 1. Introduction

Observations of cosmological probes such as the cosmic microwave background (CMB) or Type Ia supernovae have established a standard model of cosmology, resting on the concepts of a cosmological constant $\Lambda$ driving the observed accelerated expansion of the universe at late times and a cold dark matter (CDM) component accounting for about 80% of the matter content in the universe. To learn more about these mysterious dark components, it is crucial to test the $\Lambda$CDM model with observations from different astrophysical probes and independent experiments.

We summarize a technique for measuring the progress in constraining the parameters of a cosmological model from different datasets by quantifying the variation between the constraints using the Kullback-Leibler divergence that we proposed in earlier work (Seehars *et al.* 2014). This technique measures both shifts in parameter space and improvements in the precision of the constraints and is hence able to determine the constraining power of a dataset and its consistency with other probes. After formally introducing the method in section 2, we apply it to joint analyses of CMB data from the WMAP mission and measurements of small scale CMB experiments, baryon acoustic oscillations (BAO), and the expansion rate of the universe as published by the WMAP team in Hinshaw *et al.* (2013) as well as to other large scale CMB experiments in section 3. We conclude with a discussion of our findings in section 4.

## 2. Parameter Estimation and Kullback-Leibler Divergence

We wish to compare two posteriors on the same parameter space in order to measure gains in precision and consistency of the constraints on a given cosmological model. For simplicity, we consider the special case of two complementary datasets which can be analyzed in a sequential manner, i.e. where constraints from one dataset can be used as prior information for analyzing the second. In many applications of cosmology, however, two datasets can be correlated due to cosmic variance, which is the fact that we can observe only a single realization of an inherently stochastic cosmological model. The concepts presented next can also be applied to those scenarios and we refer the reader to Seehars *et al.* (2014) for more details. Returning to the case of two uncorrelated observables $\mathcal{D}_1$ and $\mathcal{D}_2$, we consider two posteriors $p_1(\Theta) \equiv p(\Theta|\mathcal{D}_1)$ and

$$p_2(\Theta) \equiv p(\Theta|\mathcal{D}_1, \mathcal{D}_2) = \frac{\mathcal{L}(\Theta; \mathcal{D}_2)p_1(\Theta)}{\int d\Theta\, \mathcal{L}(\Theta; \mathcal{D}_2)p_1(\Theta)}. \tag{2.1}$$

To measure the difference between $p_1$ and $p_2$, we use the Kullback-Leibler divergence (KL-divergence, Kullback & Leibler (1951)):

$$D(p_2||p_1) \equiv \int d\Theta\, p_2(\Theta) \log \frac{p_2(\Theta)}{p_1(\Theta)}. \tag{2.2}$$

The KL-divergence is well suited for our purposes: It is a positive quantity which is zero if and only if $p_1$ is equal to $p_2$ almost everywhere and can be interpreted as a pseudo-distance between probability densities (it is not a distance because of the asymmetry in exchanging $p_1$ and $p_2$). The KL-divergence is furthermore invariant under invertible parameter transformations and consequently measures differences between the constraints on the model and not only a particular parametrization of the model.

We will distinguish two contributions to the KL-divergence between two posteriors: A contribution from the improvements in precision that are expected even before the actual data is gathered and a second contribution from the shifts in parameter space induced by the observed data. The former can be defined by the expectation value of $D(p_2||p_1)$ when averaging over the prior knowledge on the data:

$$\langle D \rangle \equiv \int d\mathcal{D}_2\, p(\mathcal{D}_2) D(p_2||p_1), \tag{2.3}$$

where $p(\mathcal{D}_2) \equiv \int d\Theta\, \mathcal{L}(\Theta; \mathcal{D}_2)p_1(\Theta)$. The latter is given by the difference between observed and expected KL-divergence $S \equiv D(p_2||p_1) - \langle D \rangle$ and called *surprise* in the following. Similarly, we can consider the expected variations of $D$ around $\langle D \rangle$

$$\sigma^2(D) \equiv \int d\mathcal{D}_2\, p(\mathcal{D}_2)(D(p_2||p_1) - \langle D \rangle)^2. \tag{2.4}$$

In general, we can think of the observed KL-divergence as a realization from the distribution of KL-divergences induced by the prior distribution on $\mathcal{D}_2$ and our knowledge of the likelihood.

Equation (2.2) could be estimated with a numerical Monte Carlo integrator such as nested sampling (Skilling 2004). Equation (2.3), however, is much harder to evaluate numerically. Yet, as the expected KL-divergence is a well-known quantity in Bayesian experimental design (see e.g. Chaloner & Verdinelli (1995)), computational approaches exist. For the applications to constraints on a flat $\Lambda$CDM cosmology in section 3 we will take a different approach: As the $\Lambda$CDM parameters are tightly constrained by CMB data alone, it is a reasonable approximation to assume a linear model and normal distributions for prior and likelihood when updating these constraints. In this case, equations (2.2)

https://doi.org/10.1017/S1743921314013623 Published online by Cambridge University Press

**Table 1.** KL-divergence estimates in bits for considered combinations of datasets. In the *data combination* column, WMAP refers to the full WMAP 9 data (Bennett *et al.* 2013). The BAO, $H_0$, and small scale CMB (eCMB) data are described in Hinshaw *et al.* (2013). The other CMB datasets are from the Boomerang (Jones *et al.* 2006) and Planck (Planck collaboration 2013) teams. The updating column refers to *add* if complementary data is added, *replace* if the dataset is completely replaced, and *part* if parts of the data are replaced. The p-value is an estimate for the prior probability of observing a surprise that is greater or equal (less or equal) than $S$ if $S$ is greater (smaller) than zero. It is an approximation when data is partially replaced.

| Data combination | Updating | $D$ | $\langle D\rangle$ | $S$ | $\lvert S/\sigma(D)\rvert$ | p-value |
|---|---|---|---|---|---|---|
| BOOMERANG → WMAP | replace | 22.5 | 18.4 | 4.1 | 1.6 | 0.07 |
| WMAP → WMAP + eCMB | add | 2.1 | 1.7 | 0.4 | 0.5 | 0.2 |
| WMAP + eCMB → WMAP + eCMB + BAO | add | 1.3 | 1.0 | 0.3 | 0.8 | 0.2 |
| WMAP + eCMB → WMAP + eCMB + $H_0$ | add | 0.4 | 0.3 | 0.1 | 0.1 | 0.3 |
| WMAP + eCMB → WMAP + eCMB + BAO + $H_0$ | add | 0.9 | 1.1 | −0.2 | 0.2 | 0.6 |
| WMAP → Planck + WP | part | 29.8 | 7.9 | 21.9 | 6.5 | 0.0002 |

and (2.3) are analytic and given by (Seehars *et al.* 2014):

$$D(p_2||p_1) = \frac{1}{2}\left((\Theta_1-\Theta_2)^T\Sigma_1^{-1}(\Theta_1-\Theta_2)+\mathrm{tr}(\Sigma_2\Sigma_1^{-1})-d-\log\frac{\det\Sigma_2}{\det\Sigma_1}\right), \tag{2.5}$$

$$\langle D\rangle = -\frac{1}{2}\log\frac{\det\Sigma_2}{\det\Sigma_1}, \tag{2.6}$$

$$\sigma^2(D) = \frac{1}{2}\mathrm{tr}\left((\Sigma_1^{-1}\Sigma_2-\mathbb{1})^2\right), \tag{2.7}$$

with $\Theta_i$ and $\Sigma_i$ being mean and covariance of distribution $p_i$, $d$ being the dimensionality of the parameter space, and $\mathbb{1}$ being the $d$-dimensional identity matrix. Equation 2.5 shows that the KL-divergence depends on the ratio of covariance matrices as well as the significance of the shifts in the means $\Theta_1$ and $\Theta_2$. While the former also governs the expected relative entropy (2.6), the latter is driving the surprise contribution. In order to estimate equations (2.5) to (2.7), it is left to estimate mean and covariance matrix of the posteriors from Monte Carlo Markov chains, for example. Note furthermore that the distribution of the KL-divergence induced by prior knowledge on $\mathcal{D}_2$ is a generalized $\chi^2$-distribution (Seehars *et al.* 2014).

## 3. Application to Data

We apply the results from section 2 to two scenarios: The parameter constraints on a flat ΛCDM cosmology from the joint analyses of the final CMB data release by the WMAP team with other cosmological probes as published by Hinshaw *et al.* (2013) and the constraints from a historical series of CMB experiments. While we use the official Monte Carlo Markov chains from the WMAP team to estimate the information gains in the former application, we use the CosmoHammer framework (Akeret *et al.* 2013) to generate samples for the comparison between CMB datasets. When estimating the p-values of the observed KL-divergences on its generalized $\chi^2$ distribution, we use the R-package `CompQuadForm` by Duchesne & De Micheaux (2010). The results in bits, i.e. when taking the logarithm in equation (2.2) to base two, are shown in Table 1.

Adding external data from small scale CMB experiments, BAO, and $H_0$ measurements to the WMAP constraints results in an information gain between 0.4 bits for the $H_0$ prior and 2.1 bits for small scale CMB data. We generally find small surprise compared to the expected variations in $D$. Hence, a flat ΛCDM model is consistent with all datasets and changes in the constraints are coming from the expected statistical variations in the data.

When analyzing the differences between the constraints from different CMB experiments, we must consider the effects of cosmic variance. This changes the form of equations 2.6 and 2.7, but similar results can be derived when considering the case of replacing correlated data. Results when comparing different CMB experiments are extensively discussed in Seehars *et al.* (2014). Here we focus on replacing CMB data from Boomerang data with WMAP 9 data and replacing the WMAP temperature power spectrum with the results of the Planck team. Both data updates generate large information gains in parameter space as measured by KL-divergences of 22.5 and 29.8 bits, respectively. In these cases, however, it is important to look at the decomposition of $D$ into the expected KL-divergence and the surprise: Comparing the Boomerang constraints with the WMAP 9 results, we find that the KL-divergence is dominated by the increased precision in the constraints as measured by $\langle D \rangle = 18.4$ bits. The Planck update also improves the precision of the constraints ($\langle D \rangle = 7.9$) but furthermore introduces shifts in parameter space that are significantly larger than expected a priori with a surprise of 21.9 ($p = 0.0002$). This significant surprise implies that the model is not able to consistently fit WMAP and Planck temperature data within the errors and hints towards either systematics in the data or physics beyond a flat $\Lambda$CDM cosmology.

## 4. Conclusions

We described a technique for measuring variations in the constraints on cosmological parameters coming from different cosmological probes and datasets that we proposed in earlier work (Seehars *et al.* 2014). It is based on applying the Kullback-Leibler divergence to the posteriors of two separately or jointly analyzed measurements. With this technique, we are able to separate contributions from gains in precision and shifts in parameter space by comparing the observed KL-divergence to a-priori expectations.

By applying these concepts to the constraints on a flat $\Lambda$CDM cosmology from CMB, BAO, and $H_0$ measurements, we show that this technique is able to quantify inconsistencies between the full posteriors that are not immediately apparent from the distributions of the individual parameters. In particular, we find no inconsistencies between the constraints on $\Lambda$CDM parameters from joint analyses of WMAP data with small scale CMB, BAO, and $H_0$ data as published by Hinshaw *et al.* (2013). Comparing the constraints from a joint analysis of Planck temperature and WMAP polarization data with prior expectations from the constraints of WMAP data alone, however, we find significant surprise contributions hinting towards tensions between the datasets. The described method may thus be a valuable tool for quantifying inconsistencies between data from different experiments in order to detect systematics or even signs of new physics.

## References

Akeret, J. , Seehars, S., Amara, A., Refregier, A., & Csillaghy, A. 2013 *Astron. Comput.*, 2, 27

C L Bennett, D Larson, J L Weiland, & N Jarosik 2013 *ApJS*, 208(2), 54

K Chaloner, & I Verdinelli 1995 *Statist. Sci.*, 10, 273

P Duchesne, & P L De Micheaux 2010 *Comput. Stat. Data Anal.*, 54(4), 858

G Hinshaw, D Larson, E Komatsu, D N Spergel, C L Bennett, J Dunkley, M R Nolta, M Halpern, R S Hill, N Odegard, L Page, K M Smith, J L Weiland, B Gold, N Jarosik, A Kogut, M Limon, S S Meyer, G S Tucker, E Wollack, & E L Wright 2013 *ApJS*, 208(2), 19

William C Jones, PAR Ade, J J Bock, J R Bond, J Borrill, A Boscaleri, P Cabella, C R Contaldi, B P Crill, & P De Bernardis 2006 *ApJ*, 647(2), 823

S Kullback, & R A Leibler 1951 *Ann. Math. Stat.*, 22(1), 79

Planck collaboration, P A R Ade, N Aghanim, C Armitage-Caplan, M Arnaud, M Ashdown, F Atrio-Barandela, J Aumont, C Baccigalupi, A J Banday, *et al.* 2013 *arXiv*, 1303.5075

S Seehars, A Amara, A Refregier, A Paranjape, & J Akeret 2014 *arXiv*, 1402.3593

J Skilling 2004 *AIP Conference Proceedings*, 735(1), 395