

Precision cosmology, Accuracy cosmology and Statistical cosmology

Licia Verde^{1,2}

¹ ICREA & Instituto de ciencias del Cosmos, Iniversitat de Barcelona ICC-UB IEEC

² Institute of Theoretical Astrophysics, University of Oslo, 0315 Oslo, Norway
email: liciaverde@icc.ub.edu

Abstract. The avalanche of data over the past 10-20 years has propelled cosmology into the “precision era”. The next challenge cosmology has to meet is to enter the era of accuracy. Because of the intrinsic nature of studying the Cosmos and the sheer amount of data available now and coming soon, the only way to meet this challenge is by developing suitable and specific statistical techniques. The road from precision Cosmology to accurate Cosmology goes through statistical Cosmology. I will outline some open challenges and discuss some specific examples.

Keywords. methods: data analysis, methods: statistical, Cosmology: cosmological parameters, cosmology: large-scale structure of universe

1. Introduction

Cosmology in the past twenty years or so has made the transition to *precision cosmology*. Cosmological parameters that were known only within an order of magnitude are now measured with percent precision. This transition was brought about by the avalanche of data provided by massive large-scale structure surveys and ambitious mapping of the Cosmic Microwave Background (CMB) radiation. Thanks to this, over the past twenty years, cosmology has made the transition from a data-starved science to a data driven science. As a result, cosmology has now a standard (base, or Λ CDM) model. The standard model for cosmology requires only a handful of parameters to describe the origin, composition and evolution of the entire Universe. However, there is a big difference between modelling and understanding. While the standard cosmological model works extremely well it is highly unsatisfactory, as it has many ingredients we do not understand and we know it is incomplete. This is driving virtually all the experimental efforts in the field of the near future. From the modelling point of view, there are many plausible extensions of the base model, where in practice one or more parameters are being added. In these extended models the “precision” even of the base parameters gets significantly degraded.

In this Cosmology is special among the experimental sciences, but the peculiarity of Cosmology goes deeper than that.

2. Cosmology is special

As an experimental science Cosmology has a peculiarity in that we cannot make controlled experiments. When we mention experiments we should always bear in mind that we can only make observations[†]. And we only have one observable Universe. This is what I call the curse of cosmology. There are two important consequences of this limitation. One is that all we can do is to fit models to observations. To be more precise we constrain

[†] I have seen this limitation only in disciplines like the study of the effect of nuclear fallout from “unlikely” (and unlucky) events such as accidents with power plants and bombs.

numerical values of the model's parameters using observations; very few quantities are direct measurements: therefore any statement is model-dependent. To make matters worst, non-linearities (which are very hard to model) and poorly known astrophysical processes (*gastrophysics*, for aficionados) get in the way. As a result, different observations are more or less “clean” and robust and more or less trustable. It is however somewhat a question of personal taste which I like to compare to the Standard & Poor's credit rating for countries (enough said).

Cosmological results (and constraints on cosmological parameters) therefore depend not only on the adopted cosmological model but also on the data sets one is (willing to) consider. Therefore there is not one, unique, determination from cosmology of a given quantity (cosmological parameter). This can be quite confusing at first for those not working in the field.

Like every cloud has a silver lining, the curse of cosmology is also a blessing: we might have only one observable Universe and we might not perform controlled experiments on it, but we can –at least in principle– observe all there is to see. This is what I call “ultimate experiment” in cosmology. The big development for the near future (which has however already started in the past few years) is that advances in observational techniques have made possible to perform ultimate experiments and the next generation of large-scale structure surveys will be a collection of “ultimate experiments”. For example the *Planck* satellite (Planck collaboration, 2013a, Planck collaboration, 2013b) has provided us with the “ultimate experiment” for the Cosmic Microwave Background *primary* temperature anisotropies. A mission like the proposed *COrE+* (COrE collaboration, 2012; PRISM collaboration, 2014) will provide the ultimate experiment for CMB lensing and polarisation – both E modes and B modes– signal.

The other important consequence of having only one observable Universe is more subtle. In Cosmology we believe that the observed Universe is only one part of the entire Universe, which is seen as an *ensemble* of all possible observ-ed/able Universes of which the observed one is a random draw. Ambitiously, from the limited observations we can gather, we want to make *inferences* about the properties of the entire Universe. In this sense, statistics is very much ingrained in cosmology: any theoretical model will mostly predict the statistical properties of the Universe. The observations are intrinsically a statistical quantity. There is therefore a fundamental error floor for any measurement given by cosmic variance. No wonder many cosmologists are Bayesian!

3. Challenges

The avalanche of data of the past twenty years has brought about challenges, which the community was able to address and solve. The next generation of surveys will push the field in to the “big data” era. Undoubtedly this will bring in “big challenges” but, given that so far the community has a good track record on this, I will not focus on this aspect here.

I will instead dwell on the fact that with more data available, making the statistical error shrink, the systematic errors must be kept under exquisite control. Here we are entering uncharted territory. While there is a at least well defined framework to deal with statistical errors, there is no systematic way to address systematic errors.

There are several types of systematic errors, and here is my personal view (paraphrasing Rumsfeld).

- There are things that we know that we know. But what if what we think we know isn't true? Recall that any statement in cosmology is model dependent. “Essentially, all models are wrong, but some are useful” (Box and Draper, 1987).

- Known unknowns. These are the systematic effects that are probably easier to deal with: at least we know what we have to watch out for. Below I will give an example of how to deal with these.

- Unknown unknowns. Here the general wisdom is to try to measure the same thing in different, independent ways (e.g., different data) and compare. It might be a very expensive route, but at the moment it is the only option.

To summarise, after precision cosmology, cosmology should strive to become accurate. This is not news, see Peebles (2002). Systematic effects may be in the data but may also be in the model used for their interpretation. In what follows I will present a small selection of examples where the application of statistical techniques can help in the transition from precision to accurate cosmology. This selection is not meant to be exhaustive or representative, it just cover some of the problems I have been working on with my collaborators over the past couple of years.

3.1. *Example 1: Being Bayesian with non-Gaussianity*

The search for deviations from primordial non-Gaussianity is a very active research subject: any deviation from the simplest implementation of slow-roll inflation implies primordial non-Gaussianity, possibly at a detectable level. In addition, even if the initial conditions were Gaussian, non-linear gravitational evolution creates non-Gaussianity. Despite the fact that Bayesian methods are routinely employed in cosmology, the state-of-the-art non-Gaussianity analyses are done in the Frequentist framework: a bispectrum estimator is employed which has been shown to be optimal but only in the limit of vanishing non-Gaussianity.

The problem in applying the Bayesian method in this context is that a full, analytic probability distribution function does not exist even for the simplest non-Gaussian model. The available literature on this has concentrated on the *local* type of non-Gaussianity where the field of interest Φ is expressed in terms of a Gaussian auxiliary field ϕ and a dimensionless non-Gaussianity parameter f_{NL} :

$$\Phi(\mathbf{x}) = \phi(\mathbf{x}) + f_{NL} [\phi^2(\mathbf{x}) - \langle \phi^2(\mathbf{x}) \rangle]. \tag{3.1}$$

Recall that $\langle \phi^2(\mathbf{x}) \rangle = \sigma_\phi^2$. Here we also consider this model. The goal is to find the posterior distribution of the amplitude of non-Gaussianities given the data, $P(f_{NL}|d)$; this includes information from all correlation orders not just the bispectrum. This has been attempted by Elsner and Wandelt (2010), Elsner *et al.* (2010), Ensslin *et al.* (2008), here I follow the approach of Verde *et al.* (2013). I will only outline the philosophy of the approach and explain the implications of the findings, the detailed expressions and equations may be found in Verde *et al.* (2013).

To understand the set up, let us warm up by re-deriving known results in the Gaussian case.

Let us define a field A , which is the observable field and can be the temperature of the CMB or the a_l^m or the matter overdensity field etc. The value of the observed field at any spatial point i is related to the underlying potential Φ via

$$A(\mathbf{x}) = A_i = \int d^3y \mathcal{M}(\mathbf{x}, \mathbf{y}) \Phi(\mathbf{y}) \equiv (\mathcal{M}, \Phi). \tag{3.2}$$

The full information about the Gaussian field ϕ is given by the Gaussian generating functional $\mathcal{P}[\phi]$; for example the joint multivariate probability of A_1, \dots, A_n is

$$\mathcal{P}(A_1, A_2, \dots, A_n) = \int [\mathcal{D}\phi] \mathcal{P}(\phi) \prod_{i=1}^n \delta^D [A_i - (\mathcal{M}, \Phi(\phi))]. \tag{3.3}$$

Here,

$$\mathcal{P}(\phi) = \frac{\exp[-1/2(\phi, K_0, \phi)]}{\int D[\phi] \exp[-1/2(\phi, K_0, \phi)]} \quad (3.4)$$

and we have used the shorthand notation

$$\int d^3y d^3z \phi(\mathbf{y}) K(\mathbf{y} - \mathbf{z}) \phi(\mathbf{z}) \equiv (\phi, K, \phi), \quad (3.5)$$

with K_0 defined by the relation to the 2-point correlation function of ϕ , ξ^ϕ :

$$\int d^3y K_0(|\mathbf{x} - \mathbf{y}|) \xi^\phi(|\mathbf{y} - \mathbf{z}|) = \delta(|\mathbf{x} - \mathbf{z}|). \quad (3.6)$$

This is a very powerful approach although sometimes it is tricky to deal with the Dirac delta function which Fourier representation involves highly oscillatory functions. On the other hand, observations are often discretized/pixelized, so we discretize these integrals to simplify their analytic calculation, then, if needed, one take the continuous limit at the end. All the above convolutions then become just matrix operations.

For example it is easy to see that if \mathcal{M} is invertible and Φ is Gaussian then A is also Gaussian with covariance given by the covariance of A : $\hat{K}_0 = (\mathcal{M}^{-1})^T K_0 \mathcal{M}^{-1}$ (matrix multiplication) and $[\hat{K}_0^{-1}]_{ij} = \xi_{ij}^A$.

If \mathcal{M} is not invertible (say Φ is the three dimensional gravitational potential and the observable field A is the CMB temperature spherical harmonic coefficients a_ℓ^m) it is still possible to split it in two blocks $\mathcal{M} = (\mathfrak{N} \ \mathcal{N})$ where the \mathfrak{N} block is invertible. Then standard linear algebra yields the known results i.e., that the A (i.e., a_ℓ^m) field is Gaussian with correlation function given by $\mathcal{M}^T \xi^\Phi \mathcal{M}$.

The moment a nonzero f_{NL} is introduced, things complicate. It is instructive to first consider the case where \mathcal{M} is the identity matrix. Then it is well known that in this case one can solve the Dirac delta function in Eq. (3.3) by finding its roots. Neglecting the exponentially suppressed root (which is an excellent approximation over virtually all the support of the Probability Density Function, PDF) it is possible to write down the exact multi-variate PDF, $\mathcal{P}(\Phi)$.

Of course in the more general case for \mathcal{M} , once one has the expression for $\mathcal{P}(\Phi)$, one could simply say:

$$\mathcal{P}(\vec{A}|f_{\text{NL}}) = \int \mathcal{D}[\Phi] \mathcal{P}(\Phi) \delta^D(A - \int M\Phi). \quad (3.7)$$

This then must be integrated numerically via Monte-Carlo methods, as for example explored in Elsner *et al.* (2010), Elsner and Wandelt (2010). This is computationally extremely intensive, as the dimensionality of the integrand is given by Φ it is not that of the observable field. We can attempt to proceed further analytically. Of course if \mathcal{M} is invertible (e.g., A is the matter over density field) one could do a suitable transformation on the observables:

$$A_b \longrightarrow \sum_a (\mathcal{M}^{-1})_{ba} A_a \equiv \tilde{A}_b, \quad (3.8)$$

and get a nice closed expression. In the general case when \mathcal{M} is not invertible it is possible to further simplify the resulting expressions, to yield a (still) highly-dimensional integral but where the dimensionality is that of the observed field A . What remains to perform numerically is still computationally intensive especially for applications such as mega-pixel CMB maps. Still, compared to Eq. 3.7, the dimensionality has been dramatically reduced, for example, by a factor $\mathcal{O}(10^6)$, for a full sky CMB map from current experiments!

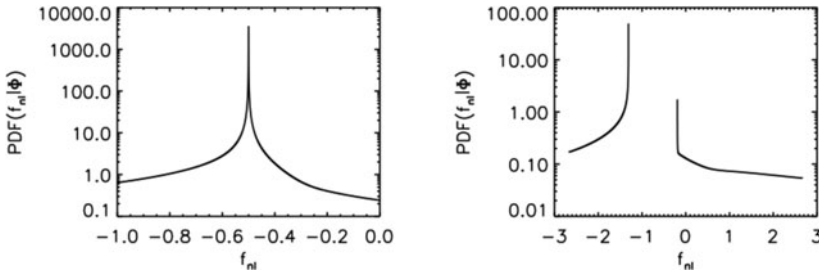


Figure 1. PDF for the non-Gaussianity parameter $f_{nl} = f_{NL}\sigma_\phi$ given a value of the field Φ : left panel $\Phi = 1$, right panel $\Phi = 1.5$. Note the discontinuity on the left panel and, on the right panel, the sharp divergencies at the edge of a region where the PDF is zero.

Before proceeding further it is important to understand the following warning and its consequences. So far we have been aiming at constructing $\mathcal{P}(\Phi|f_{NL})$. Assuming a uniform prior on f_{NL} , this can be interpreted as $\mathcal{P}(f_{NL}|\Phi)$. However $\mathcal{P}(f_{NL}|\Phi)$ has discontinuities which prevent any perturbative expansion. Only well away from these these discontinuities the moments are defined and the central limit theorem applies. In particular while in $\mathcal{P}(\Phi|f_{NL})$ f_{NL} is fixed and can be taken to be small, and maybe there one might find that perturbative expansions work well, in $\mathcal{P}(f_{NL}|\Phi)$ the values of f_{NL} are not bounded and can (and do) get very large. Fig. 1 shows an example of a 1-dimensional PDF for $f_{nl} = f_{NL}\sigma_\phi$, for different values of Φ .

Discontinuities are clearly visible. The support of the PDF depends on the value of the parameter f_{NL} : the PDF is zero for $\Phi < -f_{nl} - 1/(4f_{nl})$. This creates many problems. For example, one could try to find a maximum likelihood estimator (rather than working with the full PDF). But because of this the Cramer-Rao bound is invalid. This sharp discontinuity will also prevent any truncated expansion of the PDF being a good approximation.

The above discussion is valid for a single pixel, but it has consequences when many pixels (measurements) are combined. The full posterior when n independent pixels can be computed analytically and is shown in the left panel of Fig. 2: the PDF is zero for $f_{nl} > f_{nl}^{true}$. This sharp discontinuity will also prevent any truncated expansion of the PDF being a good approximation.

Under some conditions however, the chances of reaching the “excluded” regions can be made vanishingly small, especially for small f_{nl} , for example because of sampling or in the presence of noise as shown in the right panel of Fig. 2. The cutoff disappears also if we now consider that for a sample of finite size the tails of the distribution are not well sampled. For example for for $f_{nl} = 0.05$ the “excluded” region is not sampled if the sample is smaller than about 1.7×10^6 . This explains why for small values of f_{NL} and for realistic surveys the popular Non-Gaussianity estimator (which is a maximum likelihood estimator only if f_{NL} is zero) works. This also explains why, at the end of the day, perturbative expansions could work. Note however that a truncation to first order will yield to a PDF that does not have a maximum, which therefore cannot be a good approximation. At the minimum to have a maximum the approximated PDF must be truncated at second order or higher in f_{NL} .

Encouraged by this finding, we proceed in finding a useful approximation to the full PDF based on a second-order Edgeworth expansion (i.e., for small deviations from Gaussianity). The full expression can be found in Eqs. 4.28 and 5.2 of Verde *et al.* (2013), and we will not report it here in full. It will suffice to say that it involves the three- and four-point functions and that, when written explicitly for the CMB a_ℓ^m , one can recog-

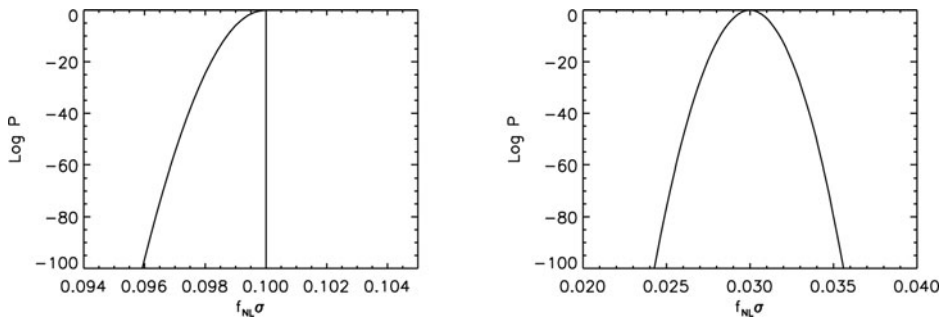


Figure 2. Left panel: $\text{Log}(\text{Posterior})$ for f_{nl} , with true $f_{nl}^{true} = 0.1$ and 10^6 pixels but ignoring sampling issues and noise. $f_{nl} > f_{nl}^{true}$ is excluded, as the true PDF is non-zero for some values of Φ where the trial PDF is zero. Right panel: $\text{Log}(\text{Posterior})$ for f_{nl} given 10^6 pixels, and $f_{nl}^{true} = 0.03$. A Gaussian noise of rms $0.01\sigma_\phi$ is added. The noise makes all values of Φ reachable in principle, and so the posterior is always non-zero. This removes the strict cutoff apparent in the left panel.

nise the optimal bispectrum estimator (see e.g., Mangilli and Matarrese’s contributions) and the trispectrum estimator of Regan *et al.* (2010). The expression is valid for non-Gaussianities more general than the local form as long as departures from Gaussianity is small. This opens up the possibility to extent the work beyond the CMB and possibly to galaxy surveys where, however, non-Gaussianity is not strictly small so the applicability of the expansion would need to be verified.

3.2. Example 2: Are two (or more) measurements in agreement?

Let us imagine we have performed two measurements (A and B) of cosmologically interesting quantities in the form of a two or higher dimensional posterior distribution. In the Bayesian framework, how would one quantify whether these two measurements are or not in agreement (tension)? (see Fig.3).

In other words, if the null hypothesis is that the two measurements are sampled from the base model adopted, when should the null hypothesis be abandoned? If the answer is that the two measurements are in tension, Bayesian model selection can be used to study extensions to the base model adopted and select which is the favoured model. Alternatively the detected tension might indicate the presence of unaccounted for, residual systematic errors (e.g., “unknown unknowns”). Possible options at this point are: discredit the measurement most likely affected by systematics or artificially increase its errors. If instead no tension is detected, the measurements can be combined to perform, for example, joint parameter estimation. This has been investigated for example in Verde *et al.* (2013b) which is the approach I will outline here but see also Marshall *et al.* (2016), March *et al.* (2011). For each measurement A, B, we produce a posterior $P_{A,B}(\theta|D_{A,B})$ where θ represents the parameters of the model and $D_{A,B}$ represents the data from experiments A, B respectively. Let us also assume that for producing both posteriors we have used the same, uniform priors, π , over the same support, x , i.e., $\pi_A = \pi_B = \pi$, $\pi = 1$ or 0 and therefore $\pi_A\pi_B = \pi$. Let H_1 be the (null) hypothesis that both experiments measure the same quantity, the models are correct and there are no unaccountable errors. In this case, the two experiments will produce two posteriors, which, although can have different (co)variances, and different distributions, have means that are in agreement. The alternative hypothesis, H_{-1} is when the two experiments, for some unknown reason, do not agree, either because of systematic errors or because they are effectively measuring different things or the model (parameterization) is incorrect. In

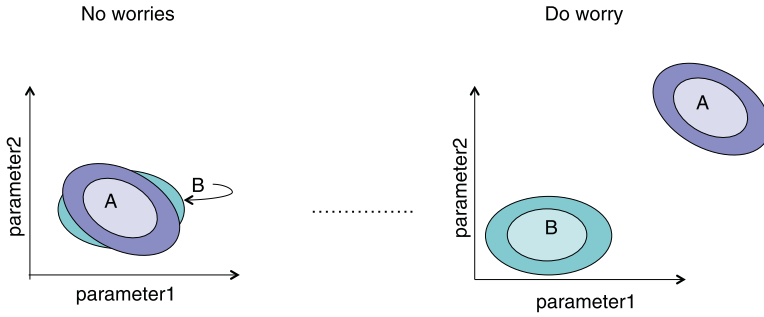


Figure 3. Schematic representation of the set up. Although the contours are reminiscent of Gaussians 1 and 2 σ confidence regions, the argument applies to any form for the distributions. The left hand side situation can be taken as the reference (“straw man”) null hypothesis, a “just so” scenario. On the right hand side a situation where clearly the two measurements are not in agreement, but there is a continuum between these two cases, and we would like to find a quantitative scale for it (and know at what point the two measurements are not consistent).

this case, the two experiments will produce two posteriors with two different means and different variances.

To distinguish the two hypothesis we use the Bayes factor, that is the ratio of the Bayesian Evidences (see D. Mortlock contribution and references there). In any practical application, the absolute normalization of the posteriors is often unknown, but we can still work as follows. We define:

$$\int P_A P_B dx = \lambda \int \mathcal{L}_A \mathcal{L}_B \pi_A \pi_B dx = \lambda \int \mathcal{L}_A \mathcal{L}_B \pi dx = \lambda E = \mathcal{E}, \tag{3.9}$$

where \mathcal{L} denotes the likelihood and $\lambda^{-1} = \int \mathcal{L}_A \pi_A dx \int \mathcal{L}_B \pi_B dx'$. E is the Bayesian Evidence for the joint distribution, thus \mathcal{E} is akin to an unnormalized Evidence.

However we can consider a “straw man” null hypothesis where the maxima of the two distributions coincide. For example, imagine that we perform a *translation* (shift) of (one or both of) the distributions in x and let us define \bar{P}_A the shifted distribution. Eq. (3.9) becomes

$$\int \bar{P}_A \bar{P}_B dx = \bar{\mathcal{E}}|_{\max A = \max B}. \tag{3.10}$$

This translation changes the location of the maximum but does not change the shape or the width of the distribution. Clearly the Evidence ratio for the (null) hypothesis E_1 is $\mathcal{E}/\bar{\mathcal{E}}|_{\max A = \max B}$, as the normalization factors λ cancel out, and the Evidence ratio for the alternative H_{-1} is its reciprocal. We therefore introduce:

$$\mathcal{T} = \frac{\bar{\mathcal{E}}|_{\max A = \max B}}{\mathcal{E}}, \tag{3.11}$$

which denotes the degree of tension that can be interpreted in the widely used Jeffrey’s (Jeffreys, 1973; Kass and Raftery, 1995) scale. \mathcal{T} indicates the odds: 1 : \mathcal{T} are the chances for the null hypothesis. In other words, a large tension mens that the null hypothesis ($\max A = \max B$) is unlikely. \mathcal{E} can be seen as the evidence for the joint distribution where we interpret one data set as the prior. Then this is normalised by a “just so” scenario: $\bar{\mathcal{E}}|_{\max A = \max B}$.

Of course, and as already mentioned, the obvious application of this is in searching for either new physics or systematics of the “unknown unknowns” type. But, is there a way to distinguish new physics from systematics? As far as I know there is no systematic treatment for that but some insights can be garnered by considering more than two

measurements. Let us take for example the recent collection of claims of non-zero neutrino mass that have appeared in the literature and follow the approach of Leistedt *et al.* (2014). The authors argue that the need for extra parameters describing new physics beyond the base cosmological model, yielding a new cosmological concordance can only be convincing if the combined datasets are in tension in the minimal model, and in agreement in extended model. If the tension remains in the extended model it is due to systematic effects (or to new physics different from the one considered). In my view more work remains to be done to put all this into a consistent framework.

3.3. Example 3: Bias (clustering of peaks of a Gaussian field)

Most theoretical models predict the statistical properties of the distribution of dark matter. Unfortunately most of the observations rely on objects that “light up” (such as galaxies) which may not be faithful tracers of the dark matter distribution. This is known as (galaxy) bias: the clustering properties of the observed (tracers) field are not the same as those of the dark matter. For most applications the bias is assumed to be scale-independent or to be very slowly varying with scale. A strongly scale-dependent bias would be very problematic for the interpretation of large-scale structure data. But galaxies are expected inhabit dark matter halos which are believed to correspond to high peaks of the *initial* density field. The initial distribution is expected to be very close to Gaussian, thus modelling the clustering properties of peaks of a Gaussian distribution is of great interest. Attempts to address this go back to Otto Politzer and Wise (1986). We can start by considering the expression for the (joint) N points probability distribution of peaks above a threshold t of a Gaussian field $\phi(\mathbf{r})$ which can be written using the Gaussian path integral:

$$P(\mathbf{r}_1, \dots, \mathbf{r}_N) = \int [d\phi(\mathbf{r})] P[\phi(\mathbf{r})] \times \prod_{j=1}^N \left[\int dw_{(j)} |\det w_{(j)}| \delta^3(\nabla\phi(\mathbf{r}_j)) \delta^6(\nabla\nabla\phi(r_j) - w_{(j)}) \theta(\phi(\mathbf{r}_j) - t) \right]. \quad (3.12)$$

In this equation $P[\phi]$ is the Gaussian probability distribution function, θ denotes the Heaviside step function and $w_{(j)}$ in three spatial dimensions is the symmetric 3×3 matrix of the second derivatives of ϕ at position \mathbf{r}_j and δ^3 , δ^6 denote Dirac delta functions. In Eq. (3.13) the integration on $dw_{(j)}$ has to be extended only over negative definite eigenvalues, in order to identify local maxima. Then the correlation function ξ_N is given by $1 + \xi_n = P(\mathbf{r}_1.. \mathbf{r}_N)/P^N(\mathbf{r})$ where $P(\mathbf{r})$ is known (e.g., Bardeen *et al.*, 1986). Eq. 3.13 might look innocuous but it is extremely complicated, no exact analytic solution has been found beyond one spatial dimension, despite extensive studies in the literature over a number of years, starting with the pioneering works of Rice(1944), Rice(1945), Adler(1981), Bardeen *et al.* (1986), Peacock and Heavens (1985), Kaiser(1984), Jensen & Szalay(1986). Here I follow the approach of Verde *et al.* (2014) and report their findings. They argue (and show on simulations) that virtually all extrema are peaks above a not too high threshold, therefore one could simply compute the properties of extrema rather than peaks without introducing significant systematics. This gets rid of the (immensely complicated to deal with) six dimensional Dirac delta function.

For a high threshold the value of the integral above the threshold can be approximated by the value of the function at the threshold. Thus the integral involving the threshold can be dropped (or performed eventually numerically if needed).

At this point the residual hurdle is the presence of the $|\det w|$ in the integrand. However this could be regarded as a weight and so one could simplify it (or drop it altogether) if

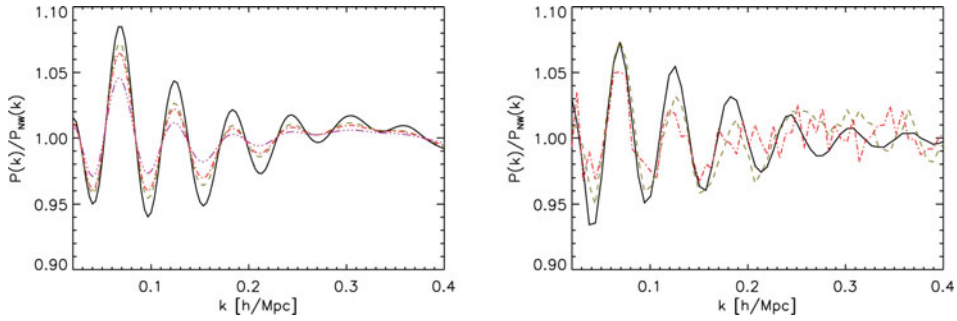


Figure 4. The BAO feature as $P(k)/P_{NW}$. The left panel shows the theory prediction obtained by Fourier transforming eq.3.13, the right panel shows the mean of 20 Gaussian realisations. Solid/Black for matter, green/dashed for 2σ peaks, red/dot-dashed for 3σ peaks; purple/dashed-dot-dot-dot for 4σ peaks.

the data could be suitably weighted to compensate. To begin with, $|\det w|$ was simply substituted by unity, leaving to forthcoming work to find a weighting scheme for the observations that can compensate for this.

With these simplifications it is possible to find an analytic expression at least for the two-point function:

$$P'(\mathbf{r}_1, \mathbf{r}_2, m_1, m_2) = \frac{2\pi}{\det X'} \frac{(2\pi)^{1/2}}{(\det H')^{1/2}} \times \frac{2\pi}{\det(\xi - Q^T H'^{-1} Q)} \exp\left[-\frac{1}{2} m^T (\xi - Q^T H'^{-1} Q)^{-1} m\right], \quad (3.13)$$

where m_1 and m_2 denote the thresholds, and Q , H' and X' are matrices that can be computed from the two point function of the dark matter, see Verde *et al.* (2014) for the explicit expressions. This might look ugly, but it is not. An important implication of this result is that, since the matter power spectrum is not a featureless power law the tracers' bias as function of scale $b(r)$, defined from the ratio of the correlation functions of tracers to that of matter, is not scale independent, even at linear scales. In particular, a new scale-dependent feature is found in the bias (a ‘‘bump’’) which is located very near the Baryon Acoustic Oscillation feature. The BAO feature corresponds to the (local) minimum of $b(r)$ at around $r = 110 \text{ Mpc}/h$. The presence of the BAO signal in the matter power spectrum introduces a changing first and second derivative of the correlation function which are responsible for non-negligible effects. In other words, selecting peaks of a Gaussian field is a highly non-linear operation which creates a highly non-Gaussian peaks field. The essence of non-Gaussianity is mode coupling which, by moving power across scales, tends to move around, distort or even erase localized features. In the BAO case the localised scale-dependence of the bias does not move the BAO feature but reduces its amplitude. The smoothing is more marked for higher thresholds as seen in Fig. 4.

The figure shows the ratio of the power spectrum divided by a power spectrum without the BAO feature as a function of the threshold. Note that the BAO feature is smoothed and the smoothing increases with the threshold height. While it is reassuring that this effect does not change the location of the BAO feature, it might have nevertheless important practical implications. The signal-to-noise for measurements that depend on the BAO location is usually computed adopting a model with a fixed BAO smoothing parameter interpreted as the one generated by non-linearities. Should this be underestimated, as the effective one is a combination of the effect of non-linearities and this new effect, then the signal-to-noise would be overestimated.

There are consequences also for survey selection considerations: highly biased tracers are preferentially targeted to beat shot noise when designing a BAO survey –in technical terms, to maximise nP where n denotes the tracer number density–. In fact given a finite amount of observing time and a finite aperture telescope the selection of bright, highly biased objects yields the best signal-to-noise. However highly biased tracers will have a reduced BAO feature: it may be advantageous to select less rare and less biased tracers if they carry a more pronounced signal.

3.4. Example 4: Cancelling out systematics

Suppose we have N observables (data points) $O_i, i = 1, \dots, N$ that depend on m interesting quantities μ_i and n nuisance quantities ν_i which can possibly introduce systematic errors on μ_i if fixed at incorrect values. The usual approach is to treat systematics as nuisance parameters and marginalise. However this might not be satisfactory especially if we are ignorant about the mean values and the errors of the nuisance parameters. There are some example in the literature on how to deal with this situation but not a lot of work on a systematic approach. Here I outline the approach proposed by Noreña *et al.* (2012).

Our goal should be to find combinations of the observables O_i that are insensitive to the nuisance parameters ν_i . In other words we must find a set of $M = N - n$ functions

$$f_k = f_k(O_1, \dots, O_N) \quad (3.14)$$

where $k = 1, \dots, M$, so that

$$\frac{df_k}{d\nu_i} = \sum_{j=1}^N \frac{\partial f_k}{\partial O_j} \frac{\partial O_j}{\partial \nu_i} = 0 \text{ for } i = 1, \dots, n. \quad (3.15)$$

A natural interpretation of this is the *renormalisation group* equation (Wilson & Kogut, 1974). Of course one must know how O_i depend on ν_i so this can apply only to “known unknowns”. Let us consider for example a power-law dependence on nuisance parameters ν_i :

$$O_i - \widehat{O}_i = g(\vec{\mu}) \prod_{j=1}^n (\nu_j - \widehat{\nu}_j)^{\alpha_{ij}} \quad (3.16)$$

where the $\widehat{}$ symbol denotes the true value and g is some function of $\vec{\mu}$, a vector containing all other quantities on which the observables depend. Then the solution for the system of differential equations, Eq. 3.15, is of the form:

$$f_k = \prod_{i=1}^N (O_i - \widehat{O}_i)^{b_i^k} \quad (3.17)$$

which gives a system of linear algebraic equations for the unknown b_i^k . If $M = N - n > 0$, i.e., there are more data than nuisance parameters, then there are m non-trivial solutions:

$$\sum_{i=1}^M \alpha_{ij} b_i^k = 0. \quad (3.18)$$

There are therefore $N - n$ combinations of data that cancel out the effect of the systematics (nuisance parameters). A similar solution can be found in other cases, e.g. linear dependence on the ν_i etc.

Obviously it follows that it is not possible to do better than marginalisation, but with marginalisation the true value and the dispersion of the nuisance parameters must be known or assumed. Also one needs more data than nuisance parameters otherwise exact

solutions do not exist. However, even for more nuisance parameters than data it is still possible in some cases to find *approximate* solutions. A particular case of interest is when observables have similar, but not identical, dependences on some of the nuisance parameters. In this case it is possible to minimise (rather than cancel completely) the effects of systematics. It turns out that one must minimise this:

$$\mathcal{L}_k = \sum_{j=1}^n \left(\frac{df_k}{d\nu_j} \right)^2 \Delta\nu_j^2 - \lambda_k \left(\sum_i (b_i^k)^2 - A_k^2 \right) \tag{3.19}$$

where λ_k is a Lagrange multiplier to be solved for, A_k is the norm of the vector b_i^k , and $\Delta\nu_j$ is the uncertainty on the j nuisance parameter. For example for the power law case discussed above we obtain the eigenvalue equations:

$$\sum_{l=1}^N \mathcal{M}_{il} = \sum_{l=1}^N \left[f_k^2 \sum_{j=1}^n \left(\frac{\Delta\nu_j}{\nu_j} \right) \alpha_{ij} \alpha_{lj} \right] b_l^k = \lambda_k b_i^k. \tag{3.20}$$

This formulation is general and include the case discussed above. Note that the eigenvalues of the matrix \mathcal{M} measure how much the solution is affected by the nuisance parameters. The eigenvectors correspond to independent combinations of observables, and, if we are interested in minimizing the impact of nuisance parameters, we should choose those eigenvectors corresponding to eigenvalues which are small with respect to typical entries of the matrix. If an eigenvalue is zero, there is an independent non-trivial solution which is unaffected by changes of the nuisance parameters, as in the previous set-up.

For an example of a possible application, let us consider the case of Baryon Acoustic Oscillations which rely on the CMB to determine the “standard ruler”: the sound horizon at radiation drag r_s . Since the determination of r_s is model-dependent, using an incorrect model (e.g., adiabatic model in the presence of isocurvature or non-standard neutrino properties in the presence of new physics in the neutrino sector) can introduce unwanted systematic errors in the interpretation of the results. For N redshift bins and therefore $2N$ measurements (radial and tangential BAO scale) there are $2N - 1$ independent combinations (i.e., $N - 1$ tangential, $N - 1$ radial and one tangential vs radial relative BAO measurements) which cancel out completely the dependence on r_s . For more than ~ 10 redshift bins this does not increase significantly the statistical errors on the recovered cosmological parameters, but makes the measurement much more robust.

4. Summary and conclusions

I hope I have motivated why one cannot study Cosmology without being fluent in statistical techniques. Moreover, the forthcoming challenges that Cosmology faces cannot be addressed without the development of suitable statistical techniques. I have concentrated on dealing with systematic errors (rather than statistical errors), because systematic errors will likely be the major limitation in the near future and there is no systematic framework to deal with them. I have presented four specific examples, which are not representative of exhaustive, but I hope they serve to illustrate the point. The route from precision cosmology to accurate cosmology goes through statistical cosmology.

Acknowledgments

I am supported by the European Research Council under the European Communitys Seventh Framework Programme grant FP7- IDEAS-Phys.LSS and I acknowledge Mineco grant FPA2011-29678- C02-02. I would like to thank all my collaborators in the works mentioned here: Luis Alvarez-Gaumé, Cesar Gomez, Alan Heavens, Raul Jimenez, Boris Leistedt, Sabino Matarrese, Jorge Noreña, Hiranya Peiris, Carlos Peña-Garay, Pavlos Protopapas, Fergus Simpson.

References

- Planck Collaboration, *et al.*, 2013, arXiv, arXiv:1303.5076
 Planck Collaboration, *et al.*, 2013, arXiv, arXiv:1303.5062
 The CORe Collaboration, Armitage-Caplan, C., Avillez, M., *et al.* 2011, arXiv:1102.2181
 André, P., Baccigalupi, C., Banday, A., *et al.* 2014, JCAP, 2, 6
 Box, George, E. P., Norman, R. Draper (1987). *Empirical Model-Building and Response Surfaces*, p. 424, Wiley.
 Peebles, P. J. E. 2002, arXiv:astro-ph/0208037
 Elsner, F. & Wandelt, B. D. , *Astrophys. J.* **724**, 1262 (2010) [arXiv:1010.1254 [astro-ph.CO]].
 Elsner, F., Wandelt, B. D. & Schneider, M. D. , *Astron. Astrophys.* **513**, A59 (2010) [arXiv:1002.1713 [astro-ph.CO]].
 Ensslin, T. A. , Frommert, M., & Kitaura, F. S., *Phys. Rev. D* **80** (2009) 105005 [arXiv:0806.3474 [astro-ph]].
 Verde, L., Jimenez, R., Alvarez-Gaume, L., Heavens, A. F., & Matarrese, S. 2013a, *JCAP*, 6, 23
 Regan, D. M., Shellard, E. P. S., & Fergusson, J. R. 2010, *PRD*, 82, 023520
 Marshall, P., Rajguru, N., & Slosar, A., *Phys. Rev. D* **73**, 067302 (2006)
 March, M. C. , Trotta, R., Amendola, L., & Huterer, D., *Mon. Not. Roy. Astron. Soc.* **415**, 143 (2011)
 Verde, L., Protopapas, P., & Jimenez, R. 2013b, *Physics of the Dark Universe*, 2, 166
 Jeffreys, H., 1973. *Scientific Inference*. Cambridge University Press.
 Kass, R. E. & Raftery, A. E., 1995, *Bayes factors*. *JASA* 90, 430, 773-795.
 Leistedt, B., Peiris, H. V., & Verde, L., 2014, arXiv:1404.5950
 Otto, S., Politzer, H. D., & Wise, M. B., 1986, *Physical Review Letters*, 56, 1878
 Rice, S. O., 1944, *Bell Systems Tech. J.*, Volume 23, p. 282-332, 23, 282
 Rice, S. O., 1945, *Bell Systems Tech. J.*, Volume 24, p. 46-156, 24, 46
 Adler, R. J., 1981, *The Geometry of Random Fields*. Chichester: Wiley
 Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S., 1986, *ApJ*, 304, 15
 Jensen, L. G. & Szalay, A. S., 1986, *ApJLett*, 305, L5
 Kaiser, N., 1984, *ApJL*, 284, L9
 Peacock, J. A. & Heavens, A. F., 1985, *MNRAS*, 217, 805
 Noreña, J., Verde, L., Jimenez, R., Peña-Garay, C., & Gomez, C. 2012, *MNRAS*, 419, 1040
 Wilson, K. G. & Kogut, J., 1974, *PhR*, 12, 75