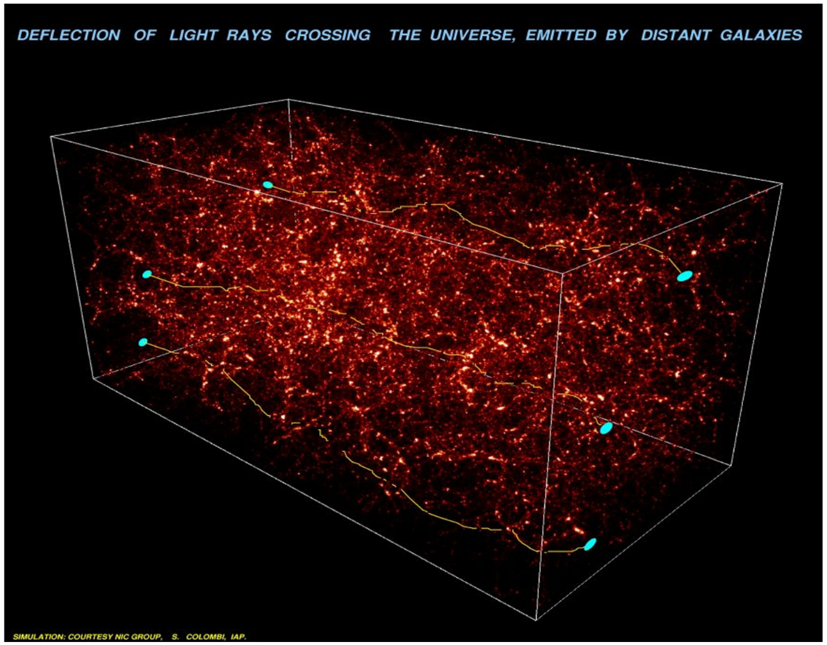
Cosmic Weak Lensing



Photons emitted by distant galaxies is deflected by tidal field along line of sight.

The shape distorition of galaxies is called (gravitational) shear:
$\epsilon \approx \epsilon_{\rm int} + \gamma$

Statistical properties of the shear reflect statistical properties of the density field.

Deflection angle
$$\alpha^i(\vec{\theta}) = 2\int_0^{\chi_{\rm s}} d\chi' \Phi_{,i}(\mathbf{x}(\vec{\theta},\chi'))\left(1-\frac{\chi'}{\chi_{\rm s}}\right)$$

Lensing potential
$$\psi(\vec{\theta}) \equiv 2\int_0^{\chi_{\rm s}} \frac{d\chi'}{\chi'} \Phi(\mathbf{x}(\vec{\theta},\chi'))\left(1-\frac{\chi'}{\chi_{\rm s}}\right)$$

Born approx: evaluate along unperturbed path

Image Distortions

Image distortions occur when the deflection angle varies with position/across the image.
Consider Jacobian of transformation from source to image plane

$$A_{ij}(\boldsymbol{\theta}) \equiv \frac{\partial \beta_i}{\partial \theta_j} = \delta_{ij} - \frac{\partial \alpha_i}{\partial \theta_j}$$

Most general form: $$A_{ij} = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}$$

$\gamma_{1,2}$ Cartesian components of **shear**

Image Distortions

Light deflection does not involve any emission or absorption processes, hence

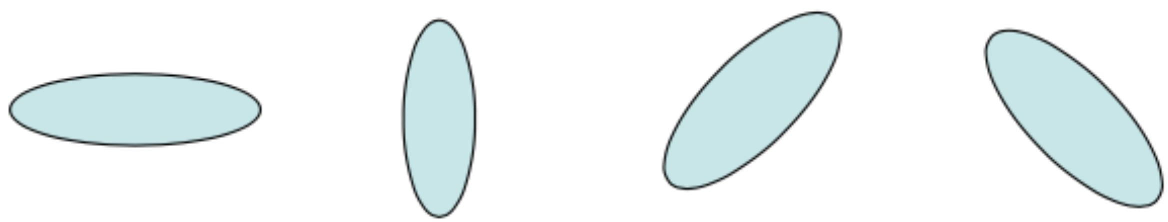
$$I(\boldsymbol{\theta}) = I^s[\boldsymbol{\beta}(\boldsymbol{\theta})]$$

For a small source, centered on $\boldsymbol{\beta}_0 = \boldsymbol{\theta}_0 - \boldsymbol{\alpha}_0$

$$I(\boldsymbol{\theta}) = I^s[\boldsymbol{\beta}_0 + A(\boldsymbol{\theta}_0) \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}_0)]$$

Hence the image of a small circular source with radius r is an **ellipse with semi-axes r $\boldsymbol{\lambda}_{1,2}$**, with $\boldsymbol{\lambda}_{1,2}$ the eigenvalues of A, and **orientation determined by the shear components $\gamma_{1,2}$**

Shear is a Spin 2 Field



$\gamma_1>0$ $\gamma_1<0$ $\gamma_2>0$ $\gamma_2<0$

The Relation Between Shear and Convergence

$$A_{ij}(\boldsymbol{\theta}) \equiv \frac{\partial \beta_i}{\partial \theta_j} = \delta_{ij} - \frac{\partial \alpha_i}{\partial \theta_j} = \delta_{ij} - \frac{\partial^2 \psi}{\partial \theta_i \theta_j}$$

$$= \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}$$

Write γ_1 and γ_2 in terms of the lensing potential.

$$\left.\begin{aligned} \gamma_1 &= \frac{1}{2}\left(\partial_1^2 - \partial_2^2\right)\psi \\ \gamma_2 &= \partial_1 \partial_2 \psi \end{aligned}\right\} \quad \begin{aligned} \gamma_1 &= -\frac{1}{2}(k_1^2 - k_2^2)\psi \\ \gamma_2 &= -k_1 k_2 \psi \end{aligned}$$

The Relation Between Shear and Convergence

$$\left.\begin{aligned} \gamma_1 &= -\frac{1}{2}(k_1^2 - k_2^2)\psi \\ \gamma_2 &= -k_1 k_2 \psi \end{aligned}\right\} \quad 4\gamma_1^2 + 4\gamma_2^2 = (k_1^4 + 2k_1^2 k_2^2 + k_2^4)\psi^2$$

Solving for γ:

$$\gamma = -\frac{1}{2}(k_1^2 + k_2^2)\psi = \kappa$$

The power spectrum of κ and γ will be identical!

Calculating Convergence

For single source plane at z_s

$$\kappa(\boldsymbol{\theta}) = \frac{3H_0^2\Omega_m}{2}\int_0^{\chi_s} d\chi' \frac{\chi'}{a(\chi)}\left(1-\frac{\chi'}{\chi_s}\right)\delta_m(\mathbf{x}(\boldsymbol{\theta},\chi')) \equiv \int_0^{\chi_s} d\chi' W(\chi')\delta_m(\mathbf{x}(\boldsymbol{\theta},\chi'))$$

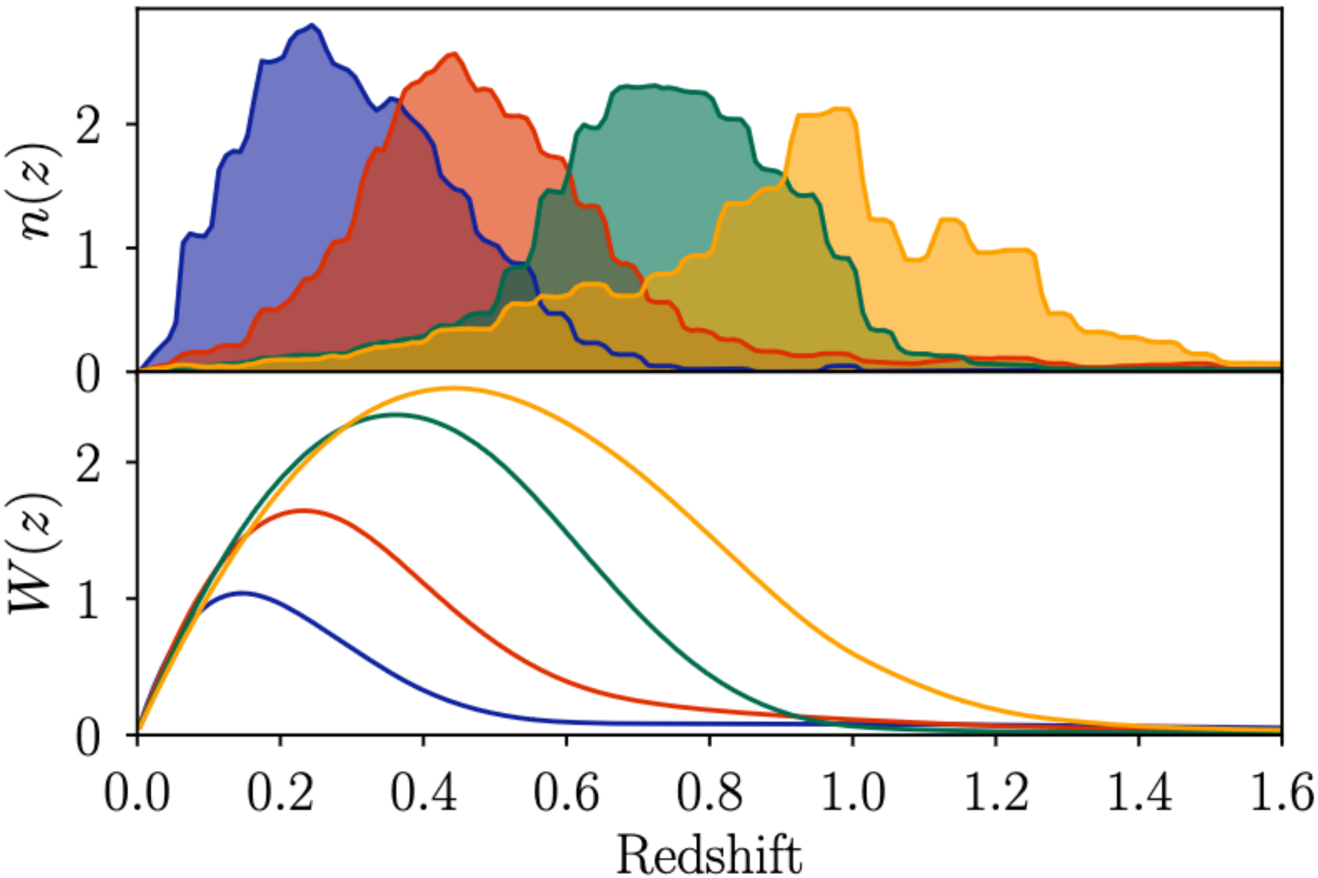
Average over tomographic source redshift distribution

$$W^i(\chi) = \frac{3H_0^2\Omega_\mathrm{m}}{2c^2}\frac{\chi}{a(\chi)}\int_\chi^{\chi_\mathrm{H}} d\chi'\, n^i\left(z(\chi')\right)\frac{dz}{d\chi'}\frac{\chi'-\chi}{\chi'}$$

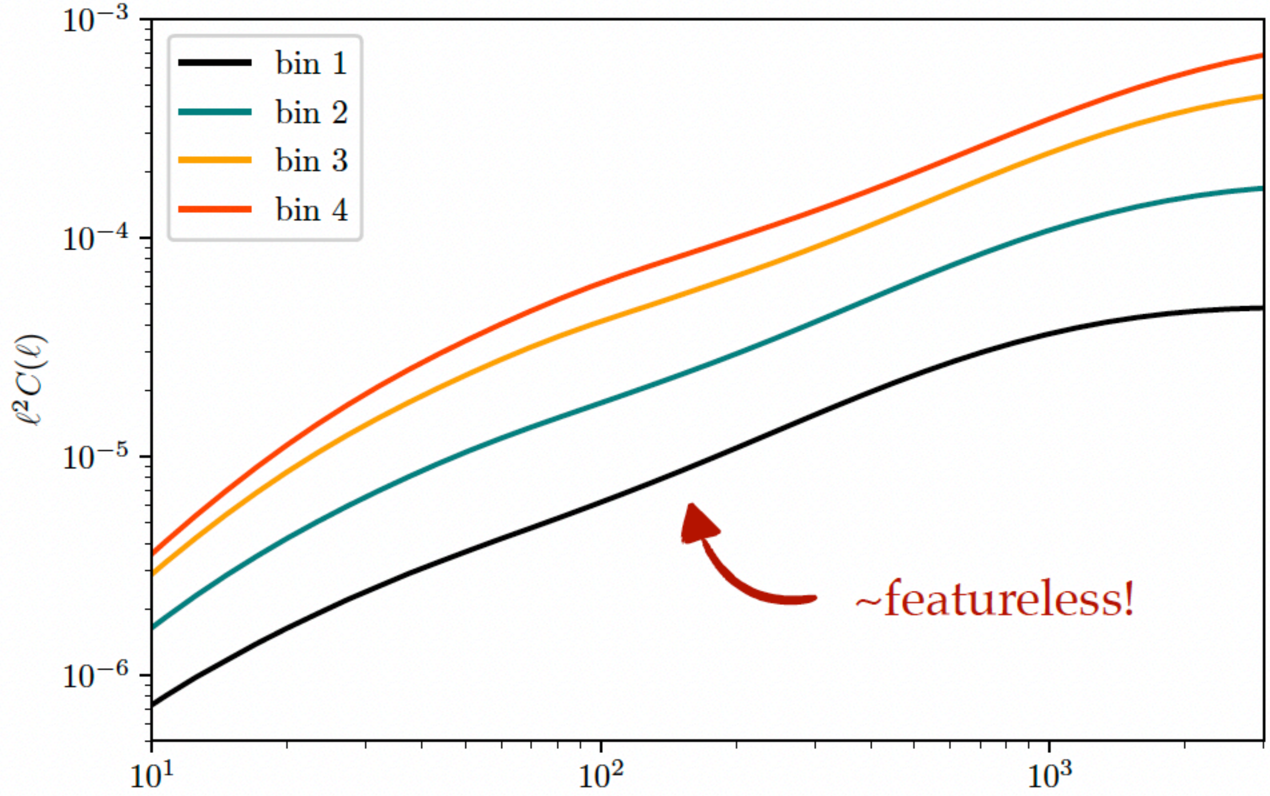
Calculate power spectrum with Limber approximation

$$C_\kappa^{ij}(\ell) = \int_0^{\chi(z_\mathrm{max})} d\chi \frac{W^i(\chi)W^j(\chi)}{\chi^2} P_\delta\left(\frac{\ell+1/2}{\chi}, z(\chi)\right)$$

Tomographic Lensing



Tomographic Lensing Power Spectra



$$C_{\kappa}^{ij}(\ell) = \int_0^{\chi(z_{\max})} d\chi \frac{W^i(\chi) W^j(\chi)}{\chi^2} P_\delta \left(\frac{\ell + 1/2}{\chi}, z(\chi) \right)$$

Integrate over lots of redshift slices of P(k) to get the projected 2D lensing power spectrum.

Magnification

$$A_{ij} = \begin{pmatrix} 1-\kappa-\gamma & 0 \\ 0 & 1-\kappa+\gamma \end{pmatrix}$$

$$a_{\rm source} = \det(A) a_{\rm image} \qquad \det(A) = (1-\kappa)^2 - \gamma^2 \approx 1 - 2\kappa$$

$$a_{\rm image} = \mu a_{\rm source} \qquad \mu = \frac{1}{(1-\kappa)^2 - \gamma^2} \approx 1 + 2\kappa$$

- Sources are **magnified** by gravitational lensing.

Magnification

- Sources are **magnified** by gravitational lensing.

Note lensing also works in reverse: we will be magnified from the point of view of a source.

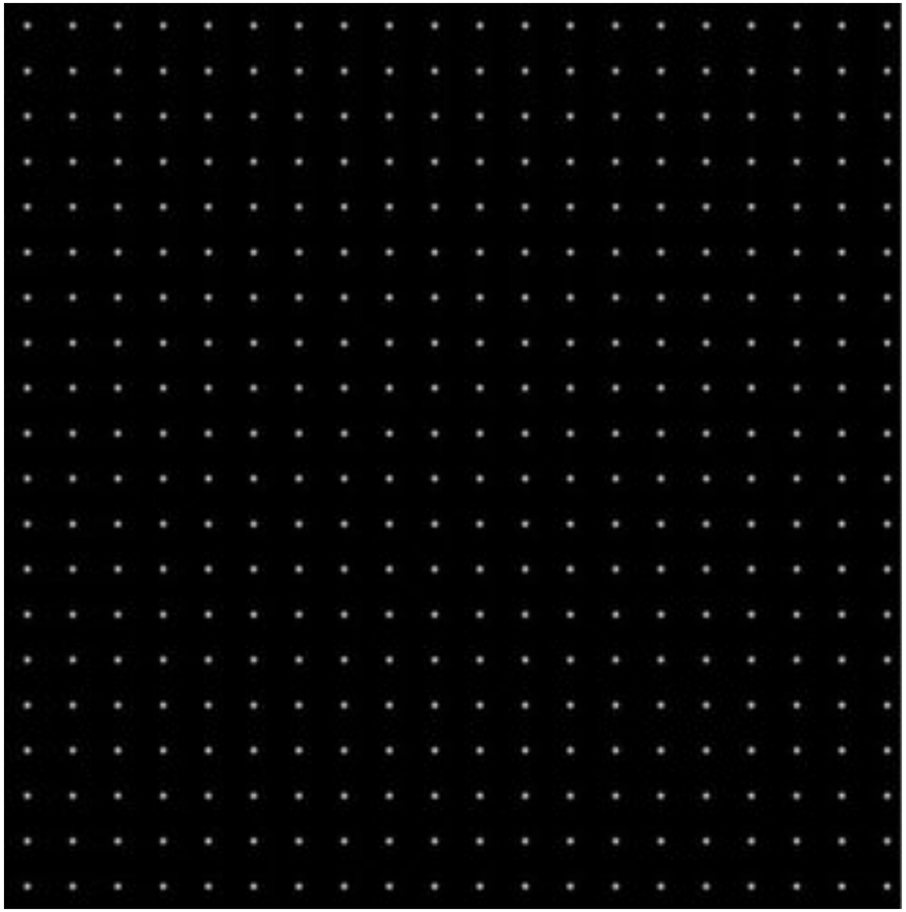
- Our telescope is effectively larger by a factor μ.
- The source will appear ***brighter*** by a factor of μ. However, <u>surface brightness is preserved</u>.

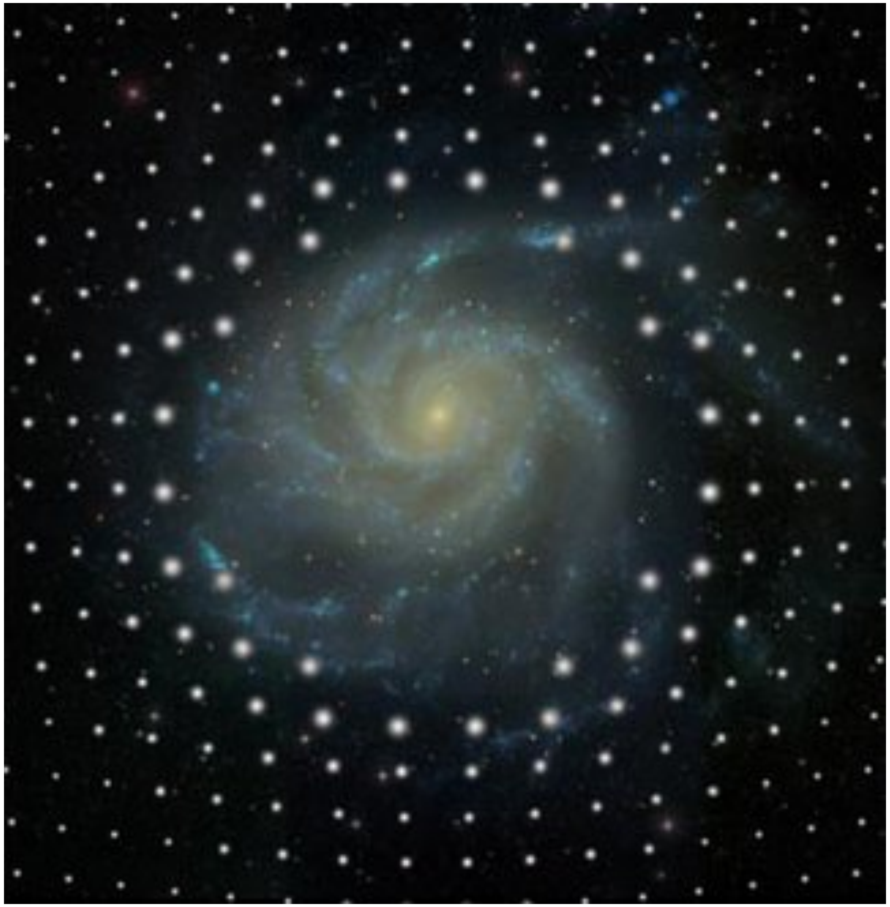
Lensing Distortions Summary

Lensing distorts a source in three ways:

- It uniformly expands the image of source (convergence).
- It shears the image, expanding one axis while contracting the other.
- It magnifies the image, making it appear larger and brighter.
- It pushes everything outwards.

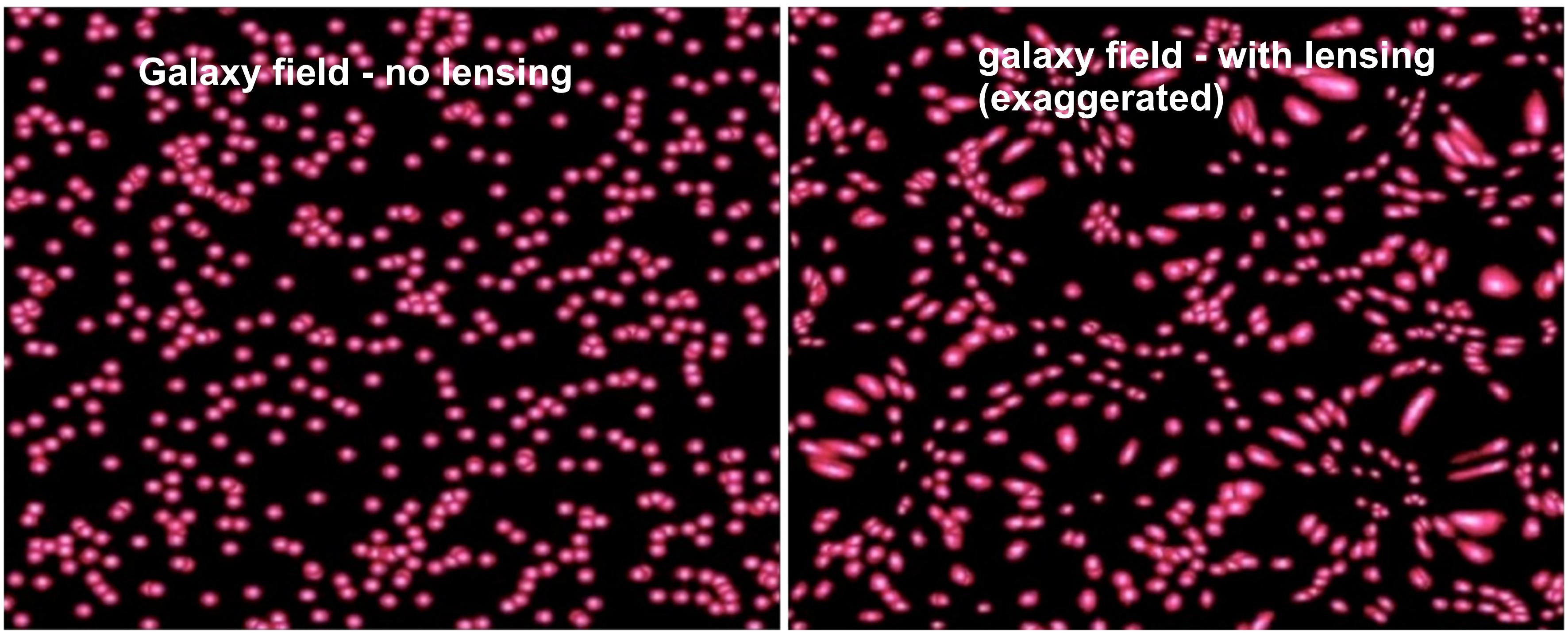
Note: lensing always preserved surface brightness.





Measuring Lensing

We could try to measure lensing using the various observational signatures.



What do we know about the unlensed galaxy properties?

Measuring Lensing

We could try to measure lensing using the various observational signatures:

- Magnification makes sources larger and brighter.
- Magnification changes the density of sources.
- Shear changes the ellipticity of sources.

Of these, the only one we "know" *a priori* is ellipticity: on average, galaxies have random ellipticities*!

Coherent distortions must be due to lensing*.

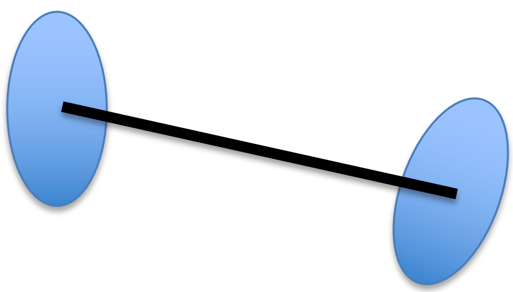
*Exception: intrinsic alignments

Tangential and Cross Shear

x-axis is defined by line connecting the two galaxies.

$$\gamma_{\mathrm{T}} = \text{Shear along } x-\text{axis}$$

$$\gamma_{\times} = \text{Shear perpendicular to } x-\text{axis}$$



Shear correlation function:

$$\xi_{\pm}(\theta) \equiv \langle \gamma_{\mathrm{T}} \gamma'_{\mathrm{T}} \rangle \pm \langle \gamma_{\times} \gamma'_{\times} \rangle$$

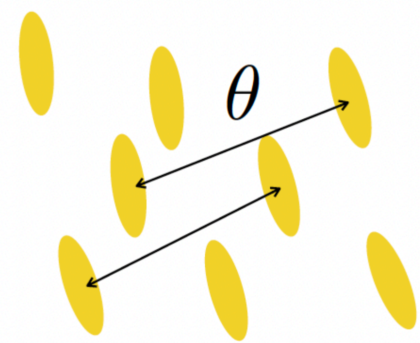
$\langle \gamma_{\mathrm{T}} \gamma'_{\times} \rangle = 0$ Parity

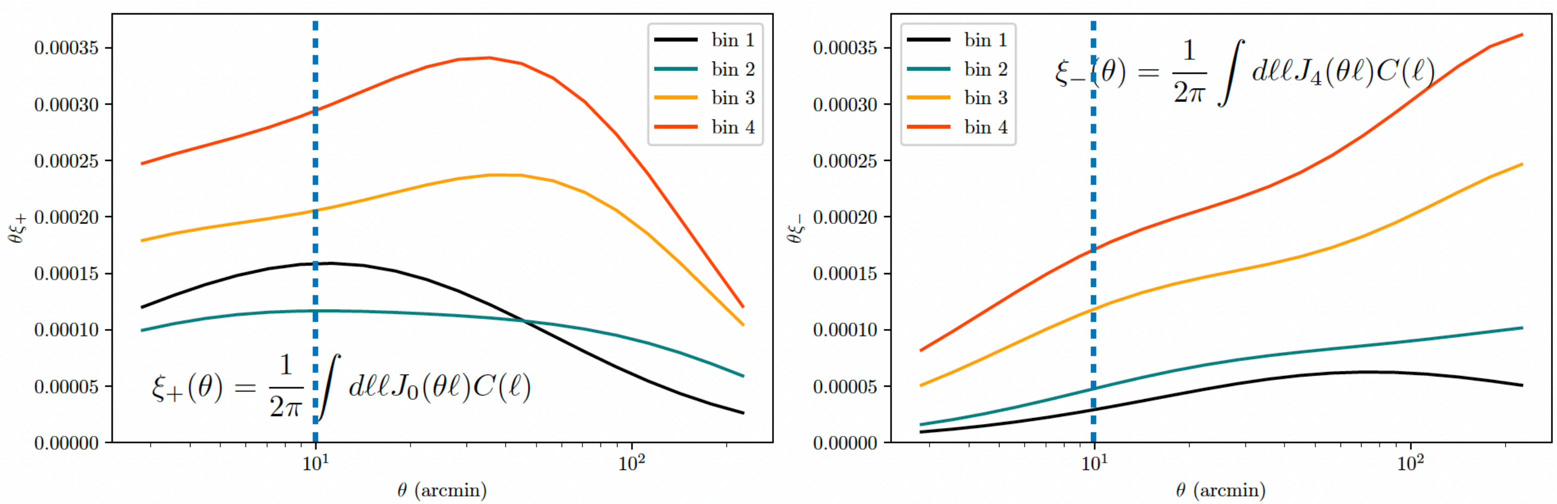
The shear correlation function is a well-defined observable!

Shear Correlation Functions

$$\xi_\pm(\theta) = \frac{\sum w^i w^j \left(\gamma_t^i(\theta_0)\gamma_t^j(\theta_0+\theta) \pm \gamma_\times^i(\theta_0)\gamma_\times^j(\theta_0+\theta)\right)}{\sum w_i w_j}$$

$$= \frac{1}{2\pi}\int d\ell\,\ell\,J_{0/4}(\theta\ell)C_\gamma(\ell)$$





Shear 2pt Statistics

- Shear correlation functions: $\xi_\pm(\theta) = \int_0^\infty \frac{\mathrm{d}\ell\,\ell}{2\pi} \mathrm{J}_{0/4}(\ell\theta)\,[C_{\mathrm{EE}}(\ell) \pm C_{\mathrm{BB}}(\ell)]$

- Band power:

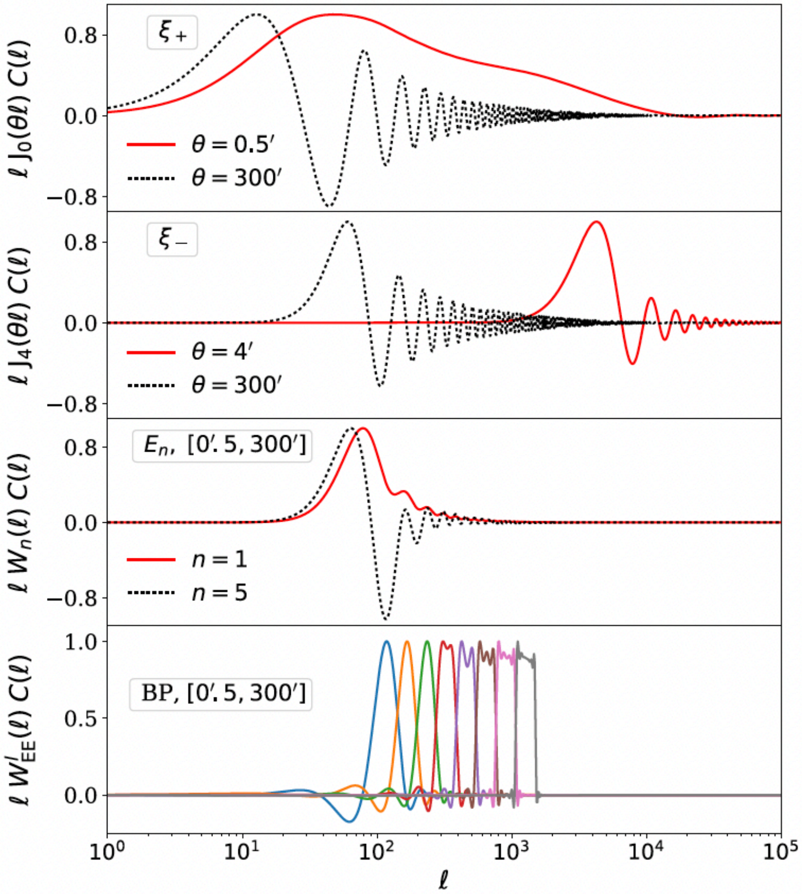
$$C_{\mathrm{E},l} = \frac{1}{2\mathcal{N}_l} \int_0^\infty \mathrm{d}\ell\,\ell\left[W^l_{\mathrm{EE}}(\ell)\,C_{\mathrm{EE}}(\ell) + W^l_{\mathrm{EB}}(\ell)\,C_{\mathrm{BB}}(\ell)\right]$$
$$C_{\mathrm{B},l} = \frac{1}{2\mathcal{N}_l} \int_0^\infty \mathrm{d}\ell\,\ell\left[W^l_{\mathrm{BE}}(\ell)\,C_{\mathrm{EE}}(\ell) + W^l_{\mathrm{BB}}(\ell)\,C_{\mathrm{BB}}(\ell)\right]$$

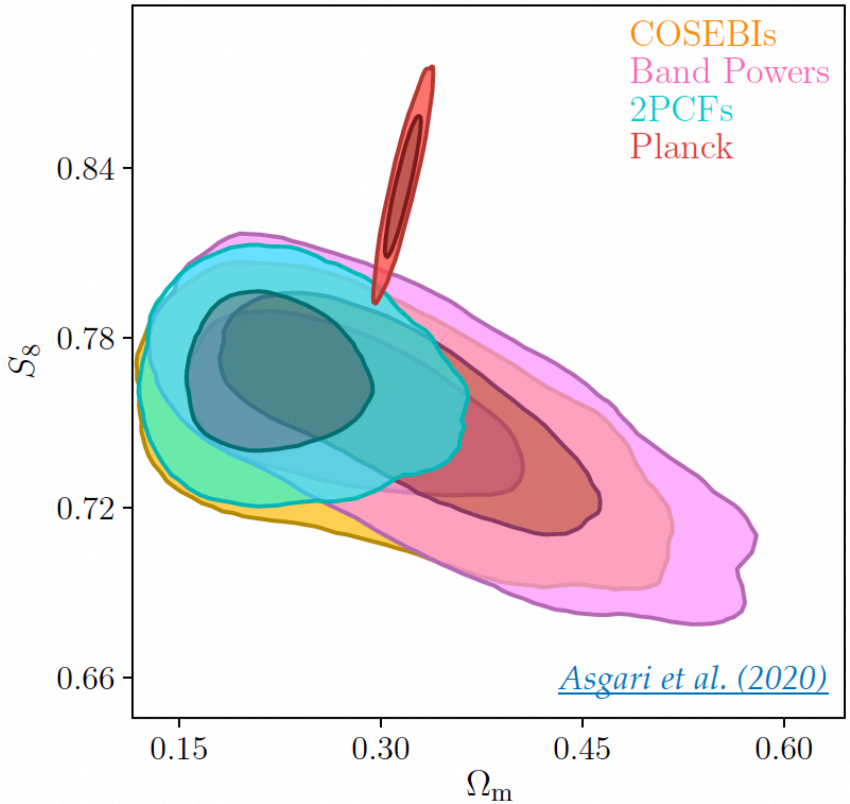
$$\begin{aligned} W^l_{\mathrm{EE}}(\ell) = W^l_{\mathrm{BB}}(\ell) \\ &= \int_0^\infty \mathrm{d}\theta\,\theta\,T(\theta)\left[\mathrm{J}_0(\ell\theta)\;g^l_+(\theta) + \mathrm{J}_4(\ell\theta)\;g^l_-(\theta)\right] \\ W^l_{\mathrm{EB}}(\ell) = W^l_{\mathrm{BE}}(\ell) \\ &= \int_0^\infty \mathrm{d}\theta\,\theta\,T(\theta)\left[\mathrm{J}_0(\ell\theta)\;g^l_+(\theta) - \mathrm{J}_4(\ell\theta)\;g^l_-(\theta)\right] \end{aligned}$$

- COSEBI's:

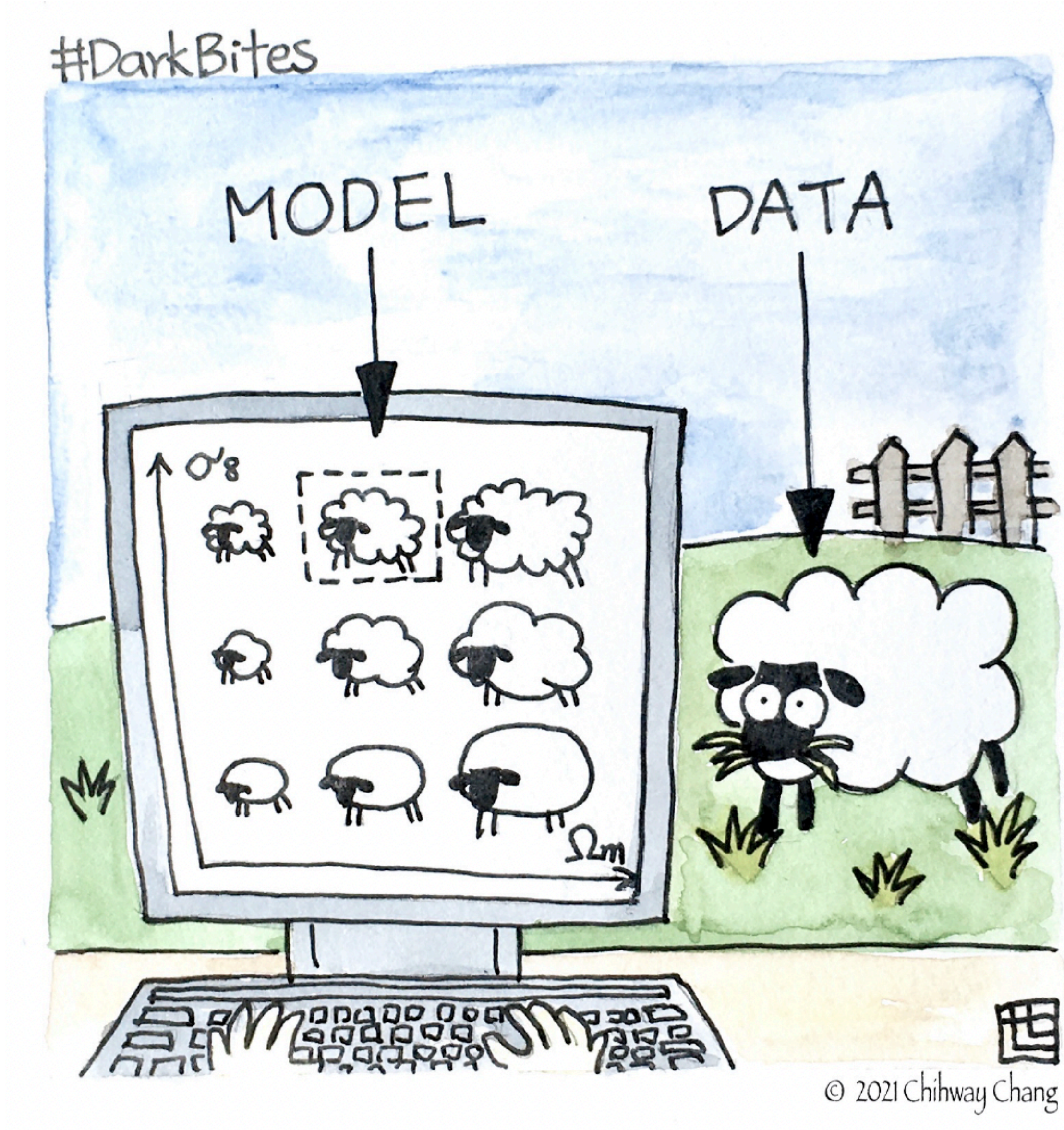
$$E_n = \int_0^\infty \frac{\mathrm{d}\ell\,\ell}{2\pi} C_{\mathrm{EE}}(\ell)\,W_n(\ell) \qquad W_n(\ell) = \int_{\theta_{\min}}^{\theta_{\max}} \mathrm{d}\theta\,\theta\,T_{+n}(\theta)\mathrm{J}_0(\ell\theta)$$
$$B_n = \int_0^\infty \frac{\mathrm{d}\ell\,\ell}{2\pi} C_{\mathrm{BB}}(\ell)\,W_n(\ell) \qquad = \int_{\theta_{\min}}^{\theta_{\max}} \mathrm{d}\theta\,\theta\,T_{-n}(\theta)\mathrm{J}_4(\ell\theta)$$

Choosing a 2pt Statistic





From Summary Statistics to Parameters



What is Probability?

Classical: Probability as frequency.

Probability of an event := *the number of times the event occurs over the total number of trials, in the limit of an infinite series of equiprobable repetitions.*

model is fixed, data are repeatable

Bayesian: Probability as degree of belief.

Probability is a measure of the degree of belief about a proposition.

data are fixed, model is repeatable

Trotta: Bayes in the sky, 0803.4089

Bayesian and Frequentist statistics

Frequentist: model is fixed, data are repeatable
Bayesian: data are fixed, model is repeatable

Say H0 = (72 ± 2) km/s/Mpc. Then:

Frequentist: Performing the same procedure with independent data will cover the real value of H0 within the limits 68% of the time.
(Limited practicability in cosmology…)

Bayesian: the posterior distribution for H0 has 68% if its integral between 70 and 74 km/s/Mpc. The posterior can be used as a prior for future analyses of independent data.

Bayesian Parameter Inference

Bayes Law:

Posterior

$$P(p|dM) = \frac{P(d|pM)P(p|M)}{P(d|M)}$$

Prior

Likelihood

$$\propto P(d|pM)P(p|M)$$

Model

Observed data

Parameters

What you know after the experiment (posterior)
= what you knew before (prior)+ what you learn (likelihood)

Priors

- Priors quantify what you knew about the parameters before the experiment

    Theoretical limits, preferences, things that must be true (e.g., from previous experiments)

- In regions where the likelihood is zero your prior doesn't matter for parameter estimation, *but can for more advanced model selection*

- It is common practice in cosmology to use uniform priors for most parameters

    easy to write down, hard to justify

→**Sensitivity analysis**: change priors, check how your conclusions change!

Transformed Priors

- In particular note that a uniform prior in one parametrisation may not be uniform in another - probability mass is conserved, not density:

$$\int P(x)\mathrm{d}x = \int P(u)\mathrm{d}u$$

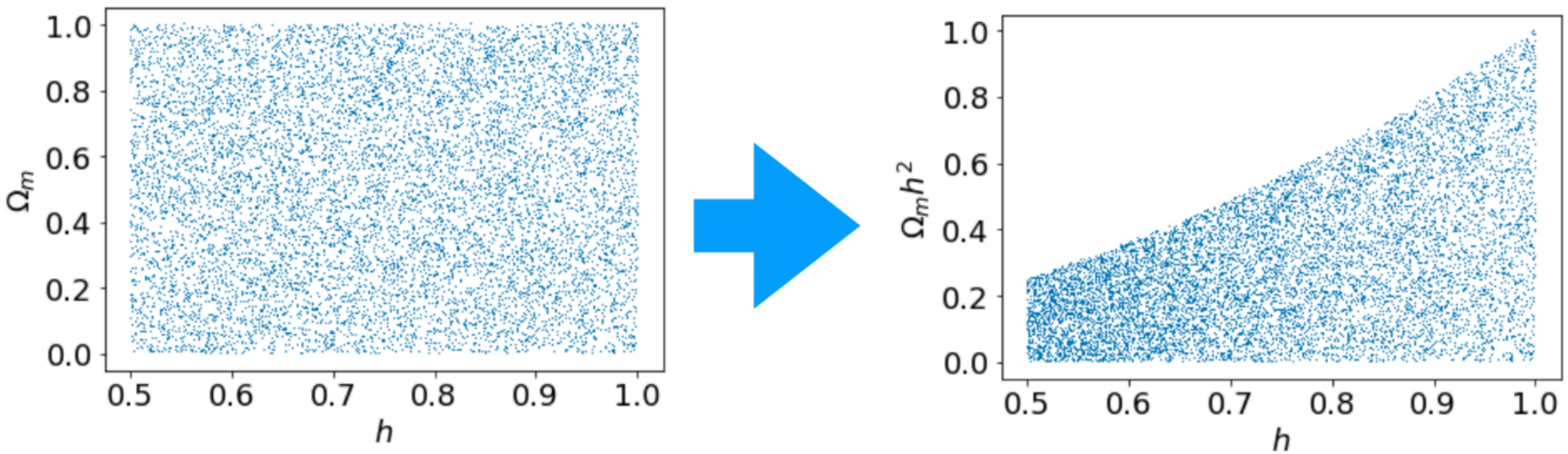
- So when we transform to new variables:

$$P(u) = P(x) \;/\; \left|\frac{\mathrm{d}u}{\mathrm{d}x}\right|$$

- or in higher dimensions:

$$P(\mathbf{u}) = P(\mathbf{x})/|J| \qquad\qquad J_{ij} = \frac{\partial u_i}{\partial x_j}$$

Transformed Priors



Jointly uniform priors on Ω_m - h

Implied priors on $\Omega_m h^2$ - h

Likelihoods

Most existing cosmological analyses assume Gaussian likelihood

$$\ln \mathcal{L}(\mathbf{D}|\mathbf{p}) \propto -\frac{1}{2}\left[(\mathbf{D}-\mathbf{M}(\mathbf{p}))^{\tau}\,\mathbf{C}^{-1}\,(\mathbf{D}-\mathbf{M}(\mathbf{p}))\right]$$

Assumes data points are Gaussian-distributed around the truth – reasonableness depends on type of measurement and sources of noise.

Alternatives:

- non-Gaussian likelihood (explored in e.g. Lin et al. 2019, Hall & Taylor 2022) – low on the priority list *for 2pt statistics.*
- Likelihood-free Inference (LFI), Simulation-base Inference (SBI)

Sampling the Likelihood

For most data sets, likelihoods cannot be written in a simple closed form equation.

We cannot just evaluate/plot posteriors directly, but instead musts indirect methods.

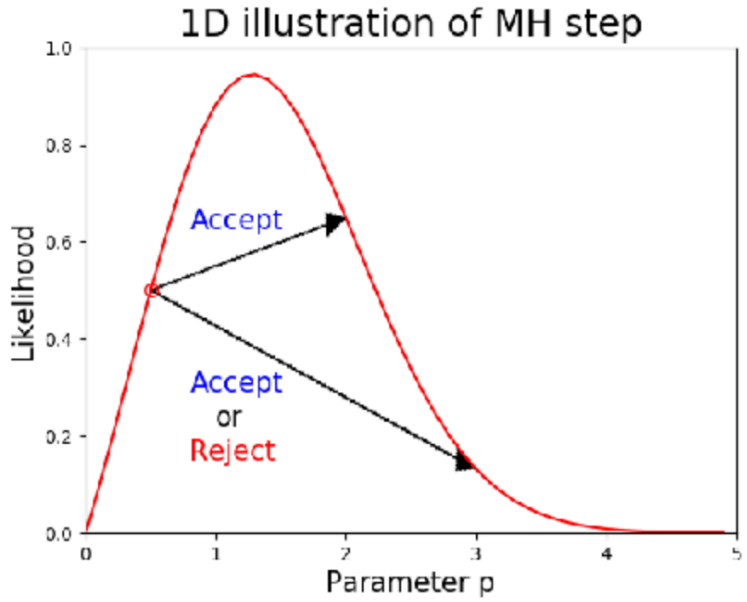
Most obvious solution is to evaluate at every point in the space, or a grid. Impossible for high-dimensional parameter spaces!

→ sampling methods like Monte-Carlo Markov Chains.

- each element of Markov Chain depends only on the previous one
    - basic algorithm: Metropolis–Hastings
        - improved in widely used packages `Emcee`, `Zeus`
    - limitations: lack of definitive the chain has converged

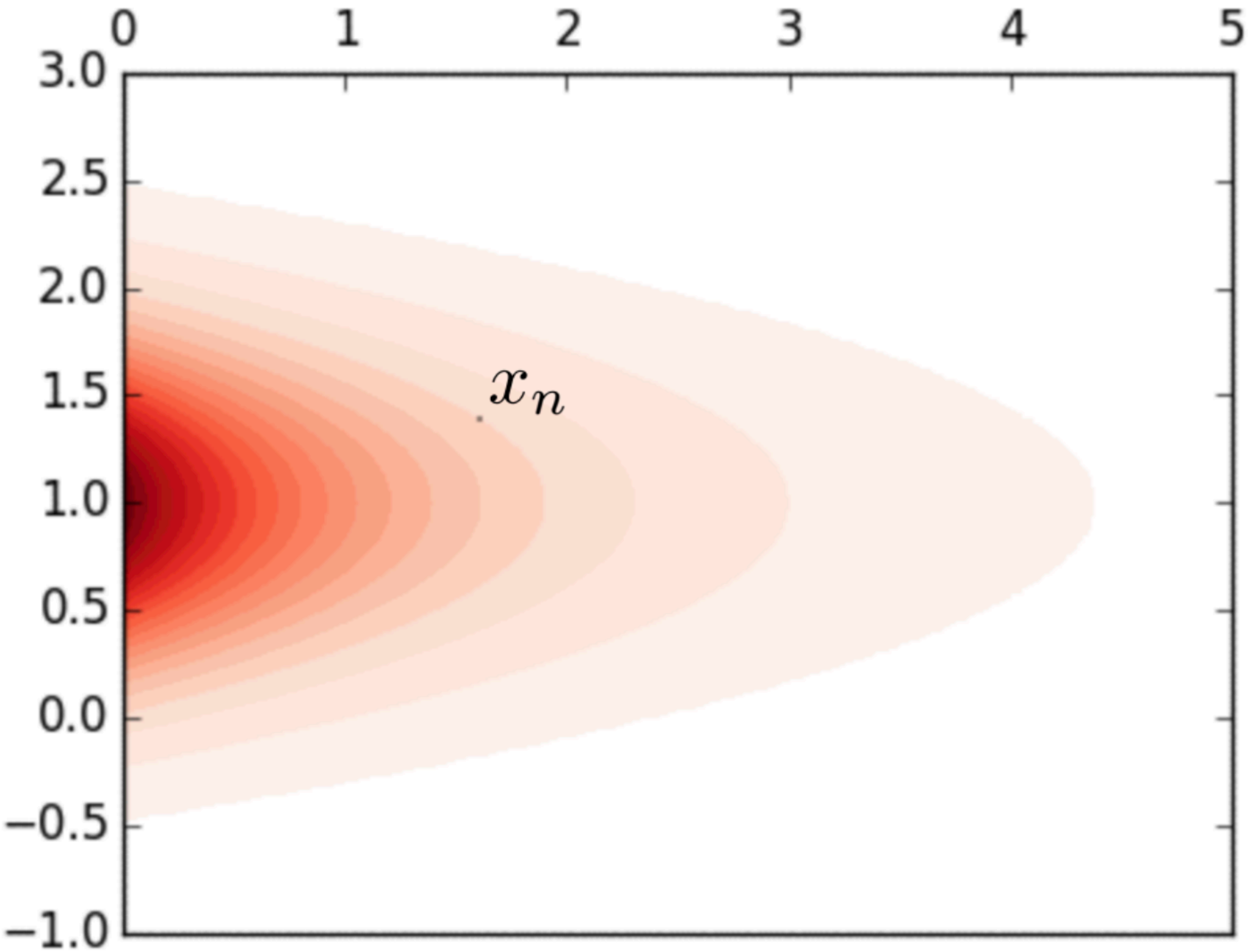
The Metropolis-Hastings algorithm

- at step t, at some parameters p_t
- propose move to $p_t' = p_t + \Delta p_t$ (randomly draw Δp_t)
- evaluate $r = L(p_t')/L(p_t)$
- MH step:
  - if r > 1 accept move
  - if r < 1 generate a random number $\alpha \in [0, 1]$
    - if $\alpha < r$, accept move
    - if $\alpha > r$, reject move
- t=t+1

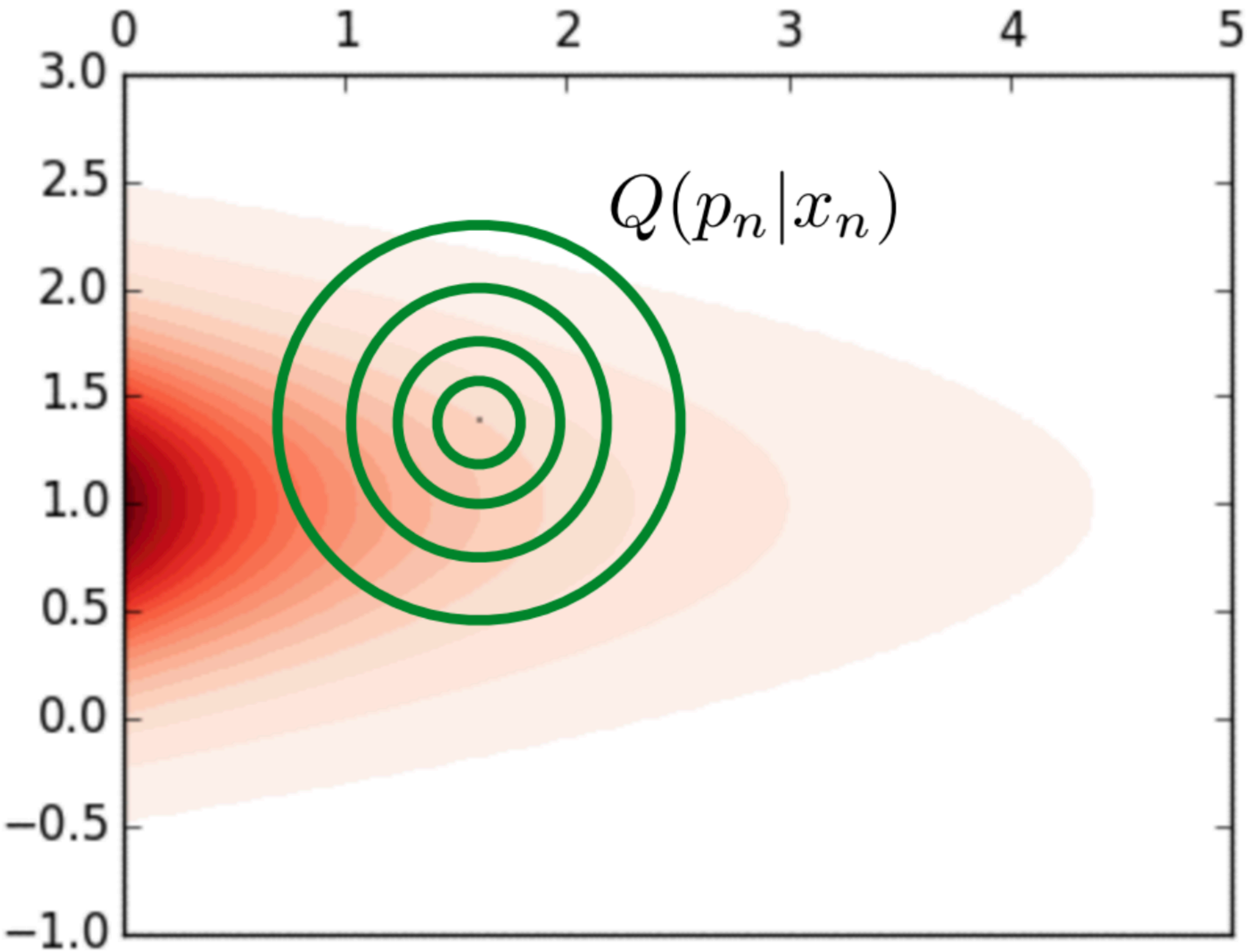


r is chosen to fulfill detailed balance → algorithm asymptotically recovers the true posterior

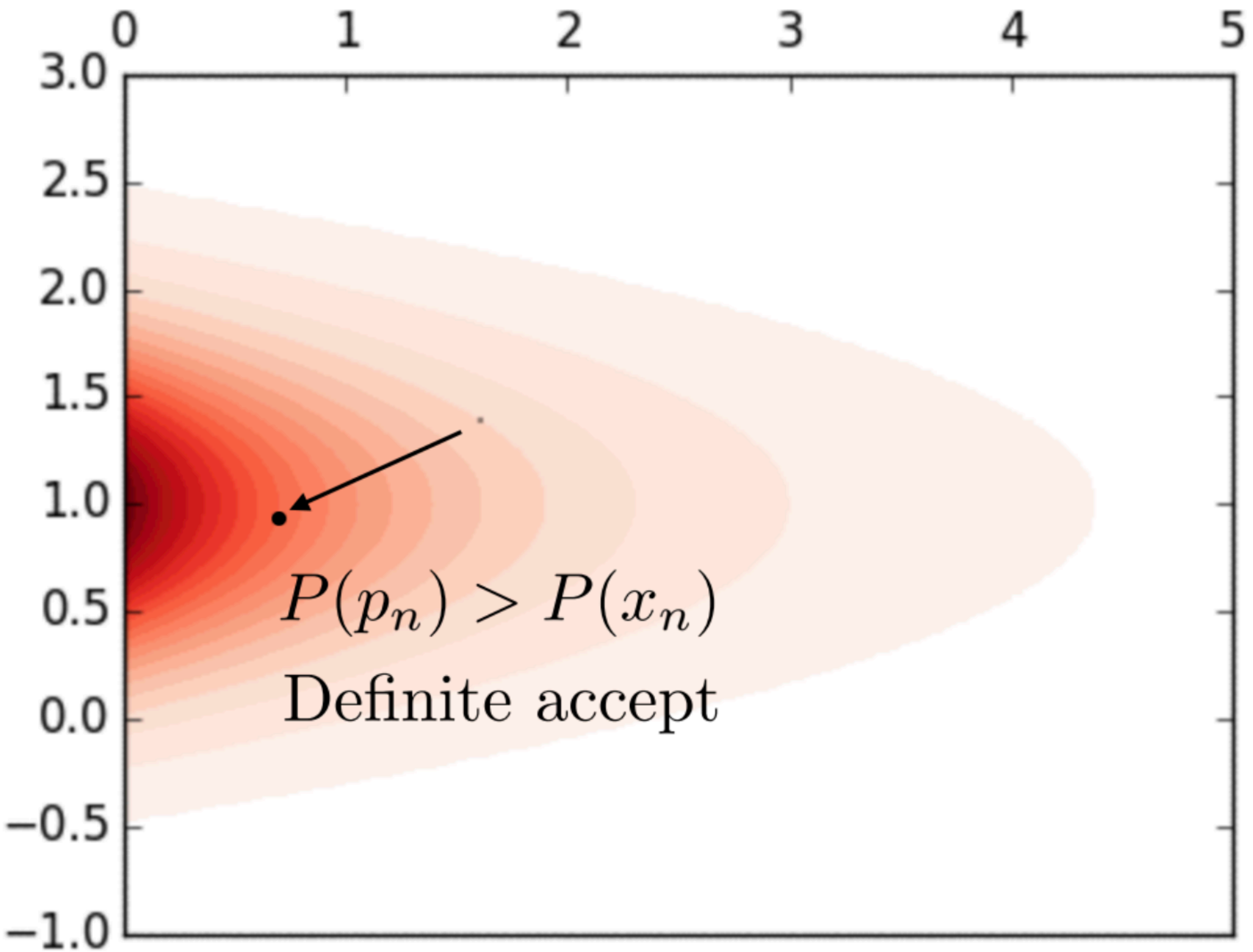
The Metropolis-Hastings algorithm



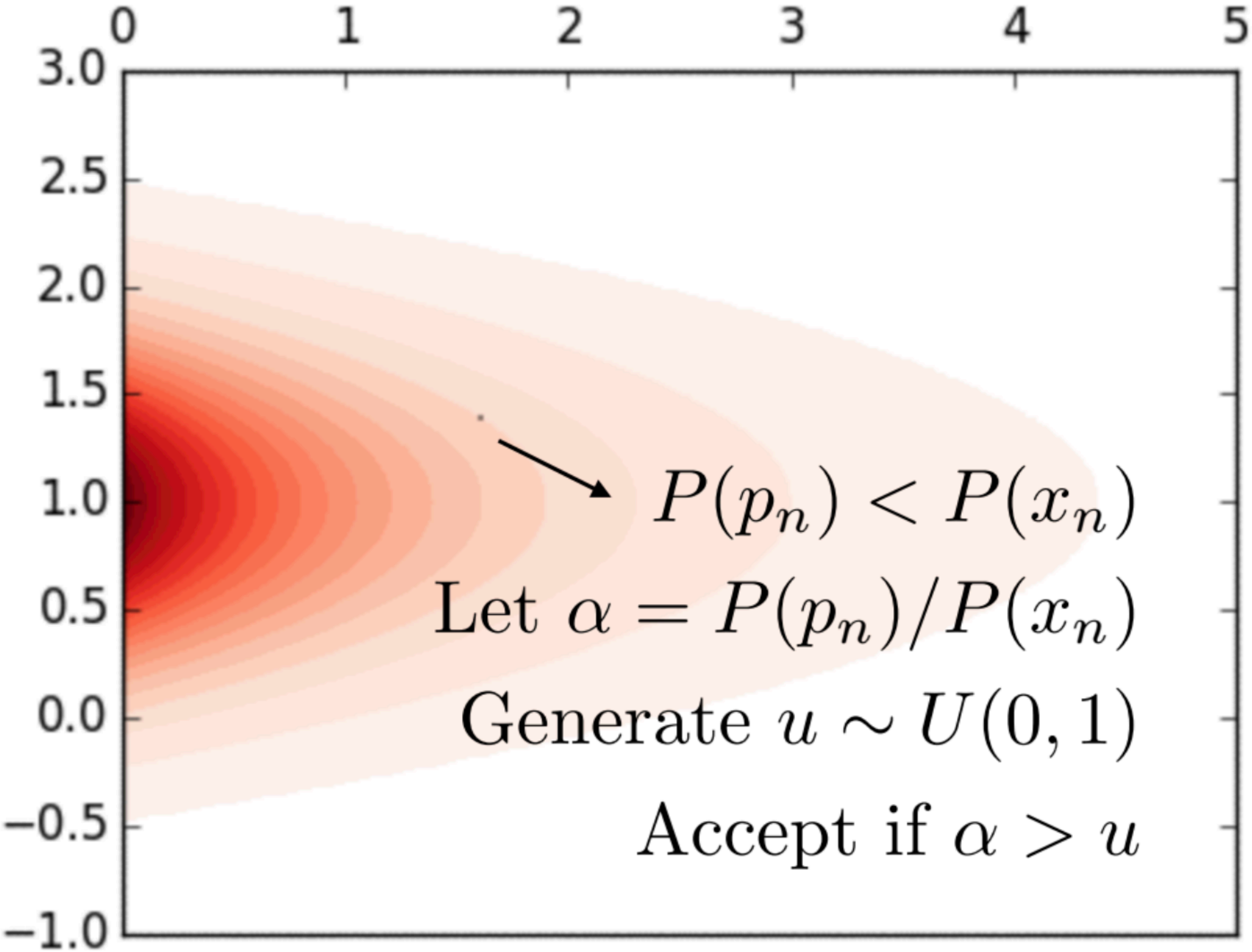
The Metropolis-Hastings algorithm



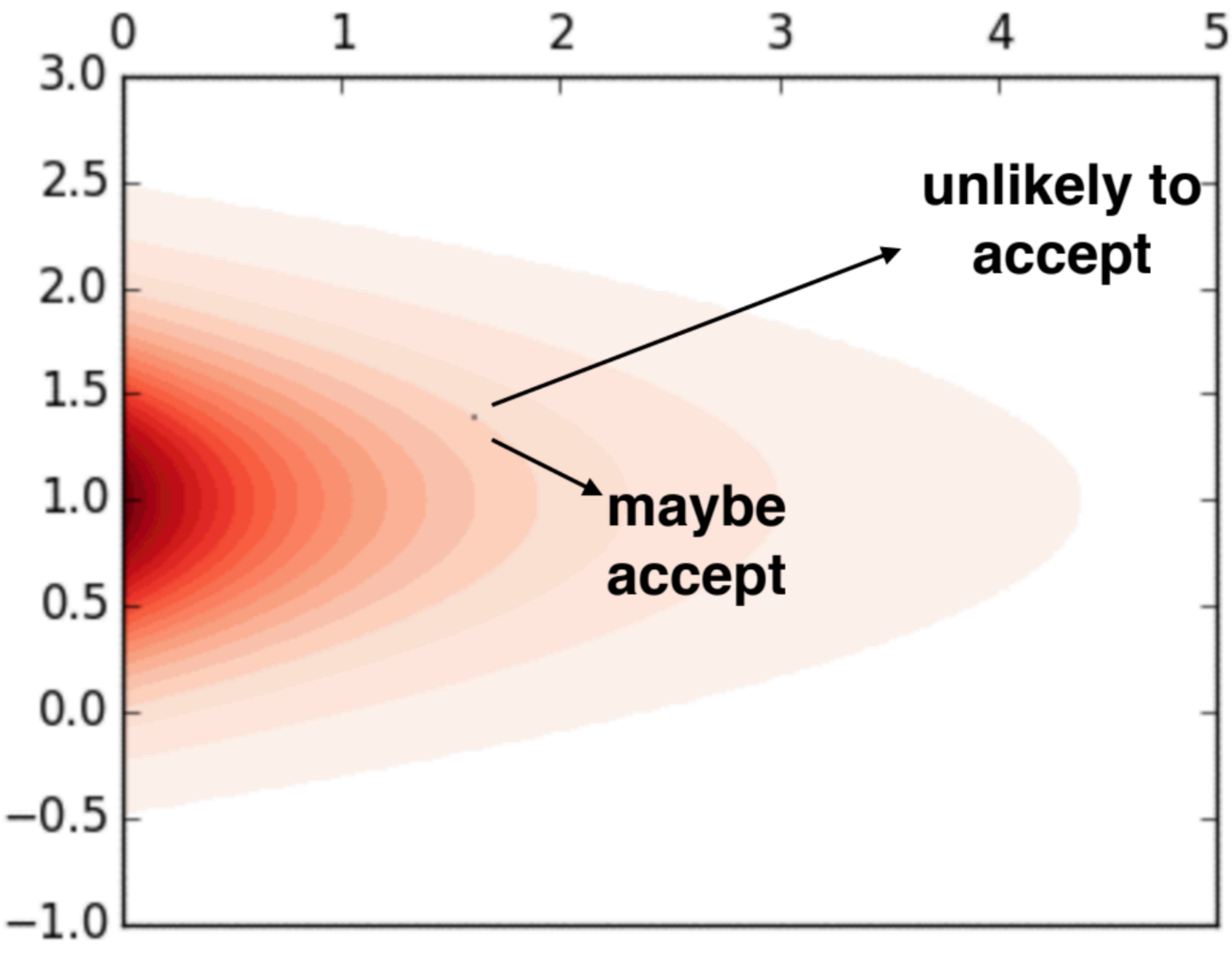
The Metropolis-Hastings algorithm



The Metropolis-Hastings algorithm

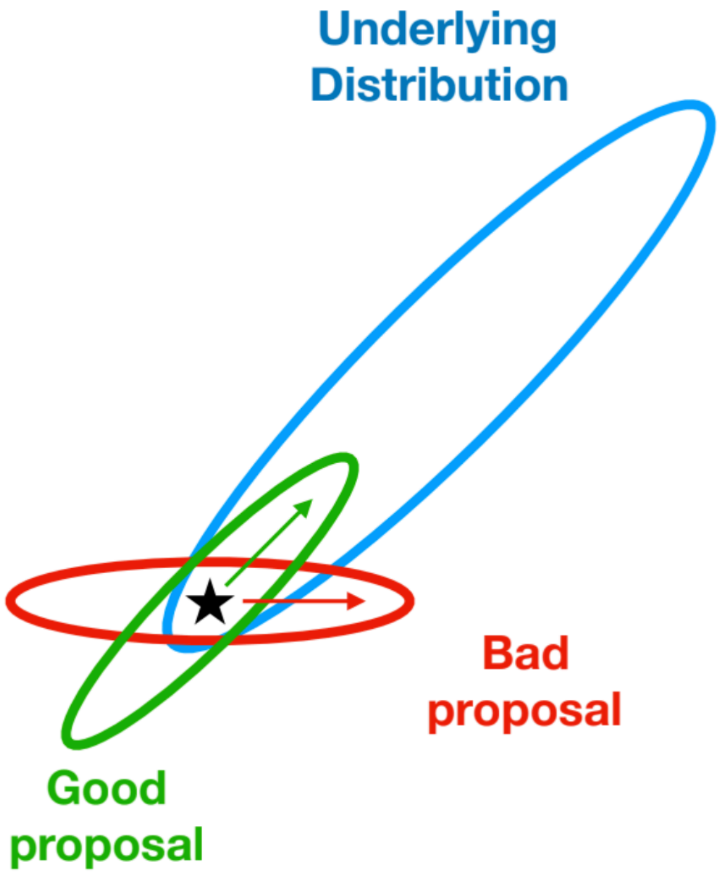


The Metropolis-Hastings algorithm



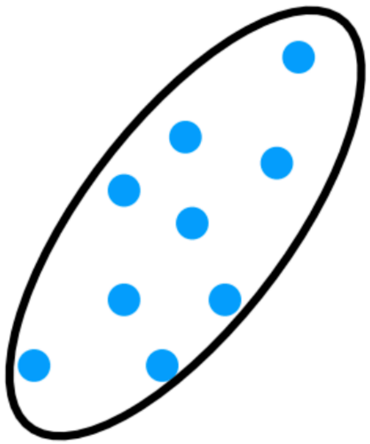
The Metropolis-Hastings algorithm

- Efficiency of MH depends dramatically on how good the proposal is
- A bad proposal will not converge in any practical length of time
- The ideal proposal matches the shape of the underlying distribution
    - We don't know this, but can look for best approximation



The Metropolis-Hastings algorithm

- One way to get a good proposal is by tuning
    - Run a short initial chain to estimate covariance
    - Use this covariance to initialise the next iteration
- You have to throw away the first chain, and only use samples from when your tuning was finished
    - Detailed balance broken
    - There are specific algorithms that do let you do a variant of this, but not standard MH



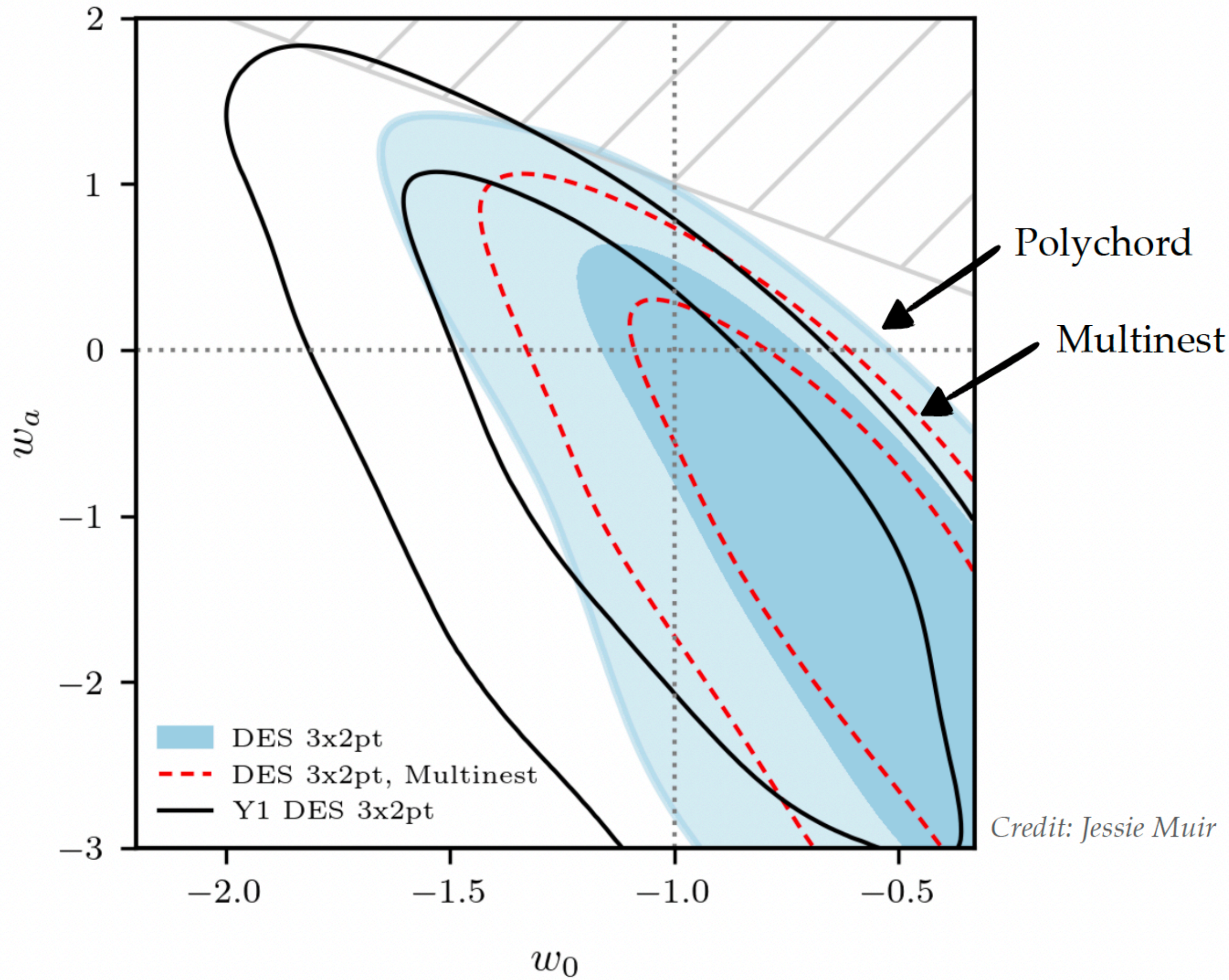
Intricacies of High-Dimensional Sampling

Nested sampling: starts with a large number of points, and repeatedly eliminates and find new replacement points

- e.g. Multinest, PolyChord
- calculate Bayesian evidence simultaneously

Choosing the right sampler to accurate sample your parameter space is an art - *and hard validation work.*

Intricacies of High-Dimensional Sampling



Credit: Jessie Muir

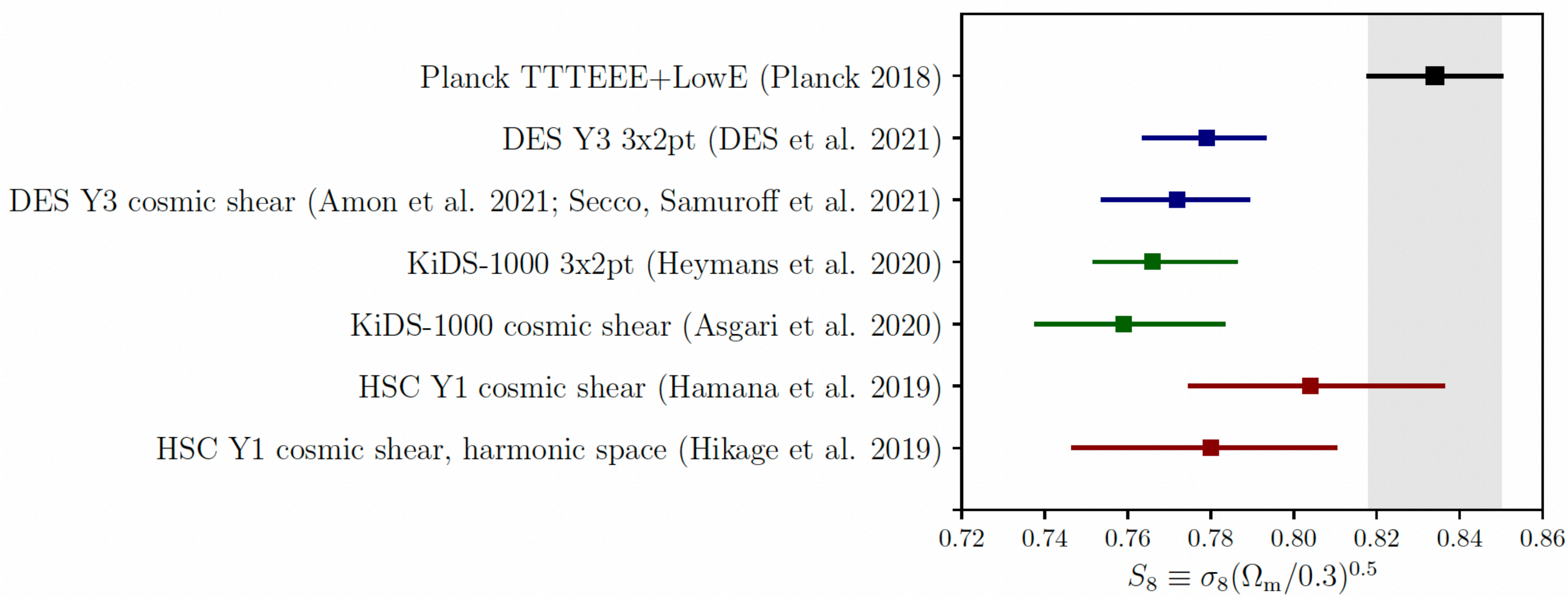
Interpreting chains

- Check to see if we actually found a good fit
- Quote the cosmological constraints, check to see if we've broken ΛCDM yet
- Compare with other similar measurements
- Compare with other independent measurements

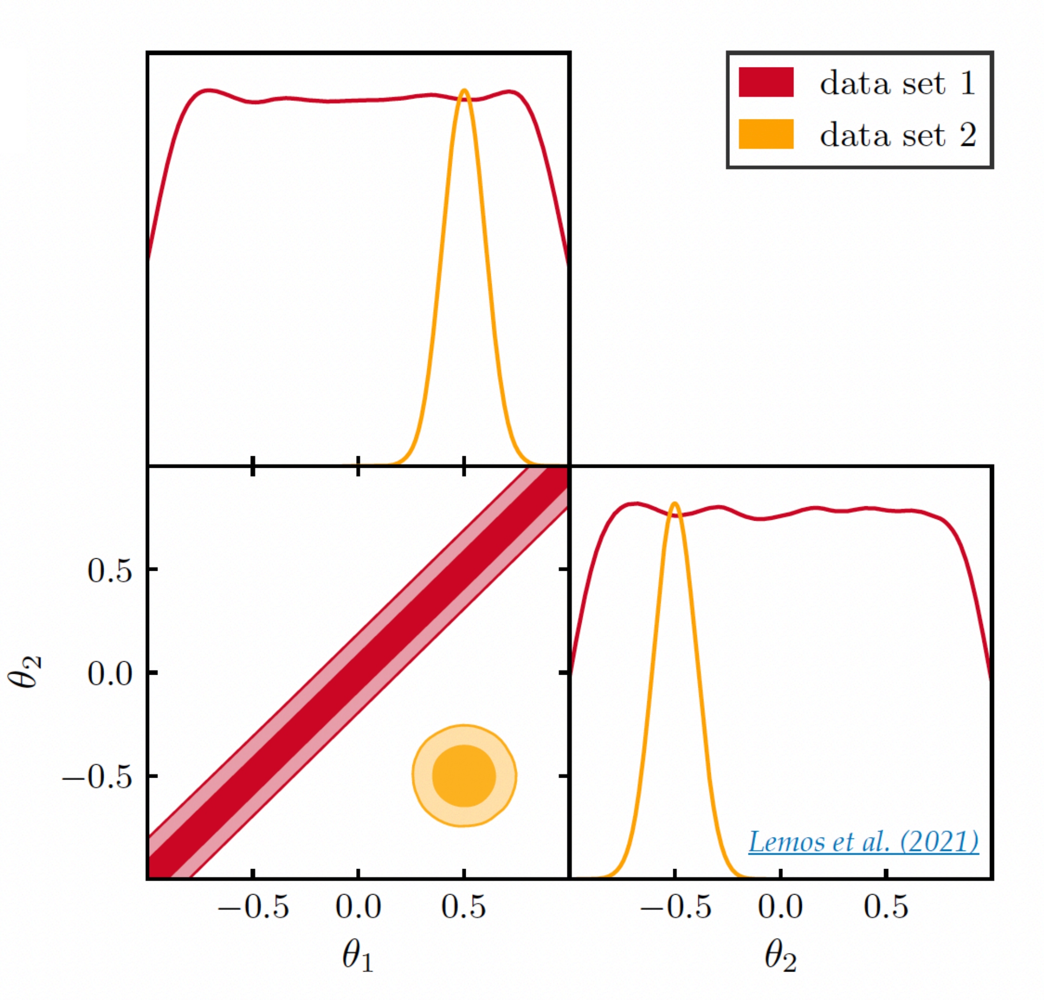
Interpreting chains

Can only plot 1D/2D results - report marginalized constraints.

$$P(\theta_1|d) = \int d^{n-1}\theta_{2..n} P(\boldsymbol{\theta}|d)$$



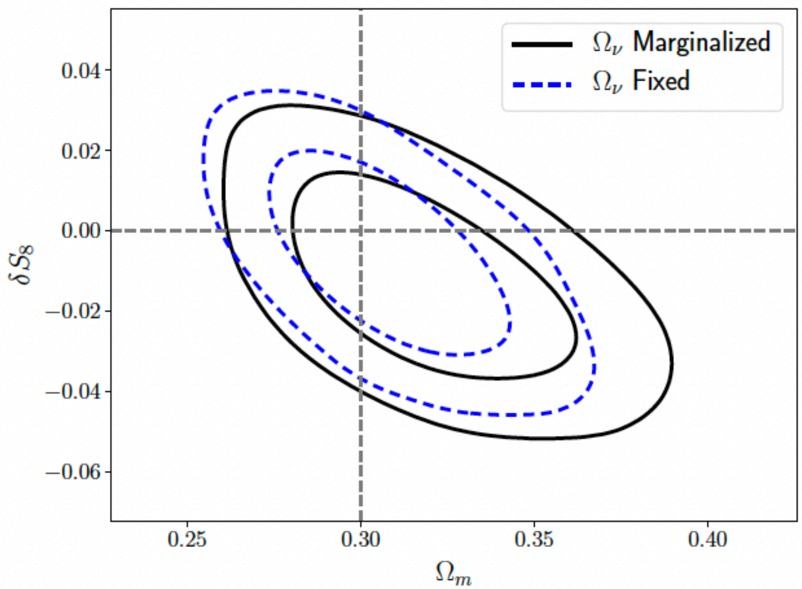
Marginalized Parameters

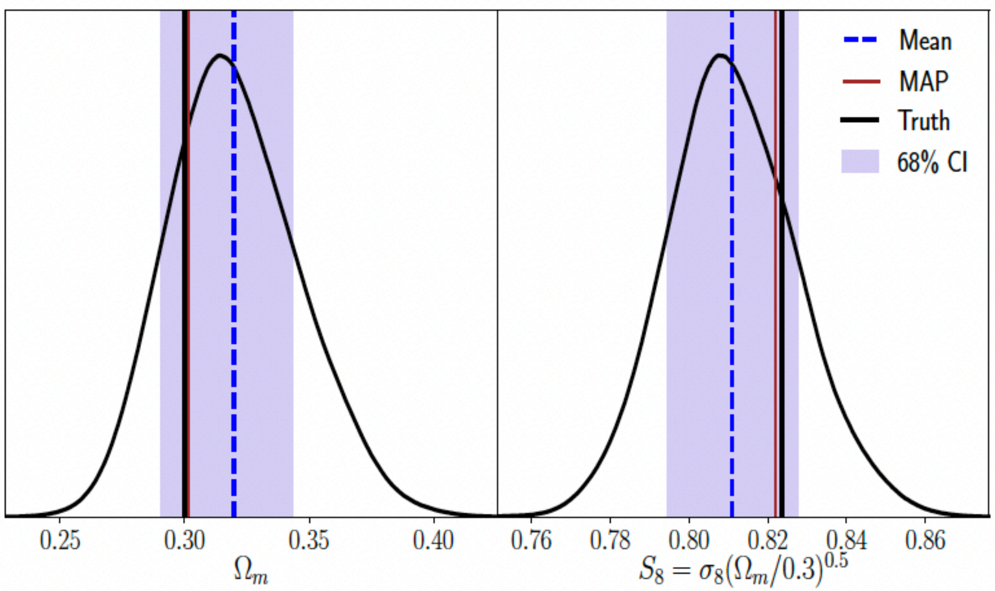


Beware of Projection/Prior Volume Effects!

Parameters of interest may be correlated with poorly constrained "nuisance parameters".

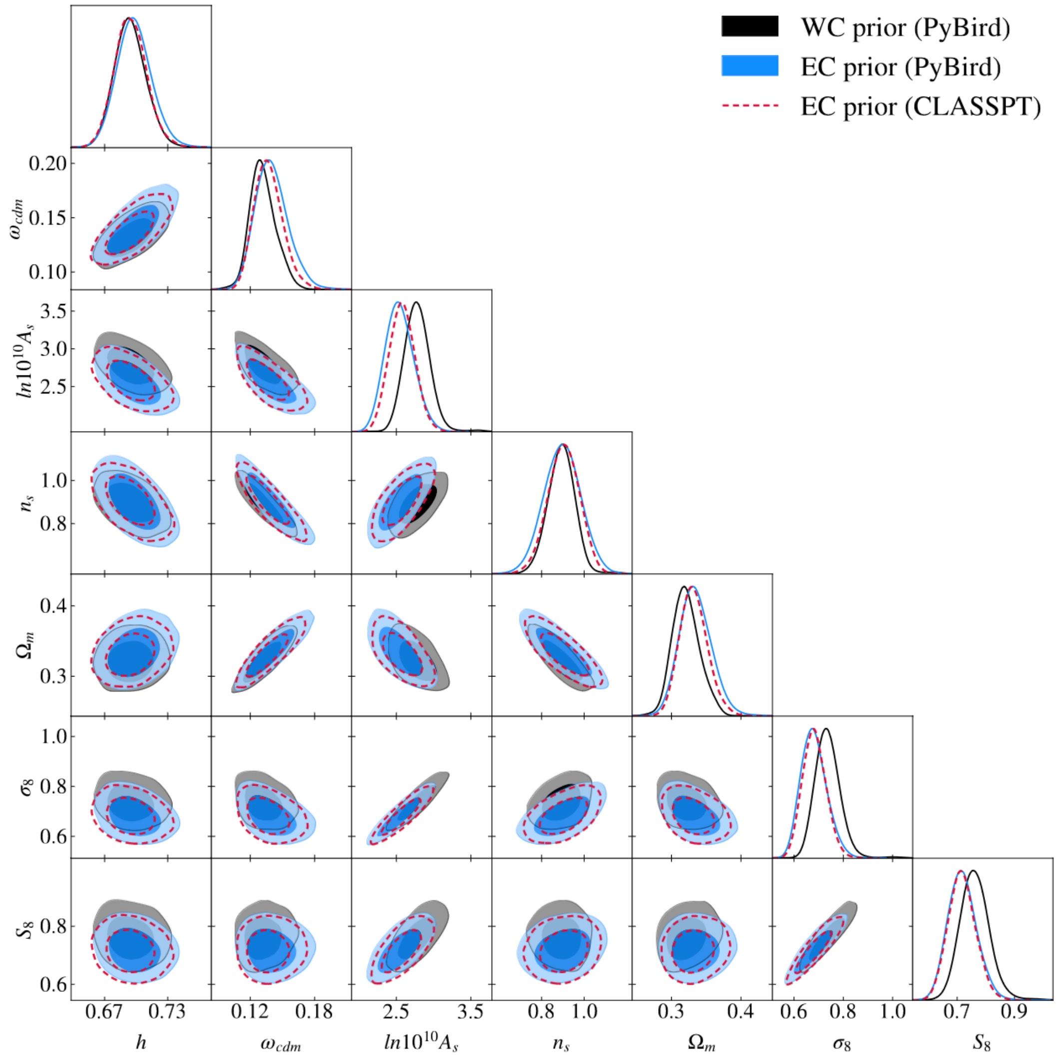
Marginalization may introduce projection effects, skew marginalized posteriors away from best fit.





This effect can be characterized on synthetic data!

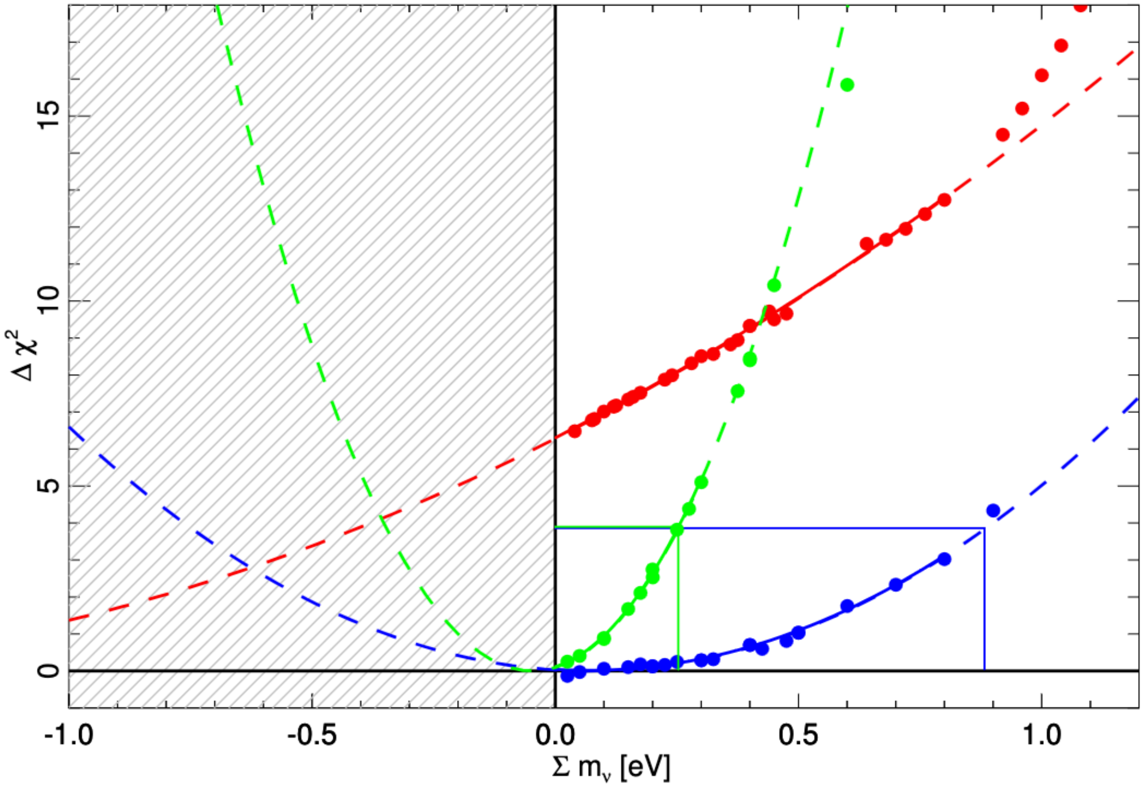
Beware of Projection/Prior Volume Effects!



Simon et al. 2023: EFTofLSS analyses of BOSS data with different nuisance parameter priors.

Profile Likelihoods

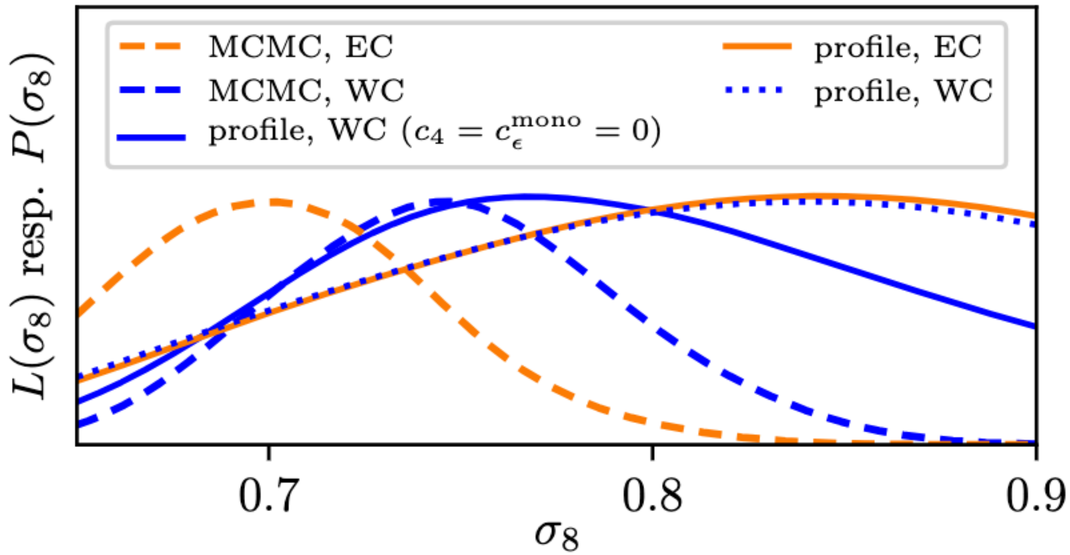
Frequentists' way to treat nuisance parameters ν $L(\boldsymbol{\theta}) = \max_{\boldsymbol{\nu}} L(\boldsymbol{\theta}, \boldsymbol{\nu})$



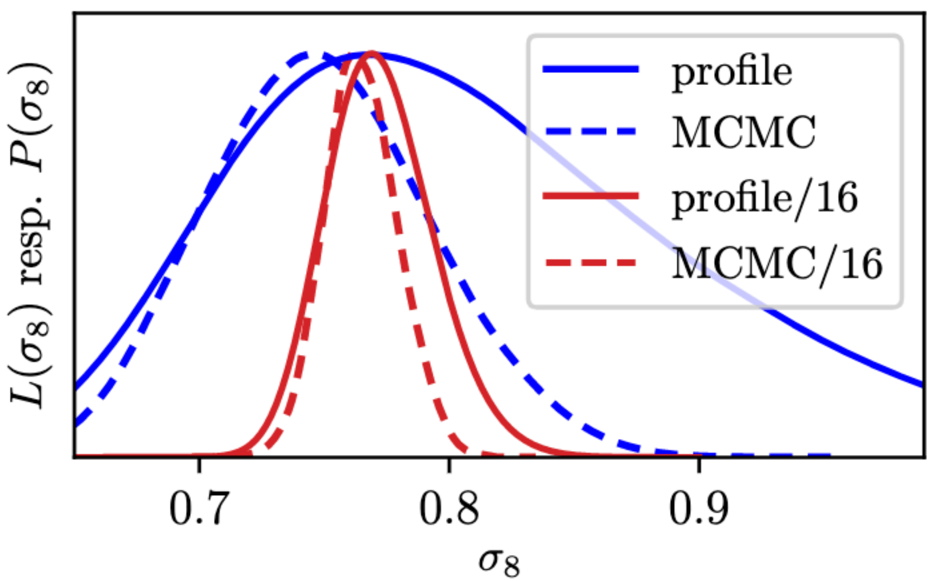
CMB
CMB+lensing
CMB+lensing+BAO

Planck Profile Likelihood (37 parameters!) 1311.1657

Profile Likelihoods



Holm+ 2023
Comparison of MCMC, profile likelihoods for EFTofLSS BOSS analyses



Improved constraining power will reduce difference between frequentist and Bayesian statistics.

Model Comparison/Selection

- Given two models, how can we decide which fits the data better, overall?
- Simplest approach: compare best fit points
    - Does not include uncertainty or Occam's Razor
- Recall that all our probabilities have been conditional on the model, as in Bayes:

$$P(p|M) = \frac{P(d|pM)P(p|M)}{P(d|M)}$$

Model Comparison/Selection

- Evidence is the bit we ignored before when doing parameter estimation

- Given by an integral over prior space

$$P(d|M) = \int P(d|pM)P(p|M)\mathrm{d}p$$

- Hard to evaluate - posterior usually small compared to prior

Model Comparison/Selection

- Can use Bayes Theorem again, on model level:

$$P(M|d) = \frac{P(d|M)P(M)}{P(d)}$$

- Only really meaningful when comparing models:

Bayesian Evidence Ratio

$$R = \frac{P(M_1|d)}{P(M_2|d)} = \frac{P(d|M_1)}{P(d|M_2)} \frac{P(M_1)}{P(M_2)}$$

Model Priors

Bayesian Evidence Values

Bayesian Model Selection

Nominally, M_1 is favored with R:1 odds over H_2 .

Jeffreys scale R > 3.2 substantial evidence, R > 10 strong evidence

But: need to recalibrate for prior volume (⟶ numerous analyses of simulated noisy realizations)

For combining data sets:

M_1 = 'data sets described by same model parameters'

M_2 = 'data sets described by different model parameters

$$R = \frac{P(D_1, D_2|M)}{P(D_1)|M)P(D_2|M)}$$

combine if R>[threshold for inconsistency]*

*agree on this before the analysis to avoid confirmation biases

Comparing Experiments

We like to quantify to what extend our results are consistent with other experiments.

Complicated since we are comparing two chains in very high dimension, and the effect of priors are non-trivial.

In the past few years, many have devised certain statistics ("tension metrics") to quantify how likely the two experiments are realizations drawn from the same underlying universe.