Statistical Challenges in 21st Century Cosmology
Proceedings IAU Symposium No. 306, 2014
A. F. Heavens, J.-L. Starck & A. Krone-Martins, eds.

doi:10.1017/S174392131401360X

The probability distribution of ellipticity: implications for weak lensing measurement

Massimo Viola

Leiden Observatory, Leiden University,
Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
email: viola@strw.leidenuniv.nl

Abstract. The weak lensing effect generates spin-2 distortions, referred to as shear, on the observable shape of distant galaxies, induced by intervening gravitational tidal fields. Traditionally, the spin-2 distortion in the light distribution of distant galaxies is measured in terms of a galaxy ellipticity. This is a very good unbiased estimator of the shear field in the limit that a galaxy is measured at infinite signal-to-noise. However, the ellipticity is always defined as a ratio between two quantities (for example, between the polarisation and measurement of the galaxy size, or between the semi-major and semi-minor axis of the galaxy) and therefore requires some non-linear combination of the image pixels. This means, in any realistic case, this would lead to biases in the measurement of the shear (and hence in the cosmological parameters) whenever noise is present in the image. This type of bias can be understood from the particular shape of the 2D probability distribution of the ellipticity of an object measured from data. Moreover this probability distribution can be used to explore strategies for calibration of noise biases in present and future weak lensing surveys (e.g. KiDS, DES, HSC,Euclid, LSST...)

Keywords. methods: data analysis, cosmology, weak-lensing

1. Introduction

Weak gravitational lensing is a very powerful tool to study properties of dark matter halos (e.g Hoekstra *et al.* 2013) as well to investigate the growth-rate of structures in the Universe (e.g. Schrabback *et al.* 2010, Kilbinger *et al.* 2013). However it is quite challenging from a practical point of view since it relies on measurements of tiny choerent distortions (shear) in the shapes of background galaxies.

One of the reasons why measuring galaxy shapes (ellipticities) is notoriously challenging is because the measurements have to be done from noisy pixels.

Since 2007 it has become common practice in the weak-lensing community to test and validate different algorithms used to infer the gravitational shear from measurements of the shape of galaxies, on some sets of common image simulations (Heymans *et al.* (2007), Massey *et al.* (2007), Bridle *et al.* (2010), Kitching *et al.* (2012)).

Some important lessons were learned: the bias in shear measurements is a strong function of the galaxy signal-to-noise and of the galaxy size with respect to the size of the point spread function (PSF): galaxies with sizes closer to the size of the PSF have more bias then larger objects at equal signal-to-noise. This behaviour is common to all algorithms tested on those simulations and usable on real data.

The bias in the shear is generally parameterised in terms of a multiplicative term m and an additive term c: $\mathrm{g^{obs} = (1 + m)g^{true} + c}$.

The additive bias c can be estimated from the data itself, making use of the fact that over the full survey area, the average of each component of the ellipticity must vanish.

For galaxies with low signal-to-noise (15 or below) the multiplicative bias in the shear is typically very large, of the order of 20% or larger. If not accounted for, this bias propagates directly into potentially large biases in the cosmological parameters or in derived properties of dark matter halos.

Ideally, any bias in the shear measurements should be smaller than the measurement statistical error, $\sigma_\gamma \simeq \sigma_\epsilon/\sqrt{\mathrm{N}}$, where $\sigma_\epsilon \simeq 0.3$ is the intrinsic ellipticity dispertion and N the number of galaxies used to infer the shear.

Even for existing surveys, like CFHTLenS, the amplitude of the bias is too large at low signal noise for not being corrected, and some calibration needs to be applied (Miller *et al.* 2013). Ongoing larger surveys, like Dark Energy Survey (DES),Kilo Degree Survey (KiDS),Hyper Suprime-Cam (HSC), have even more stringent requirements on the amplitude of the shear bias, which poses greater challenges about calibration of existing shape measurements methods.

2. The Marsaglia-Tin distribution

At a very fundamental level, the bias in shear measurements is a consequence of the fact that it is not possible to measure an unbiased ellipticity in presence of noise (if the effect of the noise is not properly accounted for). The reason is that an ellipticity measurement involves some non-linear transformation on the noisy pixels of an astronomical image.

In order to further investigate this problem, we derive the probability distribution of the observed ellipicity in presence of noise given a true ellipticity.

Since the shear is derived as an average over an ensemble of measured ellipticities in a region of the sky, this probability lies behind any weak lensing analysis.

This particular probability distribution was first derived by Marsaglia (1965) and Tin (1965) and re-derived in a weak lensing context by Melchior & Viola (2012) and Viola, Kitching, Joachimi (2014).

Here we summarise the derivation of the Marsaglia-Tin distribution.

2.1. *Some definitions*

We start by defining the $i+j$ order moments of the object surface brightness $I(x,y)$:

$$\{Q\}_{i,j} = \int I(x,y) x^i y^j \,\mathrm{d}x\mathrm{d}y \tag{2.1}$$

Note that in reality what are measured are weighted moments of the convolved surface brightness. A weighting function has to be employed in order to suppress the pixel noise at large distances from the galaxy centre, and its effect has to be accounted for (this involves measuring higher order moments of the surface brightness). Moreover, the contribution of the PSF has to be removed.

The second-order moments can be used to characterise the object's normalised polarisation χ and the ϵ-ellipticity of the object:

$$\chi := \frac{\{Q\}_{20} - \{Q\}_{02} + 2\mathrm{i}\{Q\}_{11}}{\{Q\}_{20} + \{Q\}_{02}} \quad \text{and} \quad \epsilon := \frac{\{Q\}_{20} - \{Q\}_{02} + 2\mathrm{i}\{Q\}_{11}}{\{Q\}_{20} + \{Q\}_{02} + 2\sqrt{\{Q\}_{20}\{Q\}_{02} - \{Q\}_{11}^2}}. \tag{2.2}$$

The two definitions are related through:

$$\chi = \frac{2\epsilon}{1+|\epsilon|^2}. \tag{2.3}$$

The shear is then derived by averaging many galaxies' ellipticities under the assumption that the the intrinsic orientation of galaxies in the universe is random.

Both definitions are used in literature. However only an unbiased measurement of ϵ is indeed an unbiased estimate of the shear (Seitz & Schneider (1997)).

https://doi.org/10.1017/S174392131401360X Published online by Cambridge University Press

We start focusing on what is the probability distribution for χ, which can also be written as a ratio of the Stokes parameters $u = \{Q\}_{20} - \{Q\}_{02}$,$v = 2\{Q\}_{11}$ and s=$\{Q\}_{20} + \{Q\}_{02}$.

The probability distribution function of the Stokes parameters in presence of homoscedastic noise (i.e. uncorrelated and gaussian) can be described in terms of a trivariate Gaussian with correlation coefficients ρ_{ij} between each of the variables.

The 2-dimensional probability distribution for the normalised polarisation χ defined as $(u/s, v/s)$ can be derived starting from the three-dimensional probability distribution for the Stokes parameters.

First of all we transform the distribution $p_{u,v,s}(u, v, s)$ into $p(\chi_1, \chi_2, s)$ by a change of variable and then we marginalise over s

$$p_\chi(\chi_1, \chi_2) = \int_{-\infty}^{\infty} \mathrm{d}s s^2 p_{u,v,s}(\chi_1 s, \chi_2 s, s) \; ; \tag{2.4}$$

this is the form of a two dimensional quotient distribution. The result of this integration is the so-called Marsaglia-Tin distribution. It has an analytical (even though not simple) expression, that interested readers can find in Section 3.1 of Viola, Kitching, Joachimi (2014).

The probability of measuring a χ polarisation can be transformed into the probability of measuring an ϵ-ellipticity using Equation 2.3.

We note here that in the case of uncorrelated variables which are gaussian distributed with zero-mean the Marsaglia-Tin distribution reduces to the Cauchy distribution.

2.2. *Properties of the Marsaglia-Tin distribution*

We summarise here the main properties of the Marsaglia-Tin distribution:

• Only the amplitude of the ellipticity is generally biased, while the angle $(1/2)\tan^{-1}(\chi_2/\chi_2)$ is always unbiased Wardle & Kronberg (1974);

• In the case an 'optimal' weighting function (i.e. matching exactly the radial profile, ellipticity and size of the object) and χ is used as a definition for the ellipticity, both the mean and the maximum of the Marsaglia-Tin distribution are biased, while in the case that ϵ is used only the maximum is biased while the mean is unbiased *independent* of the signal-to-noise level;

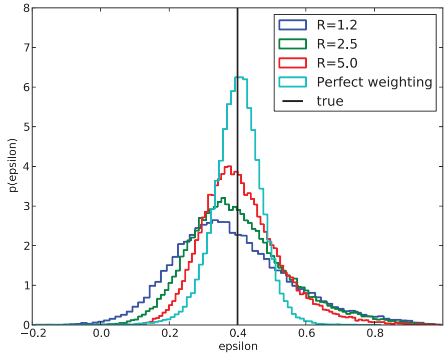
• Truncation of the ϵ-ellipticity distribution, for example by removing very elliptical objects, introduces a bias in the measurements of the mean ϵ even in the ideal case of a weighting function that perfectly matches the galaxy profile;

• In the case where a circular weighting is employed in the moment measurements, the correlation between the Stokes parameters deviates from the true one. The larger this deviation, the larger is the bias;

• For a fixed value of signal-to-noise, the amplitude of the bias is determined by two factors: the correlation between the Stokes parameters, and the signal-to-noise on the quadrupole moments (or the ellipticity of the object);

• If the size of the object becomes comparable to the size of the PSF then the probability distribution of the convolved ellipticity is the convolution of the Marsaglia-Tin distribution with the probability distribution of the ratio of the size of the galaxy and the size of the PSF (the so-called resolution): the lower the resolution, the larger the bias.

The probability distribution of the absolute value of the ϵ-ellipticity is shown in left panel of Figure 1 for different choices of the weighting function.



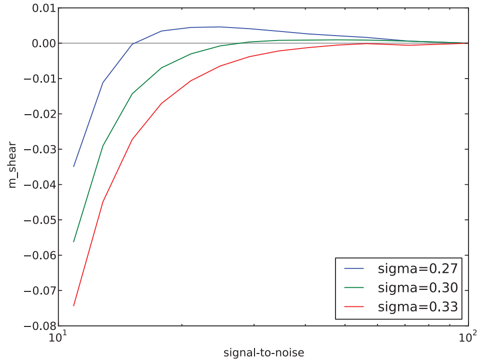


Figure 1. *Left panel:* Probability distribution of the absolute value of the ϵ-ellipticity given a true galaxy ellipticity of $\epsilon = (0.4, 0.0)$ and zero PSF ellipticity. The cyan line corresponds to the case of using a weighting function which is matched perfectly to the galaxy profile and no PSF convolution, the red line to the case of using a circular weighting function with size 1.2 times the galaxy semi-major axis and a resolution of $R = 5.0$, the green line to a resolution of 2.5 and the blue line to a resolution of 1.2. Note that in all cases we removed objects having an unphysical combination of second-order moments ($Q_{20}Q_{02} - Q_{11}^2 < 0$). *Right panel:* Shear multiplicative Marsaglia bias as a function of signal-to-noise. The three curves represent the case of galaxies with intrinsic ellipticities following a Rayleigh distribution with $\sigma_\epsilon = 0.27$ (blue), $\sigma_\epsilon = 0.3$ (green) and $\sigma_\epsilon = 0.33$ (red). The width of the weighting function has been chosen to be 1.2 times the object semi-major axis. This plot highlights the importance of knowing the ellipticity distribution in order to calibrate the shear bias. These figures are adapted from Viola, Kitching & Joachimi 2014.

3. Implication for current and future surveys

If we assume that the galaxy profile is known (i.e. we neglect the so-called model bias), the amplitude of the multiplicative bias in the shear measurements depends essentially on the galaxy resolution, the intrinsic ellipticity distribution (since the bias in the ellipticity measurements is a function of the ellipticity) and on the object signal-to-noise and it can be numerically computed starting from the Marsaglia-Tin distribution (for details we refer to Section 3 of Viola, Kitching & Joachimi.)

Therefore any attempt to characterise and calibrate the noise-bias by means of image simulations requires knowledge of these three quantities.

How well those three quantities need to be known depends on the statistical power of a survey (given by its area and its depth) which sets requirements on the knowledge of the shear multiplicative bias $\sigma_{\rm m}$.

For a surveys like CFHTLenS this number is of order $\sigma_m \simeq 10^{-2}$, for current surveys like KiDS, DES, HSC, $\sigma_m \simeq 3 \times 10^{-3}$, and for a Euclid-like survey $\sigma_m \simeq 5 \times 10^{-4}$.

Hence the requirements on the knowledge of a quantity $\vec{x} = (\sigma_\epsilon, \nu, R, ..)$ can be computed as:

$$\sigma_{\vec{x_i}_0} = \sigma_m \left[\left| \frac{\partial {\rm m}}{\partial \vec{x_i}} \right|_{\vec{x_i}_0} \right]^{-1} \tag{3.1}$$

from which it is clear that the requirements on the knowlege of the intrinsic ellipticity distribution, noise level and resolution are driven by the steepness of the multiplicative-bias as a function of this quantity and not by its amplitude.

In other words among methods with similar amplitude of the multiplicative bias, it is preferred, in the sense that it is more calibratable, the one with the shallower derivative of m as a function of $\vec{\rm x}$

https://doi.org/10.1017/S174392131401360X Published online by Cambridge University Press

The effect of the intrnsic ellipticity distribution on the shear multiplicative bias as a function of signal to noise is shown in the right panel of Figure 1.

In Viola, Kitching & Joachimi we investigated the requirements on the knowledge of the intrinsic ellipticity distribution and we found that it has to be known with a precision of $\sim 5\%$ in order to properly calibrate shear estimates for current surveys, for upcoming surveys with a precision of $\sim 1\%$ and for future surveys with a precision of $\sim 0.3\%$.

4. Conclusions

We showed in this work how the bias in shear measurements, affecting all methods applied to data so far, can be understood studying the properties of the Marsaglia-Tin distribution (which the probability distribution of measuring an ellipticity in presence of noise).

In particular we show how the amplitude of the multiplicative shear bias strongly depends on the intrinsic ellipticity distribution, the resolution and the signal-to-noise of the objects.

Hence these properties of galaxies need to be known with great precision and accuracy in order to calibrate the bias using image simulation.

One way to avoid the noise-bias in shear measurements would be using avoiding taking ratios, for example using the un-normalised stokes parameters (i.e. not normalised by the galaxy flux). However, it has been shown (Viola, Kitching & Joachimi), that the price paid for this is an increased variance in the shear estimate. This can be understood by the fact that no information about the object flux is used.

References

Bridle, S., Balan, S. T., Bethge, M., Gentile, M., Harmeling, S., Heymans, C., Hirsch, M., Hosseini, R., Jarvis, M., & Kirk, D. 2009, ArXiv e-prints 0908.0945

Heymans, C., Van Waerbeke, L., Bacon, D., Berge, J., Bernstein, G., Bertin, E., Bridle, S., Brown, M. L., Clowe, D., *et al.* 2006, *MNRAS*, 368, 1323

Hoekstra, H., Bartelmann, M., Dahle, H., Israel, H., Limousin, M., & Meneghetti, M. 2013, *SSR*, 177, 75

Kilbinger, M., Fu, L., Heymans, C., Simpson, F., Benjamin, J., Erben, T., Harnois-Déraps, J., Hoekstra, H., *et al.* 2013, *MNRAS*, p. 735

Kitching, T. D., Balan, S. T., Bridle, S., Cantale, N., Courbin, F., Eifler, T., Gentile, M., Gill, M. S. S., Harmeling, S., Heymans, C., *et al.* 2012, *MNRAS*, 423, 3163

Marsaglia, G., 1965, j-J-AM-STAT-ASSOC, 60, 193

Massey, R., Heymans, C., Bergé, J., Bernstein, G., Bridle, S., Clowe, D., Dahle, H., Ellis, R., Erben, T., *et al.* 2007, *MNRAS*, 376, 13

Melchior, P. & Viola, M. 2012, *MNRAS*, 424, 2757

Miller, L., Heymans, C., Kitching, T. D., van Waerbeke, L., Erben, T., & Hildebrandt, H. 2013, *MNRAS*, 429, 2858

Schneider, P. & Seitz, C. 1995, *A&A*, 294, 411

Schrabback, T., Hartlap, J., Joachimi, B., Kilbinger, M., Simon, P., Benabed, K., Bradač, M., Eifler, T., Erben, T., *et al.* 2010, *A&A*, 516, A63+

Seitz, C. & Schneider, P. 1997, *A&A*, 318, 687

Tin, M., 1965, j-J-AM-STAT-ASSOC, 60, 294

Wardle, J. F. C. & Kronberg, P. P. 1974, *ApJ*, 194, 249

Viola, M., Kitching, T. D., & Joachimi, B. 2014, *MNRAS*, 439, 1909