Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **485,** 2956–2969 (2019) doi:10.1093/mnras/stz558
Advance Access publication 2019 February 26

Likelihood non-Gaussianity in large-scale structure analyses

ChangHoon Hahn [1,2]★ Florian Beutler [1,3] Manodeep Sinha [4,5,6] Andreas Berlind,[6] Shirley Ho[1,2,7,8] and David W. Hogg[9,10]

[1]*Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, CA 94720, USA*
[2]*Berkeley Center for Cosmological Physics, University of California, Berkeley, CA 94720, USA*
[3]*Institute of Cosmology & Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth PO1 3FX, UK*
[4]*Centre for Astrophysics & Supercomputing, Swinburne University of Technology, 1 Alfred St., Hawthorn, VIC 3122, Australia*
[5]*ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D)*
[6]*Department of Physics and Astronomy, Vanderbilt University, Nashville, TN 37235, USA*
[7]*Department of Physics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA*
[8]*Department of Physics, University of California, Berkeley, CA 94720, USA*
[9]*Center for Cosmology and Particle Physics, New York University, New York, NY 10003, USA*
[10]*Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA*

Accepted 2019 February 22. Received 2019 February 20; in original form 2018 March 16

ABSTRACT

Standard present-day large-scale structure (LSS) analyses make a major assumption in their Bayesian parameter inference – that the likelihood has a Gaussian form. For summary statistics currently used in LSS, this assumption, even if the underlying density field is Gaussian, cannot be correct in detail. We investigate the impact of this assumption on two recent LSS analyses: the Beutler et al. power spectrum multipole (P_ℓ) analysis and the Sinha et al. group multiplicity function (ζ) analysis. Using non-parametric divergence estimators on mock catalogues originally constructed for covariance matrix estimation, we identify significant non-Gaussianity in both the P_ℓ and ζ likelihoods. We then use Gaussian mixture density estimation and independent component analysis on the same mocks to construct likelihood estimates that approximate the true likelihood better than the Gaussian *pseudo*-likelihood. Using these likelihood estimates, we accurately estimate the true posterior probability distribution of the Beutler et al. and Sinha et al. parameters. Likelihood non-Gaussianity shifts the $f\sigma_8$ constraint by -0.44σ, but otherwise does not significantly impact the overall parameter constraints of Beutler et al. For the ζ analysis, using the pseudo-likelihood significantly underestimates the uncertainties and biases the constraints of the Sinha et al. halo occupation parameters. For $\log M_1$ and α, the posteriors are shifted by $+0.43\sigma$ and -0.51σ and broadened by 42 per cent and 66 per cent, respectively. The divergence and likelihood estimation methods we present provide a straightforward framework for quantifying the impact of likelihood non-Gaussianity and deriving more accurate parameter constraints.

Key words: methods: data analysis – methods: statistical – galaxies: statistics – cosmology: observations – cosmological parameters – large-scale structure of Universe.

1 INTRODUCTION

Bayesian parameter inference provides the standard framework for deriving cosmological parameters from observation of large-scale structure (LSS) studies. Using Bayes's rule,

$$p(\theta \mid x) \propto p(x \mid \theta)\, p(\theta), \tag{1}$$

the posterior probability distributions of cosmological parameters can be derived from observed measurements such as the galaxy power spectrum. All that is required is the prior distribution of the parameters, $p(\theta)$, and the likelihood, $p(x \mid \theta)$ – probability of the data (observation) given the theoretical model. Priors are selected in analyses; so parameter inference ultimately reduces to evaluating the likelihood. Analyses can *only* yield unbiased constraints if the likelihood evaluation is correct.

In present-day LSS analyses, two major assumptions go into evaluating the likelihood. First, the likelihood is assumed to have a

★ E-mail: chh327@nyu.edu

Gaussian functional form:

$$p(\boldsymbol{x} \mid \theta) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\mathbf{C})}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - m(\theta))^t \mathbf{C}^{-1}(\boldsymbol{x} - m(\theta))\right], \quad (2)$$

where d is the dimension of the data vector $\boldsymbol{x}$, $m(\theta)$ is the theoretical predictions given the model parameters θ, and $\mathbf{C}$ is the covariance matrix. Second, the covariance matrix used in evaluating the Gaussian pseudo-likelihood is assumed to be independent of cosmology or the model parameters. The covariance matrix is evaluated only at a selected fiducial cosmology with fiducial model parameters and is assumed to be fixed throughout the analysis. In principle, the covariance matrix depends on θ, and the dependence has been shown to have a significant effect on parameter constraints (e.g. Eifler, Schneider & Hartlap 2009; Morrison & Schneider 2013; White & Padmanabhan 2015). In this paper, we focus on the first, the Gaussian pseudo-likelihood assumption. Even when analyses use covariance matrices that account for non-Gaussian covariance (e.g. Scoccimarro, Couchman & Frieman 1999; Hu & White 2001; O'Connell et al. 2016), the likelihood is still assumed to have a Gaussian functional form (equation 2). They therefore still employ a Gaussian *pseudo*-likelihood. We will test the assumption and quantify the impact of this Gaussian *pseudo*-likelihood assumption on cosmological parameter constraints.

The motivation for the Gaussian pseudo-likelihood ultimately stems from the 'central limit theorem'. Take the power spectrum of the density field, for example. On large scales, the density field is approximately a Gaussian random field and the power spectrum of a specific Fourier mode would follow a chi-squared distribution, *not* a Gaussian. However, with sufficiently many independent modes contributing, the likelihood of the power spectrum would *approach* a Gaussian distribution by the central limit theorem. In practice, we expect the Gaussian assumption to fail in low-signal-to-noise regimes. The assumption is also further invalidated by correlations among different modes caused by finite survey volume, shot noise, and systematic effects. The breakdown of Gaussianity is clearly illustrated in earlier surveys such as *IRAS*, where limited survey volume and sparse sampling cause the probability distribution function of the galaxy power spectrum to deviate significantly from Gaussian (see fig. 9 in Scoccimarro 2000). Hartlap et al. (2009) and Sellentin & Heavens (2018) similarly illustrate the breakdown of the Gaussian likelihood assumption for the cosmic shear correlation function likelihood.

Even if the likelihood is Gaussian, Sellentin & Heavens (2016) argue that since an estimate of the covariance matrix is used for the likelihood, for accurate parameter inference the true covariance matrix must be marginalized over. This marginalization leads to a likelihood that is no longer Gaussian, but rather a multivariate t-distribution. Fortunately, the Gaussian pseudo-likelihood assumption is not necessary for parameter inference. Outside of LSS, in cosmic microwave background power spectrum analyses, for instance, the Planck collaboration uses a hybrid likelihood, which only assumes a Gaussian pseudo-likelihood for C_ℓ on small scales (Ade et al. 2014; Aghanim et al. 2016; see also Efstathiou 2004, 2006). On large scales (low ℓ), the likelihood is instead computed directly in pixel space and extensively validated. Testing for likelihood non-Gaussianity and non-Gaussian likelihoods in general are not currently part of the standard practice in LSS studies. For more precise parameter constraints from LSS, however, analyses must go beyond the Gaussian pseudo-likelihood.

In this paper we investigate the impact of the likelihood Gaussianity assumption on the two recent LSS analyses of Beutler et al. (2017) (hereafter B2017) and Sinha et al. (2017) (hereafter S2017). B2017 analyse the power spectrum multipoles (P_ℓ; monopole, quadrupole, and hexadecapole) to measure redshift-space distortions along with the Alcock–Paczynski effect and baryon acoustic oscillation scale. Meanwhile S2017 analyse the group multiplicity function (ζ) in order to constrain parameters of the halo model. Using the B2017 and S2017 analyses, we show in this paper that the assumption of likelihood Gaussianity in LSS is not necessary. We will also show that the mock catalogues used in standard LSS analyses for covariance matrix estimation can be used to quantify the non-Gaussianity. More importantly, we will directly use the mocks to estimate the 'true' non-Gaussian likelihood.

We begin in Section 2 by describing the mock catalogues that we use throughout the paper, constructed originally for covariance matrix estimation in B2017 and S2017. Next in Section 3, we present non-parametric divergence estimators and quantify the non-Gaussianity of the P_ℓ and ζ likelihoods using them. Then in Section 4, we introduce two methods for estimating the 'true' likelihood using the mock catalogues. We then use the likelihood estimates to quantify the impact of likelihood non-Gaussianity on the posterior parameter constraints of B2017 and S2017 in Section 5. We discuss and conclude the paper in Section 6.

2 MOCK CATALOGUES

Mock catalogues are indispensable for standard cosmological analyses of LSS studies. They are used for testing analysis pipelines (Beutler et al. 2017; Grieb et al. 2017; Tinker et al. in preparation), testing the effect of systematics (Guo, Zehavi & Zheng 2012; Vargas-Magaña et al. 2014; Hahn et al. 2017a; Pinol et al. 2017; Ross et al. 2017), and, most relevantly for this paper, estimating the covariance matrix (Parkinson et al. 2012; Kazin et al. 2014; Alam et al. 2017; Beutler et al. 2017; Grieb et al. 2017; Sinha et al. 2017). In fact, nearly all current state-of-the-art LSS analyses use covariance matrices estimated from mocks to evaluate the likelihood.

While some argue for analytic estimates of the covariance matrix (e.g. Mohammed, Seljak & Vlah 2017) or estimates directly from data by subsampling (e.g. Norberg et al. 2009), covariance matrices from mocks have a number of advantages. Mocks allow us to incorporate detailed systematic errors present in the data as well as variance beyond the survey volume. Even for analytic estimates, a large ensemble of mocks is crucial for validation (e.g. Slepian et al. 2017). Moreover, as we show later in this paper, mocks present an additional advantage: They allow us to quantify the non-Gaussianity of the likelihood and more accurately estimate the true likelihood distribution.

In this paper, we focus on two LSS analyses: the power spectrum multipole (P_ℓ) analysis of B2017 and the group multiplicity function (ζ) analysis of S2017. Throughout the paper we will make extensive use of the mock catalogues used in these analyses. In this section, we give a brief description of these mocks and how the observables used in the analysis – $P_\ell(k)$ and $\zeta(N)$ – are calculated from them. Afterwards, we will describe how we compute the covariance matrix from the mocks and pre-process the mock observable data.

2.1 MultiDark-PATCHY mock catalogue

B2017 use the MultiDark-PATCHY mock catalogues from Kitaura et al. (2016) mocks generated using the PATCHY code (Kitaura, Yepes & Prada 2014; Kitaura et al. 2015). These mocks rely on large-scale density fields generated using augmented Lagrangian perturbation theory (ALPT; Kitaura & Heß 2013) on a mesh,

which are then populated with galaxies based on combined non-linear deterministic and stochastic biases. The mocks from the PATCHY code are calibrated to reproduce the galaxy clustering in the high-fidelity BigMultiDark N-body simulation (Klypin et al. 2016; Rodríguez-Torres et al. 2016). Afterwards, stellar masses are assigned to galaxies using the HADRON code (Zhao et al. 2015). Finally, the SUGAR code (Rodríguez-Torres et al. 2016) combines different boxes, incorporates selection effects, and masks to produce mock light-cone galaxy catalogues. The statistics of the resulting mocks are then compared to observations and the process is iterated to reach the desired accuracy. We refer readers to Kitaura et al. (2016) for further details.

In total, Kitaura et al. (2016) generated 12 228 mock light-cone galaxy catalogues for BOSS Data Release 12. In B2017, they use 2045 and 2048 for the Northern Galactic Cap (NGC) and Southern Galactic Cap (SGC) of the LOWZ + CMASS combined sample. B2017 excluded three mock realizations due to notable issues. These issues have since been addressed, so in our analysis we use all 2048 mocks for both the NGC and SGC of the LOWZ + CMASS combined sample. In B2017, they conduct multiple analyses, some using only the power spectrum monopole and quadrupole and others using the monopole, quadrupole, and hexadecapole. They also separately analyse three redshift bins: $0.2 < z < 0.5$, $0.4 < z < 0.6$, and $0.5 < z < 0.75$. In this paper, for simplicity, we focus on one of these analyses: the analysis of the power spectrum monopole, quadrupole, and hexadecapole for the $0.2 < z < 0.5$ bin.

2.2 Sinha et al. (2017) mocks

The simulations used in the Sinha et al. (2017) analysis are from the Large Suite of Dark Matter Simulations (LasDamas) project (McBride et al. 2009), which were designed to model galaxy samples from SDSS DR7. The initial conditions are generated with the 2LTPIC code (Scoccimarro 1998; Crocce, Pueblas & Scoccimarro 2006), and evolved using the N-body GADGET-2 code (Springel 2005). Haloes are identified from the dark matter distribution outputs using the `ntropy-fofsv` code (Gardner, Connolly & McBride 2007), which uses a friend-of-friends (FoF) algorithm (Davis et al. 1985) with a linking length of 0.2 times the mean interparticle separation. S2017 use two configurations of the LasDamas simulations for the SDSS DR7 samples with absolute magnitude limits $M_r < -19$ and $M_r < -21$. The 'Consuelo' simulation contains 1400^3 dark matter particles with mass of $1.87 \times 10^9\, h^{-1}\,\mathrm{M}_\odot$ in a cubic volume of $420\, h^{-1}$Mpc per side evolved from $z_{\rm init} = 99$. The 'Carmen' simulation contains 1120^3 dark matter particles with mass $4.938 \times 10^{10}\, h^{-1}\,\mathrm{M}_\odot$ in a cubic volume of $1000\, h^{-1}$ Mpc per side evolved from $z_{\rm init} = 49$.

The FoF halo catalogues are populated with galaxies using the 'Halo Occupation Distribution' (HOD) framework. The number, positions, and velocities of galaxies are described statistically by an HOD model. S2017 adopt the 'vanilla' HOD model of Zheng & Weinberg (2007), where the mean number of central and satellite galaxies is described by the halo mass and five HOD parameters: $M_{\rm min}$, $\sigma_{\log M}$, M_0, M_1, and α. Lastly, once the simulation boxes are populated with galaxies, observational systematic effects are imposed. The peculiar velocities of galaxies are used to impose redshift-space distortions. Galaxies that lie outside the redshift limits or sky footprint of the SDSS sample are removed. For further details regarding the mocks, we refer readers to S2017.

To calculate their covariance matrix, S2017 produced 200 independent mock catalogues from 50 simulations using a single set of HOD model parameters. The methods we propose in this paper rely on a large number of mocks to accurately sample high-dimensional distributions. We utilize an additional 99 sets of HOD parameters, sampled from the Monte Carlo Markov Chain (MCMC) in S2017, with 200 mocks each. Thus, we have a total of 20 000 mocks for this work. In this paper we focus on the GMF analysis of the SDSS DR7 $M_r < -19$ sample presented in S2017.

2.3 Mock observable $\mathbf{X}^{\rm mock}$ and covariance matrix C

To get from the mock catalogues described above to the covariance matrices used in B2017 and S2017, the observables were measured for each mock in the *same* way as the observations. We briefly describe how $P_\ell(k)$ and $\zeta(N)$ and the corresponding covariance matrices are measured in B2017 and S2017. We then describe how we pre-process the mock observables for the methods we describe in the next sections.

To measure the power spectrum multipoles of the BOSS DR12 galaxies and the MutliDark-PATCHY mocks (Section 2.1), B2017 use a fast Fourier transform (FFT)-based anisotropic power spectrum estimator based on Bianchi et al. (2015) and Scoccimarro (2015). This estimator estimates the monopole, quadrupole, and hexadecapole ($\ell = 0, 2, 4$) of the power spectrum using FFTs of the overdensity field multipoles for a given survey geometry. For further details on the estimator we refer readers to Section 3 of B2017. The power spectrum is computed in bins of $\Delta k = 0.01\, h\,\mathrm{Mpc}^{-1}$ over the range $k = 0.01 - 0.15\, h\,\mathrm{Mpc}^{-1}$ for $\ell = 0$ and 2 and $k = 0.01 - 0.10\, h\,\mathrm{Mpc}^{-1}$ for $\ell = 4$. From the $\vec{P}^{(n)} = \left[P_0^{(n)}(k), P_2^{(n)}(k), P_4^{(n)}(k)\right]$ of the MultiDark-PATCHY mocks, B2017 compute the (i, j) element of the covariance matrix of all multipoles as

$$C_{i,j} = \frac{1}{N_{\rm mock} - 1} \sum_{n=1}^{N_{\rm mock}} \left[\vec{P}_i^{(n)} - \overline{P}_i\right] \times \left[\vec{P}_j^{(n)} - \overline{P}_j\right]. \quad (3)$$

$N_{\rm mock} = 2048$ is the number of mocks and $\overline{P}_i$ is the mean of the mock power spectra: $\overline{P}_i = \frac{1}{N_{\rm mock}} \sum_{n=1}^{N_{\rm mock}} \vec{P}_i^{(n)}$. Since P_0 and P_2 each have 14 bins and P_4 has 9 bins, $\mathbf{C}$ is a 37×37 matrix. In this work, we compute the $P_\ell(k)$ using a similar FFT-based estimator of Hand et al. (2017b) instead of the B2017 estimator. Our choice is purely based on computational convenience. A PYTHON implementation of the Hand et al. (2017b) estimator is publicly available in the NBODYKIT package[1] (Hand et al. 2017a). We confirm that the resulting $P_\ell(k)s$ and covariance matrices from the Hand et al. (2017b) and B2017 estimators are consistent with one another.

Next, the S2017 group multiplicity function analysis starts with the Berlind et al. (2006) FoF algorithm to identify groups in the SDSS and mock data. S2017 adopt the Berlind et al. (2006) linking lengths in units of mean intergalaxy separation: $b_\perp = 0.14$ and $b_\parallel = 0.75$. In comoving lengths, the linking lengths for the SDSS DR7 $M_r < -19$ sample correspond to $(r_\perp, r_\parallel) = (0.57, 3.05)\, h^{-1}$ Mpc. Once both the SDSS galaxy and mock galaxy groups are identified, $\zeta(N)$ is derived by calculating the comoving number density of groups in bins of richness N – the number of galaxies in a galaxy group. For the $M_r < -19$ sample, S2017 use eight N bins: $(5 - 6)$, $(7 - 9)$, $(10 - 13)$, $(14 - 19)$, $(20 - 32)$, $(33 - 52)$, $(53 - 84)$, $(85 - 220)$. For further details on the GMF calculation, we refer readers to Section 4.2 of S2017. From the $\zeta^{(n)}(N)s$ of each mock,

[1] http://nbodykit.readthedocs.io/en/latest/index.html

S2017 compute the (i, j) element of the covariance matrix as

$$C_{i,j} = \frac{1}{N_{\rm mock}-1}\sum_{n=1}^{N_{\rm mock}}\left[\zeta^{(n)}(N_i)-\bar{\zeta}(N_i)\right]\times\left[\zeta^{(n)}(N_j)-\bar{\zeta}(N_j)\right]. \quad (4)$$

S2017 compute the covariance matrix using 200 mocks generated using a single fiducial set of HOD parameters. As we describe in Section 2.2, in this paper we use 20 000 mocks from 100 different sets of HOD parameters sampled from the MCMC chain. The GMF covariance matrix we use in this paper is computed with $N_{\rm mock} =$ 20 000 mocks. For the rest of the paper, in order to discuss the two separate analyses of B2017 and S2017 in a consistent manner, we define the matrix $\mathbf{D}^{\rm mock}$ of the mock observables (P_ℓ and ζ) as

$$\mathbf{D}^{\rm mock} = \left\{\mathbf{D}_n^{\rm mock}\right\} \quad \text{where } \mathbf{D}_n^{\rm mock}\begin{cases}\vec{P}^{(n)} & \text{for B2017,}\\ \zeta^{(n)} & \text{for S2017.}\end{cases} \quad (5)$$

$\mathbf{D}^{\rm mock}$ has dimensions of 2048×37 and $20\,000\times 8$ for B2017 and S2017, respectively.

For the methods in Sections 4.1 and 4.2, the mock observable data ($\mathbf{D}^{\rm mock}$) need to be pre-processed. This pre-processing involves two steps: mean-subtraction (centring) and whitening. For mean subtraction, the mean of the observable is subtracted from $\mathbf{D}^{\rm mock}$. Then $\mathbf{D}^{\rm mock}-\bar{\mathbf{D}}^{\rm mock}$ is whitened using a linear transformation to remove the Gaussian correlation between the bins of $\mathbf{D}^{\rm mock}$:

$$\mathbf{X}^{\rm mock} = L\,(\mathbf{D}^{\rm mock}-\bar{\mathbf{D}}^{\rm mock}). \quad (6)$$

This linear transformation is derived such that the covariance matrix of the whitened data, $\mathbf{X}^{\rm mock}$, is the identity matrix $\mathbf{I}$. Such a whitening linear transformation can be derived in infinite ways. One way to derive the linear transformation is through the eigen-decomposition of the covariance matrix (e.g. Hartlap et al. 2009; Sellentin & Heavens 2018). We, alternatively, derive the linear transformation $\mathbf{L}$ using Cholesky decomposition of the inverse covariance matrix (Press et al. 1992): $\mathbf{C}^{-1} = \mathbf{L}\,\mathbf{L}^T$. We have checked that different methods for whitening do not impact the results of the paper. With this pre-processed mock observable data, we proceed to quantifying the non-Gaussianity of the P_ℓ and ζ likelihoods in the next section.

3 QUANTIFYING THE LIKELIHOOD NON-GAUSSIANITY

The standard approach to parameter inference in LSS studies does not account for likelihood non-Gaussianity. However, we are not the first to investigate likelihood non-Gaussianity in LSS analyses. Nearly two decades ago, Scoccimarro (2000) examined the likelihood non-Gaussianity for the power spectrum and reduced bispectrum using mock catalogues of the *IRAS* redshift catalogues. More recently, Hartlap et al. (2009) and Sellentin & Heavens (2018) examined the non-Gaussianity of the cosmic shear correlation function likelihood using simulations of the *Chandra* Deep Field South and CFHTLenS, respectively.

While these works present different methods for identifying likelihood non-Gaussianity, they do not present a concrete way of quantifying it. Hartlap et al. (2009), for instance, identify the non-Gaussianity of the cosmic shear likelihood by looking at the statistical independence/dependence of principal components of the mock observable. In Sellentin & Heavens (2017), they use the mean integrated squared error (MISE) as a distance metric between Gaussian random variables and the whitened mock observable data vector to characterize non-Gaussian correlations between elements of the data vector. These indirect measures of non-Gaussianity are challenging to interpret or apply more generally to LSS studies.

A more direct approach can be taken to quantify the non-Gaussianity of the likelihood. We can calculate the divergence between the distribution of our observable, $p(x)$, and $q(x)$ a multivariate Gaussian described by the average of the mocks and the covariance matrix – i.e. the pseudo-likelihood. The following are two of the most commonly used divergences: the Kullback–Leibler (KL) divergence

$$D_{\rm KL}(p\parallel q) = \int p(x)\,\log\frac{p(x)}{q(x)}\,{\rm d}x \quad (7)$$

and the Rényi-α divergence

$$D_{R-\alpha}(p\parallel q) = \frac{1}{\alpha-1}\log\int p^{\alpha}(x)\,q^{1-\alpha}(x)\,{\rm d}x. \quad (8)$$

In the limit as α approaches 1, the Rényi-α divergence is equivalent to the KL divergence.

Of course, in our case, we do not know $p(x)$ – i.e. the probability distribution function of our observable. If we did, we would simply use that instead of bothering with the covariance matrix or this paper. We can, however, still estimate the divergence using non-parametric divergence estimators (Wang, Sanjeev & Sergio 2009; Póczos, Xiong & Schneider 2012a; Krishnamurthy et al. 2014). These estimators allow us to estimate the divergence, $\widehat{D}(X_{1:n}\parallel Y_{1:m})$, directly from samples $X_{1:\,n}=\{X_1,\,...X_n\}$ and $Y_{1:\,m}=\{Y_1,\,...Y_m\}$ drawn from p and q, respectively. For instance, the estimator presented in Póczos et al. (2012a) allows us to estimate the kernel function of the Rényi-α divergence,

$$D_{\alpha}(p\parallel q) = \int p^{\alpha}(x)q^{1-\alpha}(x)\,{\rm d}x, \quad (9)$$

using the kth nearest neighbour density estimators. Let $\rho_k(x)$ denote the Euclidean distance of the kth nearest neighbour of x in the sample $X_{1:\,n}$ and $\nu_k(x)$ denote the Euclidean distance of the kth nearest neighbour of x in the sample $Y_{1:\,m}$. Then

$$D_{\alpha}(p\parallel q)\approx\widehat{D}_{\alpha}(X_{1:n}\parallel Y_{1:m}) = \frac{B_{k,\alpha}}{n}\left(\frac{n-1}{m}\right)^{1-\alpha}\sum_{i=1}^{n}\left(\frac{\rho_k^d(X_i)}{\nu_k^d(X_i)}\right)^{1-\alpha}, \quad (10)$$

where $B_{k,\alpha} = \dfrac{(\Gamma(k))^2}{\Gamma(k-\alpha+1)\Gamma(k+\alpha-1)}$. Póczos et al. (2012a) prove that this estimator is asymptotically unbiased:

$$\lim_{n,m\to\infty}\mathbf{E}\left[\widehat{D}_{\alpha}(X_{1:n}\parallel Y_{1:m})\right] = D_{\alpha}(p\parallel q). \quad (11)$$

Plugging $\widehat{D}_{\alpha}(X_{1:n}\parallel Y_{1:m})$ into equation (8), we get an estimator for the Rényi-α divergence. Wang et al. (2009) derive a similar estimator for the KL divergence (equation 7). These divergence estimates have been applied to support distribution machines and used in the machine learning and astronomical literature with great success (e.g. Póczos, Szabó & Schneider 2011; Póczos et al. 2012a,b; Xu et al. 2013; Ntampaka et al. 2015, 2016; Ravanbakhsh et al. 2017). For more details on the non-parametric divergence estimators, we refer readers to Póczos et al. (2012a) and Krishnamurthy et al. (2014).

With these estimators, we can now explicitly quantify the non-Gaussianity of the likelihood by computing the divergence between the likelihood distribution and the Gaussian pseudo-likelihood distribution, $\mathcal{L}^{\rm pseudo}$. $\mathbf{X}^{\rm mock}$ is in principle sampled from $p(x)$. Then with a reference sample $\boldsymbol{Y}^{\rm ref}$ drawn from $\mathcal{L}^{\rm pseudo}$, we can use the estimators to compute $D(\,p(x)\parallel\mathcal{L}^{\rm pseudo})\approx\widehat{D}(\mathbf{X}^{\rm mock}\parallel\boldsymbol{Y}^{\rm ref})$. Similar to the experiments detailed in Póczos et al. (2012a), we construct $Y^{\rm ref}$ with a comparable sample size as $\mathbf{X}^{\rm mock}$: 2000 and 10 000 for the P_ℓ and ζ analyses, respectively. For a sample size of 1000, Sutherland et al. (2012) use $k=5$. Based on the larger sample size of $\mathbf{X}^{\rm mock}$, we calculate the divergences using the $k=10$

nearest neighbours. We note that the divergence estimates are not significantly impacted by our choice of k within the range $5 < k < 20$.

In Fig. 1, we present the resulting Rényi-α (left) and KL (right) divergences (orange) between the likelihood and the Gaussian pseudo-likelihood for the B2017 P_ℓ (top) and S2017 ζ (bottom) analyses: $\widehat{D}_{R\alpha}$ and $\widehat{D}_{\rm KL}$. For reference, we also include (in blue) divergence estimates of the pseudo-likelihood on to itself, which we calculate as $\widehat{D}(\mathbf{X}^{\rm ref} \parallel \mathbf{Y}^{\rm ref})$. $\boldsymbol{X}^{\rm ref}$ is a data vector with the same dimension as $\mathbf{X}^{\rm mock}$ sampled from the pseudo-likelihood. $\widehat{D}$ are *estimates* of the true divergence; therefore, we resample $\boldsymbol{Y}^{\rm ref}$ and compute each $\widehat{D}$ estimate 100 times. In Fig. 1, we present the resulting distributions of $\widehat{D}$, which illustrate the uncertainty of $\widehat{D}$. We also note that for the B2017 P_ℓ analysis, some of the $\widehat{D}_{\rm KL}$ estimates are negative, which violates Gibb's inequality. This is due to the limited number of $p(x)$ samples (only $N_{\rm mock} = 2048$ samples of a 37-dimensional distribution) that results in non-uniformity of the distribution near each sample point and biases the $\widehat{D}_{\rm KL}$ estimates (Kraskov, Stögbauer & Grassberger 2004; Wang et al. 2009). To account for this bias, instead of using $\widehat{D}(\mathbf{X}^{\rm mock} \parallel \boldsymbol{Y}^{\rm ref})$ to directly quantify the discrepancy between the likelihood and the pseudo-likelihood, we use the difference between the $\widehat{D}(\mathbf{X}^{\rm mock} \parallel \boldsymbol{Y}^{\rm ref})$ distributions and the reference $\widehat{D}(\boldsymbol{X}^{\rm ref} \parallel \boldsymbol{Y}^{\rm ref})$ distributions ($\Delta\widehat{D}$). Since $\boldsymbol{X}^{\rm ref}$ has the same dimensions as $\mathbf{X}^{\rm mock}$, $\Delta\widehat{D}$ more accurately reflects the discrepancy between the likelihood and the pseudo-likelihood. Each panel of Fig. 1 shows significant discrepancy between the two distributions – *both the* $P_\ell(k)$ and $\zeta(N)$ *likelihoods are significantly non-Gaussian*.

The Gaussian pseudo-likelihood assumption for P_ℓ is motivated by the central limit theorem. If enough modes contribute to the power spectrum, then the likelihood approaches a Gaussian. Given the survey volume of BOSS DR12 and the restrictive k range of the B2017 analysis ($0.01 < k < 0.15$ for $\ell = 0$ and 2; $0.01 < k < 0.10$ for $\ell = 4$), one would expect this to be mostly true. Although relatively small, we find significant $\Delta\widehat{D}$ and therefore likelihood non-Gaussianity. In order to better understand the source of this non-Gaussianity, we repeat the divergence comparisons for different k ranges. If we exclude the largest scales and set $k_{\rm min} = 0.05$, $\Delta\widehat{D}$ decreases. Meanwhile, if we exclude the smallest scales and set $k_{\rm max} = 0.1$ for all multipoles, $\Delta\widehat{D}$ increases. This suggests that the largest scales (low k) contribute most to the P_ℓ likelihood non-Gaussianity. Furthermore, when we compare the divergences for just the monopole and quadrupole, $\Delta\widehat{D}$ decreases. Among the multipole, the hexadecapole contributes most to the non-Gaussianity of the P_ℓ likelihood. In both the low-k regimes and the hexadecapole, the contribution to the non-Gaussianity is likely caused by low signal-to-noise and failure to satisfy the central limit theorem.

For ζ, the discrepancies between the $\widehat{D}$ distributions are consistent with the fact that the true ζ likelihood distribution is likely Poisson – not Gaussian – similar to the likelihood of observed cluster counts (Cash 1979; Collaboration et al. 2014; Ade et al. 2016). Although the groups identified with an FoF algorithm do not correspond to clusters, we nevertheless expect the likelihood to be non-Gaussian. We again repeat the divergence comparison for different N ranges to better understand the source of non-Gaussianity. Excluding the lowest N bin does not significantly impact $\Delta\widehat{D}$. However, when we exclude the highest N bin, $\Delta\widehat{D}$ decreases significantly. We therefore find that the high-richness end of ζ contibutes most to the non-Gaussianity of the ζ likelihood. The contribution to the non-Gaussianity, similar to the P_ℓ case, comes most from the low-signal-to-noise regime. Besides likelihood non-Gaussianity, biases that arise from estimating the covariance matrix from a limit number of mocks may also contribute to $\Delta\widehat{D}$. With >2000 mocks, however, this bias is likely unimportant for the P_ℓ analysis and even less so for the ζ analysis where we use 20 000 mocks (Hartlap et al. 2009). None the less, this underlines another limitation of using pseudo-likelihoods for parameter inference in LSS studies.

4 ESTIMATING THE NON-GAUSSIAN LIKELIHOOD

In the previous section, we estimate the divergence between the P_ℓ and ζ likelihoods and their respective Gaussian pseudo-likelihoods. These divergences identify and quantify the significant non-Gaussianity in the likelihoods of LSS studies. Our ultimate goals, however, are to quantify the impact of likelihood non-Gaussianity on the final cosmological parameter constraints and to develop more accurate methods for parameter inference in LSS. From the divergence estimates alone, it is not obvious how they propagate on to the final parameter constraints. Therefore, in this section, we present two methods for more accurately estimating the true non-Gaussian likelihoods of P_ℓ and ζ from their corresponding mocks. These methods provide more accurate estimates of the likelihood than the Gaussian pseudo-likelihood. Moreover, we will use them later to quantify the impact of likelihood non-Gaussianity on the B2017 and S2017 parameter constraints.

4.1 Gaussian mixture likelihood estimation

When mock catalogues are used for parameter inference in LSS analyses, they essentially serve as data points sampling the likelihood distribution. For the pseudo-likelihood, this distribution is assumed to have a Gaussian functional form, which is why we estimate the covariance matrix from mocks. However, the Gaussian functional form, or any functional form for that matter, is *not* necessary to estimate the likelihood distribution. Instead, the multidimensional likelihood distribution can be directly estimated from the set of mock catalogues – for instance using Gaussian mixture density estimation (Press et al. 1992; McLachlan & Peel 2000). Besides its extensive use in machine learning and statistics, in astronomy, Gaussian mixture density estimation has been used for inferring the velocity distribution of stars from the *Hipparcos* satellite (Bovy, Hogg & Roweis 2011), classifying galaxies in the Galaxy And Mass Assembly Survey (Taylor et al. 2015), classifying pulsars (Lee et al. 2012), and much more (see also Hogg, Bovy & Lang 2010; Kuhn & Feigelson 2017).

Gaussian mixture density estimation is a 'semi-parametric' method that uses a weighted sum of k Gaussian component densities, a Gaussian mixture model (hereafter GMM)

$$\widehat{p}(x;\boldsymbol{\theta}) = \sum_{i=1}^{k} \pi_i \mathcal{N}(x;\boldsymbol{\theta}_i), \tag{12}$$

to estimate the density. The component weights (π_i; also known as mixing weights) and the component parameters $\boldsymbol{\theta}_i$ are free parameters of the mixture model. Given some data set $\mathbf{X}_N = \{x_1, ..., x_N\}$, these free GMM parameters are, most popularly, estimated through an expectation-maximization (EM) algorithm (Dempster, Laird & Rubin 1977; Neal & Hinton 1998). The EM algorithm begins by randomly assigning $\boldsymbol{\theta}_i^0$ to the k Gaussian components. The algorithm then iterates between two steps. In the first step, the algorithm computes $\boldsymbol{x}_n$, a probability of being generated by each component

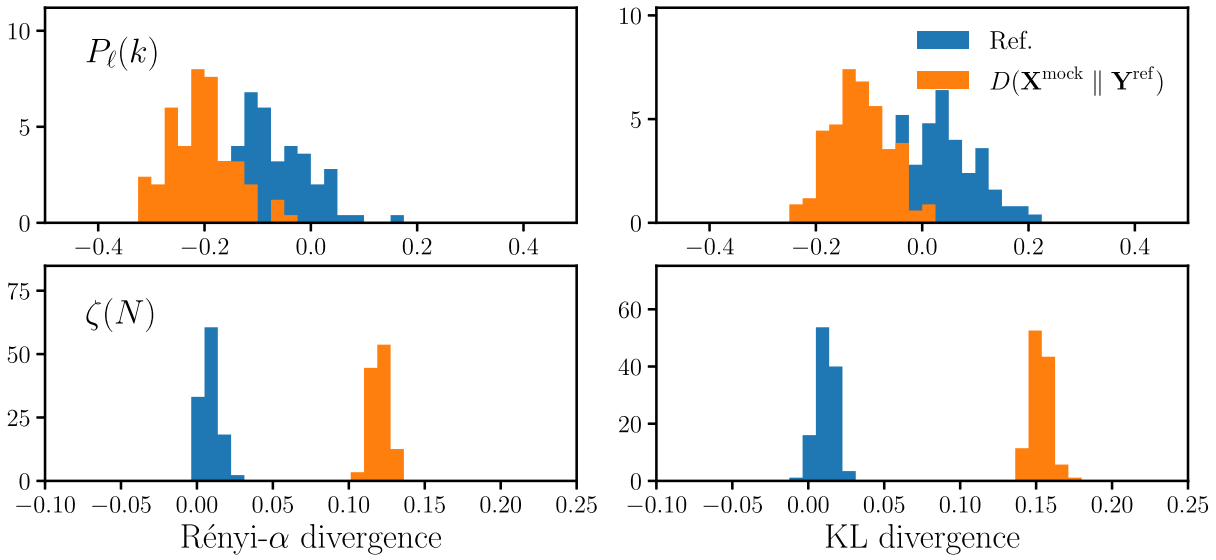

Figure 1. Rényi-α and KL divergence estimates ($\widehat{D}_{R\alpha}$ and $\widehat{D}_{KL}$; orange) between the likelihood distribution and the Gaussian pseudo-likelihood for the B2017 P_ℓ (top) and S2017 ζ (bottom) analyses. We include in blue, as reference, the divergence estimates of the pseudo-likelihood on to itself. $\widehat{D}_{R\alpha}$ and $\widehat{D}_{\rm KL}$ are computed using the non-parametric k-NN estimator (Section 3) on the mock data $\mathbf{X}^{\rm mock}$ and a reference sample $\mathbf{Y}^{\rm ref}$ drawn from the pseudo-likelihood. We compute $\widehat{D}_{R\alpha}$ and $\widehat{D}_{\rm KL}$ 100 times and plot their distribution in order to illustrate the uncertainty of the $\widehat{D}$ estimator. The significant discrepancy between the two divergence distributions in each of the panels identifies the *significant non-Gaussianity of the* $P_\ell(k)$ and $\zeta(N)$ *likelihoods*.

of the model, for every data point. These probabilities can be thought of as weighted assignments of the points to the components. Next, given the x_n assignment to the components at some step t, $\boldsymbol{\theta}_i^t$ of each component are updated to $\boldsymbol{\theta}_i^{t+1}$ to maximize the likelihood of the assigned points. At this point, π_i can also be updated by summing up the assignment weights and normalizing it by the total number of data points, N. This entire process is repeated until convergence – i.e. when the log-likelihood of the mixture model log $p(\mathbf{X}_N;\boldsymbol{\theta}^t)$ converges. As demonstrated in Wu (1983), the EM algorithm is guaranteed to converge to a local maximum of log $p(\mathbf{X}_N;\boldsymbol{\theta})$. In practice, instead of arbitrarily assigning the initial condition, $\boldsymbol{\theta}_i^0$ is derived from a `k-means` clustering algorithm (Lloyd 1982). The `k-means` algorithm clusters a data set, $\mathbf{X}_N$, into k clusters, each described by the mean (or centroid) μ_i of the samples in the cluster. The algorithm then iteratively chooses centroids that minimize the average squared distance between points in the same cluster. For our GMMs, we initialize the EM algorithm using the `k-means++` algorithm of Arthur & Vassilvitskii (2007).

So far in our description of GMMs, we have kept the number of components k fixed. k, however, is a free parameter and selecting k is a crucial step in Gaussian mixture density estimation. With too many components the model may overfit the data, while with too few components the model may not be flexible enough to approximate the true underlying distribution. In order to address this model selection problem when selecting k, we make use of the Bayesian information criterion (BIC; Schwarz 1978). BIC has been widely used for determining the number of components in mixture modelling (e.g. Leroux 1992; Roeder & Wasserman 1997; Fraley & Raftery 1998; Steele & Raftery 2010) and for model selection in general in astronomy (e.g. Liddle 2007; Broderick et al. 2011; Wilkinson et al. 2015; Vakili & Hahn 2016). According to BIC, models with higher likelihood are preferred; however, to address the concern of overfitting, BIC introduces a *penalty* term for the number of parameters in the model:

$$\mathrm{BIC} = -2\ln\mathcal{L} + N_{\rm par}\ln N_{\rm data}. \tag{13}$$

We select k based on the number of components in the model with the lowest BIC.

With Gaussian mixture density estimation we can directly estimate the likelihood distribution using the mock catalogues. We first fit GMMs with $k < 30$ components to the whitened mock data $\mathbf{X}^{\rm mock}$ using the EM algorithm for each model. For each of the converged GMMs, we calculate the BIC and then select the model with the lowest BIC as the best density estimate of the likelihood distribution: $\widehat{p}_{\rm GMM}(x)$. For the B2017 and S2017 analyses, we find GMMs with $k = 1$ and 5 components, respectively, have the lowest BIC. These selected density estimates can then be used to calculate the likelihood and quantify the impact of likelihood non-Gaussianity on the parameter constraints of B2017 and S2017. But first, we test whether $\widehat{p}_{\rm GMM}$ provides a better estimate of the non-Gaussian likelihoods over Gaussian pseudo-likelihoods by repeating the divergence estimates from Section 3.

To estimate the divergence between our Gaussian mixture density estimate, $\widehat{p}_{\rm GMM}$, and the likelihood distribution, we take the same approach as our $\widehat{D}(\mathbf{X}^{\rm mock}|\boldsymbol{Y}^{\rm ref})$ calculation in Section 3. Instead of $\boldsymbol{Y}^{\rm ref}$ drawn from the pseudo-likelihood, we draw samples from $\widehat{p}_{\rm GMM}(x)$ with the same dimensions. Then we calculate k-NN Rényi-α and KL divergence estimates between this sample and $\mathbf{X}^{\rm mock}$. To get a distribution of divergence estimates that reflects the scatter in the estimator, we repeat the estimates 100 times resampling $\widehat{p}_{\rm GMM}$ each time (exactly the same method as for Fig. 1). In Fig. 2, we present the resulting distribution of divergences between $\widehat{p}_{\rm GMM}$ and the likelihood distribution in purple for the $P_\ell(k)$ (top) and $\zeta(N)$ (bottom) analyses. For comparison, we include the $\widehat{D}$ distributions for Gaussian pseudo-likelihoods from Fig. 1.

From Fig. 2, we see that the Gaussian mixture density estimate significantly improves the divergence discrepancy compared to the pseudo-likelihood for the $\zeta(N)$ analysis of S2017. In other words, *our Gaussian mixture density estimate is a significant better estimate of the* ζ *likelihood distribution than the pseudo-likelihood*. On the other hand, the Gaussian mixture density estimate

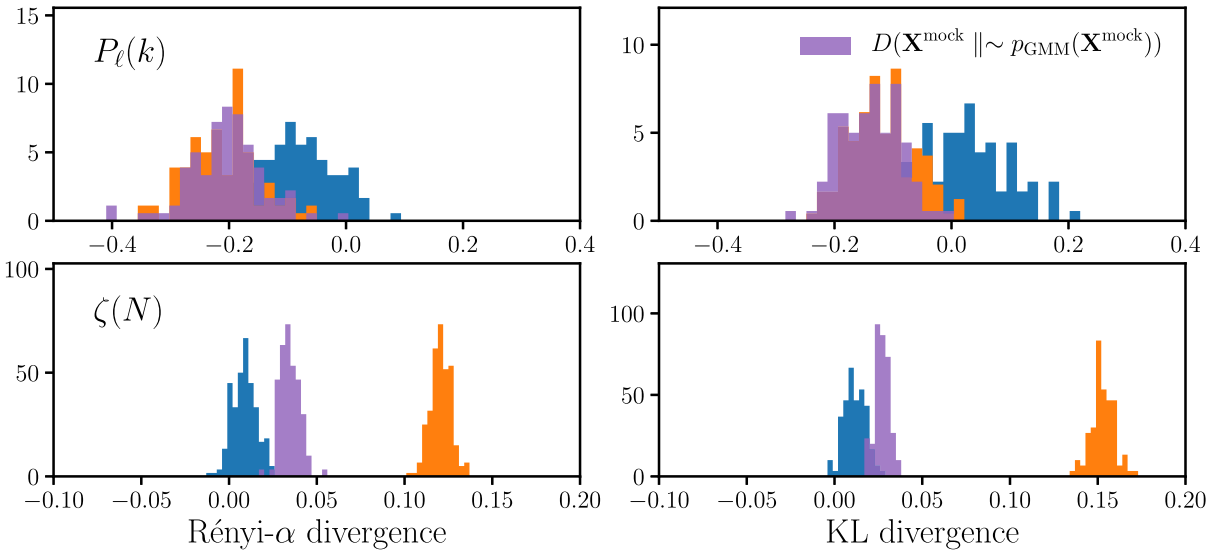

Figure 2. Rényi-α and KL divergence estimates ($\widehat{D}_{R\alpha}$ and $\widehat{D}_{KL}$; purple) between the likelihood distribution and the Section 4.1 GMM likelihood estimate for the B2017 P_ℓ (top) and S2017 ζ (bottom) analyses. We include the divergence estimates for the Gaussian pseudo-likelihood from Fig. 1 (blue) for comparison. The Gaussian mixture likelihood does not significantly improve the discrepancy in divergence for the P_ℓ analysis. This is due to the high dimensionality (37 dimensions) of the P_ℓ likelihood. For the ζ analysis, *our Gaussian mixture likelihood estimate is a significantly better estimate of the likelihood than the pseudo-likelihood*.

for the $P_\ell(k)$ analysis of B2017 does not improve the divergence discrepancy. This is expected since the P_ℓ Gaussian mixture density estimate that we select based on BIC only has one component and therefore is equivalent to the pseudo-likelihood. Also, one would expect a direct density estimation to be more effective for the S2017 case, where we estimate an eight-dimensional distribution with $N_{\rm mock} = 20\,000$ samples, compared to the B2017 case where we estimate a 37-dimensional distribution with only $N_{\rm mock} = 2048$ samples. Given the unconvincing accuracy of the Gaussian mixture density estimate of the P_ℓ likelihood, in the next section we present an alternative method for estimating the non-Gaussian likelihood.

4.2 Independent component analysis

Gaussian mixture density estimation fails to accurately estimate the 37-dimensional P_ℓ likelihood distribution of B2017. Rather than estimating the likelihood distribution directly, if we can transform the observable $\mathbf{x}$ (e.g. P_ℓ) into statistically independent components $\boldsymbol{x}^{\rm IC}$ the problem becomes considerably simpler. Since $\boldsymbol{x}^{\rm IC}$ is statistically independent, the likelihood distribution becomes

$$p(x) = \prod_{n=1}^{N_{\rm bin}} p_{x_n^{\rm IC}}(x) \tag{14}$$

where $N_{\rm bin}$ is the number of bins in the observable and the number of independent components. For the B2017 case, this reduces the problem of estimating a 37-dimensional distribution with 2048 samples to a problem of estimating 37 one-dimensional distributions with 2048 samples each. The challenge is in *finding* such a transformation.

Efforts in the past have attempted to tackle this sort of high-dimensional problem (e.g. Scoccimarro 2000; Eisenstein & Zaldarriaga 2001; Gaztañaga & Scoccimarro 2005; Norberg et al. 2009; Sinha et al. 2017). They typically use singular value decomposition or principal component analysis (PCA; Press et al. 1992). For a Gaussian likelihood, the PCA components of it are statistically independent. However, when the likelihood is *not* Gaussian, the PCA components are uncorrelated but *not necessarily statistically independent* (Hartlap et al. 2009). Since the P_ℓ and ζ likelihoods are non-Gaussian, we cannot use PCA. Instead, we follow Hartlap et al. (2009) and use independent component analysis (ICA; Hérault & Ans 1984; Comon 1994; Hyvärinen & Oja 2000; Hyvarinen 2001).

In order to find the transformation of $\mathbf{x}$ to $\boldsymbol{x}^{\rm IC}$ we first assume that $\mathbf{x}$ is generated by some linear transformation $\mathbf{x} = \mathbf{M}\,\boldsymbol{x}^{\rm IC}$. Then the goal of ICA is to invert this problem, $\mathbf{y} = \mathbf{W}\,\mathbf{x}$, and find $\mathbf{W}$ and $\mathbf{y}$ that best estimate $\boldsymbol{x}^{\rm IC} \approx \mathbf{y}$. The basic premise of ICA is simple, *maximizing non-Gaussianity maximizes the statistical independence*. Consider a single component of $\mathbf{y}$:

$$\boldsymbol{y}_n = \boldsymbol{w}_n^t\,\mathbf{x} = \boldsymbol{w}_n^t\,\mathbf{M}\,\boldsymbol{x}^{\rm IC} \tag{15}$$

where $\boldsymbol{w}_n^t$ is the nth row of $\mathbf{W}$. Since $\boldsymbol{y}_n$ is a linear combination of the independent components $\boldsymbol{x}^{\rm IC}$, from the central limit theorem $\boldsymbol{y}_n$ is necessarily more Gaussian than any of the components *unless* $\boldsymbol{y}_n$ is equal to one of the $\boldsymbol{x}^{\rm IC}$ components. In other words, we can achieve $x^{\rm IC} \approx \mathbf{y}$ by finding $\mathbf{W}$ that maximizes the non-Gaussianity of $\mathbf{y}$. For a more rigorous justification of ICA we refer readers to Hyvarinen (2001). In practice, non-Gaussianity is commonly measured using differential entropy – 'negentropy'. For $\boldsymbol{y}_n$ with density function p_{y_n} the entropy is defined as

$$H_{y_n} = -\int p_{y_n}(y)\,\log p_{y_n}(y)\,{\rm d}y. \tag{16}$$

Since the Gaussian distribution has the largest entropy among all distributions with a given variance, the negentropy can be defined as

$$J_{y_n} = H_{y_n^{\rm Gauss}} - H_{y_n}. \tag{17}$$

Finding the statistically independent components is now a matter of finding the $\mathbf{W}$ that maximizes $\sum_n J_{\mathbf{y_n}}$ – the negentropy of $\mathbf{y}$. In this paper, we make use of the `FastICA` fixed-point iteration algorithm (Hyvarinen 1999). The algorithm starts with randomly selected $\boldsymbol{w}_n$; then it uses approximations of negentropy

from Hyvärinen (1998) and Newton's method to iteratively solve for $\mathbf{W}$ that maximizes negentropy. For details on the `FastICA` algorithm, we refer readers to Hyvarinen (1999).

Performing ICA on the whitened observable data $\mathbf{X}^{\rm mock}$, we derive the matrix $\mathbf{W}$ that transforms $\mathbf{X}^{\rm mock}$ into $N_{\rm bin}$ approximately independent components:

$$\mathbf{X}^{\rm ICA} = \mathbf{W}\,\mathbf{X}^{\rm mock} = \{X_1^{\rm ICA}, ..., X_{N_{\rm bin}}^{\rm ICA}\}. \tag{18}$$

From these statistically independent components and equation (14), we can estimate the likelihood distribution. $p_{x_n^{\rm IC}}(x)$, from equation (14), is the one-dimensional distribution function of the nth ICA component. This distribution is sampled by $\mathbf{X}_n^{\rm ICA}$, the transformed mock data. That means $\mathbf{X}_n^{\rm ICA}$ can be used to estimate $p_{x_n^{\rm ICA}}$ using a method like kernel density estimation (KDE; Hastie, Tibshirani & Friedman 2009; Feigelson & Babu 2012). With KDE, the density estimate, $\widehat{p}_{x_n^{\rm ICA}}$, is constructed by smoothing the empirical distribution of the ICA component $x_n^{\rm ICA}$ using a smooth kernel:

$$\widehat{p}_{x_n^{\rm ICA}}(x) = \frac{1}{b\,N_{\rm mock}} \sum_{j=1}^{N_{\rm mock}} K\left(\frac{x - X_n^{(j),{\rm ICA}}}{b}\right), \tag{19}$$

where b is the bandwidth and K is the kernel function. Following the choices of Hartlap et al. (2009), we use a Gaussian distribution for K and the 'rule of thumb' bandwidth (also known as Scott's rule; Scott 1992; Davison 2008) for b. Combining the $\widehat{p}_{x_n^{\rm ICA}}$ estimates for all $n = 1, ..., N_{\rm bin}$ into equation (14), we can estimate the likelihood distribution $p(x) \approx \prod_n \widehat{p}_{x_n^{\rm ICA}}(x)$

We again check whether the likelihood estimate from the ICA is actually a better estimate of the true likelihood distribution compared to the Gaussian pseudo-likelihood. Following the same procedure as we did for the Gaussian mixture likelihood in Section 4.1, we calculate the divergence between our ICA likelihood, $\prod \widehat{p}_{x_n^{\rm ICA}}(x)$, and the likelihood distribution, $p(x)$. We draw a sample from $\prod \widehat{p}_{x_n^{\rm ICA}}$ with the same dimensions as $\mathbf{Y}^{\rm ref}$ (Section 3), apply the mixing matrix (undoing the ICA transformation), and then calculate the k-NN Rényi-α and KL divergence estimates between the sample and $\mathbf{X}^{\rm mock}$. We repeat these steps 100 times to get the distribution of estimates that reflects the scatter in the estimator. In Fig. 3, we present the resulting distribution of $\widehat{D}\left(\mathbf{X}^{\rm mock} \,\|\sim \prod \widehat{p}_{x_n^{\rm ICA}}\right)$ in green for the $P_\ell(k)$ (top) and $\zeta(N)$ (bottom) analyses. For comparison, we include the distributions for the Gaussian pseudo-likelihood from Fig. 1.

For *both* B2017 and S2017, our ICA likelihood significantly improves the divergence discrepancy compared to the pseudo-likelihood. For S2017, however, the ICA likelihood proves to be less accurate than the Gaussian mixture likelihood in Section 4.1. More importantly, for B2017 where the Gaussian mixture likelihood did not improve upon the pseudo-likelihood, the ICA method provides a significantly more accurate likelihood estimate. This demonstrates that the ICA method is an effective alternative to the more direct Gaussian mixture method. The effectiveness of the ICA method in estimating higher dimensional likelihoods with fewer samples (mocks) is particularly appealing for LSS, since analyses continue to increase the size of their observable data vector. In Hartlap et al. (2009), they suggest that a low $N_{\rm mock}$ may bias the ICA likelihood estimate. By examining the divergence discrepancy as we did in Figs 2 and 3, we ensure that the likelihood estimation methods provide a better estimate of the true likelihood than the Gaussian pseudo-likelihood. Multiple methods can easily be tested to construct the best estimate of the likelihood distribution for *each* specific analysis. Based on the performances of the GMM and ICA methods, we chose the ICA likelihood for the B2017 analysis and the GMM likelihood for the S2017 analysis.

5 IMPACT ON PARAMETER INFERENCE

To derive the posterior distribution of their model parameters, both B2017 and S2017 use the standard Monte Carlo Markov Chain approach with the Gaussian pseudo-likelihood. The B2017 analysis includes 11 parameters,

$$\left\{f\sigma_8,\ \alpha_\parallel,\ \alpha_\perp,\ b_1^{\rm NGC}\sigma_8,\ b_1^{\rm SGC}\sigma_8,\ b_2^{\rm NGC}\sigma_8,\ b_2^{\rm SGC}\sigma_8,\ \sigma_v^{\rm NGC},\ \sigma_v^{\rm SGC},\ N^{\rm NGC},\ \text{and}\ N^{\rm SGC}\right\},$$

while the S2017 analysis includes 5 parameters,

$$\left\{\log M_{\rm min},\ \sigma_{\log M},\ \log M_0,\ \log M_1,\ \text{and}\ \alpha\right\}.$$

Using the improved likelihood estimates of Sections 4.1 and 4.2, we can now better estimate the true posteriors for the parameters and quantify the impact of likelihood non-Gaussianity on parameter constraints. The ideal method to determine the true posterior distributions would be to run new MCMC chains with non-Gaussian likelihood estimators. While re-running MCMC chains is relatively tractable for the B2017 analysis, for S2017 this is *significantly* more involved. Rather than a perturbation theory based model from B2017, the S2017 model is a forward model, identical to their mocks (Section 2.2). Re-running the MCMC samples would involve evaluating the computationally costly forward model of S2017 $\sim 10^6$ times and is prohibitively expensive.

Without having to re-run the MCMC chains, we instead use importance sampling to derive the new posteriors from the original chains (see Wasserman 2004 for details on importance sampling). The *target* distribution we want is the new posterior. To sample this distribution, we re-weight the original posterior as the *proposal* distribution with importance weights. In our case, the importance weights are the ratio of the (non-Gaussian) likelihood estimates over the (Gaussian) pseudo-likelihood. If we let $P(\mathbf{x}|\boldsymbol{\theta})$ be the original pseudo-likelihood and $P'(\mathbf{x}|\boldsymbol{\theta})$ be our 'new' likelihood, then the new marginal likelihood can be calculated through importance sampling:

$$P'(\mathbf{x}|\theta_1) = \int P'(\mathbf{x}|\boldsymbol{\theta})\,{\rm d}\theta_2 ... {\rm d}\theta_m = \int \frac{P'(\mathbf{x}|\boldsymbol{\theta})}{P(\mathbf{x}|\boldsymbol{\theta})}\,P(\mathbf{x}|\boldsymbol{\theta})\,{\rm d}\theta_2 ... {\rm d}\theta_m. \tag{20}$$

Then through Monte Carlo integration,

$$P'(\mathbf{x}|\theta_1) \approx \sum_{\boldsymbol{\theta}^{(i)}\in S} \frac{P'(\mathbf{x}|\boldsymbol{\theta}^{(i)})}{P(\mathbf{x}|\boldsymbol{\theta}^{(i)})}, \tag{21}$$

where S is the sample drawn from $P(\mathbf{x}|\boldsymbol{\theta})$. S is simply the original MCMC chain in our case. The only calculation required is the importance weights in equation (21), $P'(\mathbf{x}|\boldsymbol{\theta}^{(i)})/P(\mathbf{x}|\boldsymbol{\theta}^{(i)})$ for each sample $\boldsymbol{\theta}^{(i)}$ of the original MCMC chain. For B2017, $P(\mathbf{x}|\boldsymbol{\theta}^{(i)})$ is the ICA likelihood; for S2017, $P(\mathbf{x}|\boldsymbol{\theta}^{(i)})$ is the GMM likelihood.

In Fig. 4 we present the resulting posterior distributions using the non-Gaussian ICA likelihood for the $\{f\sigma_8, \alpha_{||}, \alpha_\perp, b_1^{\rm NGC}\sigma_8, b_1^{\rm SGC}\sigma_8, b_2^{\rm NGC}\sigma_8, b_2^{\rm SGC}\sigma_8\}$ parameters in the B2017 P_ℓ analysis (orange). We include the original B2017 posteriors for comparison in blue. At the bottom of each panel, we also include box plots marking the confidence intervals of the updated and original posteriors. The boxes and 'whiskers' represent the 68 per cent and 95 per cent confidence intervals, respectively. The median and 68 per cent confidence intervals of the posteriors are also listed in Table 1. $f\sigma_8$ and $b_2^{\rm SGC}\sigma_8$ are the main parameters with noticeable changes in their posteriors. After accounting for the non-Gaussian likelihood, the posterior of $b_2^{\rm SGC}\sigma_8$ broadens from $0.476^{+1.262}_{-1.175}$ to $0.422^{+1.517}_{-1.377}$. More importantly, the $f\sigma_8$ posterior shifts from $0.478^{+0.053}_{-0.049}$ to $0.456^{+0.059}_{-0.049}$,

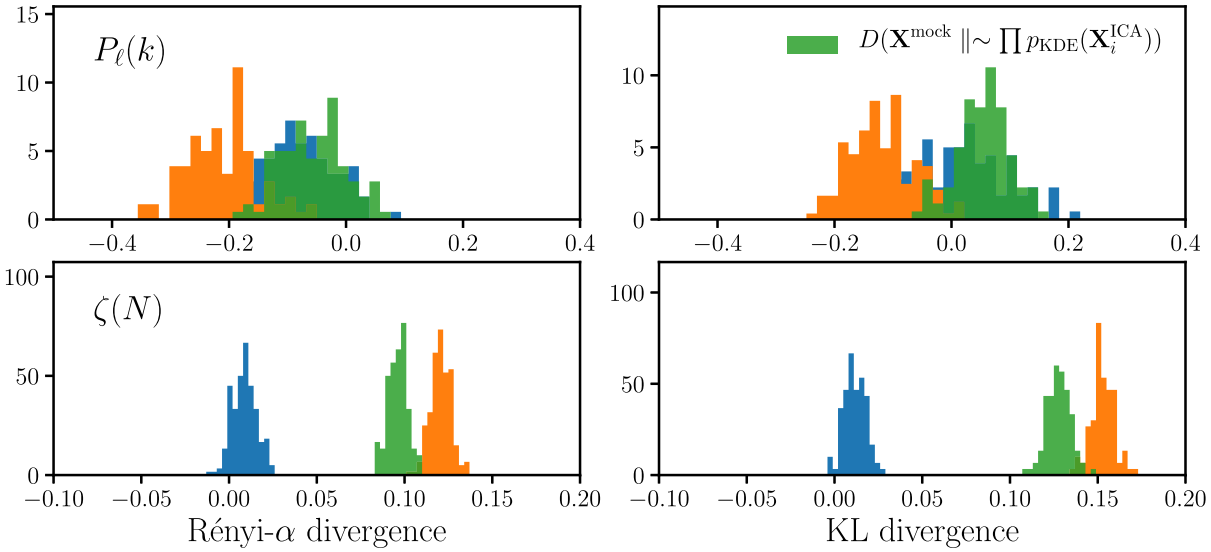

Figure 3. Rényi-α and KL divergence estimates ($\widehat{D}_{R\alpha}$ and $\widehat{D}_{\rm KL}$; green) between the likelihood distribution and the Section 4.2 ICA likelihood estimate for the B2017 P_ℓ (top) and S2017 ζ (bottom) analyses. We include the divergence estimates from Fig. 1 for comparison. The ICA likelihood significantly improves the divergence discrepancy for both the P_ℓ and ζ analyses. For ζ, the improvement of the ICA likelihood over the pseudo-likelihood is more modest than our GMM estimate from Section 4.1. However, for P_ℓ where the GMM method struggled, our ICA likelihood provides a significantly better estimate of the true P_ℓ likelihood than the pseudo-likelihood.

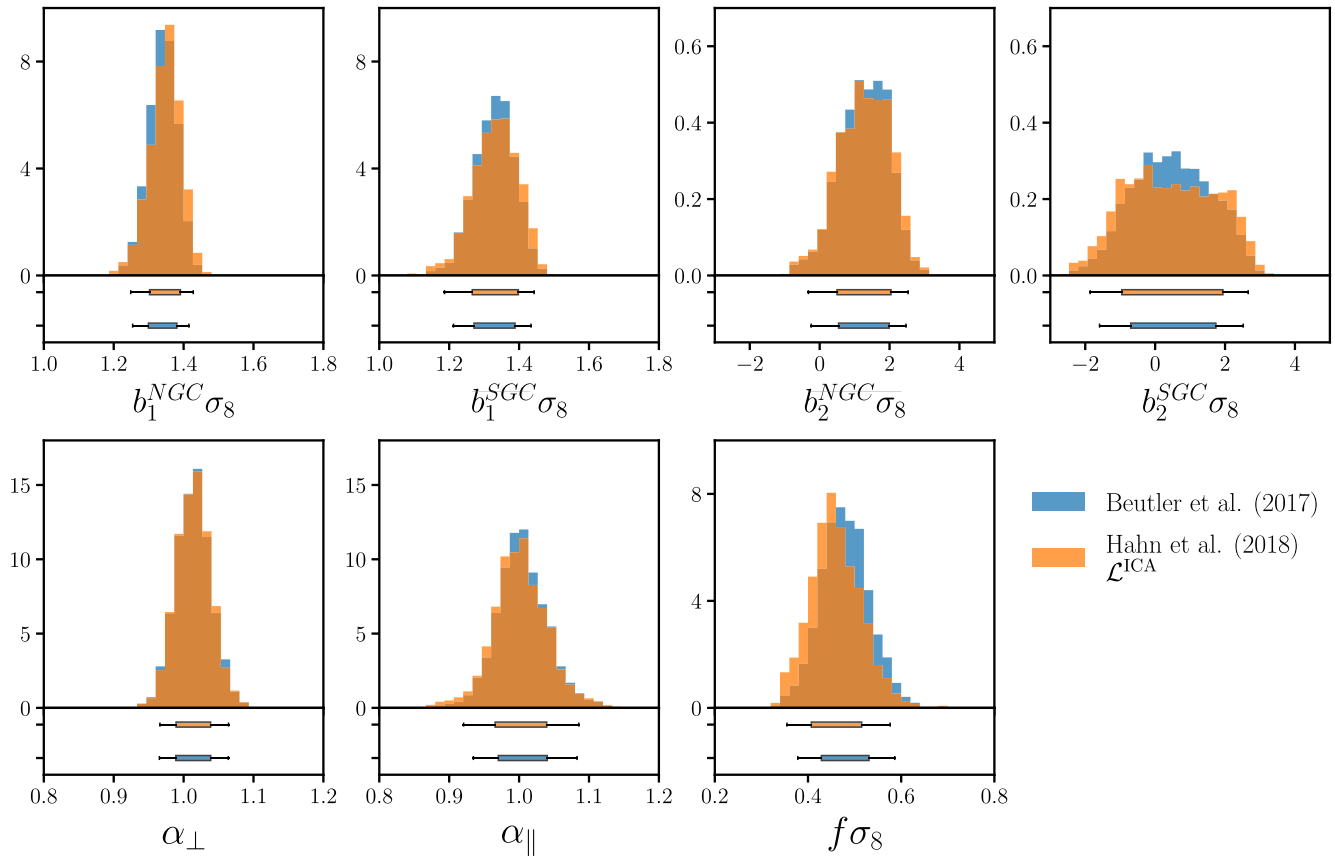

Figure 4. The posterior distribution for $\{f\sigma_8, \alpha_{\shortparallel}, \alpha_\perp, b_1^{\rm NGC}\sigma_8, b_1^{\rm SGC}\sigma_8, b_2^{\rm NGC}\sigma_8, b_2^{\rm SGC}\sigma_8, \}$ in the B2017 P_ℓ analysis using the non-Gaussian ICA likelihood (orange). We include in blue the original B2017 posteriors for comparison. At the bottom of each panel we include box plots that mark the 68 per cent and 95 per cent confidence intervals of the posterior. The discrepancies between the posteriors are most evident for the parameters $f\sigma_8$ and $b_2^{\rm SGC}\sigma_8$. The $f\sigma_8$ constraint shifts by -0.44σ. Hence, using the pseudo-likelihood in the P_ℓ analysis biases the posteriors of these parameters. However, likelihood non-Gaussianity does *not* have a significant impact on the overall parameter constraints of the P_ℓ analysis.

Table 1. Impact of likelihood non-Gaussianity on the posterior parameter constraints of B2017 and S2017.

	B2017 P_ℓ analysis			
	$b_1^{\rm NGC}\sigma_8$	$b_1^{\rm SGC}\sigma_8$	$b_2^{\rm NGC}\sigma_8$	$b_2^{\rm SGC}\sigma_8$
	$f\sigma_8$	$\alpha_{\rm ll}$	$\alpha_\perp$	
B2017	$1.341^{+0.040}_{-0.042}$	$1.333^{+0.056}_{-0.062}$	$1.293^{+0.697}_{-0.752}$	$0.476^{+1.262}_{-1.175}$
	$0.478^{+0.053}_{-0.049}$	$1.003^{+0.038}_{-0.032}$	$1.014^{+0.025}_{-0.025}$	
non-Gaussian	$1.351^{+0.040}_{-0.049}$	$1.335^{+0.063}_{-0.069}$	$1.295^{+0.746}_{-0.798}$	$0.422^{+1.517}_{-1.377}$
$\mathcal{L}_{\rm ICA}$	$0.456^{+0.059}_{-0.049}$	$1.001^{+0.039}_{-0.035}$	$1.014^{+0.024}_{-0.025}$	
	S2017 ζ analysis			
	$\log M_{\rm min}$	$\sigma_{\log M}$	$\log M_0$	$\log M_1$	α
S2017	$11.68^{+0.148}_{-0.128}$	$0.585^{+0.255}_{-0.367}$	$9.154^{+2.074}_{-2.162}$	$12.62^{+0.064}_{-0.077}$	$0.928^{+0.042}_{-0.054}$
Gaussian $\mathcal{L}_{\rm pseudo}$	$11.68^{+0.152}_{-0.131}$	$0.586^{+0.264}_{-0.369}$	$9.195^{+2.086}_{-2.180}$	$12.61^{+0.070}_{-0.074}$	$0.936^{+0.043}_{-0.049}$
non-Gaussian					
$\mathcal{L}_{\rm GMM}$	$11.69^{+0.188}_{-0.135}$	$0.554^{+0.317}_{-0.378}$	$9.159^{+2.174}_{-2.198}$	$12.64^{+0.095}_{-0.109}$	$0.909^{+0.067}_{-0.086}$

which corresponds to a shift of -0.44σ. The other parameter constraints, however, remain largely unaffected by likelihood non-Gaussianity.

Focusing on the main cosmological parameters $f\sigma_8$, $\alpha_{\rm ll}$, and $\alpha_\perp$, we present their joint posterior distributions in Fig. 5. The contours mark the 68 per cent and 95 per cent confidence intervals of the posteriors. The shift in the $f\sigma_8$ distribution is reflected in the ($f\sigma_8$, $\alpha_{\rm ll}$) and ($\alpha_\perp$, $f\sigma_8$) contours (left-hand and middle panels, respectively). The ($\alpha_{\rm ll}$, $\alpha_\perp$) distribution (right), however, shows nearly no change from the non-Gaussian likelihood. Despite its impact on $f\sigma_8$ and $b_2^{\rm SGC}\sigma_8$, likelihood non-Gaussianity does *not* significantly impact the overall parameter constraints of the P_ℓ analysis. $b_2^{\rm SGC}\sigma_8$ is a poorly constrained nuisance parameter and although using the pseudo-likelihood biases $f\sigma_8$, the impact relative to its uncertainty is small – less than 0.5σ. Furthermore, some of the impact may be from statistical fluctuation, although this is likely not an important contributor since the PATCHY mocks are calibrated so that their $\overline{P_\ell}$ is consistent with the BOSS P_ℓ. Some uncertainty is also introduced by the finite sampling of the MCMC chains. As mentioned in Section 3, some of the impact may also come from biases in covariance matrix estimation. Never the less, the fact that the P_ℓ analysis is largely unaffected by likelihood non-Gaussianity is consistent with the relatively small divergences found in Fig. 1. It also illustrates the remarkable effectiveness of the central limit theorem.

Next, in Fig. 6, we present the posterior distributions calculated using the non-Gaussian GMM likelihood for the HOD parameters in the S2017 ζ analysis (orange). We include the posteriors calculated using the pseudo-likelihood for comparison in blue. The box plots at the bottom of each plot mark the 68 per cent and 95 per cent confidence intervals of the posteriors. In the dotted lines, we plot the original S2017 posteriors, which differ slightly from the blue distribution. This discrepancy is caused by the difference in the covariance matrix we use in the pseudo-likelihood (see Section 2.3). The difference, however, is negligible and goes to illustrate that the covariance matrix of ζ does not have a strong dependence on HOD parameters. In other words, our analysis is not significantly affected by our use of mocks generated from multiple HOD parameters.

Besides the poorly constrained parameters $\sigma_{\log M}$ and $\log M_0$, likelihood non-Gaussianity significantly impacts the posterior distributions of the HOD parameters. Each of the parameter constraints for $\log M_{\rm min}$, $\log M_1$, and α is significantly broadened and shifted from the pseudo-likelihood constraints (see Table 1 for details). The $\log M_1$ constraint, for instance, is shifted by $+0.43\sigma$ and its 68 per cent confidence interval is expanded by 42 per cent. Similarly, the α constraint is shifted by -0.51σ and its 68 per cent confidence interval is expanded by 66 per cent. The impact of likelihood non-Gaussianity is further emphasized in the joint posterior distributions in Fig. 7. The $\log M_{\rm min}$ versus $\sigma_{\log M}$ and $\log M_{\rm min}$ versus α contours are both shifted and broadened compared to the $\mathcal{L}^{\rm pseudo}$ posterior. Figs 6 and 7 reveal that *using the Gaussian pseudo-likelihood significantly underestimates the uncertainty and biases the HOD parameter constraints of the S2017 ζ analysis.*

The contrast between the pseudo-likelihood posteriors and our posteriors in Figs 6 and 7 reflect the divergences in Fig. 1, which revealed significant discrepancy between the ζ likelihood and the pseudo-likelihood. These divergences and posteriors are consistent with the expectation that the true ζ likelihood distribution is likely Poisson. Although we expect the likelihood to be similar to the observed cluster count likelihood, the complicated connection between FoF groups and the underlying matter overdensity makes writing down the exact ζ likelihood function tremendously difficult. None the less, the GMM likelihood estimation method we present provides an accurate estimate of the non-Gaussian likelihood.

The updated posteriors of the S2017 ζ analysis highlight the importance of accounting for likelihood non-Gaussianity in parameter inference of LSS studies. One of the main results of the S2017 HOD analysis is that the lambda cold dark matter (ΛCDM) + HOD model can successfully fit either $\zeta(N)$ or the projected two-point correlation function $w_{\rm p}(r_{\rm p})$ separately, but struggles to jointly fit both (see fig. 10 in S2017). Such a tension suggests that the 'vanilla' HOD model is not sufficiently flexible in describing the galaxy–halo connection. Likelihood non-Gaussianity is likely to impact this result. Once the non-Gaussianity is included in the analysis, the posteriors are broadened and shifted towards relaxing the tensions. We examine the effect of likelihood non-Gaussianity for HOD parameter constraints in more detail in Hahn et al. (in preparation).

Even for the P_ℓ analysis, the impact of likelihood non-Gaussianity on the parameter constraints cannot be easily dismissed as we demand increasingly more precise constraints from future experiments. Using the pseudo-likelihood biases the $f\sigma_8$ constraints by $\sim$0.5 per cent. Meanwhile, the Dark Energy Spectroscopic Instru-

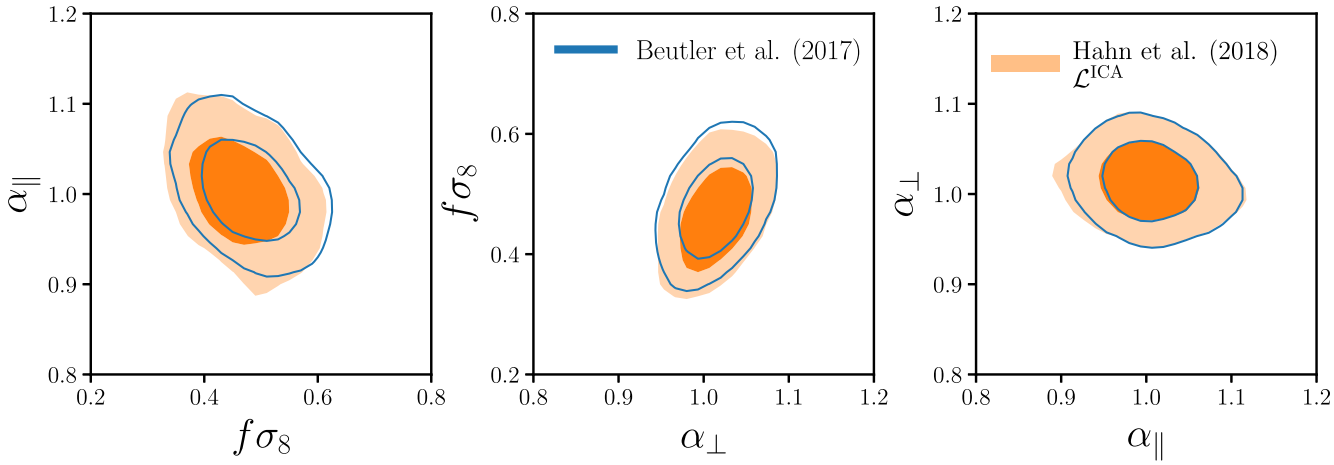

Figure 5. Joint posterior distributions of $f\sigma_8$, $\alpha_\parallel$, and $\alpha_\perp$ in the B2017 P_ℓ analysis, computed using the non-Gaussian ICA likelihood (orange). We include, in blue, the original B2017 posteriors for comparison. The contours in the left-hand and middle panels reflect the shift in $f\sigma_8$ caused by likelihood non-Gaussianity. Otherwise, the contours illustrate that likelihood non-Gaussianity has little impact on the cosmological parameters for the P_ℓ analysis.

ment (DESI; Levi et al. 2013), for instance, seeks to constrain $f\sigma_8$ to within a per cent.[2] The future, however, may be encouraging in this regard. The next surveys will expand the cosmic volumes probed by galaxies and therefore increase the number of modes on all scales. Even as they seek to extend the k range of analyses, thanks to the central limit theorem, we expect likelihood non-Gaussianity to have a smaller effect. However, without precisely quantifying the impact, as we have done in this paper, it remains to be determined whether likelihood non-Gaussianity will significantly impact future P_ℓ analyses.

Constraints on primordial non-Gaussianity ($f_{\rm NL}$) from LSS (e.g. Dalal et al. 2008; Slosar et al. 2008; Ross et al. 2013; Giannantonio et al. 2014) will likely be significantly impacted by likelihood non-Gaussianity. In fact, the constraining power for $f_{\rm NL}$ comes from the largest scales. These are the scales that we find contribute most to the likelihood non-Gaussianity in Section 3 due to the low signal-to-noise and failure to satisfy the central limit theorem. Future experiments such as Euclid (Amendola et al. 2018), which seek to measure $\sigma(f_{\rm NL}) < 5$ (Giannantonio et al. 2012; Amendola et al. 2018), will need to robustly account for likelihood non-Gaussianity for accurate parameter constraints. Fortunately, the methods we present in this paper can easily be extended to other observables and analyses.

For higher order statistics, likelihood non-Gaussianity will also have a more significant effect. Scoccimarro (2000) found that the reduced bispectrum likelihood is significantly more non-Gaussian than the power spectrum likelihood. However, these higher order statistics, and observables from future surveys in general, will also have the added challenge of higher dimensional data. Even for the B2017 P_ℓ analysis, we found significant bias in the KL divergence from sampling the 37-dimensional likelihood distribution with only 2048 samples (Section 3). In current bispectrum analyses, which exclude a significant number of triangle configurations, data vectors exceed >700 dimensions (e.g. Gil-Marín et al. 2017). Accurately estimating such high-dimensional distributions and their divergences will surely require more than 2048 samples. Fortunately a number of methods have been presented in the literature for optimal massive data compression (Tegmark, Taylor & Heavens 1997; Alsing & Wandelt 2018; Heavens et al. 2017; Charnock, Lavaux & Wandelt 2018). Some massive data compression methods have already been utilized to reduce the dimensionality of data-space for likelihood-free inference (Papamakarios & Murray 2016; Alsing, Wandelt & Feeney 2018). In a similar fashion, massive data compression can be combined with the methods we present in this paper to robustly account for likelihood non-Gaussianity in analysing high-dimensional data from future surveys.

6 SUMMARY AND DISCUSSION

Current LSS analyses make a major assumption in their parameter inference – the likelihood has a Gaussian functional form. Although this assumption is motivated by the central limit theorem, in detail the assumption cannot be true. In this paper, we investigate the impact of this Gaussian likelihood assumption on two recent LSS analyses: the B2017 power spectrum multipole ($\ell = 0$, 2, and 4) analysis and the S2017 group multiplicity function analysis. Using mock catalogues, originally constructed for covariance matrix estimation in these analyses, and non-parametric divergence estimators, used in machine learning, we measure the divergences between the P_ℓ and ζ likelihoods and the Gaussian pseudo-likelihoods from B2017 and S2017. For both the P_ℓ and ζ likelihoods, the divergences reveal significant likelihood non-Gaussianity. For the P_ℓ likelihood, large scales (low k) and the hexadecapole contribute most to the relatively small non-Gaussianity. For the ζ likelihood, the high-richness end of ζ contributes most to the non-Gaussianity. In both likelihoods, we find that the low-signal-to-noise regime contributes the most to the likelihood non-Gaussianity.

From the same mock catalogues of B2017 and S2017, we estimate the true non-Gaussian P_ℓ and ζ likelihoods with two different non-parametric density estimates – Gaussian mixture density and independent component analysis. For the ζ likelihood, we find more accurate estimates of the likelihood with the Gaussian mixture density method. For the B2017 P_ℓ analysis, which has fewer mocks and a higher dimensional likelihood, we use independent component analysis to transform the likelihood distribution into statistically independent components. By estimating the one-dimensional distribution of these independent components, we derive an estimate of the high-dimensional likelihood distribution for the B2017 P_ℓ

[2]DESI Final Design Report: http://desi.lbl.gov/wp-content/uploads/2014/04/fdr-science-biblatex.pdf

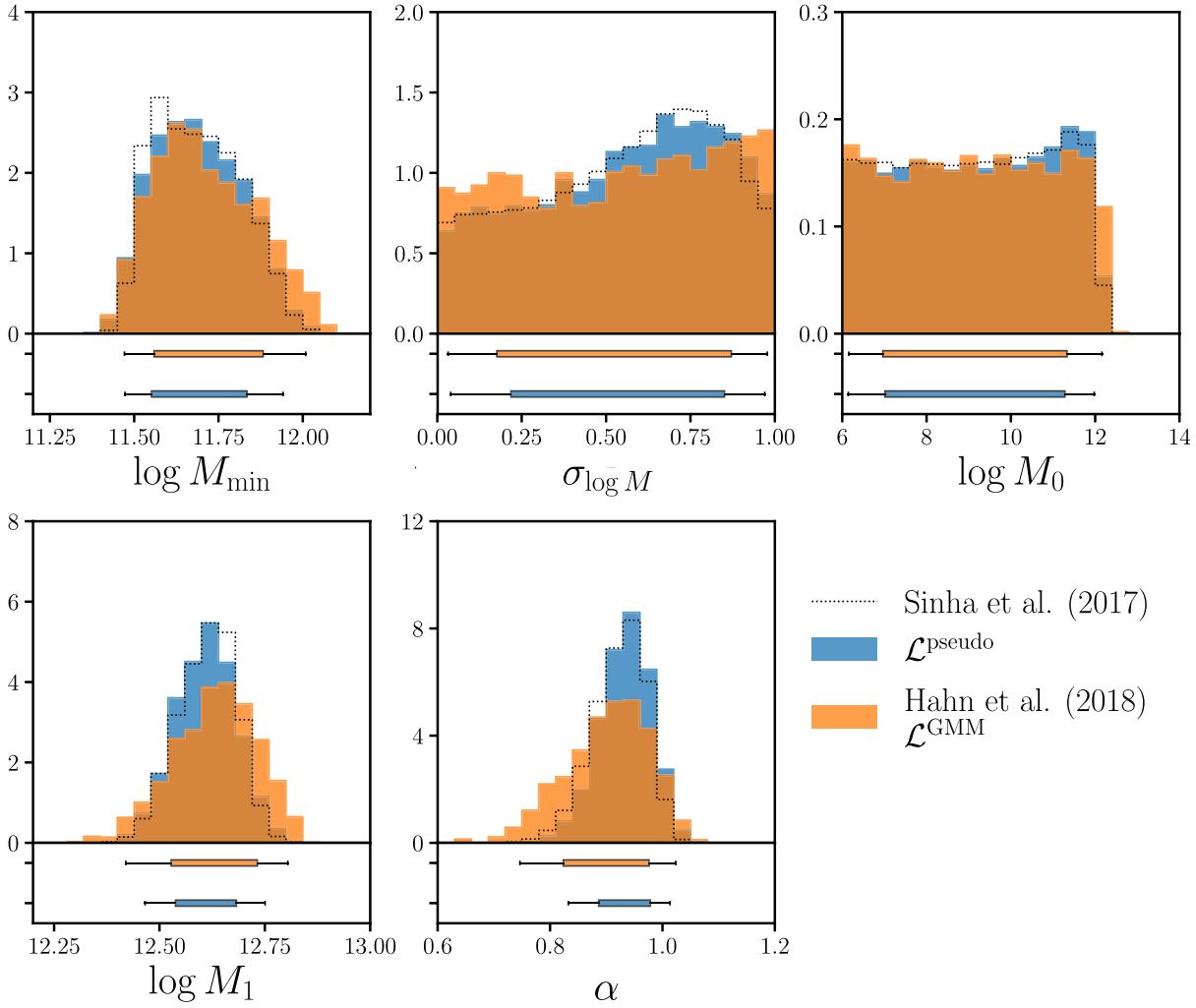

Figure 6. The posterior distribution for HOD parameters log $M_{\rm min}$, $\sigma_{\log M}$, log M_0, log M_1, and α in the S2017 ζ analysis using the non-Gaussian GMM likelihood (orange). We include in blue the posteriors calculated from the pseudo-likelihood for comparison. We also include the original S2017 posterior (dotted; see text for details). At the bottom of each panel we include box plots that mark the 68 per cent and 95 per cent confidence intervals of the posterior. Besides the poorly constrained parameters $\sigma_{\log M}$ and log M_0, the posteriors of log $M_{\rm min}$, log M_1, and α are significantly broader and shifted compared to the pseudo-likelihood constraints. Likelihood non-Gaussianity significantly impacts the parameter constraints of the *zeta* analysis. Therefore, using the pseudo-likelihood underestimates the uncertainty and biases the HOD parameter constraints.

analysis. The divergence between our two likelihood estimates and the P_ℓ and ζ likelihoods demonstrate that we derive *more accurate* estimates of the true likelihoods than the assumed Gaussian pseudo-likelihoods.

Finally, with these better estimates for the non-Gaussian P_ℓ and ζ likelihoods and importance sampling, we calculate more accurate posterior parameter constraints for the B2017 and S2017 analyses. By comparing our posteriors to the parameter constraints from B2017 and S2017, we find that likelihood non-Gaussianity significantly impacts both analyses. Among the non-nuisance parameters in the P_ℓ analysis of B2017, accounting for likelihood non-Gaussianity shifts $f\sigma_8$ constraints by -0.44σ. Meanwhile for the S2017 ζ analysis, likelihood non-Gaussianity significantly impacts the posterior distributions of the HOD parameter. Using the pseudo-likelihood significantly underestimates the width of the log $M_{\rm min}$, log M_1, and α posteriors and significantly biases the S2017 constraints. For log M_1 and α, the posteriors are broadened by 42 per cent and 66 per cent and shifted by $+0.43\sigma$ and -0.51σ, respectively. Accounting for likelihood non-Gaussianity likely eases the tension between the ζ and $w_{\rm p}(r_{\rm p})$ constraints found in S2017. Our comparisons of the posteriors highlight the importance of incorporating likelihood non-Gaussianity in parameter inference of LSS studies.

Based on our results, it is unclear whether future P_ℓ analyses will be significantly impacted by likelihood non-Gaussianity. Future surveys (e.g. DESI, Euclid, WFIRST) will expand the cosmic volumes probed by galaxies and therefore increase the number of modes included in P_ℓ analyses on all scales. Over the same k range, this will reduce likelihood non-Gaussianity due to the central limit theorem and therefore reduce the impact on parameter constraints. However, future analyses seek to extend analyses to both higher and lower k, which will introduce likelihood non-Gaussianity from these scales. Meanwhile, for ζ analyses with the same multiplicity range, we expect future surveys to reduce the impact of likelihood non-Gaussianity, since larger cosmic volumes will probe more high-multiplicity groups. For a wider multiplicity range, however,

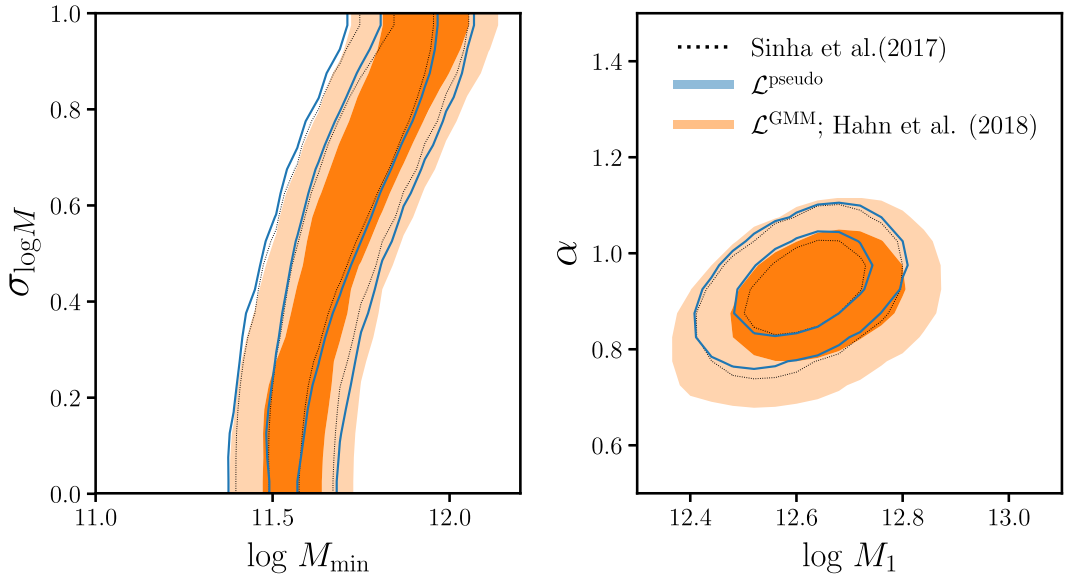

Figure 7. Joint posterior distributions of select HOD parameters in the S2017 ζ analysis, computed using the non-Gaussian GMM likelihood (orange). We include, in blue, the posteriors computed using the pseudo-likelihood; we also include the original S2017 posterior (dotted; see text for details). The contours confirm that that *due to likelihood non-Gaussianity, posteriors from the pseudo-likelihood underestimate the uncertainties and significantly bias the parameter constraints of the S2017 analysis*.

likelihood non-Gaussianity may still be a significant effect. For higher order statistics such as the galaxy bispectrum or three-point function, even for future surveys, likelihood non-Gaussianity will likely be a significant effect to consider for parameter inference. We also expect it to significantly impact primordial non-Gaussianity ($f_{\rm NL}$) constraints from LSS, which derive most of their constraining power from the largest, most non-Gaussian, scales. Regardless of our expectation, for more accurate parameter inference the Gaussian likelihood assumption must be extensively tested. The divergence and likelihood estimations we introduce in this paper provide a straightforward framework for testing and quantifying the impact of likelihood non-Gaussianity on the final parameter constraints.

Our likelihood estimation methods also allow us to go beyond the pseudo-likelihood and derive more accurate estimates of the likelihood. With a similar motivation at addressing likelihoods that are non-Gaussian or difficult to write down, methods for likelihood-free inference such as approximate Bayesian computation (ABC; Hahn et al. 2017b; Kacprzak et al. 2018) have recently been introduced to LSS studies. Although as a likelihood-free inference method ABC has the advantage of relaxing *any* assumption on the likelihood, even with smart sampling methods like population Monte Carlo, it requires an expensive generative forward model to be computed far more times than the number of mocks required for covariance matrix estimation. Our methods (especially the ICA method) do not require any more mocks than those already constructed for accurate covariance matrix estimation. For future analyses that will analyse even higher dimensional data, our method can easily be combined with optimal massive data compression methods (e.g. Heavens et al. 2017; Alsing et al. 2018). Therefore, the methods for likelihood estimation we present in this paper provide both accurate and practical methods for Bayesian parameter inference in LSS.

ACKNOWLEDGEMENTS

It's a pleasure to thank Emanuele Castorina, Yu Feng, Simone Ferraro, Daniel Foreman-Mackey, Emmanuel Schaan, Roman Scoccimarro, UrošŽ Seljak, Sukhdeep Singh, Michael Wilson, and Martin White for valuable discussions. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under contract No. DE-AC02-05CH11231. This project used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Parts of this research were conducted by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. This project also made use of the NASA Astrophysics Data System and open-source software PYTHON, NUMPY, SCIPY, MATPLOTLIB, and SCIKIT-LEARN.

REFERENCES

Ade P. A. R. et al., 2014, A&A, 571, A15
Ade P. A. R. et al., 2016, A&A, 594, A24
Aghanim N. et al., 2016, A&A, 594, A11
Alam S. et al., 2017, MNRAS, 470, 2617
Alsing J., Wandelt B., 2018, MNRASL, 476, L60
Alsing J., Wandelt B., Feeney S., 2018, MNRAS, 477, 2874
Amendola L. et al., 2018, Living Rev. Relativ., 21, 2
Arthur D., Vassilvitskii S., 2007, Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '07. Society for Industrial and Applied Mathematics, Philadelphia, PA, p. 1027
Berlind A. A. et al., 2006, ApJS, 167, 1
Beutler F. et al., 2017, MNRAS, 466, 2242
Bianchi D., Gil-Marín H., Ruggeri R., Percival W. J., 2015, MNRAS, 453, L11
Bovy J., Hogg D. W., Roweis S. T., 2011, Ann. Appl. Stat., 5, 1657
Broderick A. E., Fish V. L., Doeleman S. S., Loeb A., 2011, ApJ, 735, 110
Cash W., 1979, ApJ, 228, 939
Charnock T., Lavaux G., Wandelt B. D., 2018, Phys. Rev. D, 97, 083004
Collaboration P. et al., 2014, A&A, 571, A20
Comon P., 1994, Signal Process., 36, 287
Crocce M., Pueblas S., Scoccimarro R., 2006, MNRAS, 373, 369

Dalal N., Doré O., Huterer D., Shirokov A., 2008, Phys. Rev. D, 77, 123514
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Davison A. C., 2008, Statistical Models (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge Univ. Press, Cambridge
Dempster A. P., Laird N. M., Rubin D. B., 1977, JRSSB , 39, 1
Efstathiou G., 2004, MNRAS, 349, 603
Efstathiou G., 2006, MNRAS, 370, 343
Eifler T., Schneider P., Hartlap J., 2009, A&A, 502, 721
Eisenstein D. J., Zaldarriaga M., 2001, ApJ, 546, 2
Feigelson E. D., Babu G. J., 2012, Modern Statistical Methods for Astronomy, Cambridge University Press.
Fraley C., Raftery A. E., 1998, Comput. J., 41, 578
Gardner J. P., Connolly A., McBride C., 2007, in ASP Conf. Ser. Vol. 376, Astronomical Data Analysis Software and Systems XVI. Astron. Soc. Pac., San Francisco, p. 69
Gaztañaga E., Scoccimarro R., 2005, MNRAS, 361, 824
Giannantonio T., Porciani C., Carron J., Amara A., Pillepich A., 2012, MNRAS, 422, 2854
Giannantonio T., Ross A. J., Percival W. J., Crittenden R., Bacher D., Kilbinger M., Nichol R., Weller J., 2014, Phys. Rev. D, 89, 23511
Gil-Marín H., Percival W. J., Verde L., Brownstein J. R., Chuang C.-H., Kitaura F.-S., Rodríguez-Torres S. A., Olmstead M. D., 2017, MNRAS, 465, 1757
Grieb J. N. et al., 2017, MNRAS, 467, 2085
Guo H., Zehavi I., Zheng Z., 2012, ApJ, 756, 127
Hahn C., Scoccimarro R., Blanton M. R., Tinker J. L., Rodríguez-Torres S. A., 2017a, MNRAS, 467, 1940
Hahn C. et al., 2017b, MNRAS, 469, 2791
Hand N. et al., 2017a, AJ, 156, 160
Hand N., Li Y., Slepian Z., Seljak U., 2017b, J. Cosmol. Astropart. Phys., 07, 002
Hartlap J., Schrabback T., Simon P., Schneider P., 2009, A&A, 504, 689
Hastie T., Tibshirani R., Friedman J., 2009, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics). Springer, New York City,
Heavens A. F., Sellentin E., de Mijolla D., Vianello A., 2017, MNRAS, 472, 4244
Hérault J., Ans B., 1984, Comptes Rendus de l'Académie des Sciences Paris, Série III, Life Sciences, 299, 525
Hogg D. W., Bovy J., Lang D., 2010, preprint (arXiv:1008.4686)
Hu W., White M., 2001, ApJ, 554, 67
Hyvärinen A., 1998, in Jordan M. I., Kearns M. J., Solla S. A., eds, Advances in Neural Information Processing Systems 10. MIT Press, Cambridge MA, p. 273
Hyvarinen A., 1999, IEEE Trans. Neural Netw., 10, 626
Hyvarinen A., 2001, Independent Component Analysis. J. Wiley, New York
Hyvärinen A., Oja E., 2000, Neural Netw., 13, 411
Kacprzak T., Herbel J., Amara A., Réfrégier A., 2018, J. Cosmol. Astropart. Phys., 2018, 42
Kazin E. A. et al., 2014, MNRAS, 441, 3524
Kitaura F.-S., Heß S., 2013, MNRAS, 435, L78
Kitaura F.-S., Yepes G., Prada F., 2014, MNRAS, 439, L21
Kitaura F.-S., Gil-Marín H., Scóccola C. G., Chuang C.-H., Müller V., Yepes G., Prada F., 2015, MNRAS, 450, 1836
Kitaura F.-S. et al., 2016, MNRAS, 456, 4156
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Kraskov A., Stögbauer H., Grassberger P., 2004, Phys. Rev. E, 69, 066138
Krishnamurthy A., Kandasamy K., Poczos B., Wasserman L., 2014, preprint (arXiv:1402.2966)
Kuhn M. A., Feigelson E. D., 2017, preprint (arXiv:1711.11101)
Lee K. J., Guillemot L., Yue Y. L., Kramer M., Champion D. J., 2012, MNRAS, 424, 2832
Leroux B. G., 1992, Ann. Stat., 20, 1350
Levi M. et al., 2013, preprint (arXiv:1308.0847)
Liddle A. R., 2007, MNRAS, 377, L74
Lloyd S., 1982, IEEE Trans. Inf. Theory, 28, 129
McBride C. et al., 2009, BAAS, 213, 425.06
McLachlan G., Peel D., 2000, Finite Mixture Models. Wiley-Interscience, Hoboken New Jersey
Mohammed I., Seljak U., Vlah Z., 2017, MNRAS, 466, 780
Morrison C. B., Schneider M. D., 2013, J. Cosmol. Astropart. Phys., 11, 009
Neal R. M., Hinton G. E., 1998, in Learning in Graphical Models, NATO ASI Series. Springer, Dordrecht, p. 355
Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, MNRAS, 396, 19
Ntampaka M., Trac H., Sutherland D. J., Battaglia N., Póczos B., Schneider J., 2015, ApJ, 803, 50
Ntampaka M., Trac H., Sutherland D. J., Fromenteau S., Póczos B., Schneider J., 2016, ApJ, 831, 135
O'Connell R., Eisenstein D., Vargas M., Ho S., Padmanabhan N., 2016, MNRAS, 462, 2681
Papamakarios G., Murray I., 2016, preprint (arXiv:1605.06376)
Parkinson D. et al., 2012, Phys. Rev. D, 86, 103518
Pinol L., Cahn R. N., Hand N., Seljak U., White M., 2017, J. Cosmol. Astropart. Phys., 008
Póczos B., Szabó Z., Schneider J., 2011, 2011 19th European Signal Processing Conference. p. 1718
Póczos B., Xiong L., Schneider J., 2012a, preprint (arXiv:1202.3758)
Póczos B., Xiong L., Sutherland D. J., Schneider J., 2012b, 2012 IEEE Conference on Computer Vision and Pattern Recognition. p. 2989
Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical Recipes in C, 2nd edn. The Art of Scientific Computing. Cambridge Univ. Press, New York
Ravanbakhsh S., Lanusse F., Mandelbaum R., Schneider J., Poczos B., 2017, Thirty-First AAAI Conference on Artificial Intelligence.
Rodríguez-Torres S. A. et al., 2016, MNRAS, 460, 1173
Roeder K., Wasserman L., 1997, J. Am. Stat. Assoc., 92, 894
Ross A. J. et al., 2013, MNRAS, 428, 1116
Ross A. J. et al., 2017, MNRAS, 464, 1168
Schwarz G., 1978, Ann. Stat., 6, 461
Scoccimarro R., 1998, MNRAS, 299, 1097
Scoccimarro R., 2000, ApJ, 544, 597
Scoccimarro R., 2015, Phys. Rev. D, 92, 083532
Scoccimarro R., Couchman H. M. P., Frieman J. A., 1999, ApJ, 517, 531
Scott D. W., 1992, Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley, Hoboken, New Jersey
Sellentin E., Heavens A. F., 2016, MNRAS, 456, L132
Sellentin E., Heavens A. F., 2018, MNRAS, 473, 2355
Sinha M., Berlind A. A., McBride C. K., Scoccimarro R., Piscionere J. A., Wibking B. D., 2018, MNRAS, 478, 1042
Slepian Z. et al., 2017, MNRAS, 469, 1738
Slosar A., Hirata C., Seljak U., Ho S., Padmanabhan N., 2008, J. Cosmol. Astropart. Phys., 2008, 031
Springel V., 2005, MNRAS, 364, 1105
Steele R. J., Raftery A. E., 2010, Frontiers of Statistical Decision Making and Bayesian Analysis, Springer, New York, 113
Sutherland D. J., Xiong L., Póczos B., Schneider J., 2012, preprint (arXiv:1202.0302)
Taylor E. N. et al., 2015, MNRAS, 446, 2144
Tegmark M., Taylor A. N., Heavens A. F., 1997, ApJ, 480, 22
Vakili M., Hahn C. H., 2016, preprint (arXiv:1610.01991)
Vargas-Magaña M. et al., 2014, MNRAS, 445, 2
Wang Q., Sanjeev K., Sergio V., 2009, IEEE Trans. Inf. Theory, 55, 2392
Wasserman L., 2004, All of Statistics: A Concise Course in Statistical Inference (Springer Texts in Statistics). Springer, New York
White M., Padmanabhan N., 2015, J. Cosmol. Astropart. Phys., 12, 058
Wilkinson D. M. et al., 2015, MNRAS, 449, 328
Wu C. F. J., 1983, Ann. Stat., 11, 95
Xu X., Ho S., Trac H., Schneider J., Poczos B., Ntampaka M., 2013, ApJ, 772, 147
Zhao C., Kitaura F.-S., Chuang C.-H., Prada F., Yepes G., Tao C., 2015, MNRAS, 451, 4266
Zheng Z., Weinberg D. H., 2007, ApJ, 659, 1

This paper has been typeset from a TEX/LATEX file prepared by the author.