

Considerations in the Interpretation of Cosmological Anomalies

Hiranya V. Peiris

Department of Physics and Astronomy, University College London,
Gower Street, London WC1E 6BT, United Kingdom
email: h.peiris@ucl.ac.uk

Abstract. Anomalies drive scientific discovery – they are associated with the cutting edge of the research frontier, and thus typically exploit data in the low signal-to-noise regime. In astronomy, the prevalence of systematics — both “known unknowns” and “unknown unknowns” — combined with increasingly large datasets, the widespread use of *ad hoc* estimators for anomaly detection, and the “look-elsewhere” effect, can lead to spurious false detections. In this informal note, I argue that anomaly detection leading to discoveries of new physics requires a combination of physical understanding, careful experimental design to avoid confirmation bias, and self-consistent statistical methods. These points are illustrated with several concrete examples from cosmology.

Keywords. cosmic microwave background, cosmological parameters, early universe, cosmology: miscellaneous, methods: statistical, methods: data analysis

1. Introduction

In the next decade, Big Data will form a pivotal part of the experimental landscape in cosmology, with a multitude of large surveys, from cosmic microwave background (CMB) experiments, photometric and spectroscopic galaxy surveys, 21 cm arrays, and direct-detection gravitational wave observatories, expected to yield a deluge of information about the origin and evolution of the Universe. The associated high data rates and huge data volumes, bringing with them the curse of dimensionality, pose challenging problems for data analysis and scientific discovery. Some of the challenges include: (1) data compression (*i.e.*, the formulation of almost-sufficient statistics), filtering, sampling, associated with very large datasets; (2) making robust and accurate inferences or conclusions from such datasets; (3) the small signal-to-noise regime in which discoveries at the research frontier are often made; (4) a very large model space; (5) cosmic variance (*i.e.*, the fact that we have access to a single realisation of an inherently stochastic cosmological model).

We can imagine two qualitatively different kinds of modelling that would be required to extract the maximum information from this cornucopia of cosmological data. **Mechanistic** (physical) modelling is what has driven cosmological discovery thus far – in this case, forward modelling based on physics is feasible, leading directly to standard parameter estimation and model comparison analyses that form the bread and butter of modern cosmology. However, as cosmological data probe deeper into the non-linear regime of the evolution of structure, against complex foregrounds, **empirical** (data-driven) modelling — characterising relationships discovered in the data — will gain increasing prominence. Such models may be purely data-driven, or qualitatively based on physics but be required in cases where forward modelling is infeasible; they may be used to postulate new theories, or generate statistical predictions for new observables.

In this context, the treatment of data anomalies acquires great importance. Anomalies — unusual data configurations — can consist of statistical outliers, unusual concentrations of data points or sudden behaviour changes. They may arise from chance configurations due to random fluctuations, systematics (unmodelled astrophysics; instrument/detector artefacts; data processing distortions), or they can indicate genuinely new discoveries. Determining into which of these categories a given anomaly falls is fraught with difficulty, especially as humans have evolved in such a way that we can often spuriously identify patterns in data where none exist, a phenomenon known as *pareidolia*. When considering whether a cosmological anomaly indicates new physics, one must consider that many such anomalies are often identified using *a posteriori* estimators, which leads to spurious enhancements of detection significance. In the absence of an alternative model for comparison, one often cannot account for the “look-elsewhere effect” arising from multiple testing, or formulate model priors to compare with the standard model. In the absence of an alternative theory, how can we judge if a given anomaly represents new physics? In the rest of this article, I will summarise several case studies demonstrating different approaches to answering this question.

2. Assessing anomalies

There are two distinct steps associated with anomalies in the cosmological context. The first involves finding data anomalies in the first place — this provides an area of rich algorithmic development related to assessing measures of irregularity, unexpectedness, unusualness etc., especially in the coming “Big Data” era. The second involves drawing inferences from the results of the search for anomalies — *i.e.*, assessing whether the anomaly is due to random chance, or whether it represents an unknown mechanism which may point to new physics (or, more prosaically, systematics). In making these inferences, the particle physicists’ “look-elsewhere effect”, or multiple testing, comes into play. One must correct for any *a posteriori* choices that were made in the process of detecting an anomaly, and properly account for the possible ways in which an anomaly could have shown up (but did not).

In cases where an alternative model motivated the search for an anomaly or the formulation of the statistic in which an anomaly was discovered, the probability that the anomaly represents a new physical mechanism (versus random chance) can be assessed using Bayesian model comparison: the model and parameter priors account for the uncertainties associated with multiple-testing. Even in this case, however, encapsulating all the relevant uncertainties can be quite difficult. For example, Cruz *et al.* (2007) carried out a Bayesian model comparison analysis of the well-known cosmic microwave background (CMB) anomaly termed the *Cold Spot* and concluded that it was likely a texture (a type of spatially-localised cosmic defect). This was a highly sophisticated analysis — one of the first principled Bayesian model comparison analyses in the context of CMB anomalies. Nevertheless, it represents an incomplete attempt to account for *a posteriori* selection effects associated with basing the analysis on a single feature at a particular location which is known to be a statistical outlier.

The texture model predicts a statistically isotropic distribution of spatially-localised features in the CMB sky (with an expectation value for the sky fraction covered by textures), and textures can lead to both hot and cold spots in the CMB. Interpreted as a correction for the “look-elsewhere” effect of having analysed a patch of sky containing the Cold Spot, the Cruz *et al.* (2007) analysis accounts for the expected sky fraction covered by textures in a patch, but does not account for the fact that textures could be placed anywhere on the sky. The coordinates of the central position of the texture

template should be marginalised over in the context of statistical isotropy, whereas the texture template was centred on the Cold Spot in the analysis. There is also a factor of two associated with the fact that the feature could have been a hot spot, rather than a cold spot. In Feeney *et al.* (2012, 2013), where the texture hypothesis is formulated in terms of a hierarchical Bayesian model, the impact of these *a posteriori* selection effects becomes clear and, when fully accounting for these uncertainties in computing the marginal likelihood, there is no preference for augmenting the standard cosmological model with textures.

This example highlights how the “look-elsewhere” effect can spuriously enhance the significance of the anomaly when all the relevant uncertainties are not incorporated into the formulation of an alternative model representing new physics. Nevertheless, if one possesses alternative model(s) with well-motivated priors, the standard Bayesian model comparison framework can be brought to bear on the problem, providing the most straightforward scenario for assessing anomalies.

3. “Just-so” models

If the anomaly is detected as an unusual property of a large dataset and is not motivated by an alternative model, accounting for the “look-elsewhere” effect can be highly non-trivial. In this context, designer theories that stand in for best-possible explanations — “just-so” models — can prove useful in terms of gaining an intuition for whether the anomaly provides evidence for new physics.

In this case one would proceed as follows: (a) find a *designer theory* or “just-so” model which maximises the likelihood of the anomaly; (b) thus determine the maximum available likelihood gain for this particular anomaly with respect to the standard model, or null hypothesis; (c) judge whether this is compelling given the baroque nature of the designer theory.

Such an example can be found in the context of another well-known CMB anomaly often termed the $C(\theta)$ anomaly (Copi *et al.* 2007). It is a long-standing observation that the statistic $S_{1/2} = \int_{60^\circ}^{180^\circ} [C(\theta)]^2 \cos(\theta) d\theta$, where $C(\theta)$ is the angular correlation function of the CMB, is anomalous when evaluated outside typical Galactic sky cuts. Here, the lower integration limit is an *a posteriori* choice (Bennett *et al.* 2011). Under the standard Λ CDM model, the probability of obtaining a cut-sky $S_{1/2}$ statistic of the observed value or less is $\sim 0.03\%$; however, it is notable that the *full sky* value of $S_{1/2}$ (evaluated on reconstructed full sky maps or Galactic foreground-cleaned maps) is not anomalous.

It is possible that such an observation can arise in a cosmological model which breaks statistical isotropy, in contrast to the standard Λ CDM model; however, a specific well-motivated broken-isotropy model predicting the observed characteristics of the large-angle CMB sky is not currently available. Pontzen & Peiris (2010) used a convex optimisation algorithm to maximise the likelihood of the cut sky $S_{1/2}$ statistic subject to fixed full sky angular power spectrum multipoles C_ℓ over all anisotropic Gaussian models with zero mean[†]. Using this technique, they constructed a designer anisotropic model (Fig. 1) which gives the maximum likelihood improvement over Λ CDM: this model, which had ~ 6900 degrees of freedom (compared to the isotropic case with eight) gave an improvement in likelihood of $\ln \mathcal{L} \sim 5$. This allowed for a finite limit to be placed on the Bayesian statistical gain available under a wide class of alternative straw-man models, providing a plausible way to probe the significance of this *a posteriori* anomaly.

[†] The covariance matrix of spherical harmonic coefficients $a_{\ell m}$ can be arbitrarily correlated, as long as it is positive-definite.

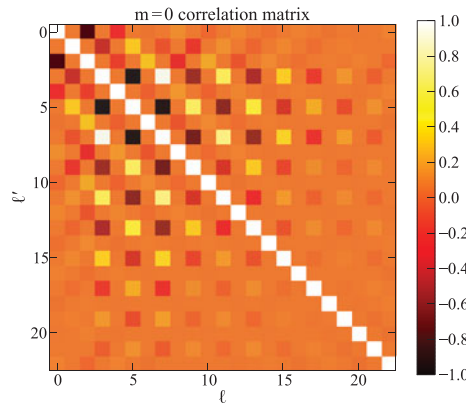


Figure 1. The $m = 0$ component of the covariance matrix defining the designer anisotropic model derived for the $S_{1/2}$ anomaly (Pontzen & Peiris 2010).

It is interesting to note that the decision-making process after computing the likelihood gain supplied by the just-so model involves a subjective judgement. For example, Pontzen & Peiris (2010) concluded that the level of fine-tuning involved in this designer model in order to provide a relatively small likelihood gain did not make for a compelling pointer towards new physics. However, a more speculative scientist may make a different decision to continue the search for an alternative theory, assuming that the number of degrees of freedom can be dramatically reduced by unknown symmetries. In this case, the designer model may provide physical intuition about the correlation properties necessary in such a theory in order to reproduce the observations.

A further example of the utility of just-so models in cosmology where physically-motivated model priors are not available is found in the context of the number of relativistic degrees of freedom N_{eff} , which may differ from the standard value of 3.046 due to, *e.g.*, extra neutrino species. The fundamental debate here focuses on whether tensions observed in cosmological datasets require the standard cosmological model to be augmented with a non-standard N_{eff} . While this may appear to be a straightforward Bayesian model comparison problem, the model space leading to N_{eff} differing from the standard value is vast, and simple priors on N_{eff} may not capture the physical prior uncertainties. In this context, Verde *et al.* (2013) show that the *profile likelihood ratio* (Wilks 1938) — the ratio of the maximum likelihood conditioned on a particular value of the parameter of interest and the overall, unconditional maximum likelihood — provides the parameter value at which a designer model would maximise the Bayesian evidence. While it is unreasonable to assess a model by tuning it to the same data being used to test it, this upper limit on the marginal likelihood provides an heuristic assessment of the possible need for an extra parameter in the absence of well-motivated parameter priors, as well as a pointer towards its most likely value.

4. Data-driven models

As mentioned above, data-driven models are often constructed *a posteriori* to fit various statistical anomalies observed in large data sets. In order to be testable, such tuned models must, of necessity, make predictions for new data beyond those they were designed to fit. In the context of CMB anomalies, several studies demonstrate how models designed to mimic apparent isotropy-breaking features in the CMB temperature field and features in the CMB temperature power spectrum can be tested with new data not used in

constructing the model, such CMB polarisation (Dvorkin *et al.* 2008; Mortonson *et al.* 2009), large-scale structure data (Pullen & Hirata 2010), and non-Gaussianity (Adshead *et al.* 2011; Peiris *et al.* 2013).

The statistical technique of cross-validation can be used as a powerful method to test the consistency of physical inferences from datasets by constructing a data-driven model describing part of the data, and checking how well it predicts the rest of the data. For example, this technique has been used to characterize the deviation from scale invariance in the primordial power spectrum in a minimally-parametric way (Verde & Peiris 2008; Peiris & Verde 2010; Bird *et al.* 2011). The idea is as follows: (1) Choose a functional form which allows a great deal of freedom in the form of the deviation from scale invariance (*e.g.*, smoothing splines)†. (2) Now perform cross-validation: set aside some of the data (validation set), fit the rest (training set), and see how well it predicts the validation set. A very good fit to the training set, which poorly predicts the validation set, indicates over-fitting of noisy data. (3) The final ingredient is a parameter that penalises fine-scale structure of the functional form. By performing cross-validation as a function of this parameter, one can judge when the structure in the smoothing spline is what the data requires without fitting the noise. The technique is also very powerful in detecting systematic issues in data analysis: Verde & Peiris (2008) identified a “kink” in the reconstructed power spectrum from WMAP at a particular scale which subsequently turned out to be a problem in point source subtraction (Huffenberger *et al.* 2008; Peiris & Verde 2010).

5. Blind analysis

The data analysis challenges of next-generation cosmological surveys require careful experimental design to minimize false detections due to experimenters’ (subconscious) bias. This has been an integral part of the particle physics and medical research culture for decades, but has yet to find broad adoption within cosmology. As cosmological constraints make the transition from *precise* to *accurate*, and the search for new physics leads to the hunt for small signals embedded within the Λ CDM cosmology, the systematic bias due to non-blind experimental design can no longer be ignored if we hope to make convincing claims for paradigm-shifting new physics from cosmological data.

Blind analysis is based on the simple idea that the *value* of a measurement does not contain any information about its *correctness*. Knowing the value of a measurement is therefore of no use in performing the analysis itself. Blind analysis is necessary because data collection, analysis and inference necessarily involves a human stage, which can lead to unquantifiable subjective inferences. Some examples of the origin of such bias include: looking for bugs when a result does not conform to expectation (and not looking for them when it does); looking for additional sources of systematic uncertainty when a result does not conform to expectation; deciding whether to publish, or wait for more data; choosing cuts while looking at the data and knowing whether it fits expectations; preferentially keeping / dropping outlier data. Examples of good experimental design which are used in cosmology, such as “double unblind” (doing two independent analyses in parallel), “mock data” analysis, and “semi-blind” (using a fraction of the data to calibrate the analysis and freezing the pipeline before doing the final analysis), while extremely helpful in their own right, nevertheless do not constitute blind analysis.

† Naively fitting this to the data will lead one to fit the fluctuations due to noise, with arbitrary improvement in the goodness-of-fit due to over-fitting.

Harrison (2002) and Roodman (2003) give examples of systematic trends in measurements (periods of surprisingly small variation, followed by jumps of several standard deviations) in the presence of experimenters' bias using examples from particle physics, and provide some excellent strategies for implementing blind analysis, including:

- **Encrypting the science result.** *e.g.*, adding a non-changing random number (not revealed to the analyst) to a numerical result or transforming a variable, in order to thwart preconceptions due to “expected” results. It is important to note that it is not necessary to blind how the result *changes* due to changes in the analysis pipeline, or to blind calibration data.

- **Hiding the “signal region”.** This is useful when the observable is a peak or localised feature in datasets.

- **Blind injection of signals into the data.** This is useful to test biases in rare event searches, and has been adopted by the gravitational wave direct detection community.

- **Mixing in an unknown fraction of simulated data** during calibration, etc.

The first of these, the so-called *hidden offset* method, is a straightforward technique to implement at the level of the parameter estimation step for essentially all cosmological data types; it works as follows. The parameter estimation code adds a fixed, unknown random number to the fitted value of the measured parameter, $x^* = x + \mathcal{R}$; x^* is returned with the true error and the likelihood value instead of x . \mathcal{R} can be set *e.g.*, by sampling from a Gaussian of mean zero and standard deviation \sim few times the experimental standard deviation. Relative changes in the result as the analysis changes can be hidden using a second offset. Informative visual aspects in plots may need to be hidden as well.

While blind analysis within observational cosmology may be conceptually more difficult than in an experimental field such as particle physics, nevertheless numerous considerations motivate us to adopt at least some practices that lead to the mitigation of experimenters' bias. Such considerations include: (i) the accuracy and reliability of scientific inferences from next-generation surveys; (ii) best return on the enormous investment of public funds in cosmological surveys, including satellite experiments and large data analysis efforts requiring huge teams, which may be very difficult — if not impossible — to replicate due to cost considerations; (iii) the substantial wasted scientific effort in going down blind alleys due to premature announcements of false detections, a side effect of which is the potential damage to the public perception of science; and (iv) missing discoveries of new physics due to overly-cautious treatment of data that do not agree with expectations.

Even simply thinking about how to blind an analysis pipeline can lead to a greater understanding of potential pitfalls. Adopting blind analysis practices requires a shift in community standards, and leadership by large data collaborations. A very welcome shift in this direction has been made by the weak lensing community, *e.g.*, in the recent CFHTLenS analyses (Heymans *et al.* 2012; Fu *et al.* 2014). The POLARBEAR *B*-mode polarisation blind analysis (Ade *et al.* 2014) represents a milestone in the CMB context, and such techniques have been in use for some time in supernova cosmology (Conley *et al.* 2006). In general, setting all the free parameters in analysis pipelines using only null-tests is a good blind analysis practice. An example of blind mitigation of systematics in constraining primordial non-Gaussianity using quasar surveys (Leistedt & Peiris 2014; Leistedt *et al.* 2014) is presented elsewhere in these proceedings.

6. Summary

I have presented an informal summary, from a practitioner's viewpoint, of techniques for evaluating cosmological anomalies. The discussion has emphasised the pitfalls

associated with multiple-testing, or the “look-elsewhere” effect. I have given practical examples of techniques that can be used in the absence of alternative models in order to gain intuition on whether or not a given cosmological anomaly represents new physics. Finally, I have discussed blind analysis as a strategy to guard against experimenters’ subconscious bias, which introduces “unknown unknowns” into physical inferences. These techniques do not represent a cure-all for data analysis problems, but rather cultivate a mindset that attempts to rise to Feynman’s challenge: “*The first principle is that you must not fool yourself — and you are the easiest person to fool.*”

Acknowledgements— I thank Stephen Feeney, David Hand, Boris Leistedt, Daniel Mortlock, Andrew Pontzen, Aaron Roodman, and Licia Verde for influencing my evolving take on this topic. I am grateful to Andrew Pontzen for generating Fig. 1, which was previously unpublished.

References

- Ade, P. *et al.* 2014, *ApJ*, 794, 171
 Adshad, P., Hu, W., Dvorkin, C., & Peiris, H. V. 2011, *Phys. Rev.*, D84, 043519
 Bennett, C. L., Hill, R. S., Hinshaw, G., *et al.* 2011, *ApJS*, 192, 17
 Bird, S., Peiris, H. V., Viel, M., & Verde, L. 2011, *MNRAS*, 413, 1717
 Conley, A. J. *et al.* 2006, *ApJ*, 644, 1
 Copi, C., Huterer, D., Schwarz, D., & Starkman, G. 2007, *Phys. Rev.*, D75, 023507
 Cruz, M., Turok, N., Vielva, P., Martinez-Gonzalez, E., & Hobson, M. 2007, *Science*, 318, 1612
 Dvorkin, C., Peiris, H. V., & Hu, W. 2008, *Phys. Rev.*, D77, 063008
 Feeney, S. M., Johnson, M. C., McEwen, J. D., Mortlock, D. J., & Peiris, H. V. 2013, *Phys. Rev.*, D88, 043012
 Feeney, S. M., Johnson, M. C., Mortlock, D. J., & Peiris, H. V. 2012, *Phys. Rev. Lett.*, 108, 241301
 Fu, L., Kilbinger, M., Erben, T., *et al.* 2014
 Harrison, P. F. 2002, *Journal of Physics G: Nuclear and Particle Physics*, 28, 2679
 Heymans, C., Van Waerbeke, L., Miller, L., *et al.* 2012, *MNRAS*, 427, 146
 Hufenberger, K. M., Eriksen, H. K., Hansen, F. K., Banday, A. J., & Górski, K. M. 2008, *ApJ*, 688, 1
 Leistedt, B. & Peiris, H. V. 2014, *MNRAS*, 444, 2
 Leistedt, B., Peiris, H. V., & Roth, N. 2014, ArXiv e-prints
 Mortonson, M. J., Dvorkin, C., Peiris, H. V., & Hu, W. 2009, *Phys. Rev.*, D79, 103519
 Peiris, H., Easther, R., & Flauger, R. 2013, *JCAP*, 1309, 018
 Peiris, H. V. & Verde, L. 2010, *Phys. Rev.*, D81, 021302
 Pontzen, A. & Peiris, H. V. 2010, *Phys. Rev.*, D81, 103008
 Pullen, A. R. & Hirata, C. M. 2010, *JCAP*, 5, 27
 Roodman, A. 2003, in *Statistical Problems in Particle Physics, Astrophysics, and Cosmology*, ed. L. Lyons, R. Mount, & R. Reitmeyer, 166
 Verde, L., Feeney, S. M., Mortlock, D. J., & Peiris, H. V. 2013, *JCAP*, 9, 13
 Verde, L. & Peiris, H. V. 2008, *JCAP*, 0807, 009
 Wilks, S. S. 1938, *The Annals of Mathematical Statistics*, 9, 60