Likelihood-Free Inference for Cosmic Shear with LSST

Miles Cranmer

Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA

In this report I review weak lensing and cosmic shear, with a focus on the ability of LSST to detect it and use cosmic shear correlation maps for measuring the dark energy equation of state. Highlighting the many systematics involved in these measurements, I then delve into a review of Approximate Bayesian Computation and Likelihood-Free Inference, and its prospect for increasing the figures of merit for LSST. I focus on the study by Alsing et al. (2019) and their demonstration of neural likelihood estimators as efficient estimators of cosmological parameters from Euclid tomographic shear maps. This type of algorithm is a very straightforward but immensely powerful technique for addressing all types of systematic errors in surveys where one possesses accurate forward simulators of observed data. True likelihoods for data as a function of astrophysical parameters are impossible to find in astrophysics, so accurate Bayesian inference must use these approximate but principled methods, and likelihood-free inference with neural estimators seems to be an optimal strategy.

I. Science Case

Light bends in a gravitational potential, so we can use the warping of background images assumed to have some standard shape to study the mass distribution in the Universe. Shearing is the anisotropic stretching of an image and is caused by an anisotropic mass distribution. Averaged on large scales, this is called "Cosmic Shear". Here I derive cosmic shear and the physics that it probes.

A. Cosmic Shear

In the following science sections, I mostly follow [3]. I assume a flat universe for simplicity ($K \equiv 0$), and based on current measurements. The core premise of cosmic shear is that, in General Relativity (GR), the path of light bends in a gravitational potential according to the distance metric (we work with $c \equiv 1$)

$$ds^2 = (1 + 2\Psi)dt^2 - a^2(t)(1 - 2\Phi)dl^2, \tag{1}$$

$$\text{where } ds^2 = 0 \text{ for light,} \tag{2}$$

$$\text{and } dl^2 = d\chi^2 + \chi^2 d\Omega^2. \tag{3}$$

Here, Ψ and Φ are approximations of the gravitational field called the Bardeen gravitational potentials (equal to each other in GR, but different in modified gravity theories), and dl^2 is the spatial line element.

Viewing a background source in front of a massive foreground body causes a distortion in the image of the background source, with the mass acting a lens. The amount of distortion and the direction can be calculated in terms of the size, proximity, and shape of the mass to the line of sight. Thus, if one has a strong prior on the shape of the background source, one can solve the inverse problem to make inferences about the parameters of the foreground source. For single sources, this is known as strong lensing. However, we are interested in large-scale distortions called weak lensing. Calculating distortions can be done in a statistical scenario on large-scale observations, in which one only has a prior on statistical properties of the background source, to calculate a posterior on statistical properties of the foreground source (i.e., Large-Scale Structure properties). This can be used to make inferences on cosmological parameters using the predicted evolution of the power spectrum.

Going back to eq. (1), we can evolve $a(t)$ with the Friedmann equations in the presence of various fluids whose relative densities we can fit for in the current epoch $a = 1$. From this, we can estimate fluctuations in density which give rise to the changing gravitational potential, and numerically solve for the change of the density contrast, δ (a field), over time. Of interest is the Fourier transform of this field $\tilde{\delta}(k)$, and the transfer function $T(k)$, which describes the growth of the density contrast at each scale k from initial conditions relative to the growth for $k = 0$.

The density contrast of an ideal fluid with zero pressure is related to the Laplacian of the gravitational potential via

$$\nabla^2 \Phi = 4\pi G a^2 \langle \rho \rangle \delta.$$

This, which in Fourier space is $k^2 \tilde{\Phi} = 4\pi G a^2 \langle \rho \rangle \tilde{\delta}$, can be numerically solved in terms of the evolution of the scale factor $a(t)$, or perturbatively solved analytically to first order. We can also take the metric, and get the geodesic equation as $dt = a\sqrt{\frac{1-2\Phi}{1+2\Psi}}dl$, which, for GR and small Φ, is just $dt = a(1-2\Phi)dl$ to first order. This can give us the path of the light in terms of a potential Φ. As in [3], we can find the deflection angle to be

$$\hat{\alpha} = -2 \int dl \nabla^p_\perp \Phi,$$

$$\text{or simply } d\hat{\alpha} = -2(\nabla^p_\perp \Phi) dl,$$

where we take the gradient perpendicular to the path of light. This will deflect the path of light through the Universe, both strongly for massive foreground masses, and weakly on a cosmological scale due to density correlations.

We can integrate this equation for two converging light rays to derive the difference between the observed angle between the light rays, θ, and the unperturbed angle β, as

$$\alpha = \beta - \theta = 2 \int_0^\chi d\chi' \left(1 - \frac{\chi'}{\chi}\right) \left(\nabla^{C_1}_\perp \Phi(\chi') - \nabla^{C_2}_\perp \Phi(\chi')\right),$$

where C_i indicates the path taken by ray i starting at χ, and being perturbed at distance χ' (see fig. 1).

All angles are assumed to be small. Thus, the difference between observed and unperturbed angle between two light rays is a measure of the integrated difference between the gravitational force field. We can further calculate $\frac{\partial \beta_i}{\partial \theta_j}$ as being equal to

$$\frac{\partial \beta_i}{\partial \theta_j} = I_{ij} - \partial_{\theta_i} \partial_{\theta_j} \psi \equiv A_{ij},$$

$$\text{for } \psi = 2 \int_0^\chi d\chi' \frac{\chi - \chi'}{\chi \chi'} \Phi(\chi' \theta, \chi'),$$

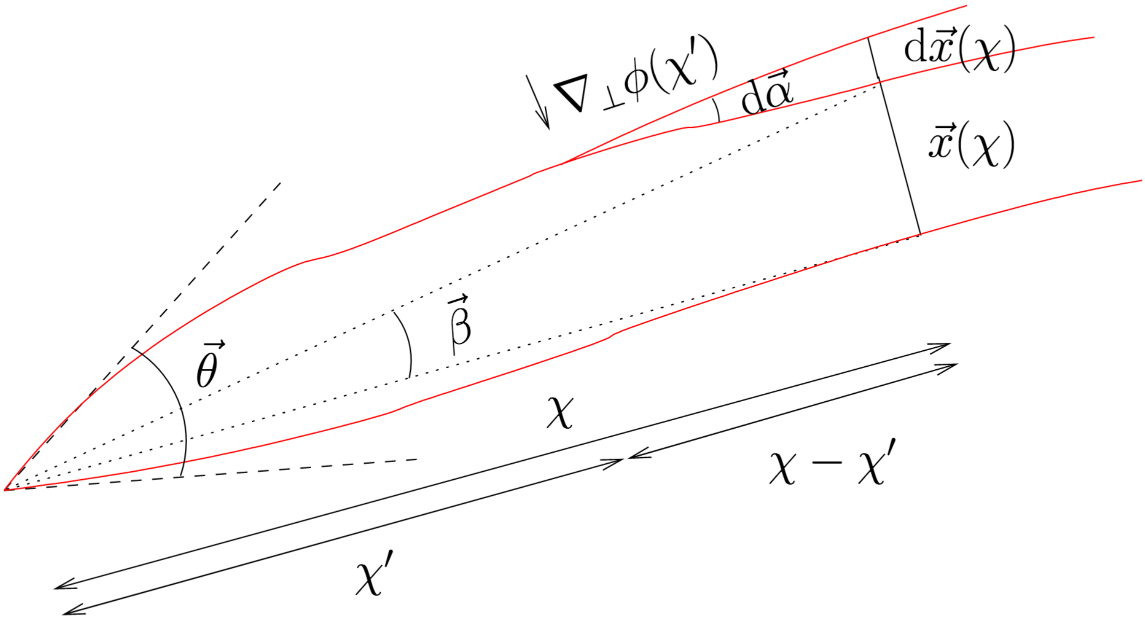

FIG. 1: The bending of background light caused by a foreground mass at distance χ'. From [3].

where A is a symmetric matrix with $A_{12} = -\gamma_2, A_{11} = 1 - \kappa - \gamma_1$, and $A_{22} = 1 - \kappa + \gamma_1$, where κ is the "convergence" and $\gamma = (\gamma_1, \gamma_2)$ is the "shear." As $\mathrm{Tr}(A)$ measures the isotropic component, we see κ measures how many light rays converge uniformly, and γ, the remaining component, measures anisotropic stretching.

B. Measurement of Cosmic Shear

In this report, we are interested in γ. To actually measure cosmic shear, we look at the shapes of galaxies. We can write down an observable, the "reduced shear," as

$$g = \frac{\gamma}{1-\kappa},$$

In the weak lensing regime, the observed ellipticity of a galaxy, $\epsilon = \frac{a-b}{a+b} e^{2\phi i}$, for ϕ its position angle and a, b the major and minor axis, is approximately $\epsilon \approx \epsilon' + \gamma$, where ϵ' is the ellipticity at the source.

Since our universe is isotropic and if we assume there is no preferred direction for galaxies we observe, then we should have that the expectation $\langle \epsilon' \rangle = 0$, so $\langle \epsilon \rangle = g$. Estimating the ellipticity over many galaxies can tell us the cosmic shear in the weak-lensing regime.

To actually make a measurement, we work on a distribution of light parametrized by F as a function of line-of-sight θ and photometric band g. The most common methods here assume some analytic or semi-analytic model for the galaxy, and fit with Bayesian inference through the PSF. From this, we combine measurements for redshift, and construct a tomographic map which we can then do analysis.

1. Biases

However, making these measurements of galaxy shapes is extremely difficult, especially in the high-redshift regime where data is most valuable for cosmological inference. As in [5] typical $r-$band magnitudes for these faint galaxies are around 24 and the shape extends over just a few pixels. To make matters worse, a variable PSF is convolved with the observation of the galaxy. Further, one needs to worry about telescope systematics. $F(\theta, g)$ is produced over multiple exposures with variable PSFs. This introduces problems like correlated pixel noise and other systematics from stacking. Outside of measurement difficulty, there are also biases related to the physics of galaxies themselves, making them difficult to model analytically. This is discussed in detail in [6] as well as [1].

II. Observational Facility

A. Survey

Here we refer to values given in [6] throughout. LSST will be a 10-year survey with state-of-the-art photometric and astrometric measurements (3 mas parallax at $r \sim 24$; 0.005 photometric accuracy) for 20 billion objects ($\sim$ 10 billion stars, $\sim$ 10 billion galaxies). LSST will use $\sim$ 90% of its observing time to do a deep-wide-fast survey up to a single visit depth of $r \sim 24.5$. The remaining time will be dedicated to deep ($r \sim 26$), sources requiring high-cadence observations, and galactic regions of interest. The coadded survey depth should reach up to $r \sim 27.5$. The main and deep survey are the most interesting for weak lensing. The main survey will cover $\sim$20,000 deg^2 of sky, with 15-second exposures in a given filter, producing a total of 1000 visits over all bands. There is also the potential for the survey to add "deep drilling fields," whereby a section of the sky is observed to $\sim$ 2 magnitudes deeper. The six bands are *ugrizy* which span 320 to 1050 nm. The effective field of view for the system is 9.6 deg^2. There is 3.2 Gpix total in the camera, with about 0.2 arcsec size per pixel.

For galaxies, about 4 billion galaxies will be measured with full six-band photometry. LSST is particularly powerful in terms of image quality, which is very important for shear measurements. The median seeing requirement of LSST is $0.7''$, which is $\sim$ 4 kpc at $z = 0.5$. Non-parametric Bayesian models will thus become very important at larger redshifts, where they are able to integrate information from many different galaxies. Current non-parametric models include metrics that quantify the distribution of galaxy pixels. Likelihood-free inference will be able to do away with these hand-defined metrics and learn information-maximizing summary statistics based on the raw pixels.

We have given the observables for shear previously. What we would now like to ask is how sensitive is LSST for these observables. We can write down the "shear noise" for a section of phase space as γ_{rms}, which encodes the shape noise on γ. The effective shear noise after incorporating many galaxies is $\gamma_{rms}/\sqrt{N_{\text{eff}}}$ ($N_{\text{eff}} = N$ if galaxies are perfectly measured). For LSST, we will have $N_{\text{eff}} = 40$ arcmin^{-2} using the standard pipeline.

Galaxy-Galaxy lensing (g-g lensing), which is weak lensing around galaxies, is a powerful probe for dark matter, marginalizing some unknown physics to give a strong estimate of the dark matter equation of state. The cross-correlation function we are interested in is:

$$\zeta_{gm}(\mathbf{r}) = \langle \delta_g(0)\delta_m(\mathbf{r}) \rangle \,,$$

(with the δ measuring the overdensities of galaxies) which is averaged over all possible universes. This can be related to shear via the surface density of galaxies at different scales:

$$\langle \Sigma(< R) \rangle - \Sigma(R) = \gamma_t(R)\Sigma_c,$$

where γ_t is the tangential shear, and Σ_c is the critical surface density. This g-g lensing term can be used to remove additive bias from the cosmic shear measurement. Finally, the two-point and three-point correlation functions for shear can also be used as powerful probes of the growth of structure, expansion of the Universe, and dark energy models. This is called "3x2-point" analysis and covers shear-shear, galaxy-shear, and galaxy-galaxy correlations.

B. Systematics

LSST's Dark Energy Science Collaboration defines various "figures of merit," or FoMs, in their paper for characterizing their different missions through detailed analysis. For Dark Energy, the 3x2-point correlation function of shear is targetted for making a measurement of the equation of state parametrized by $\omega(a) = \omega_0 + (1-a)\omega_a$. The Dark Energy Task Force paper defines FoM as the L_2 norm of the Fisher Information matrix for (ω_0, ω_a) marginalized over other parameters, which is a dimensionless number. A larger norm of the Fisher Information matrix essentially places a stronger bound on the width of the posterior over inferred parameters via the Cramer-Rao bound:

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{\mathcal{I}(\hat{\theta})},$$

where $\mathcal{I}(\hat{\theta})$ is the Fisher Information for a parameter $\hat{\theta}$.

Currently, LSST targets an FoM of 40 after year 10, though with reduced systematics, this can be increased to 87. Significant systematics, some of which are listed in fig. 2, include baryonic effects, galaxy bias, magnification of stars, shear/photo-z/detector/image processing systematics. Every single one of these is potentially able to be calibrated via likelihood-free inference. LFI has the potential to improve (increase the size of) the FoM from weak lensing measurements up from 40.

III. Methods

A. Background

Most probabilistic models that one might be used to are invertible. The normal distribution, also called Gaussian distribution, is the most commonly used in astronomy and physics. One can go forward, and sample from the model to generate data, or go backward, and calculate the explicit probability of data being drawn from that particular model. In this case, the model is invertible: we have the **likelihood**. For example, the probability of a point $x \in \mathbb{R}$ being drawn from a Gaussian

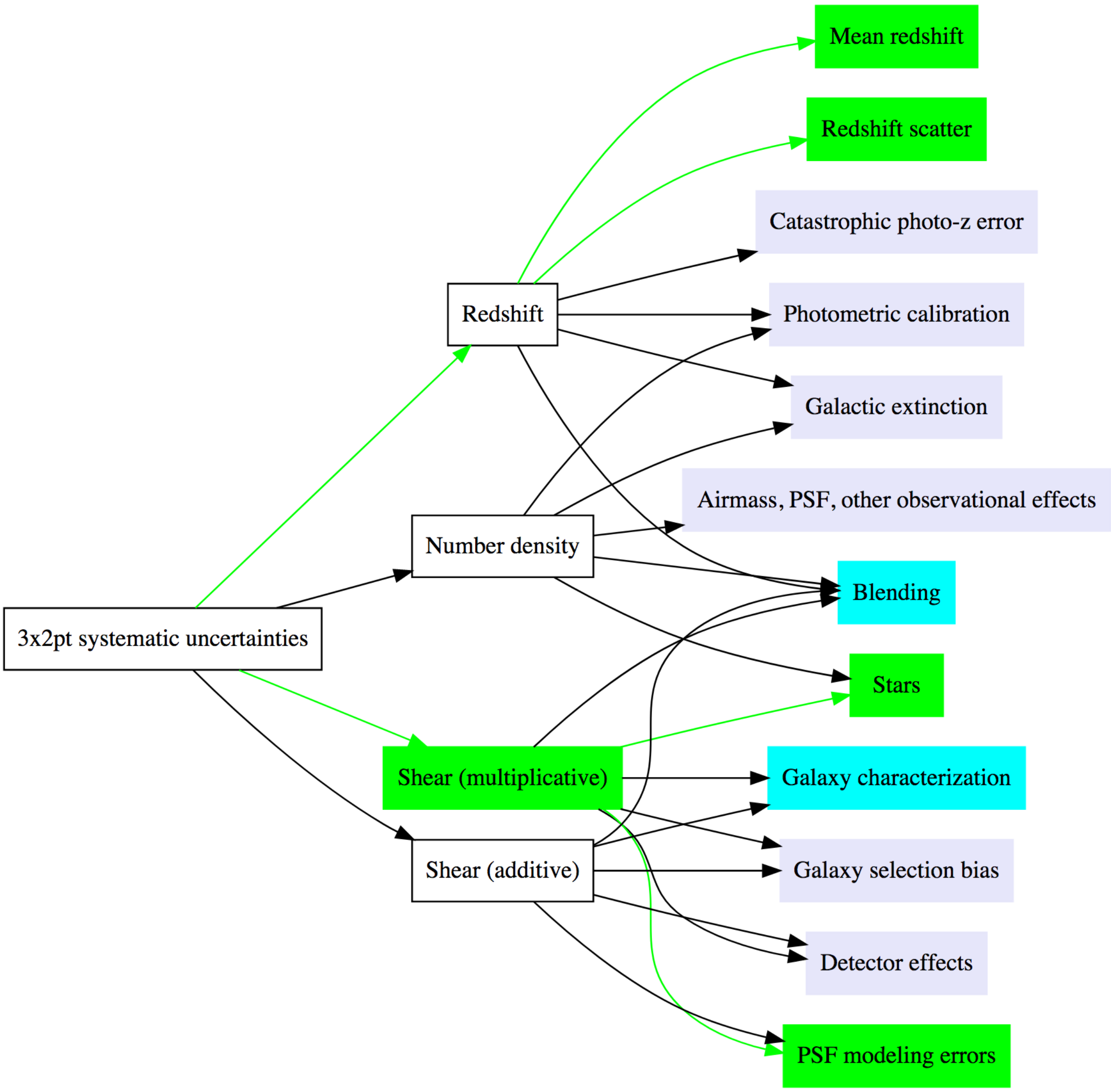

FIG. 2: Various systematic uncertainties (right-hand side) in the weak lensing analysis of LSST, taken from [5] in Figure D3. Likelihood-free inference has the potential to address and reduce the impact of all of these, via different forward simulations and learned likelihood algorithms.

with 0 mean and 1 variance is $\frac{1}{\sqrt{2\pi}}\exp\left(-x^2/2\right)$. This is the Gaussian likelihood. We can also sample from a Gaussian by calculating its cumulative distribution function $\left(\left(1+\mathrm{Erf}(x/\sqrt{2})\right)/2\right)$, inverting it (a convergent Maclaurin series), and passing uniformly random $x \in [0,1]$ values through.

However, say that I have a galaxy simulator followed by a simulator of image shearing followed by a simulator of my telescope. I can forward simulate the evolution of galaxies given some messy physics, initial conditions, and cosmological parameters, pass the image through a shearing, and then through my telescope model. What about the inverse? Even if I can forward model the systematics and noise exactly, how do I invert them? If I have an image of a Galaxy, what is the likelihood of it being found in my cosmological simulation? This is infinitesimally small, of course, but how can I tell if one image of a galaxy is more likely than another? This is where we turn to Approximate Bayesian Computation (ABC), as nicely reviewed in [4].

1. Parameter Estimation and Bayesian Inference

In parameter estimation, one is typically interested in estimating parameters θ given data D and a model M. For example, the parameter might be H_0, the data might be a tomographic map of cosmic shear, and the model might be ΛCDM.

From probability theory, we can write down the probability of an event A occurring (e.g., some random variable X being equal to a constant c) given an event B has occurred as: $P(A|B) = \frac{P(A,B)}{P(B)}$. This is something you can derive from a Venn diagram. We can also write down $P(B|A) = \frac{P(A,B)}{P(A)}$. Equating the joint probability $P(A,B)$ gives us Bayes theorem, $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$. Here, we commonly refer to $P(A), P(B)$ as the "prior" over A and B, respectively: a distribution that contains our preconceived notions about the events. $P(A|B)$ is the posterior—the distribution we would like to extract, and $P(B|A)$ is the "likelihood."

When we rewrite this in terms of parameter estimation, we write $P(\theta|D,M) = \frac{P(D|\theta,M)P(\theta)}{P(D)}$, in which case we are trying to find a posterior over our parameters given the data and a model that generates the data. Here, $P(D|\theta,M)$ is the likelihood of the data given the model and the model parameters. Often $P(D)$ isn't meaningful as we are interested in comparing parameter choices with each other, so we write $P(\theta|D,M) \propto P(D|\theta,M)P(\theta)$. Therefore, for example, if we make a guess for H_0, knowing $P(D|\theta,M)$ would allow us to evaluate the probability of observing a particular tomographic map in the context of a ΛCDM model. We can evaluate $P(\theta|D,M)$ for many different θ to fill out the distribution (Markov Chain Monte Carlo), and make statements about the expected θ value and uncertainties in that estimate.

There is a significant challenge, however. What exactly is $P(D|\theta,M)$ for such a complex model as a cosmological simulator? It is certainly not a Gaussian in terms of θ! What we do have is an extremely non-linear simulator in terms of random initial conditions and a set of cosmological parameters that can generate potential data D. The problem is that D, being a tomographic map, is extremely high-dimensional; even with all the computing power in the world, one would likely never generate a sample, $D' \sim M(\theta)$, exactly equal to the observed D. So we would be unable to estimate $P(D|\theta,M)$. Most often, users of data spend significant time designing a parametric inverse here, but this introduces additional systematics from innaccurate models. Thus, in the absence of a tractable likelihood, we should turn to Approximate Bayesian Computation (ABC).

2. Approximate Bayesian Computation

The premise of the simplest ABC model is that rather than sampling the model $M(\theta)$ until several D' have been generated that are exactly equal to D so we know the fraction, instead we use a softer threshold. Instead of equality, we define some distance threshold ϵ and distance metric d so that we accept a sample D'_i as "being equal" to D if $d(D'_i, D) \leq \epsilon$.

The simplest version of this algorithm in a Monte Carlo setting is the following rejection sampling algorithm. Sample θ from some prior $P(\theta)$. Generate $D' \sim M(\theta)$, and evaluate whether $d(D', D) \leq \epsilon$. If it is, record that θ as an approximate sample of $P(\theta|D,M)$. In the limit $\epsilon \to 0$, and $d = L_2$ (L_2 norm of the residual of data in the observed $\mathbb{R}^n$ space), this approximate posterior will approach the true posterior. However, realistically we will need to choose $\epsilon > 0$ and d in terms of summary statistics to have a tractable algorithm. To add a Markov Chain, one can use Metropolis-Hastings sampling, by defining a proposal distribution $q(\theta_{i+1}|\theta_i)$ to sample from (e.g., a Gaussian), and using the standard Metropolis-Hastings acceptance probability multiplied by the threshold $d(D', D) \leq \epsilon$ (a 1 or 0).

B. Likelihood-Free Inference for Cosmic Shear

In very high-dimensional datasets, we are often faced with the "curse of dimensionality." In other words, distances are greater in high dimensions between randomly sampled points. Selecting d and ϵ becomes more challenging here. For that reason, ABC is often done with a type of compression, or summary statistics.

One can also learn this directly, using a neural network (see [7]), which is capable of efficiently expressing high-dimensional relations, and learning them via stochastic gradient descent. This means that it can be "plugged-in" to a model to learn summary statistics via chain rule.

What we are doing with this is essentially learning the likelihood from simulated data. Thus, this technique is called "likelihood-free inference," since we approximate a likelihood by learning it from a simulator for our model. As before, we would like to efficiently find $P(\theta|D, M)$. With learned likelihood-free inference, we learn a model (conditional distribution) $q_\phi(\theta|D')$ by maximizing the value of $\sum_i \log q_\phi(\theta_i|D'_i)$ for parameter samples θ_i, and model-generated data $D'_i \sim M(\theta_i)$ over parameters ϕ. For example, this could be a parametrized distribution over Cosmological parameters given a tomographic map D', or even shear for an individual galaxy given its image. We subject $q_\phi(\theta|D')$ to the usual constraints on probability distributions: it must normalize to 1 at each conditional value. With enough examples, $q_\phi(\theta|D')$ should learn to approximate $P(\theta|D, M)$, and can be used for efficient MCMC inference on θ given real D'.

A Neural Network is a universal function approximator between vectors or grids that is differentiable and can be optimized via gradient descent. The name "neural network" is a terrible misnomer, for they are actually very much not like the brain. A neural network is nothing more than piecewise linear regression (line of best fit). The most basic type is called a multi-layer perceptron, and can be expressed as repeated matrix multiplication, vector addition, and then thresholding, i.e., (n layer MLP):

$$\text{NeuralNetwork}(\mathbf{x}) = A_1\sigma(A_2\sigma(\cdots\sigma(A_n\mathbf{x} + b_n)\cdots) + b_2) + b_1,$$

where every A_i is a matrix, and every b_i is a vector. σ is almost always a vector function that converts every negative value to 0. This function is capable of efficiently learning and approximating *any* continuous vector function with large enough matrices and dataset. Essentially what we do in likelihood-free inference is build a Neural Network that, for example, can take a pixelized image of a galaxy and convert it to an estimate of the potential distribution of parameters used to generate that image (e.g., shear). We can apply regularizations and structure on the above multi-layer perceptron to be translationally-invariant over images, which is called a convolutional neural network. Then, with a small number of simulated images of galaxies, it is possible to learn a neural network that produces estimates of the mean and variance of a distribution of parameters used to generate a given galaxy image. This new technique is known as neural likelihood-free inference, since we learn a likelihood directly from a simulation, and will be profoundly useful for unifying simulational and observational astrophysics going forward.

As discussed above and shown in fig. 2, there are many potential systematics that can enter the calculation of shear correlation functions and therefore lower the FoM for dark energy analysis. If one has a realistic simulator for data given estimated system + observation parameters, these are all capable of being heavily reduced by the optimal data compression performed by likelihood-free inference, as shown in [1]. Essentially what the authors show is that if one can generate accurate mock data (they use *Euclid* mocks) for galaxies, one can also add realistic noise over top, simulate observational systematics and masks to generate a realistic dataset for a survey. Using this, and resampling within the noise to "fill out" the training distribution, one can then learn two models: a compression and then a likelihood in a two-step process.

These are both differentiable models: the first a MOPED [2] compression to optimally compress a masked tomographic shear map into a single vector. The next: a mapping from this vector to a distribution over the cosmological parameters. Since many of the systematics can be exactly included in the forward model, these are optimally accounted for in this likelihood-free inference scheme via learning the compression and likelihood. This was studied in detail in [1] for *Euclid* and produced posteriors in agreement with simulated cosmological parameters, as shown in fig. 3. One can also potentially do the same directly for shear of a single galaxy, and use multiple learned likelihoods in a larger algorithm for full simulation-defined inference. This algorithm has great potential to be used for LSST data in working with cosmic shear data to estimate cosmological parameters, potentially helping to increase the 10-year FoM for dark energy from 40 up to approaching 87. Future studies on this type of analysis should use mock LSST datasets to study likelihood-free inference against existing pipelines.

IV. Conclusion

In this report I have reviewed weak lensing and cosmic shear. I discussed LSST and its ability for measuring cosmic shear, which can be used to characterize the dark energy equation of state. Finally, I derived and discussed Likelihood-Free Inference, and its prospect for increasing the detection efficiency of LSST, with a focus on the study by Alsing et al. (2019) and their demonstration of neural likelihood estimators as efficient estimators of cosmological parameters from Euclid tomographic shear maps. True likelihoods for data as a function of astrophysical parameters are impossible to find in astrophysics, so this type of algorithm is a very straightforward but immensely powerful technique for addressing all types of systematic errors in surveys where one possesses accurate forward simulators

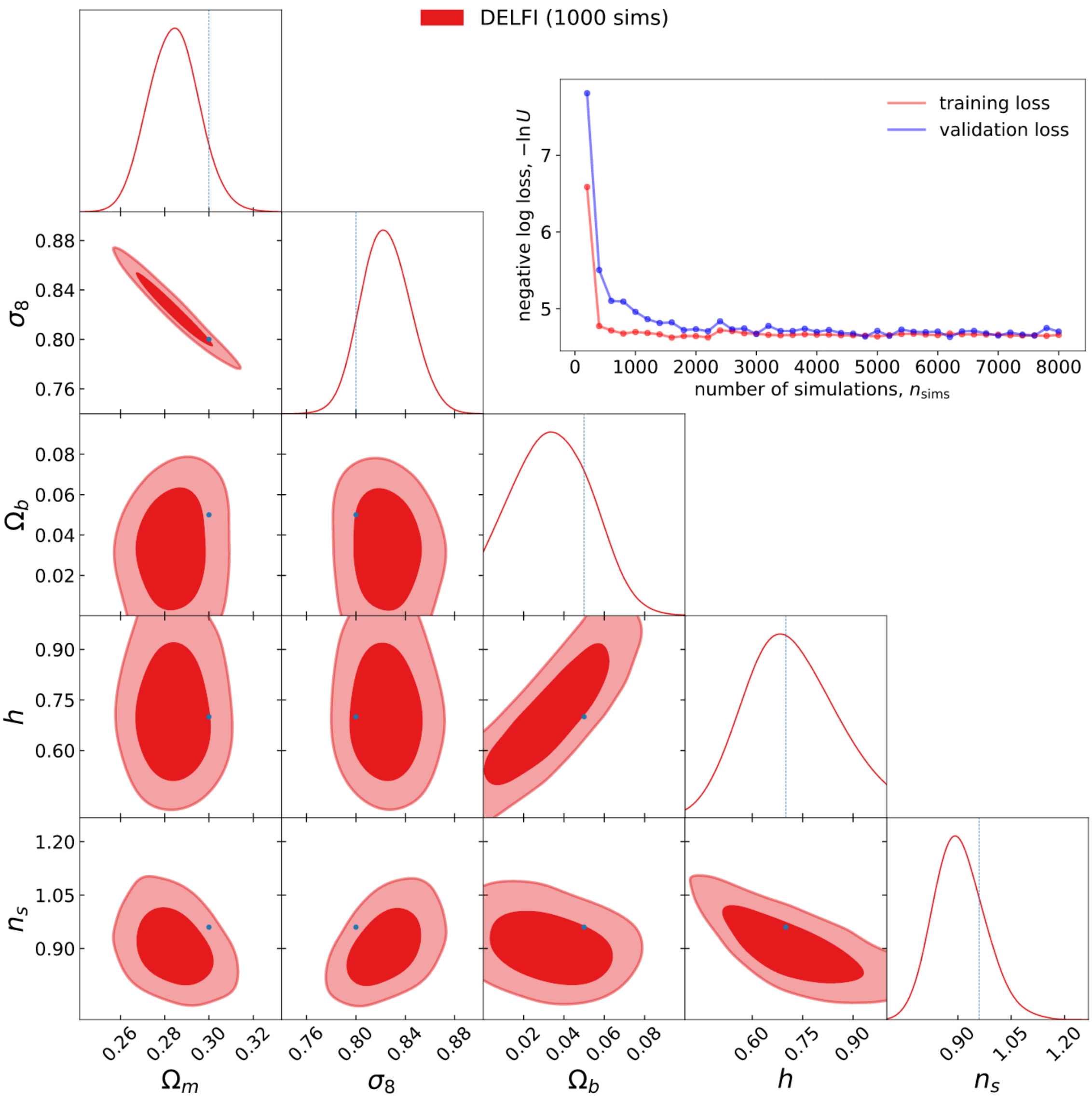

FIG. 3: Reconstruction of posteriors over Cosmological parameters using Likelihood-Free Inference on simulated *Euclid* tomographic shear maps in [1].

of observed data.

---

[1] Justin Alsing, Tom Charnock, Stephen Feeney, and Benjamin Wandelt. Fast likelihood-free cosmology with neural density estimators and active learning. *Monthly Notices of the Royal Astronomical Society*, Jul 2019.

[2] Alan F. Heavens, Raul Jimenez, and Ofer Lahav. Massive lossless data compression and multiple parameter estimation from galaxy spectra. , 317(4):965–972, Oct 2000.

[3] Martin Kilbinger. Cosmology with cosmic shear observations: a review. *Reports on Progress in Physics*, 78(8):086901, Jul 2015.

[4] Jarno Lintusaari, Michael U. Gutmann, Ritabrata Dutta, Samuel Kaski, and Jukka Corander. Fundamen-

tals and Recent Developments in Approximate Bayesian Computation. *Systematic Biology*, 66(1):e66–e82, 09 2016.

[5] LSST Dark Energy Science Collaboration. The LSST Dark Energy Science Collaboration (DESC) Science Requirements Document, 2018.

[6] LSST Science Collaboration. LSST Science Book, Version 2.0. *arXiv e-prints*, page arXiv:0912.0201, Dec 2009.

[7] George Papamakarios, David C. Sterratt, and Iain Murray. Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows, 2018.