Hierarchical Probabilistic Inference of Cosmic Shear

Astronomy in the 2020s: Synergies with WFIRST

Michael D. Schneider
with Josh Meyers and Will Dawson

June 27, 2017

Collaborators:
D. Bard, D. Hogg, D. Lang, P. Marshall, K. Ng

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

Summary

- WFIRST & LSST are ideally suited for a joint cosmic shear measurement to constrain cosmological parameters and dark energy
- New shear inference methods are required to fully exploit the sensitivities of WFIRST & LSST
- A hierarchical probabilistic forward model approach shows the most promise for meeting shear bias requirements
- These probabilistic algorithms enable both:
  - Exploitation of information in new cosmological statistics
  - More flexible computing pipelines to ingest and interpret new data

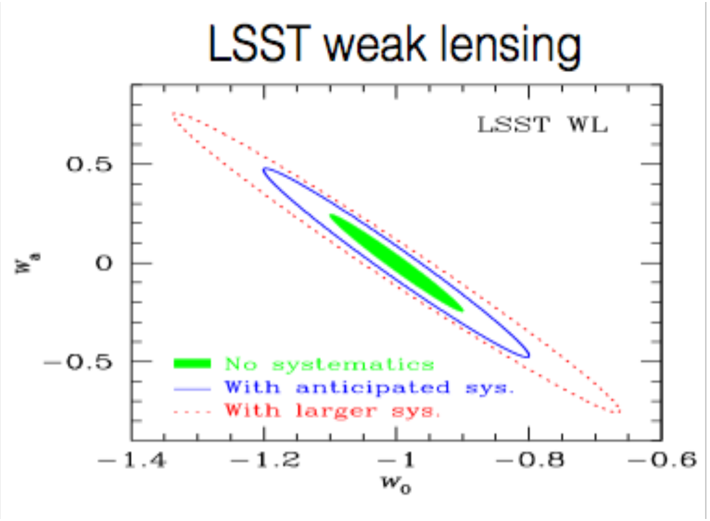
Introduction

Cosmic shear measurement

- The lensing by large scale structure
- Looking for very small signal under very large amount of noise
- We **don't know "unsheared" shapes**, but can (roughly) assume they are isotropically distributed
- Cosmic shear distorts statistical isotropy; galaxy ellipticities become correlated
- Exquisite probe of DE, if systematics can be controlled
- LSST: will measure few billion galaxy ellipticities. Excellent sensitivity to both DE and systematics!

Cosmic shear signal is comparable to ellipticity of the Earth, ~0.3%

- D. Wittman



What will the WFIRST HLS add to cosmic shear?

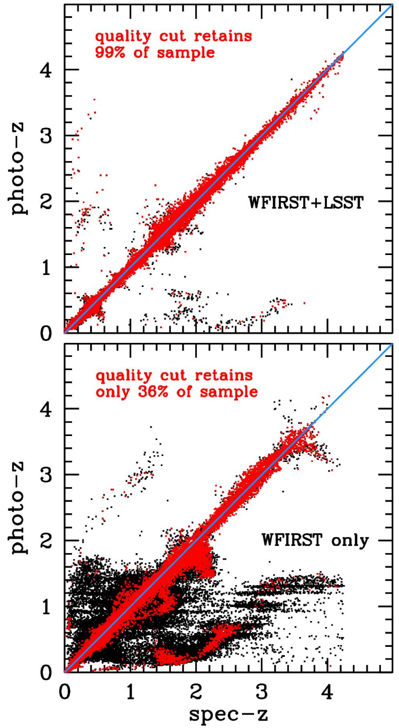
Improved tomography, reduced bias means tighter dark energy constraints

- Deblending
  - 30 – 50% of LSST galaxy detections will be multiple resolved objects as seen by WFIRST
  - Shear bias: Most blends are chance alignments of galaxies at different redshifts
  - Photo-z bias: mixed colors / biased photometry
  - Assert number & properties of blend components given HLS overlap with LSST
  - Train statistical calibration for entire LSST footprint

- Improved photo-z's from LSST optical + WFIRST NIR bands

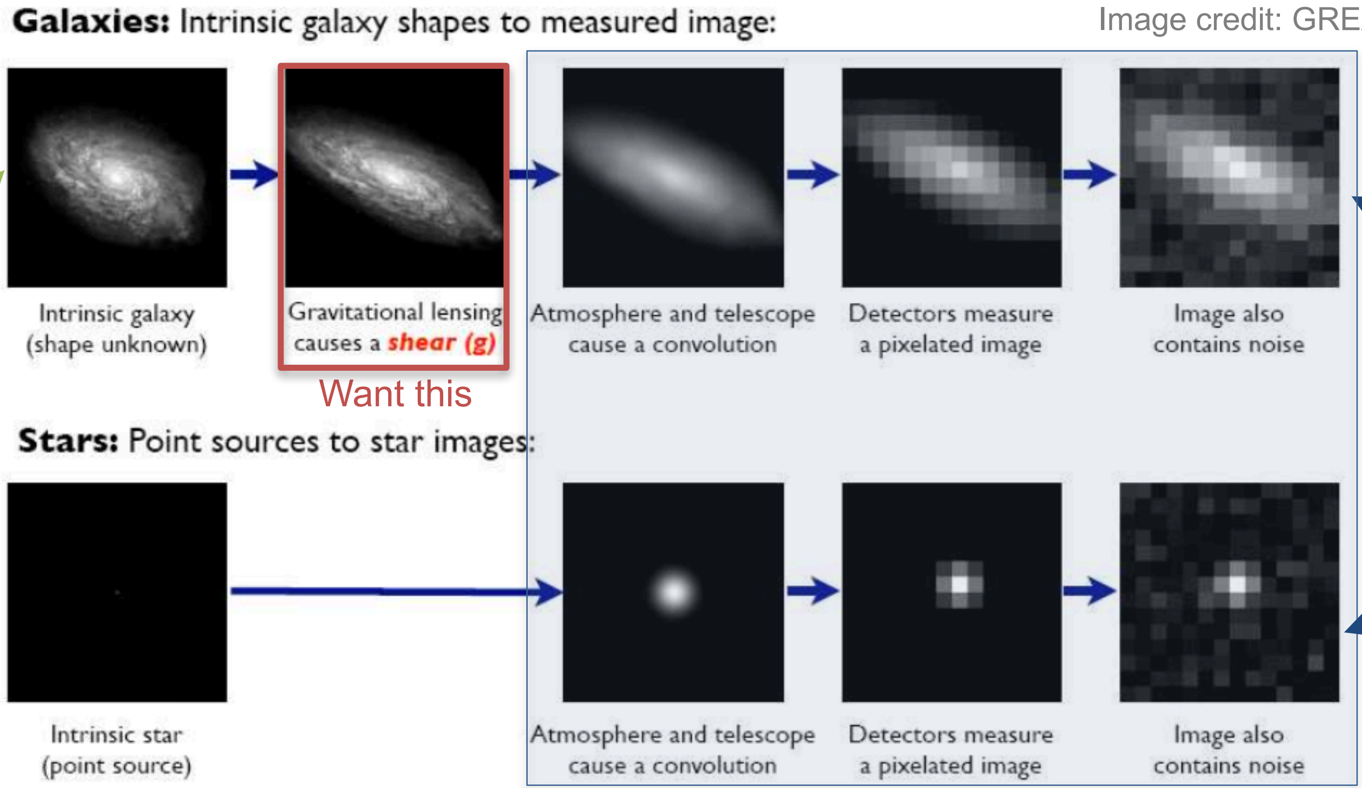
- Higher fidelity LSS cross-correlations (from grism survey)
  - Break many systematics and cosmological parameter degeneracies

- *Reduced shear bias?*



Weak lensing of galaxies: the forward model

Image credit: GREAT08, Bridle et al.



Unknown & dominates signal

Want this

Marginalize

Constrained by

We won't just have more data with 'Stage IV' surveys.
- We're in an era with qualitatively new computing capabilities

Dijkstra's Law

A quantitative difference is also a qualitative difference if the quantitative difference is greater than an order of magnitude.

Microprocessor Transistor Counts 1971-2011 & Moore's Law

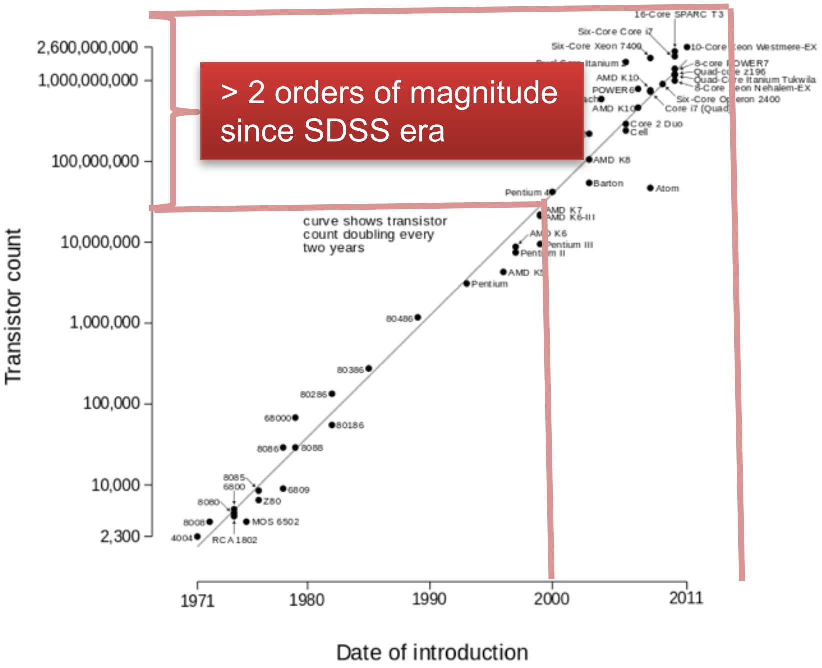


Figure credit: Wikipedia

Qualitative changes in computing enable new scientific methods

"…predictive simulation has brought together theory and experiment in such a compelling way that it's fundamentally extended the scientific method for the first time since Galileo Galilei invented the telescope in 1609…"

- Mark Seager, CTO for the HPC Ecosystem at Intel

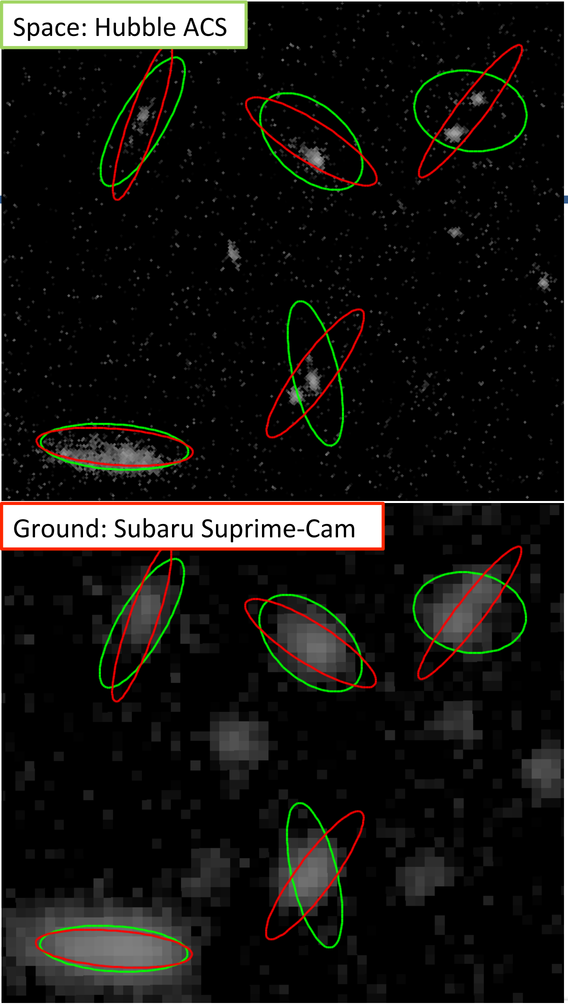
(interview in Inside HPC on June 6, 2016)

Data + Compute convergence in cosmology – DOE ASCR initiative, April 2016

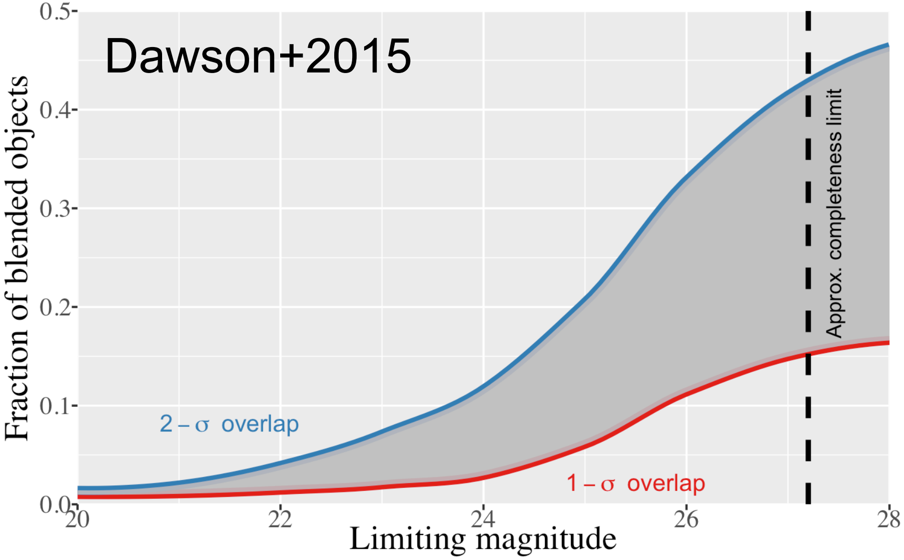
- We're facing **systematics-limited** measurements
  - End-to-end simulations of the experiment are the best approach to improve accuracy & precision
  - Ties data and simulation more intricately than in past cosmology pipelines
- Image and catalog **summary statistics** are no longer good enough to meet next generation science requirements
  - Probabilistic hierarchical models and related machine-learning approaches show promise but are much more computationally intensive
  - Potential changes to the traditional 'facility' / 'user' separate analysis stages

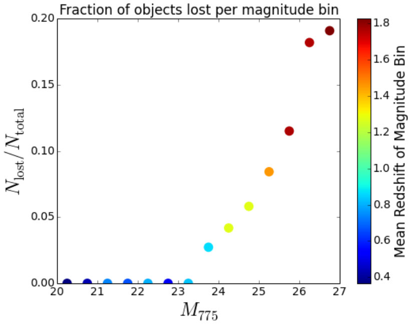
Removing the line between 'analysis' and 'simulation'.

Catalog cross-matching between space and ground is confused by significant object blending as seen by LSST



LSST blend fractions estimated from Subaru & HST overlapping imaging

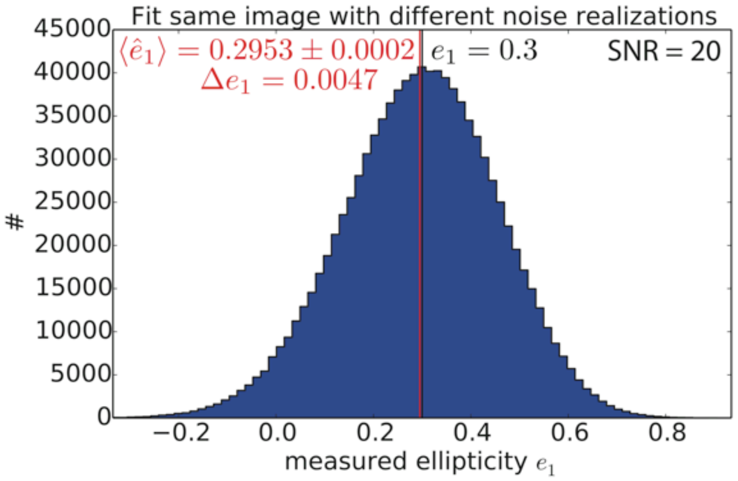


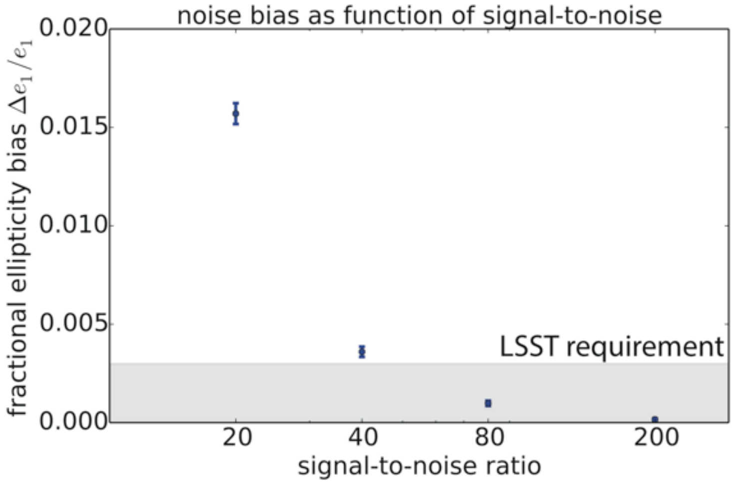


Shear bias

Shape to Shear: Noise Bias

- Ellipticity: $e = \dfrac{a-b}{a+b}\exp(2i\theta)$
- Ensemble average ellipticity is an unbiased estimator of shear.
- However, maximum likelihood ellipticity in a model fit is **not** unbiased.
- Ellipticity is a non-linear function of pixel values.





Mitigating Noise Bias – at least 2 strategies

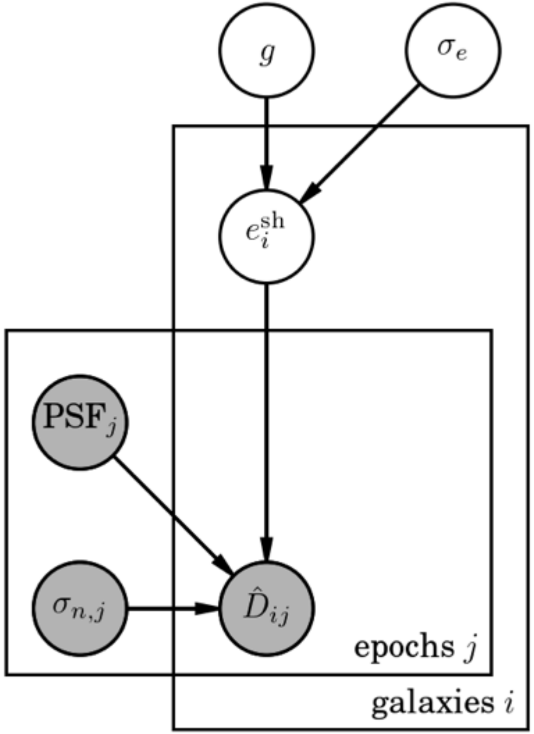
1. Calibrate using simulations. (im3shape, sfit)
   - But corrections are up to 50x larger than expected sensitivity!

2. Propagate entire ellipticity distribution function P(ellip | data)
   - Use Bayes' theorem: P(ellip | data) $\propto$ P(data | ellip) P(ellip)
   - Measure P(ellip) in deep fields. (lensfit, ngmix, FDNT).
   - Infer simultaneously with shear in a hierarchical model. (MBI).

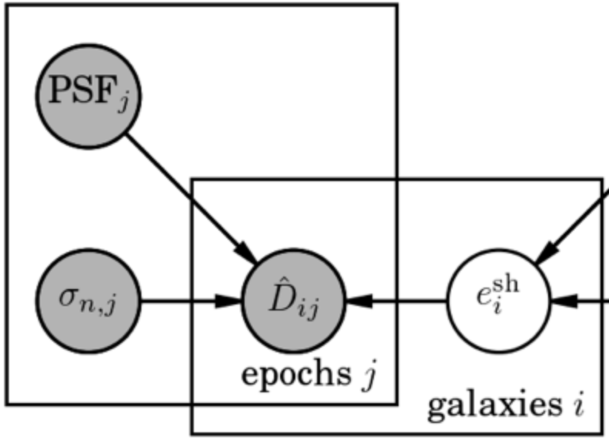
arXiv:1411.2608

A hierarchical model for the galaxy distribution

- σ_e = intrinsic ellipticity dispersion
- e^{int} = galaxy intrinsic ellipticity
- g = shear
- e^{sh} = galaxy sheared ellipticity
- PSF = point spread function
- D = model image
- σ_n = pixel noise
- D = data: observed image



Our graphical model tells us how to factor the joint likelihood



- Use a probabilistic graphical model to encode the factorization of the joint probability distribution of variables in the model.
- We don't care about e^{sh} for cosmology, so integrate it out.

$$\Pr\left(g, \sigma_e \middle| \{\mathrm{PSF}\}_j, \{\sigma_{n,j}, \{D_{ij}\}\}\right)$$

$$\propto \int d^{n_{\mathrm{gal}}}\left\{e_i^{\mathrm{sh}}\right\} \left[\prod_{ij} \Pr\left(D_{ij} | \mathrm{PSF}_j, \sigma_{n,j}, e_i^{\mathrm{sh}}\right)\right] \left[\prod_i \Pr\left(e_i^{\mathrm{sh}} | g, \sigma_e\right) \Pr(g) \Pr(\sigma_e)\right]$$

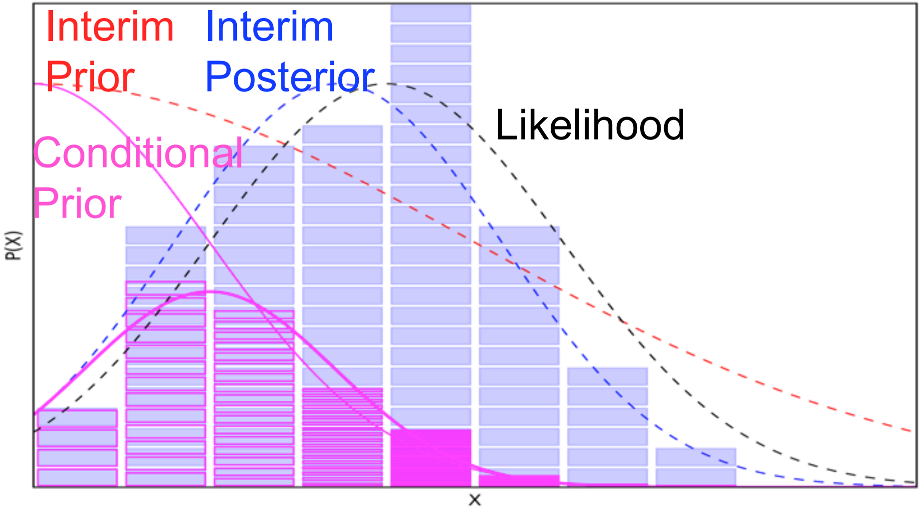
Huge complicated integral to compute for every posterior evaluation.

Importance Sampling allows tractable divide & compute

We thus estimate the *pseudo-marginal likelihood* for shear

- Don't go back to pixels for every time we sample a new g or σ_e.
- For each galaxy, draw image model parameter samples under a fixed "interim" prior. This is embarrassingly parallelizable.
- Use reweighted samples to approximate the integral via Monte Carlo.

Ongoing research question:
How many interim samples are needed?



$$\text{Draw } K \text{ samples } e^{\rm sh}_{ik} \sim P\left(e^{\rm sh}_i \middle| \hat{D}_i, I_0\right) \propto P\left(\hat{D}_i \middle| e^{\rm sh}_i\right) P\left(e^{\rm sh}_i \middle| I_0\right)$$

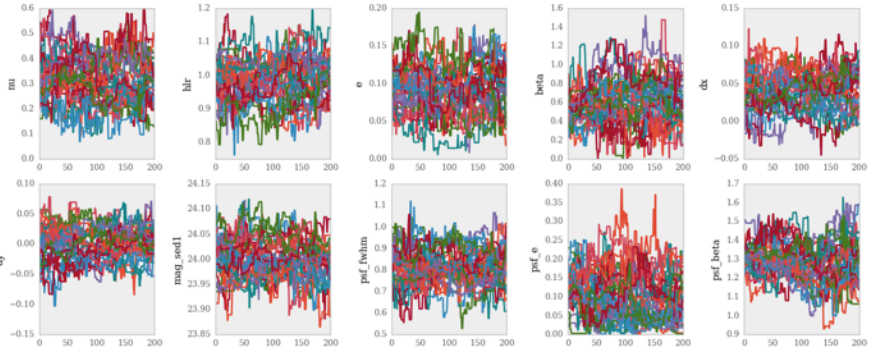
Source characterization via probabilistic image modeling

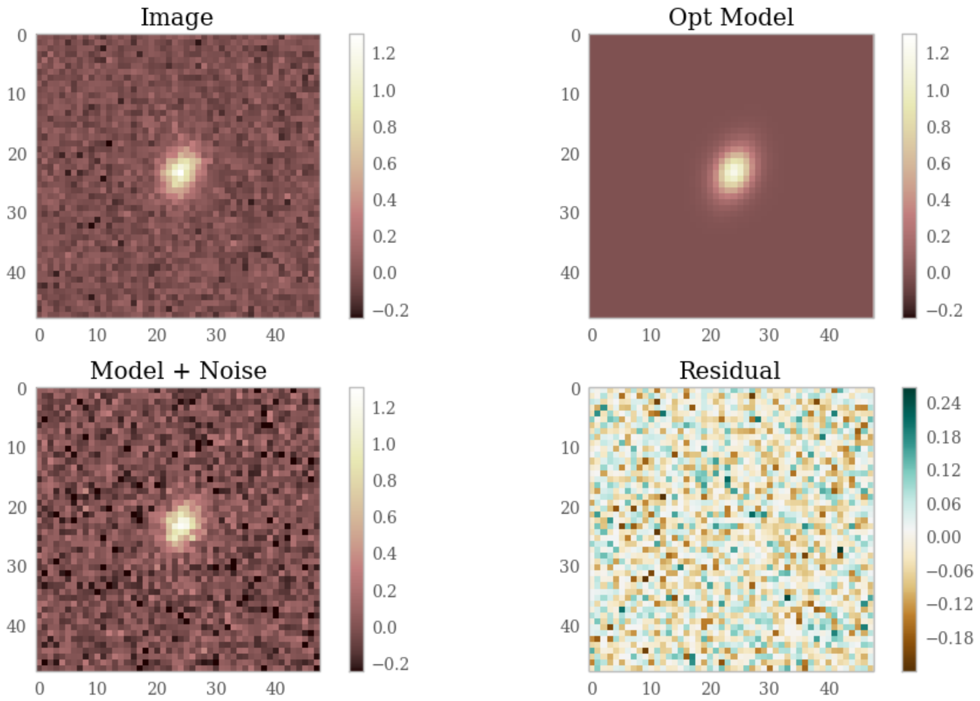
Infer image model parameters via MCMC under an interim prior distribution for the galaxy and PSF parameters.

MBI GREAT3 analysis with:

The Tractor (Lang & Hogg)

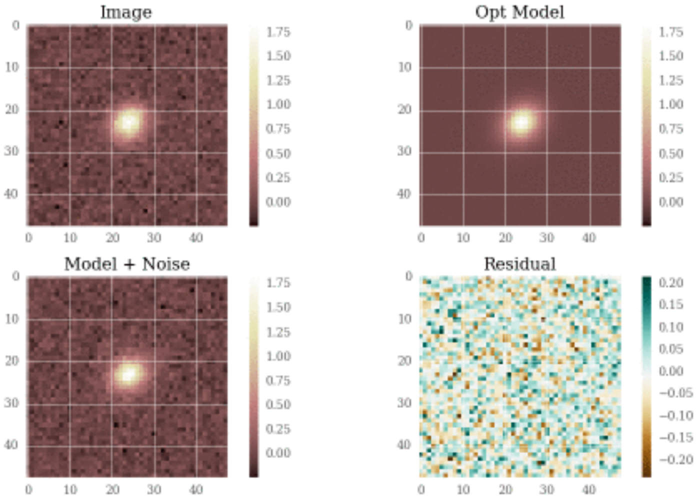
Now use GalSim + MCMC

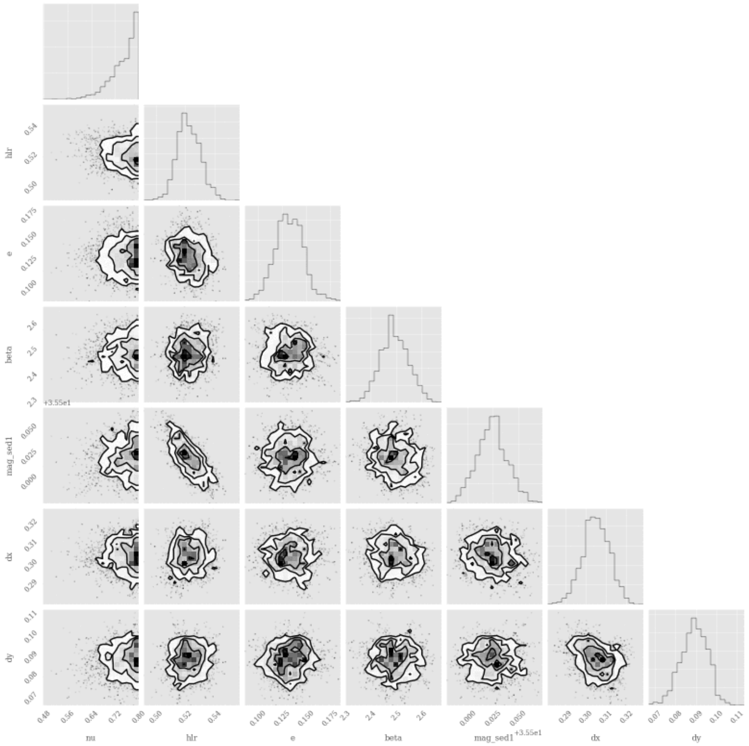




GalSim models inside an MCMC chain – **Can it be made fast enough?**

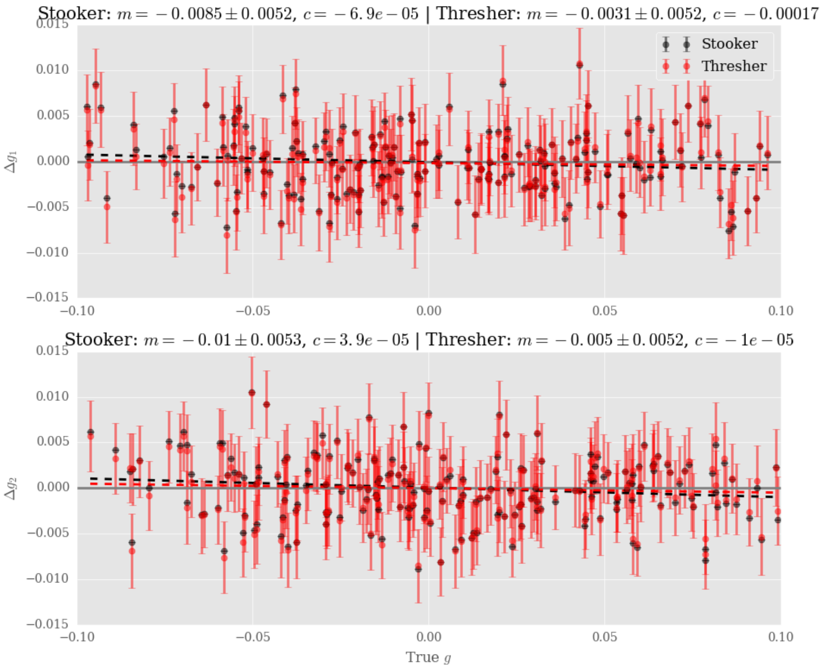
Example interim posterior inferences for galaxy stamp images





Probabilistic forward modeling can meet LSST shear bias requirements

… at least when tested on simulated images



- GREAT3 CGC-like setup
  - 200 'fields' with constant shear per field
  - 10k galaxies per field
- Marginalize 7 parameters per galaxy:
  - e1, e2, HLR, flux, dx, dy, n
  - Notable: Sersic index marginalized
- Have NOT marginalized PSF (yet!)

Immediate takeaway:
Hierarchical inference performs significantly better than ensemble average maximum likelihood ellipticity.

Multi-epoch & multi-telescope data sets

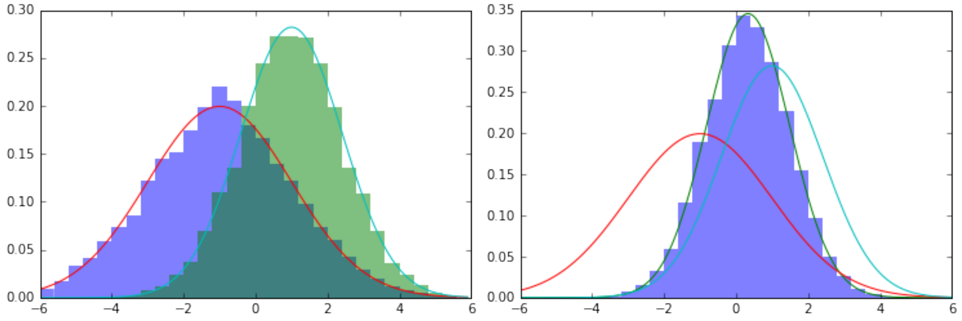
How do we combine multiple observations of the same galaxy?

Naïvely we must joint fit all epochs simultaneously

Problem: Imagine we have fit pixel data from LSST year 1.
How do we incorporate year 2 observations without redoing (expensive) calculations?

Solution: Consider single-epoch samples as draws from a multi-modal importance sampling distribution:

arXiv:1511.03095
Generalized Multiple Importance Sampling
Elvira, Martino, Luengo, & Bugallo



Multiple importance sampling (MIS) via weighted pseudo-marginals

1. Sample from the conditional posterior for each epoch individually
2. Evaluate the ratio of the conditional posterior for each epoch i to that of the MIS sampling distribution

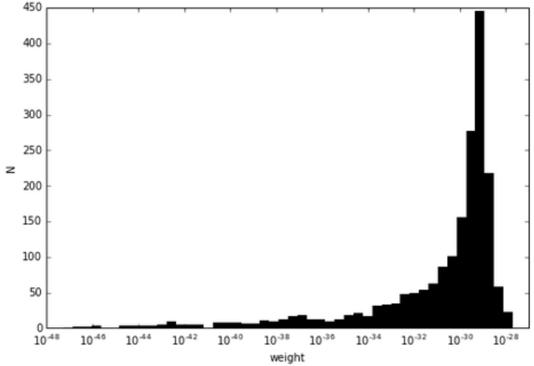
'*cross-pollination*' needed:

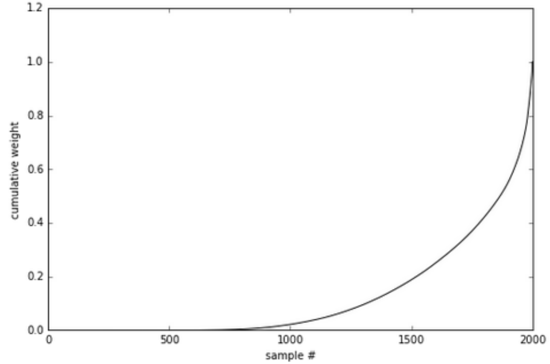
Evaluate the likelihood of epoch i given model parameter samples from epoch j, for all combinations of i, j.

A standard scatter / gather operation

Multiple importance sampling enables streaming data analysis
Efficiency is significantly enhanced by using old data as a sampling 'prior'

- Draw parameter samples from first epoch under a nominal interim prior
- Draw samples from subsequent epochs with a prior informed by previous epoch samples
- Simulation studies show:
  - ~10% of samples have significant weight when combining 200 epochs in streaming fashion



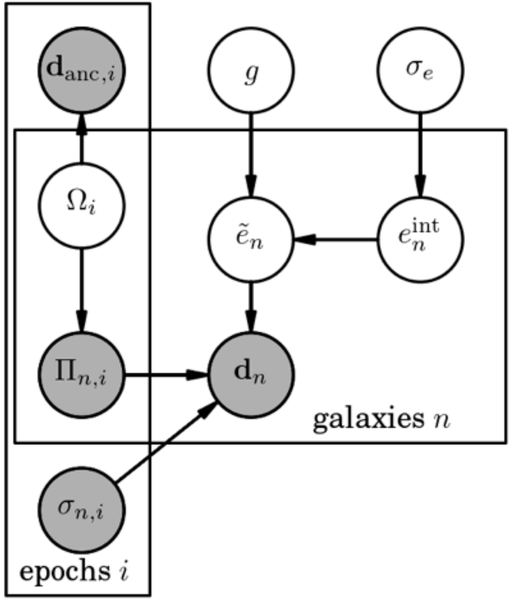


PSF marginalization

Marginalizing PSFs: MIS makes this tractable

- LSST will have ~200 epochs per object per filter
  - We aim to marginalize the PSF $\prod_{n,i}$ in every epoch
  - The marginalization is constrained by:
    - Consistency of PSF realizations over the focal plane for each epoch
    - Consistency of the underlying source model across epochs

- Simplest approach (statistically, not computationally): Infer galaxy models given all epoch imaging simultaneously
  - "Interim" samples are of size: ~10 galaxy params + 200 * ~4 PSF params = ~1k parameters!



The pipeline for PSF marginalization

1. Fit star footprints in all epochs via probabilistic forward models

2. Marginalize star image parameters to constrain the global field PSF model for each epoch
   - State of the optics aberrations, and
   - Distribution of atmosphere turbulence statistics

3. Fit all galaxy footprints in each epoch via forward models
   - Use PSF models drawn from the marginal posterior given the star images

4. Run Thresher on the interim galaxy samples for all epochs (via ‘cross-pollinator’)

One approximation needed:
Marginalize PSF model independently for each field location

Probabilistic cosmological one-point statistics

arXiv:1610.06673

Probabilistic cosmological mass mapping

Interpolate the unobserved lensing potential with GP

$$\psi_s \sim GP(0, \Sigma),$$

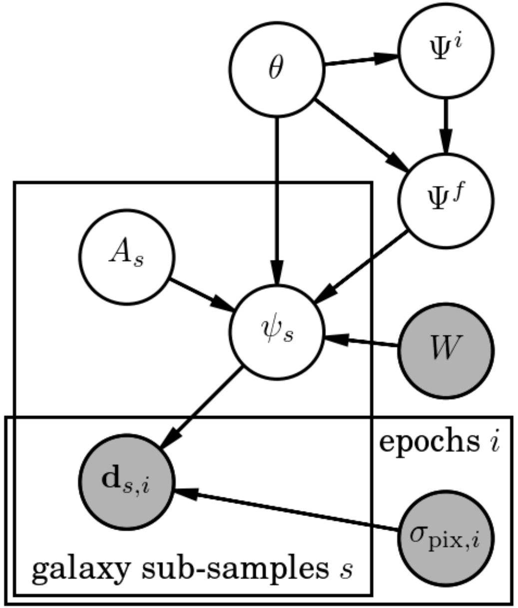
$\kappa, \gamma_1, \gamma_2$ are the second (spatial) derivatives of ψ_s

$$\mathrm{Cov}(\psi_{,ij}(\vec{x}), \psi_{,k\ell}(\vec{y})) = \Sigma_{,x_i x_j y_k y_\ell}(\vec{x}, \vec{y}).$$

GP kernels of $\kappa, \gamma_1, \gamma_2$ are linear combinations of the 4th (spatial) derivatives of the kernel of ψ_s

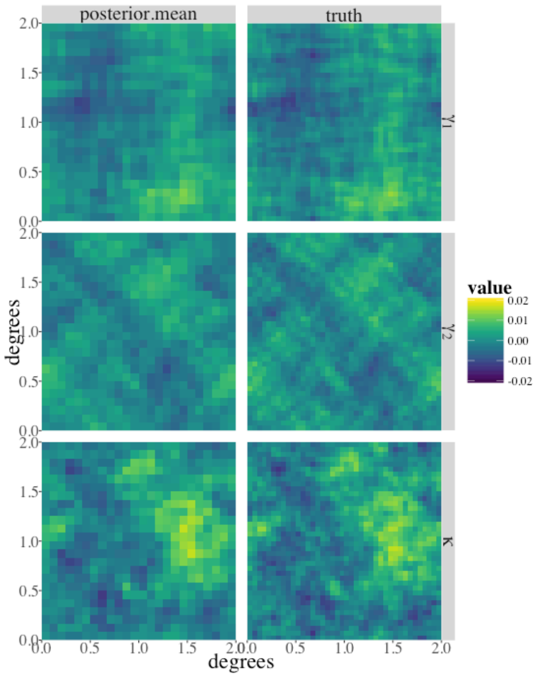
Zero E/B mode mixing by construction

Objective: infer the 3D gravitational potential of the initial conditions

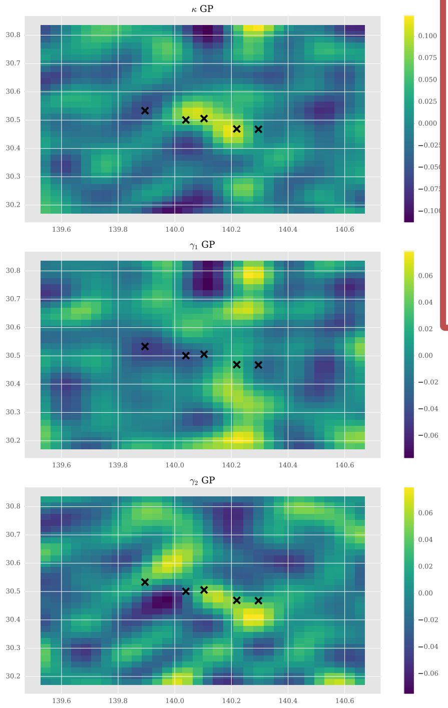


Hierarchical inference of cosmological lensing mass distributions

Validation with simulations

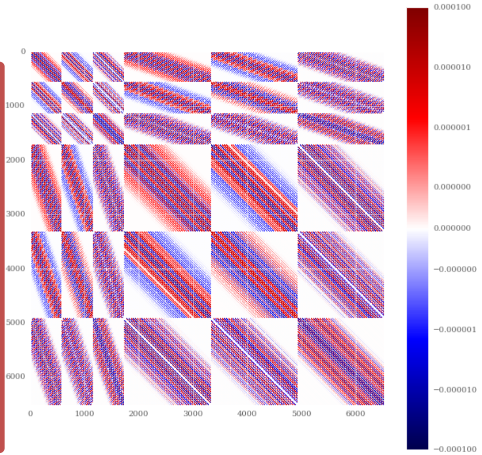


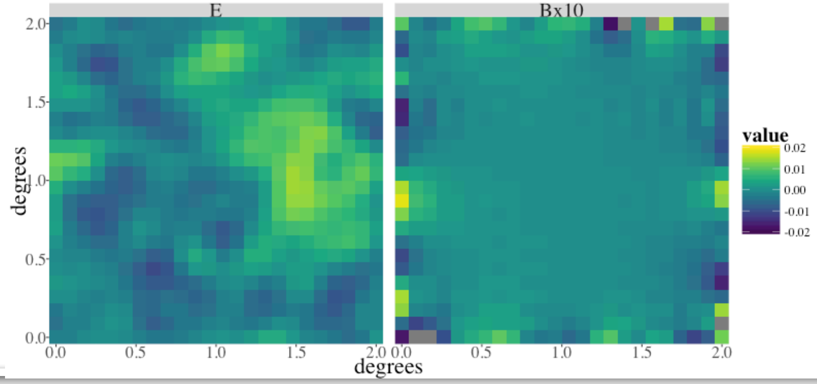
A real merging galaxy cluster



New:

- Linear and nonlinear scales reconstructed in one framework
- No E/B mode mixing by construction

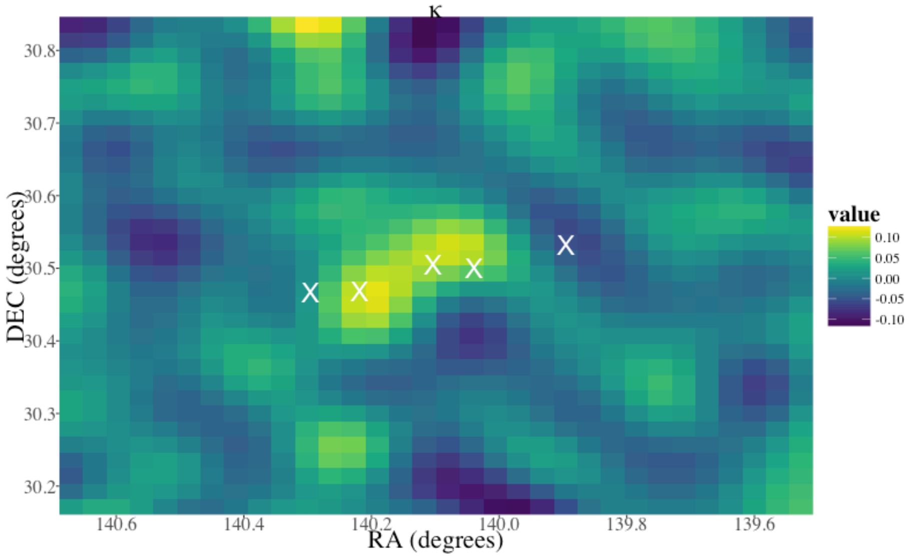




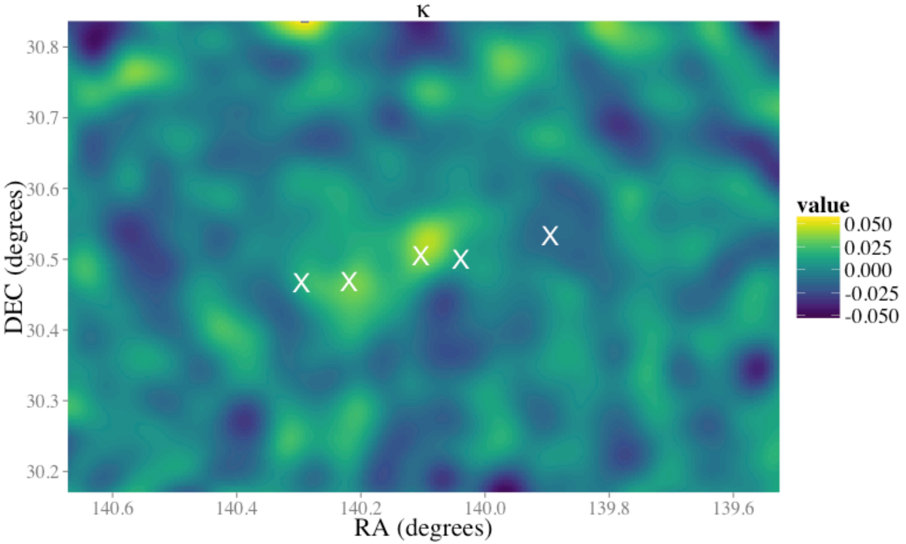
Schneider+2017, ApJ

Application to data: Weak lensing mass maps for Abel 781 merging galaxy clusters as seen by the Deep Lens Survey

Our method



Previously published method

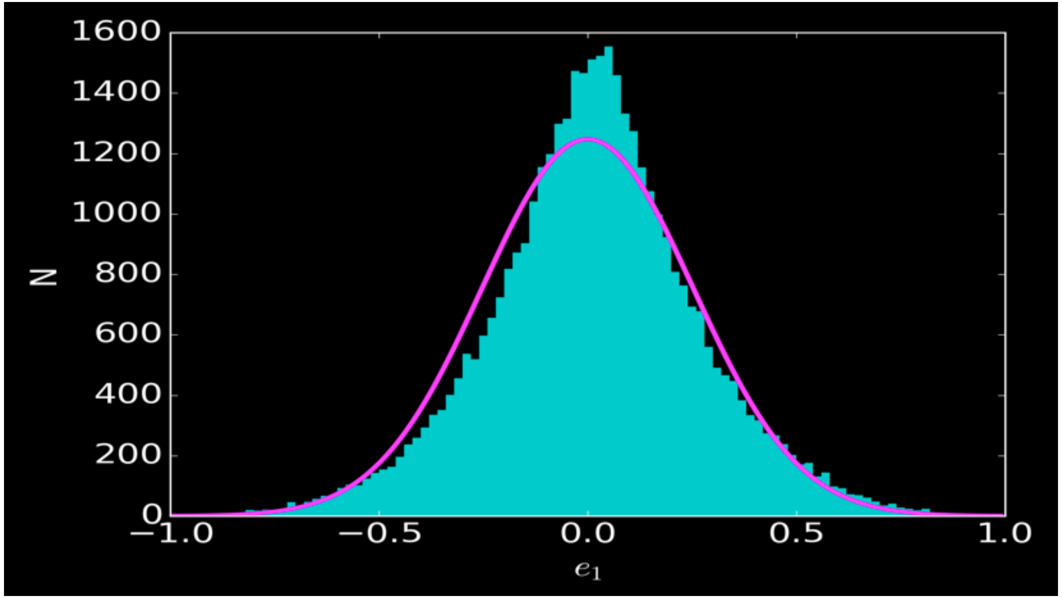


Latent features of the galaxy distribution

$Pr(e^{int})$ is not Gaussian!

- Would rather not assert a particular parametric form for $P(e^{int})$.
- Use a "non-parametric" distribution: a Dirichlet Process Mixture Model

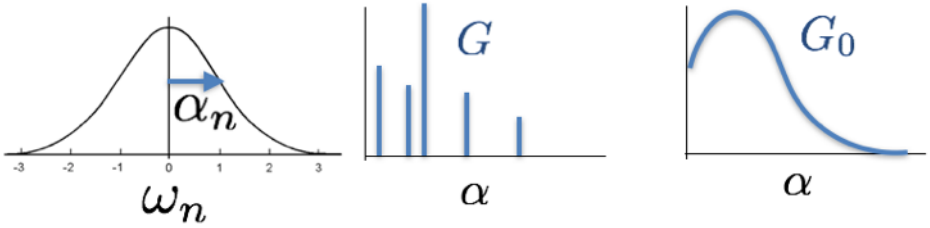
Ellipticities from COSMOS



Hierarchical inference of intrinsic galaxy properties

Specify a Dirichlet Process (DP) for the distribution of intrinsic galaxy property hyper-parameters

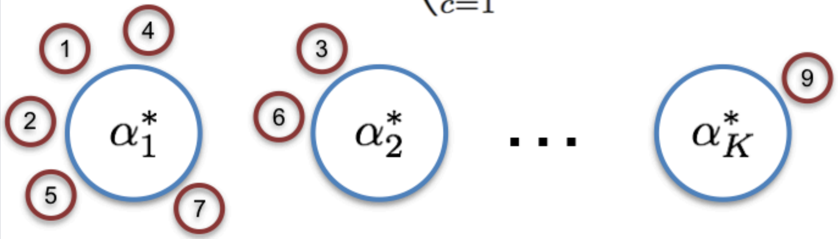
$$\omega_n \sim \mathcal{N}(0, \alpha_n), \quad \alpha_n \sim G(\alpha_n|\mathcal{A}), \quad G \sim \mathrm{DP}(\mathcal{A}, G_0)$$



The DP is a 'non-parametric' distribution with discrete support

The DP distribution allows clustering of data points (e.g., galaxies) to infer *latent structure* in the data.

$$\alpha_n|\alpha_1, \ldots, \alpha_{n-1} \sim \frac{1}{n-1+\mathcal{A}} \left(\sum_{c=1}^{K} N_c \delta_{\mathrm{D}} \left(\alpha_c^*\right) + \mathcal{A} G_0\left(\cdot\right) \right)$$



Gibbs updates in the Dirichlet Process model

Latent class assignments are updated with different conditional distributions depending on whether any other observations are assigned to the current class.

$$\Pr(c_n = c_\ell | c_{-n}, \omega_n, \alpha, \mathcal{X}) = b\, N_{-n,c} \Pr(\mathbf{d}_n | \alpha_{c_\ell}, \mathcal{X}), \qquad \forall \ell \neq n$$

$$\Pr(c_n \neq c_\ell \forall \ell \neq n | c_{-n}, \omega_n, \mathcal{X}) = b\, \kappa \int \Pr(\mathbf{d}_n | \alpha, \mathcal{X})\, G_0(\alpha)\, d\alpha,$$

The DP mixture parameters are simply updated with the posterior given all observations currently associated with the given latent class.

$$\alpha_{c_n} \sim G_0\left(\alpha_{c_n}\right) \prod_{\ell=1}^{N_{c_n}} \Pr(\mathbf{d}_\ell | \alpha_{c_n}, \mathcal{X}))$$

Neal (2000)

Highlighted integral is expensive to compute in general.

$$\Pr(c_n \neq c_\ell \forall \ell \neq n | c_{-n}, \omega_n, \mathcal{X}) = b\, \kappa \int \Pr(\mathbf{d}_n | \alpha, \mathcal{X})\, G_0(\alpha)\, d\alpha$$

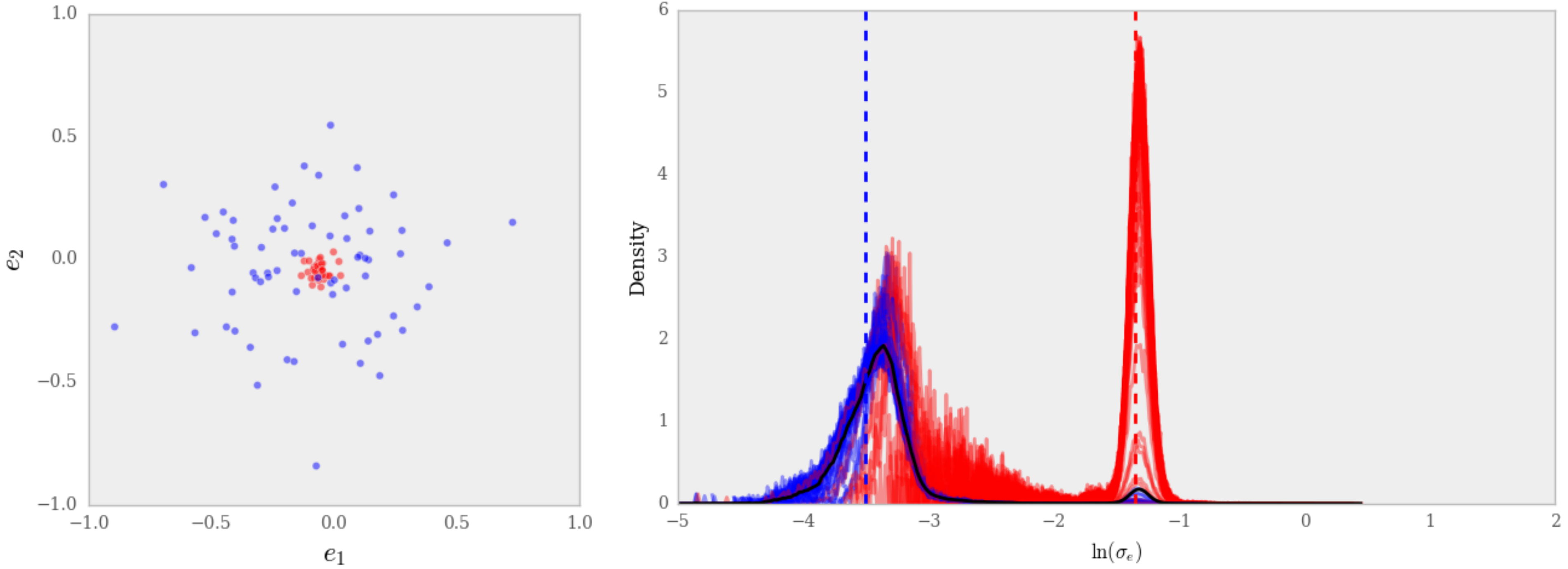
With importance sampling we only require the DP base distribution to be conjugate to the distribution of galaxy properties – *NOT* the likelihood.

$$\int \Pr(\mathbf{d}_n | \alpha, \mathcal{X})\, G_0(\alpha)\, d\alpha = \frac{Z_n}{N} \sum_{k=1}^{N} \frac{\Pr_{\mathrm{marg}}(\omega_{nk} | a)}{\Pr(\omega_{nk} | I_0)}$$

$$\Pr_{\mathrm{marg}}(\omega_{nk} | a) \equiv \int d\alpha_{c_n}\, G_0(\alpha_{c_n} | a) \Pr(\omega_{nk} | \alpha_{c_n})$$

A simulation study with 100 galaxies validates the DP model

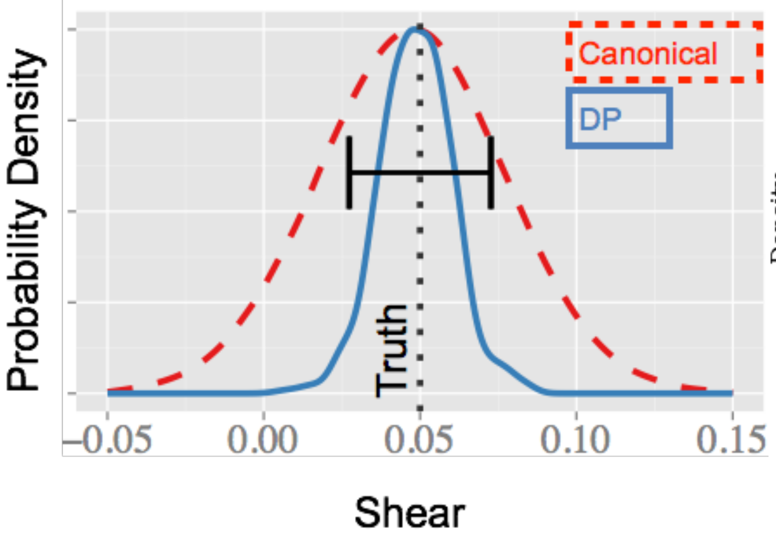
100 galaxies drawn from 1 of 2 Gaussian ellipticity distributions

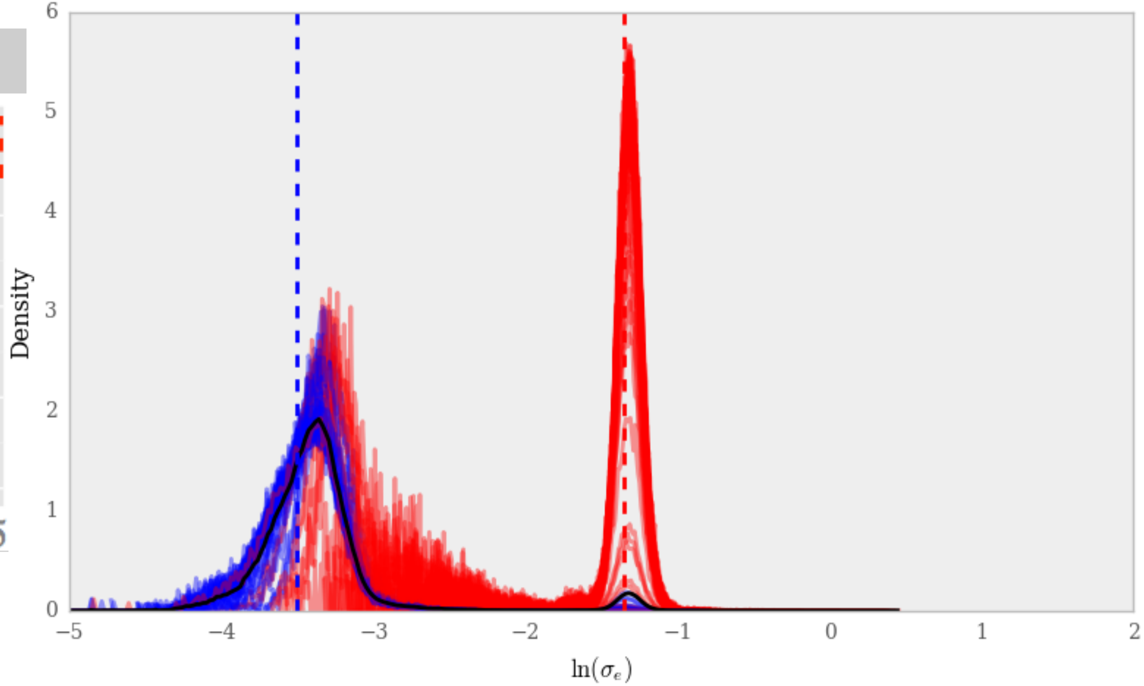


Simulation study: We can beat the traditional 'shape noise' statistical error bound by inferring latent structure in the data

100 galaxies drawn from 1 of 2 Gaussian ellipticity distributions

3x improvement in cosmic shear precision





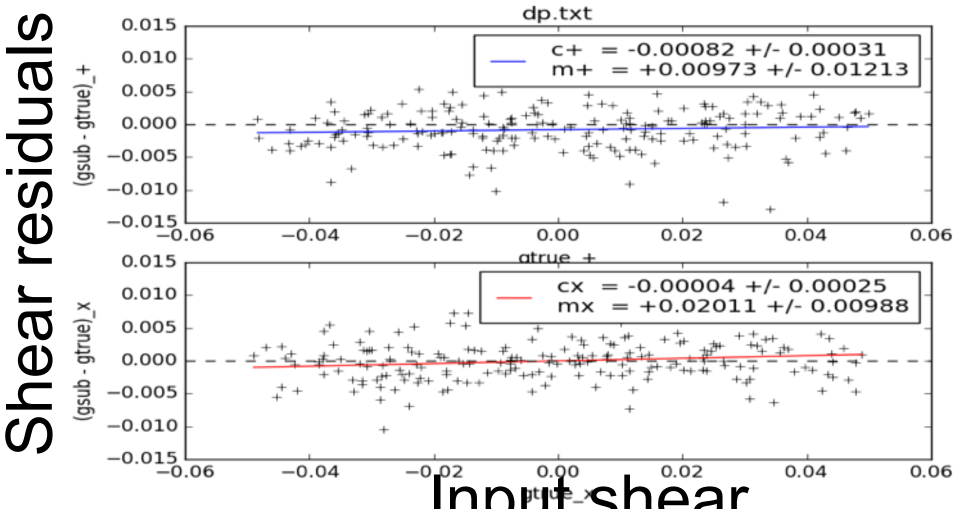
GREAT3 results

- Tested hierarchical approach using simulations from the third GRavitational lEnsing Accuracy Test (GREAT3).

- Hierarchical inference performs significantly better than ensemble average maximum likelihood ellipticity.

- The DPMM ellipticity prior performs better than the single Gaussian ellipticity prior.

Dirichlet Process Inference

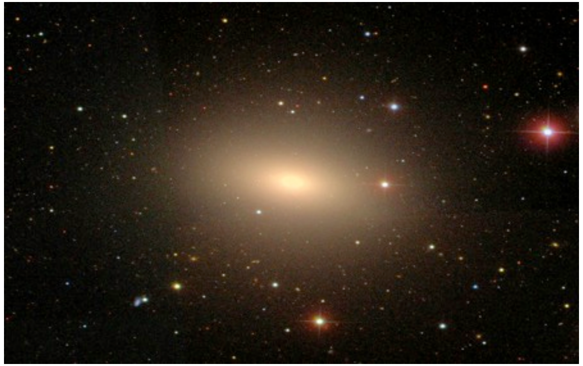


<ML> : 13% shear calibration errors

H.I. : 4% shear calibration errors

DP : 1-2% shear calibration errors

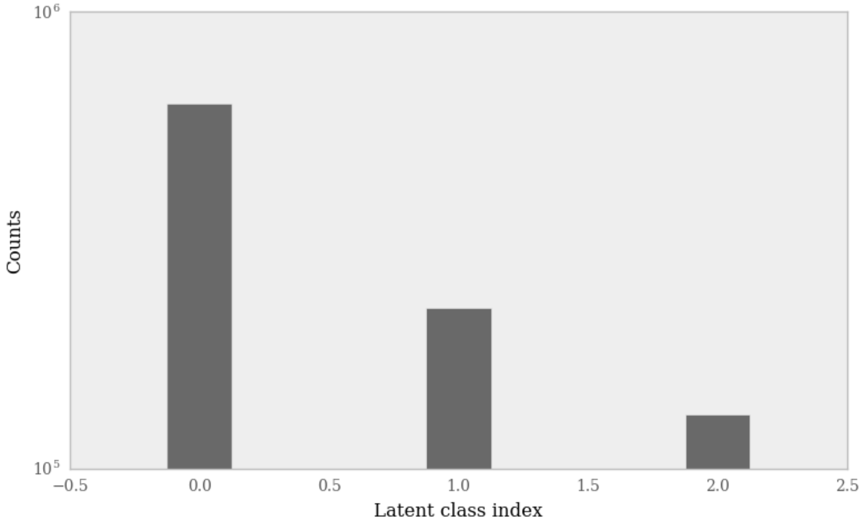
Multi-variate DP mixture model (in progress): “standardizable” ellipticities.



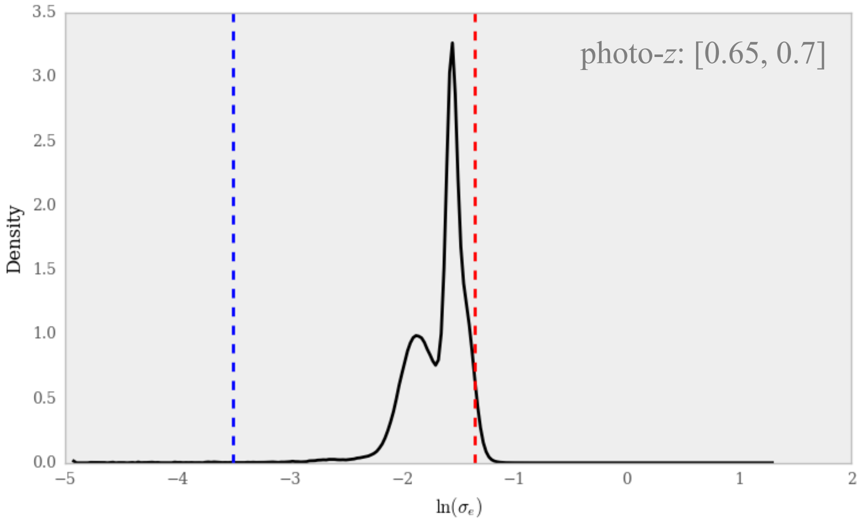
- Elliptical galaxies have a narrower intrinsic ellipticity distribution than late-type. Higher sensitivity to shear!
- Ellipticals/spirals also distinguishable by color and morphology (e.g., Sersic index, Gini coefficient, asymmetry), potentially providing additional variables with which to cluster.
- Other correlations to exploit?

Application to the Deep Lens Survey: real galaxies require at least 2 latent classes (ignoring lensing)

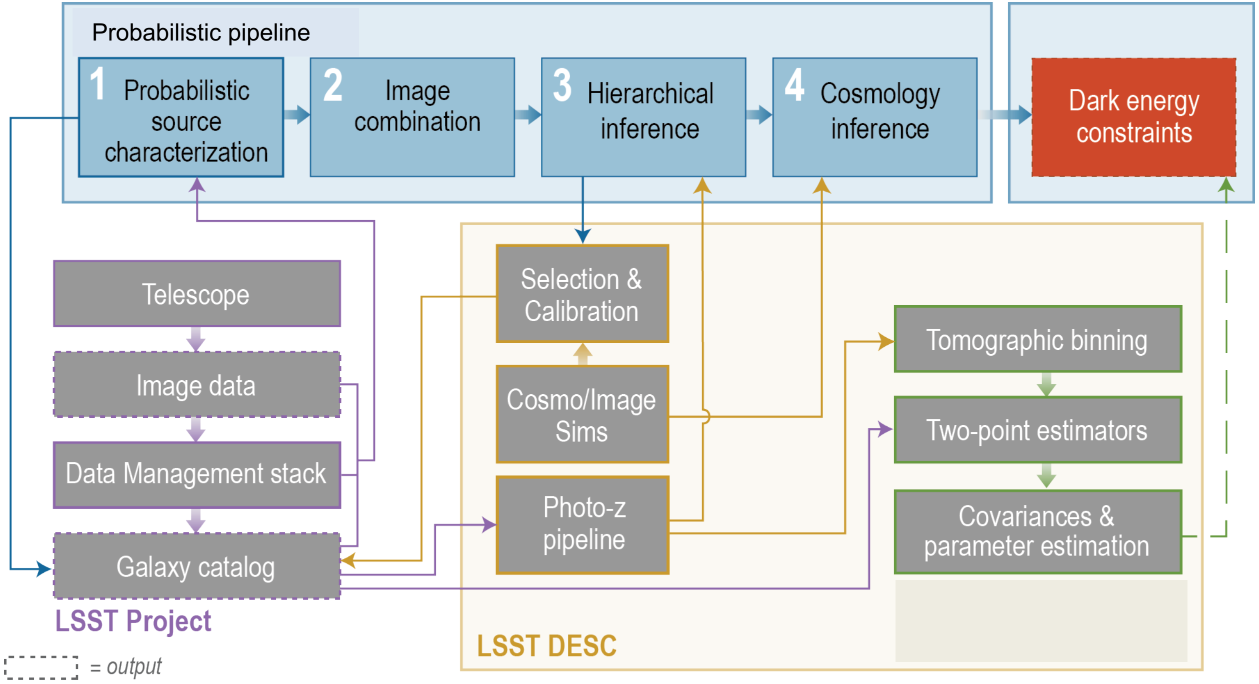
We infer 2 latent classes given only an ellipticity catalog



Preliminary: The marginal posterior distribution of ellipticity variance from the Deep Lens Survey

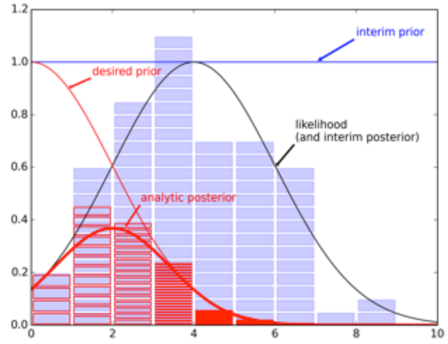


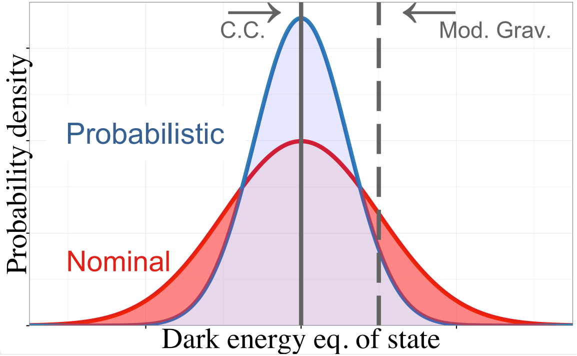
The probabilistic weak lensing workflow plan for LSST



Summary

- Cosmic shear is systematics limited & signal is dominated by PSF and astrophysics
  - A probabilistic approach is warranted to infer a small signal and mitigate biases
- A hierarchical probabilistic model for cosmic shear can trade bias for variance, but also can increase precision by learning latent structure in the galaxy distribution.
- Importance sampling methods allow tractable approaches to a probabilistic forward model of LSST & WFIRST imaging
  - With billions of galaxies and hundreds of epochs per galaxy modeling LSST or WFIRST imaging requires an approach to separating analyses of data subsets, even though statistically correlated
- We are able to sample from a probabilistic model with multiple hierarchies to marginalize both correlated image systematics and astrophysical properties of galaxies.





Lawrence Livermore National Laboratory