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/S1743921314011016

The potential of likelihood-free inference of cosmological parameters with weak lensing data

Michael Vespe

Department of Statistics, Carnegie Mellon University,
Baker Hall 132, Pittsburgh, PA 15232
email: mvespe@andrew.cmu.edu

Abstract. In the statistical framework of likelihood-free inference, the posterior distribution of model parameters is explored via simulation rather than direct evaluation of the likelihood function, permitting inference in situations where this function is analytically intractable. We consider the problem of estimating cosmological parameters using measurements of the weak gravitational lensing of galaxies; specifically, we propose the use a likelihood-free approach to investigate the posterior distribution of some parameters in the ΛCDM model upon observing a large number of sheared galaxies. The choice of summary statistic used when comparing observed data and simulated data in the likelihood-free inference framework is critical, so we work toward a principled method of choosing the summary statistic, aiming for dimension reduction while seeking a statistic that is as close as possible to being sufficient for the parameters of interest.

1. Introduction

Weak gravitational lensing, also known as cosmic shear, is the distortionary effect on images of distant galaxies by matter in between the galaxy and the observer. The ensemble behavior of this distortionary effect, which would render a circular object slightly elliptical, can yield insight into the distribution of dark matter and permit constraint of the parameters in a cosmological model. However, galaxies are not intrinsically circular; in fact, the signal from cosmic shear is very faint compared to the intrinsic variability in the ellipticity of galaxies. Thus, a large number of galaxies must be observed in order to isolate the shear signal.

Once the galaxies are observed and catalogued, weak lensing analyses traditionally proceed by summarizing these galaxies via some summary statistic, often referred to as an *observable* or a *data vector*. Common examples of observables include, among others, estimates of the two-point correlation functions $\xi_\pm$ or power spectrum modes C_ℓ (as in, e.g., Lin *et al.* 2012), among others. Then, the summary statistic is assumed to have a multivariate Gaussian distribution, so that the likelihood of a set of values $\tilde{\theta}$ for cosmological parameters is given by

$$\mathcal{L}(\tilde{\theta};\hat{d}) = \frac{1}{\sqrt{(2\pi)^p|\mathbf{C}|}} \exp\left((\hat{d}-d(\tilde{\theta}))^T \mathbf{C}^{-1}(\hat{d}-d(\tilde{\theta}))\right) \tag{1.1}$$

where $\hat{d}$ is the observable estimated from data, $d(\tilde{\theta})$ is the theoretical value of the observable given parameters $\tilde{\theta}$, and $\mathbf{C}$ is the covariance matrix of the observable, often estimated via some simulation approach. This likelihood can be evaluated either as part of a frequentist maximum likelihood analysis or, in a Bayesian framework, as one step toward deriving a posterior distribution for cosmological parameter values.

The quality of any inferences resulting from this procedure hinges on several factors: the ability of the chosen summary statistic to capture the information in the raw data relevant to the parameters; the accuracy of the estimated covariance matrix; and the validity of the assumption that the observable has a multivariate Gaussian distribution. This last assumption is not equivalent to assuming that the *parameters* have a Gaussian distribution, but it does impose some indirect constraints on their distribution.

These concerns motivate our desire to explore likelihood-free inference methods, which were introduced in contexts where methods exist for simulating data given parameter values, but evaluation of a likelihood is analytically or computationally intractable.

2. Methodology

Approximate Bayesian computation. We focus on approximate Bayesian computation (ABC), a particular likelihood-free inference method, introduced by Pritchard *et al.* (1999) in the biology literature. For full details, as well as generalizations and improvements, we refer to that paper as well as Blum *et al.* (2012) and Beaumont *et al.* (2009). Thus far, ABC has found limited use in astronomy settings, as in Cameron and Pettitt (2012) and Weyant *et al.* (2013).

The algorithm aims to generate a sample from a desired posterior distribution $\pi(\theta|\mathbf{x})$ given observed data $\mathbf{x}$; in the weak lensing setting, we aim to sample from the posterior distribution of the cosmological parameters of interest given a catalogue of galaxies with position and shear.

In its simplest form, it proceeds via the repeated execution of three steps. Step one is to generate candidate parameter values $\tilde{\theta}$ from a prior distribution. Step two is to simulate a realization $\tilde{\mathbf{x}}$ of the data set using $\tilde{\theta}$ as input parameters. Step three is to compare the simulated data $\tilde{\mathbf{x}}$ to $\mathbf{x}$ and retain $\tilde{\theta}$ if and only if $\tilde{\mathbf{x}}$ and $\mathbf{x}$ match. Formally, we retain $\tilde{\theta}$ if and only if $\rho(S(\tilde{\mathbf{x}}), S(\mathbf{x})) \leqslant \epsilon$, for some distance metric ρ, summary statistic S, and tolerance threshold ϵ.

The resulting retained parameter values constitute a sample from an approximation $\pi_\epsilon(\theta|\mathbf{x})$ to the posterior $\pi(\theta|\mathbf{x})$. It can be shown that if S is a sufficient statistic for θ, then $\pi_\epsilon(\theta|\mathbf{x}) \to \pi(\theta|\mathbf{x})$ as $\epsilon \to 0$. Any ABC analysis will be sensitive to the choice of ϵ, ρ, and S. As in the traditional analysis framework, it is desirable to choose the summary statistic S so that it captures the information from $\mathbf{x}$ relevant for inference on θ while discarding any useless information. Due to the manner in which ABC algorithms rely on accepting/rejecting simulated parameter values, there is a particular need to reduce the dimensionality of the summary statistic to the greatest extent possible, lest the simulations be burdened by the inefficiency of repeatedly rejecting parameter candidates.

Exponential family approximation. Our proposed approach to approximating sufficient statistics of low-dimension is built upon the following idea. Suppose summary statistics $\mathbf{s}$ are available that are sufficient for θ but may be high-dimensional and thus contain some redundant information. For j = 1, 2, ..., J, we seek a mapping of θ, denoted $\eta_j(\theta)$, along with a mapping of $\mathbf{s}$, denoted $T_j(\mathbf{s})$, such that $\mathbb{E}(\eta_j(\theta)|\mathbf{s}) = T_j(\mathbf{s})$. These can be thought of as compressed versions of the summary statistics $\mathbb{E}(\theta|\mathbf{s})$ that have been shown (Fearnhead and Prangle 2012) to be optimal under a reasonable choice for the loss function.

We refer to this approach as an *exponential family approximation.* The standard exponential family form for the distribution of $\mathbf{s}$ given parameter θ is

$$f(\mathbf{s}|\theta) = h(\mathbf{s}) \exp\left(\sum_{j=1}^{J} \eta_j(\theta) T_j(\mathbf{s}) - A(\theta)\right)$$

https://doi.org/10.1017/S1743921314011016 Published online by Cambridge University Press

If θ is modeled as a random variable drawn from prior distribution π, the joint distribution of $(\mathbf{s}, \theta)$ is given by

$$f(\mathbf{s}, \theta) = h(\mathbf{s}) \exp\left(\sum_{j=1}^{J} \eta_j(\theta) T_j(\mathbf{s}) - A(\theta)\right) \pi(\theta)$$

or, equivalently,

$$f(\mathbf{s}, \theta) = \exp\left(\sum_{j=1}^{J} \eta_j(\theta) T_j(\mathbf{s}) - A^*(\theta) - h^*(\mathbf{s})\right)$$

with $h^*(\mathbf{s}) = -ln(h(\mathbf{s}))$ and $A^*(\theta) = A(\theta) - ln(\pi(\theta))$. It is assumed that one can simulate pairs $(\mathbf{s}_i, \theta_i)$ for i = 1, 2, ..., n. This will be accomplished by drawing θ from a prior distribution π, simulating $\mathbf{x}$ conditional on θ, and then computing the available $\mathbf{s}(\mathbf{x})$. Then, the joint log-likelihood for this synthetic "data set" is given by

$$f(\mathbf{s}, \theta) = \sum_{i=1}^{n} \left[\sum_{j=1}^{J} \eta_j(\theta) T_j(\mathbf{s}) - A^*(\theta) - h^*(\mathbf{s})\right] \tag{2.1}$$

Heuristically, we will seek to maximize this joint log-likelihood over the space of mappings η_j and T_j. The resulting mapping $T_j(\mathbf{s})$ for j = 1, ..., J would be an approximately sufficient statistic of dimension J, where J is chosen intentionally to be closer to the dimension of the intrinsic parameter space Θ than to that of $\mathcal{S}$, the domain of $\mathbf{s}$.

3. Example application

We present an application of this method in a simple, stylized cosmic shear analysis, simulating data from known inputs. Specifically, we generate a random realization of a shear field using input cosmological parameters $\Omega_M = 0.25, \sigma_8 = 0.8$, and all other inputs (including survey redshift distribution) chosen to replicate those in the simulation exercises of Lin *et al.* (2012). In this case, we model cosmic shear as a Gaussian random field (GRF) on a grid of pixels, although future analyses would incorporate more realistic models (see Kiessling *et al.* 2011.) We add i.i.d. shape noise ($\sigma_{\text{int}} = 0.37$) to represent the effect of intrinsic galaxy ellipiticity.

For our ABC analysis, we sample candidate parameter values $(\tilde{\Omega}_M, \tilde{\sigma}_8)$ from a uniform prior distribution on the rectangle [0.1, 8] × [0.5, 1]. In Fig. 1 we compare samples from the approximate posterior distributions using two summary statistics: at left, estimates $\hat{\xi}_\pm(\theta)$ of the two-point correlation functions – evaluated in eight logarithmically spaced bins – and, at right, the first two coordinates of $\hat{T}$ learned via the exponential family method. In each case, we take ρ to be standard Euclidean distance and we choose ϵ so that 5% (blue) and 10% (green) of the candidate samples are retained for each case.

As Ω_M and σ_8 are known to be degenerate (these data can only distinguish the value of $\Omega_M^{0.7}\sigma_8$), we display the degeneracy curve corresponding to the input parameters in orange. Simple inspection suggests that using $\hat{T}$ as learned via the exponential family approximation as the summary statistic is preferable to simply using $\hat{\xi}_\pm(\theta)$, because the samples from the former assemble more tightly around the true degeneracy curve than those from the latter.

4. Discussion and future directions

The exponential family approximation-derived statistic seems to improve upon the canonical $\hat{\xi}_\pm(\theta)$ statistic, even in the simple case where shear is assumed to follow a GRF

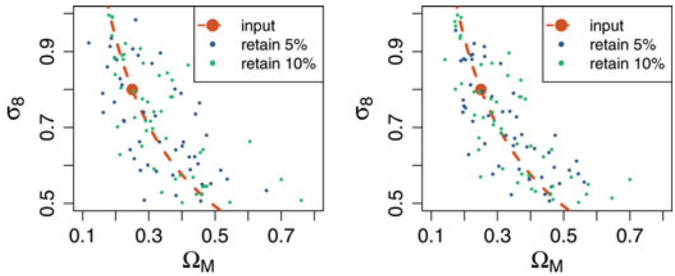

Figure 1. ABC-derived posterior samples using $\hat{\xi}_\pm(\theta)$ (left) and $\hat{T}_1, \hat{T}_2$ (right).

model. This is noteworthy because the true correlation function $\xi_\pm(\theta)$ is known to be a sufficient statistic for the GRF, so the improvement likely stems from the act of reducing the dimension of the data vector to mitigate the effect of noisy discretized estimates. Put another way, when the observable consists of noisy, somewhat redundant estimators, it is preferable to have fewer of those provided that no information is being discarded.

Future work will proceed in both theoretical and applied directions. In the former direction, we aim to better understand the theoretical properties of the summary statistic resulting from the exponential family approximation method. In addition, we hope to assess the method's feasibility under particular circumstances.

Regarding applications, we intend to apply the method in more sophisticated, realistic settings. One such setting would be that of tomographic weak lensing analysis, as in Heymans *et al.* (2013), wherein the data vector consists of auto- and cross-correlation functions $\hat{\xi}_\pm^{i,j}$ within or between various bins in redshift. In this setting, where the dimension of the data vector is inherently much larger, the motivation for principled dimension reduction is apparent. We also hope to incorporate more complex simulation models, such as the SUNGLASS pipeline of Kiessling *et al.* (2011) to better account for non-Gaussianity of the true shear field.

5. Acknowledgements

This work is done is collaboration with Peter Freeman, Rachel Mandelbaum, and Chad Schafer. Support comes from DOE Grant #ER26170.

References

Beaumont, M. A., Cornuet, J., Marin, J., & Robert, C. 2009, *Biometrika*, 96, 983
Blum, M. G. B., Nunes, M. A., Prangle, D., & Sisson, S. A. 2012, `arXiv:1202.3819`
Cameron, E. & Pettitt, A. N. 2012, *MNRAS*, 425, 44
Fearnhead, P. & Prangle, D. 2012, *J. Roy. Stat. Soc. B Met.*, 74, 419
Heymans, C., Grocutt, E., Heavens, A., Kilbinger, M., *et al.* 2013, *MNRAS*, 432, 2433
Kiessling, A., Heavens, A. F., & Taylor, A. N. 2011, *MNRAS*, 414, 2235
Lin, H., Dodelson, S., Seo, H., Soares-Santos, M., Annis, J., *et al.* 2012, *ApJ*, 761, 15
Pritchard, J. K., Seielstad, M. T., *et al.* 2012, *Mol. Biol. Evol.*, 16, 1791
Weyant, A., Schafer, C., & Wood-Vasey, W. M. 2013, *ApJ*, 764, 116