GLASS, or How to Deal with Uncertainty in Inference

Steven Gratton, DAMTP

Kavli Institute for Cosmology Cambridge

1 June 2023

Bayes' Theorem

- $p(theory|data) = \frac{p(data \mid theory)\ p(theory)}{p(data)}$

But what should one do if?

- $p(theory|data) = \frac{?????????????????? \; p(theory)}{p(data)}$

- i.e. one doesn't know what the likelihood is

When do such situations arise?

- Lacking something…
  - Knowledge about the experiment
  - Mathematical knowledge
  - CPU time
  - …

Options?

- Guess a potential likelihood

Options?

- Guess a potential likelihood
- Simulations

Options?

- Guess a potential likelihood

- Simulations

- …

Difficulties with using simulations (1)

- One's simulations will be *wrong* at some level!
- One probably doesn't have enough of them, especially in multidimensional situations with statistics that are correlated…

Difficulties with using simulations (2)

- Depending on how one uses one's simulations, one doesn't necessarily know what one's "likelihood-free" likelihood is exactly responding to…

GLASS (arXiv:1708.08479)

- Calculate/simulate what you can about that which *you most trust* and which will inform you about the models under consideration
  - This will typically be some moments of some statistics, recomputed for every model considered…

- Use the *principle of maximum entropy* to “fill in the gaps” and effectively *construct the sampling distribution* for that model that is consistent with what you’ve calculated/simulated

What is the principle of maximum entropy?

- Entropy (Shannon, 1948) measures the uncertainty of a probability distribution:

$$S = -\int p(x) \log p(x)\, dx$$

- So, if we *maximize* the entropy, subject to the constraints of what we *do* know, we generate the *broadest* or *most conservative* distribution consistent with those constraints (Jaynes, 2003 book)

- One shouldn’t go wrong if one uses this for inference! (One hopes…)

What GLASS is *not*

- *Not* just fitting a gaussian to the moments
- *Not* an Edgeworth expansion

Relations with likelihood-free approach

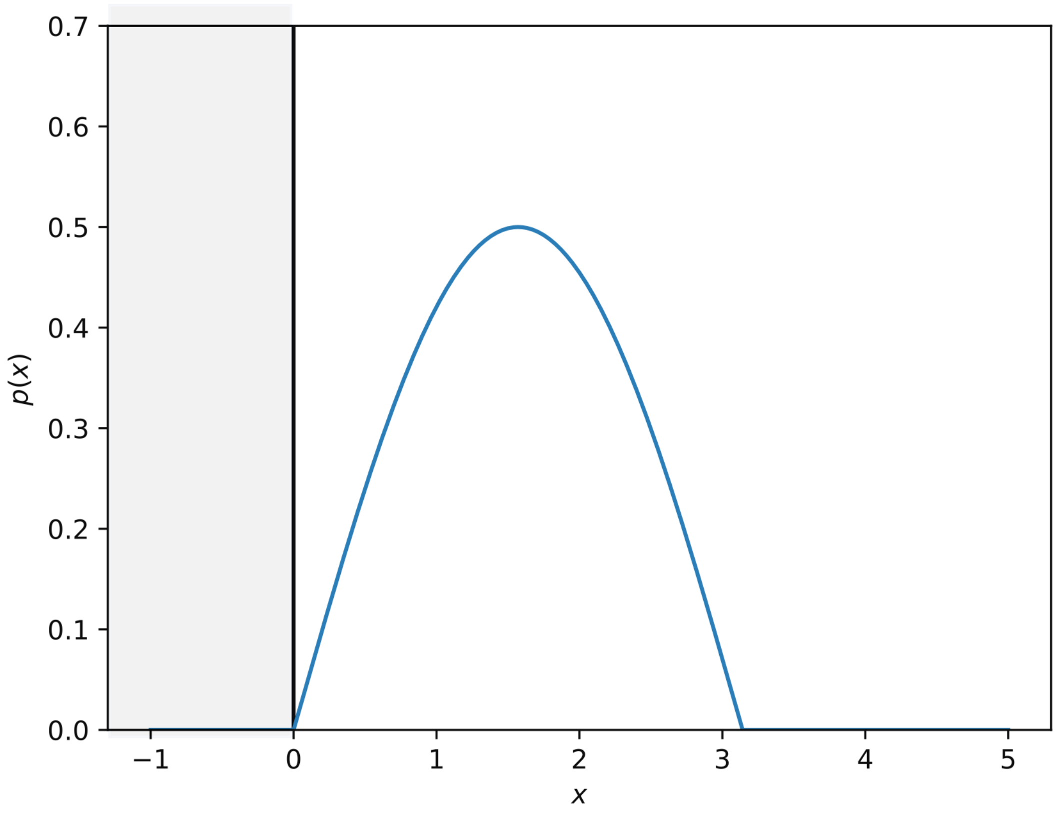
Likelihood-free

- Doesn't have an explicit likelihood
- Can simulate *data realizations*
- Combine these to compare data to model

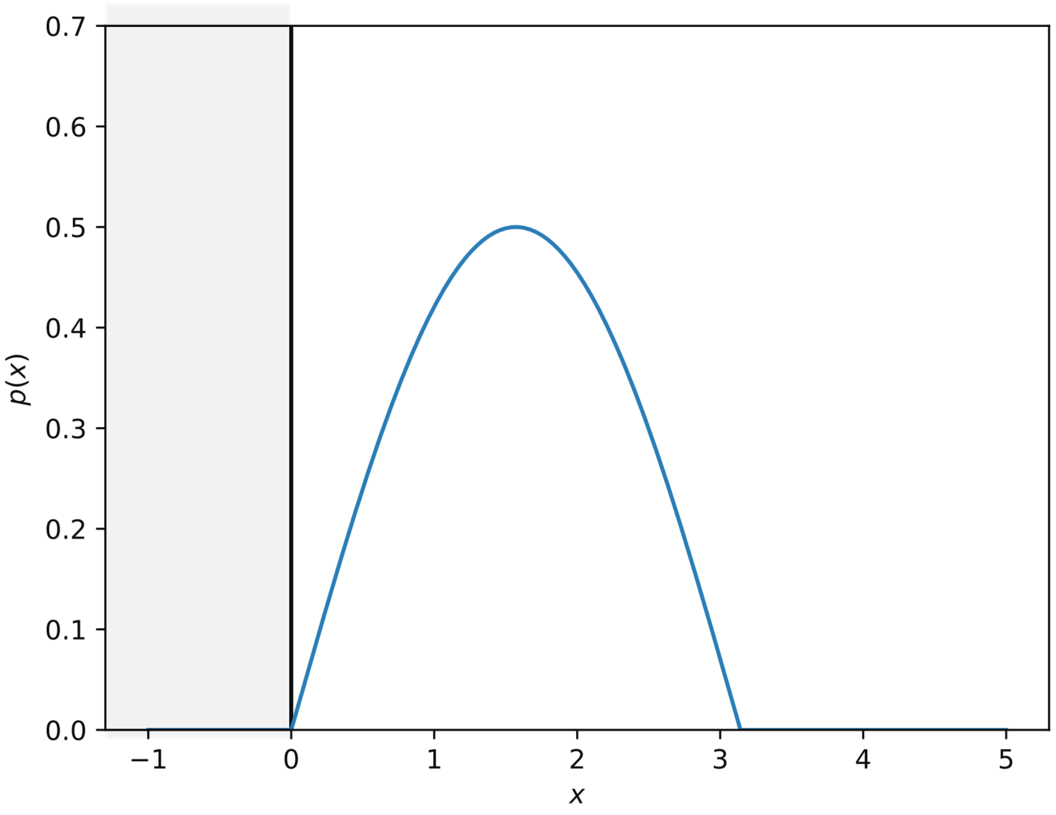
GLASS

- Doesn't have an explicit likelihood
- Can *calculate*/simulate *moments*
- Combine these to compare data to model

Example sampling distbtn.: $p(x) = \frac{a}{2}\sin(ax)$



Example sampling distbtn.: $p(x) = \frac{a}{2}\sin(ax)$

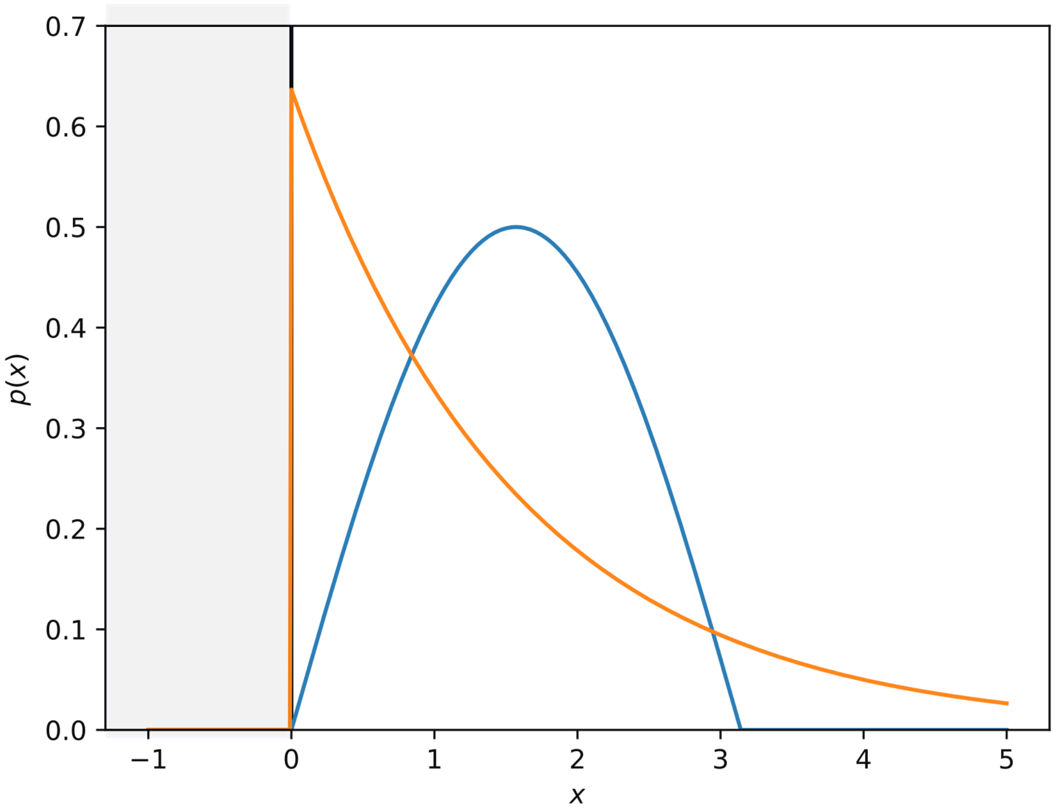


$$< x > = \frac{\pi}{2a}$$

$$< x^2 > = \frac{\pi^2}{2a^2} - \frac{2}{a^2}$$

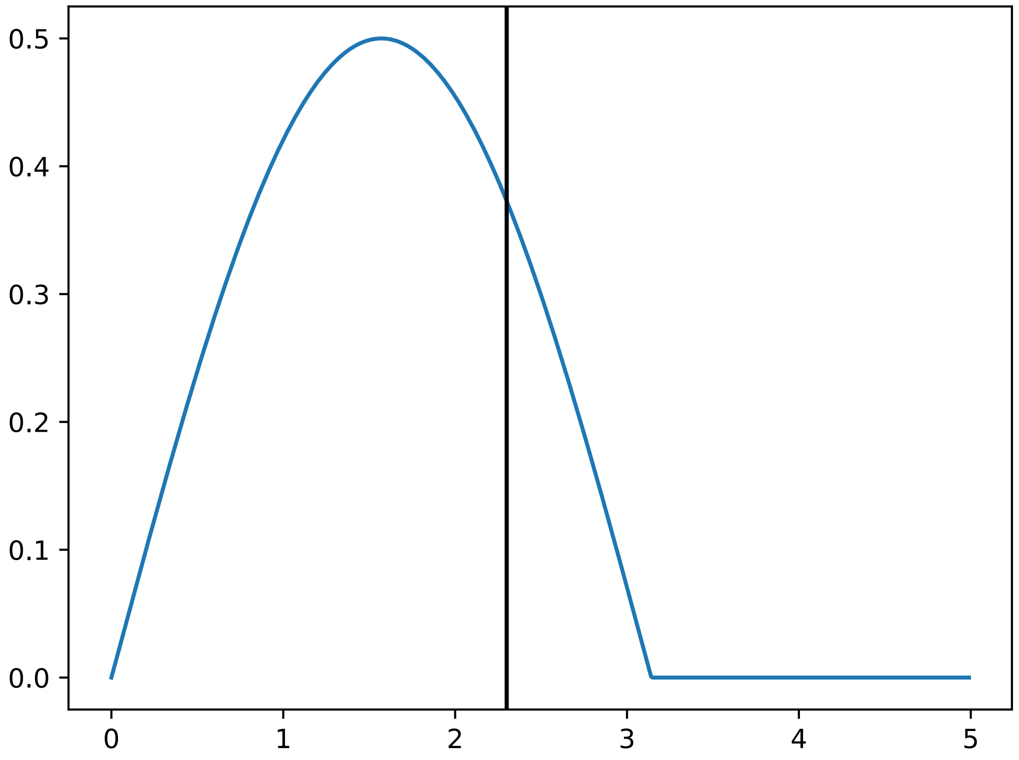
…

Fitting just the first moment:

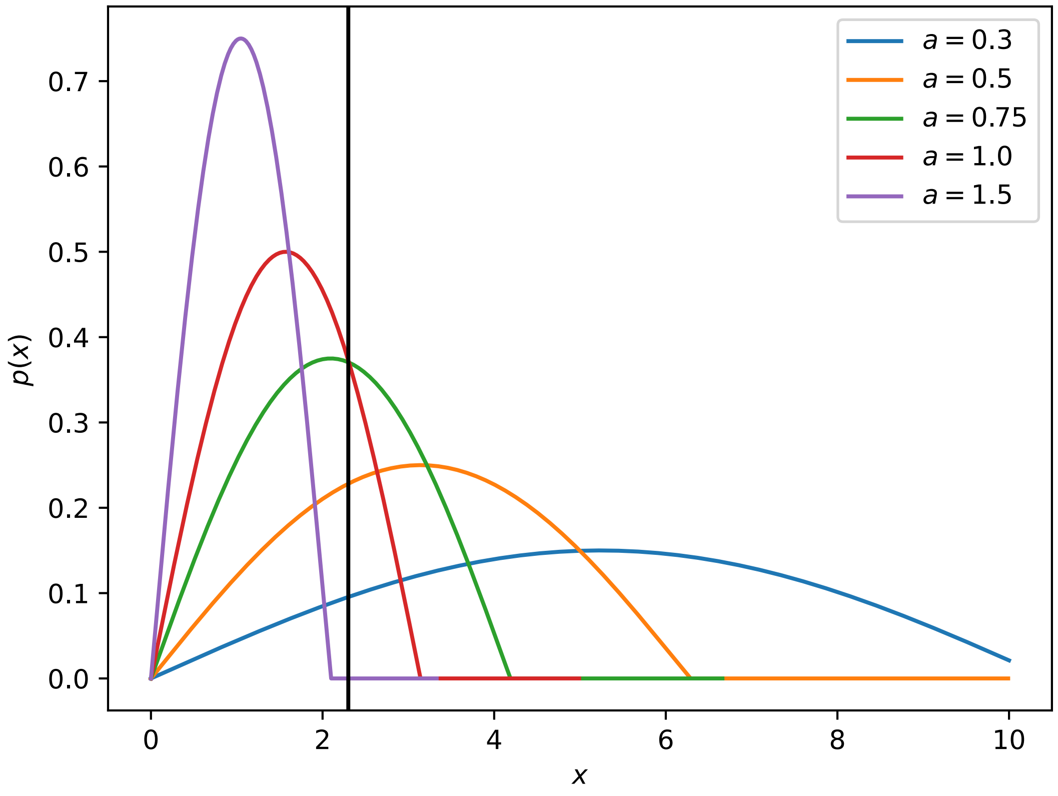


$$< x > = \frac{\pi}{2a}$$

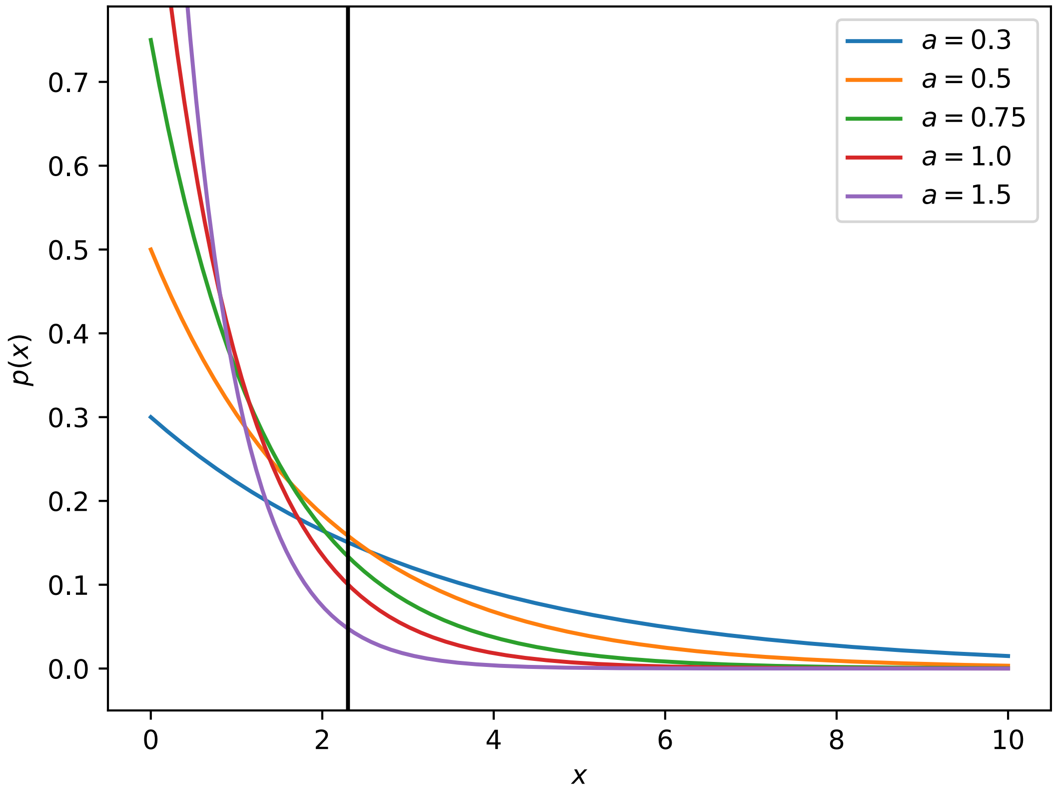
Imagine we collect some data…



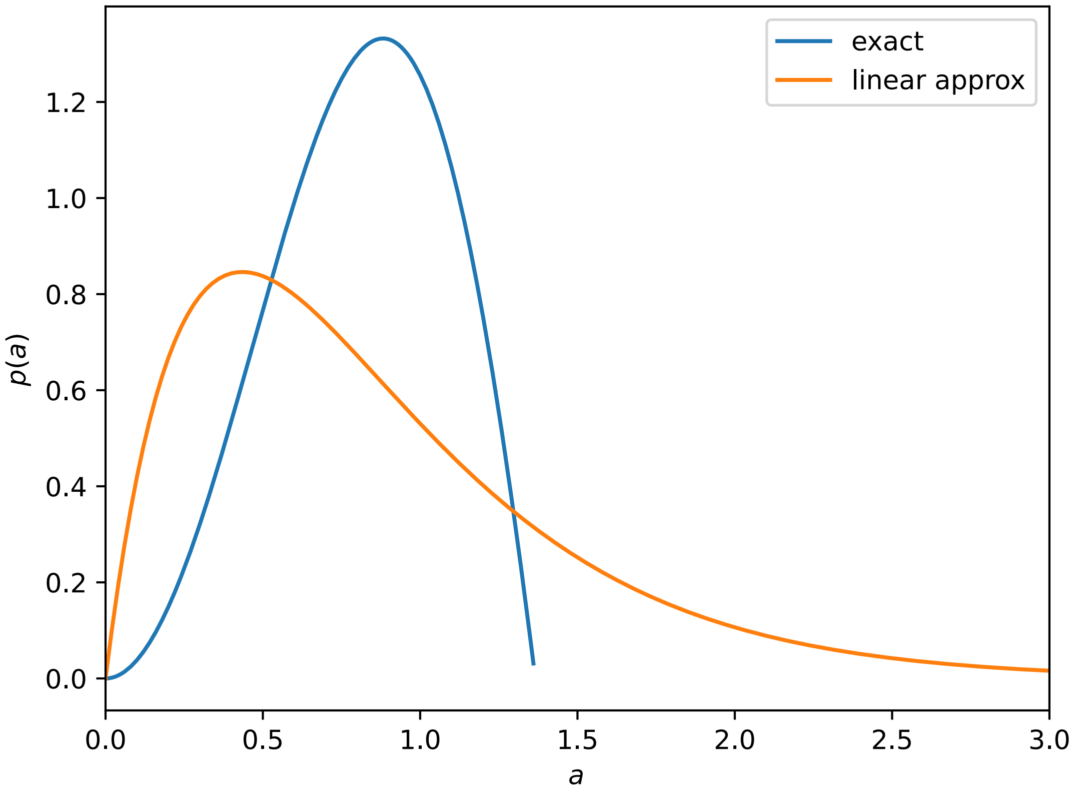
Try some sample distributions…



Try some approximate sample distributions…



Gives our posteriors…



But what about in multiple dimensions?

- When maximizing the entropy, for each trial set of lagrange multipliers, one has to numerically evaluate multidimensional integrals to see how all the moments turn out

- But multidimensional integrals are difficult/expensive

- To speed this up for parametrized models, the GLASS paper details how one may replace computing all those integrals with a one-dimensional line integral (at the cost of computing more moments…)

The key formula

- This works because one can express the gradient of the unknown likelihood in terms of the moments:

$$-\log p_{,a} = -(X - \langle X \rangle)^{\mathrm{T}} \langle\langle X X^{\mathrm{T}} \rangle\rangle^{-1} \langle X \rangle_{,a}$$

- Here, $X = (x, y, \dots, x^2, xy, y^2, \dots, x^3, x^2y, xy^2, y^3, \dots)^T$

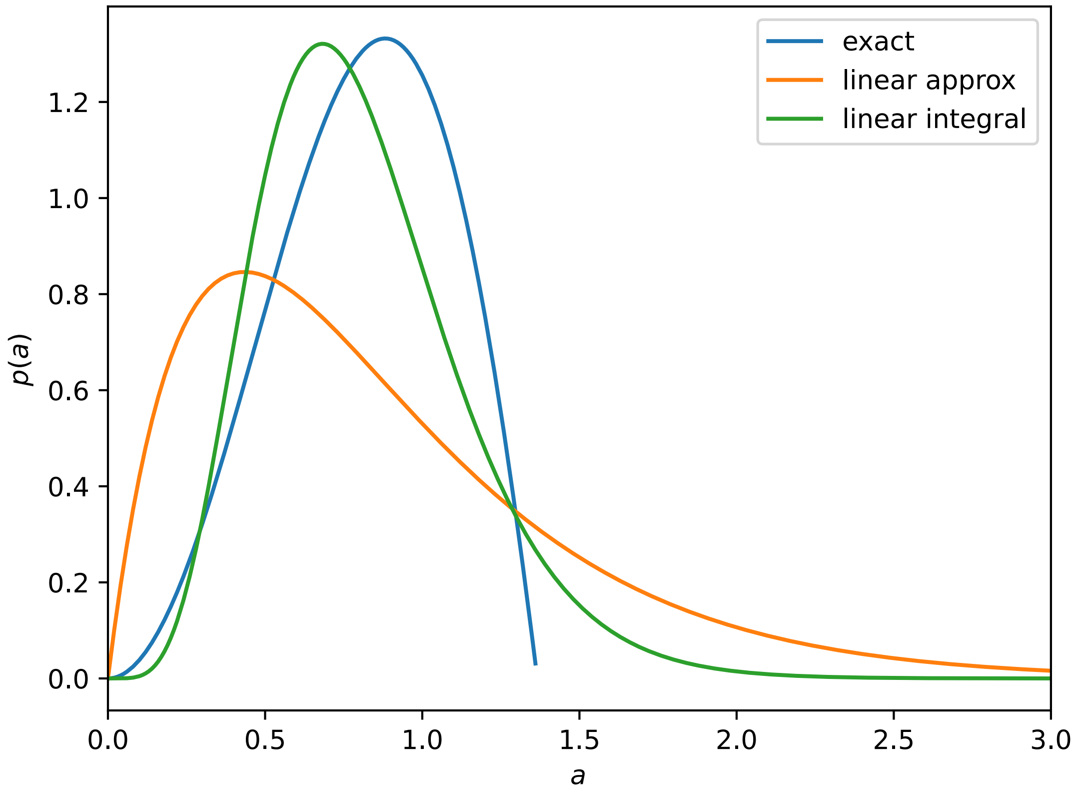
But which moments?

- Formula is derived assuming the higher moments are what one gets in the considered maximum entropy distribution

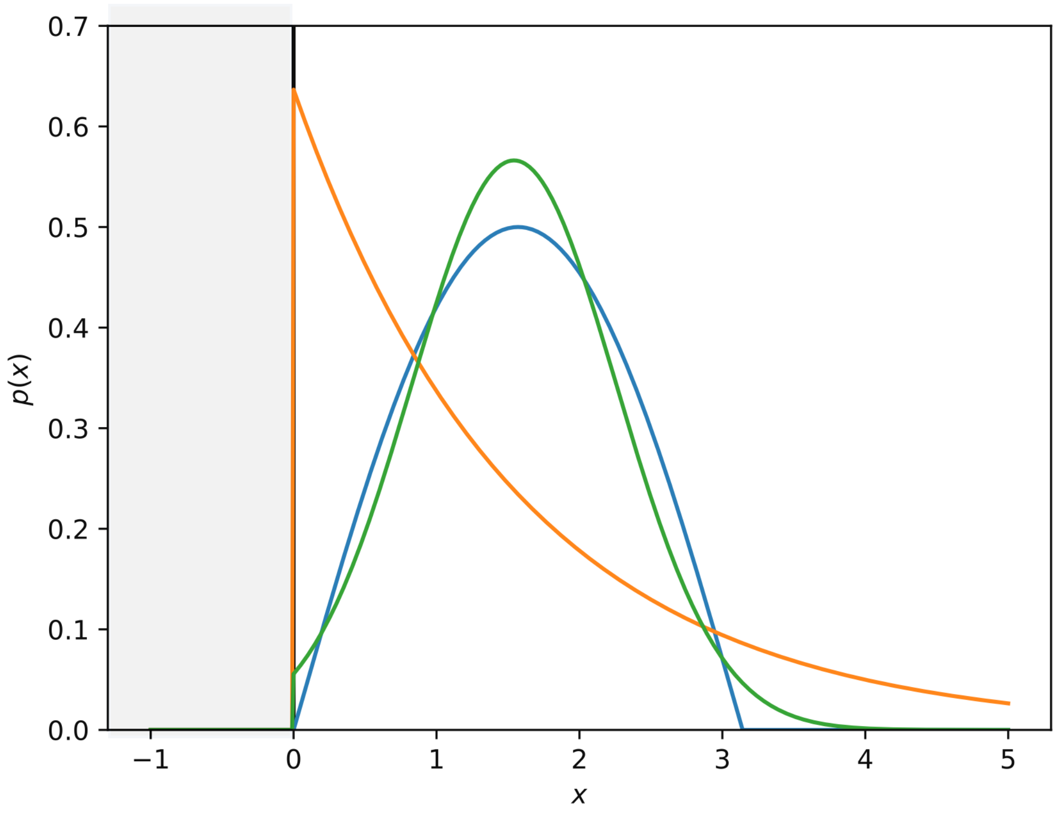
$$-\log p_{,a} = -(X - \langle X \rangle)^{\mathrm{T}} \langle\langle X X^{\mathrm{T}} \rangle\rangle^{-1} \langle X \rangle_{,a}$$

- However, one can *approximate this approximation* by using the *exact* higher moments computed/simulated from the underlying theory!...

Effect on our basic posterior!!...



Next: Fit the first and second moments…



$$< x > = \frac{\pi}{2a}$$

$$< x^2 > = \frac{\pi^2}{2a^2} - \frac{2}{a^2}$$

Application to Planck CMB (arXiv:2103.14378)

- The target: inference of the optical depth to reionization, tau, via the height of the "bump" of the low multipole EE polarization power spectrum (also affects the TE spectrum but less significantly)

- The problem: large-scale systematic residuals in the polarization maps caused by non-linearities in the onboard analogue-to-digital converters, even with the special "SROLL1" processing developed to mitigate this effect for the 2018 Legacy Release…

What did we most trust?

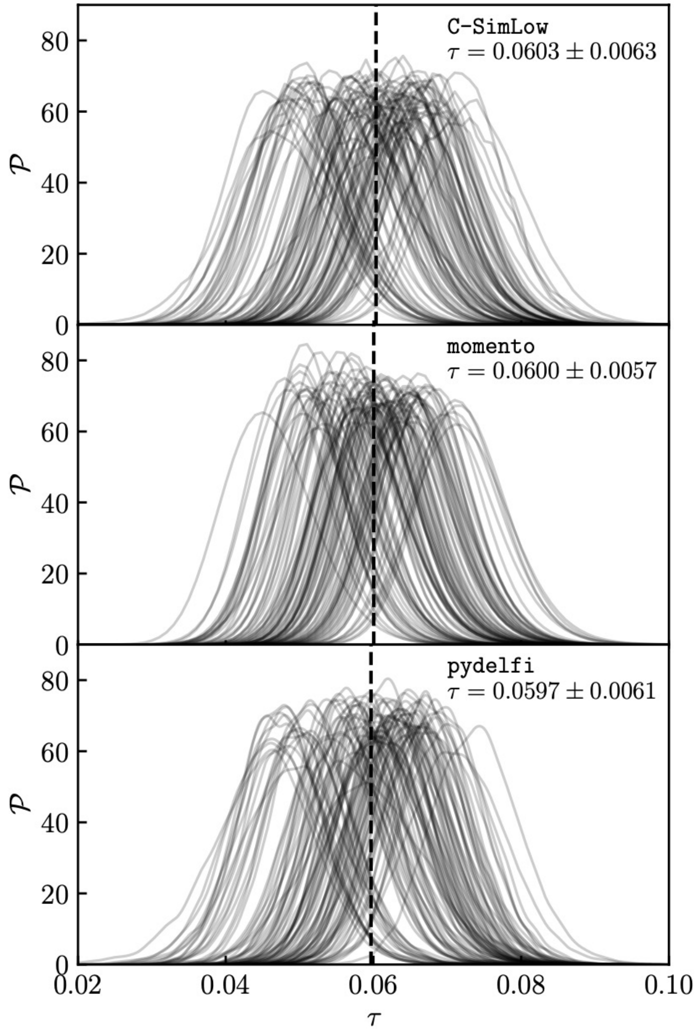
- From the data: “cross” power spectrum measurements, with each leg coming from a different frequency channel

- From end-to-end simulations: Use the limited number we had (300) to *inform* an analytic model for the noise between pixels in the map

Momento

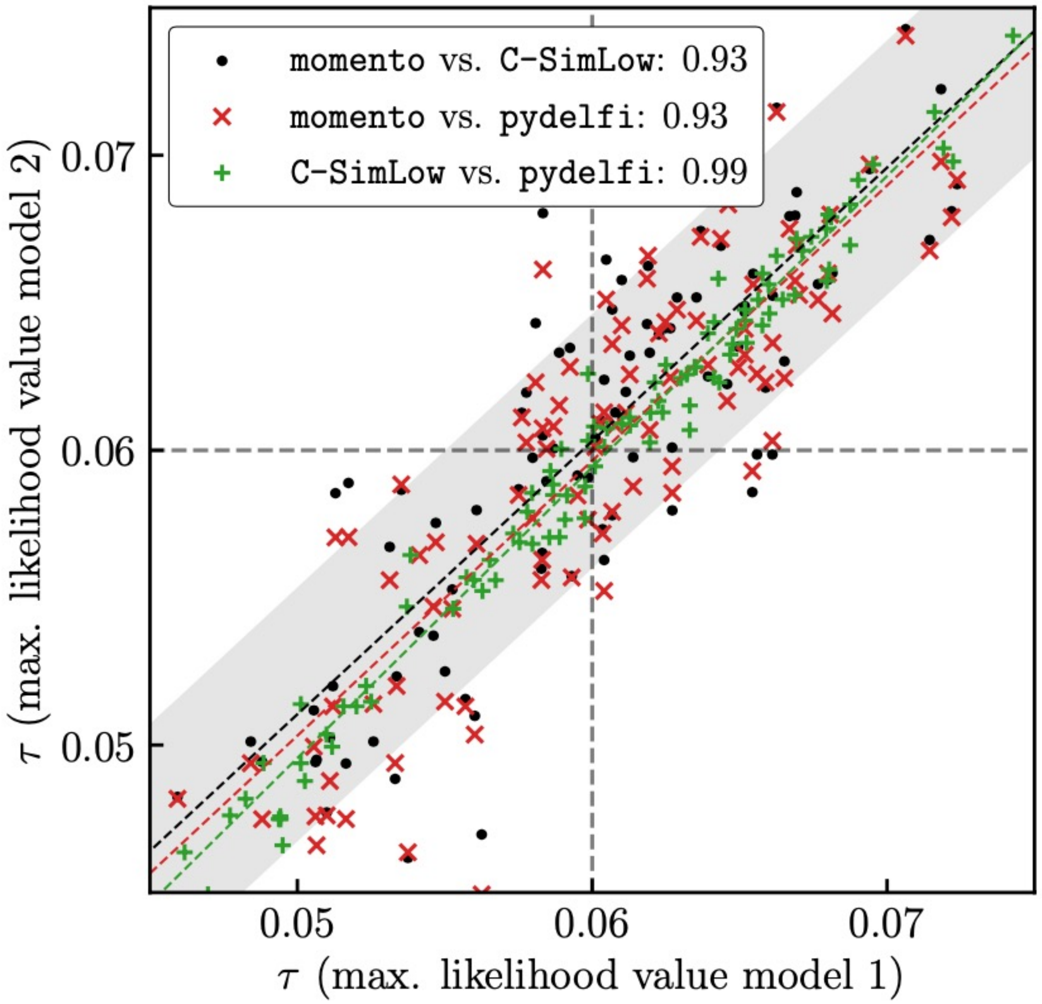
- Assuming our noise model, we quickly *calculate* on the fly moments of and between all cross spectrum elements (up to quartic order) of interest for any value of tau
- Using the GLASS procedure, we effectively build a likelihood consistent with these moments, and evaluate it at the data cross spectrum values, for each theory model we wish to consider

Tests on simulations (EE-only)



- Simulation-based, using hand-picked estimators
- Momento
- Simulation-based, using NN density estimation (pyDELFI)

Max-likelihood-value comparisons



τ constraints

Data Set	Likelihood	τ (EE)	τ $(TTTEEE)$
Planck 2018	`C-SimLow`	0.0530 ± 0.0071	...
	`momento`	0.0507 ± 0.0063	0.0527 ± 0.0058
	`pydelfi`	0.0517 ± 0.0070	0.0513 ± 0.0078
`SRoll2`	`C-SimLow`	0.0582 ± 0.0057	...
	`momento`	0.0581 ± 0.0055	0.0604 ± 0.0052
	`pydelfi`	0.0588 ± 0.0054	0.0580 ± 0.0064

Since gone on to constrain r (arXiv:2207.04903)

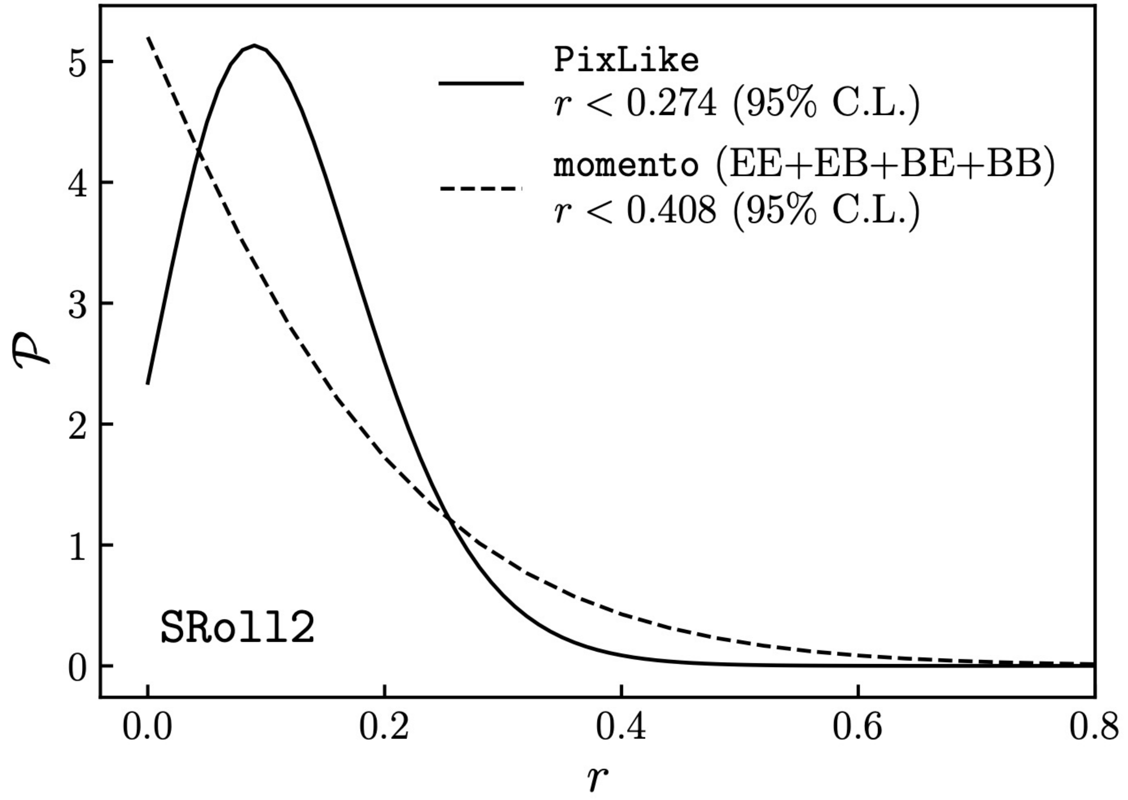


Figure 3. Posteriors for r (marginalised over τ) from two-dimensional low-ℓ scans, using a pixel-based likelihood (`pixLike`) and a QCS-based likelihood approximation scheme (`momento`), on NR-cleaned polarisation *Planck* maps. The uncertainties from the foreground cleaning have been propagated through to the final NCMs.

Conclusions and Further Thoughts

- Dealing with these sorts of issues in practice is tough!

- Relying solely on simulations is dangerous:
  - One probably won’t have enough of them
  - Even if one does, they might not be good enough in all details; one must be prepared to “guide” one’s analysis/neural network towards only using the things that are suitably trustable

- The GLASS procedure seems to have worked well and should be applicable not only to further CMB analysis but also more widely