GLASS, or How to Deal with Uncertainty in Inference

Steven Gratton, DAMTP
Kavli Institute for Cosmology Cambridge
1 June 2023
Bayes’ Theorem

\[ p(\text{theory}|\text{data}) = \frac{p(\text{data}|\text{theory}) \cdot p(\text{theory})}{p(\text{data})} \]
But what should one do if?

\[ p(\text{theory}|\text{data}) = \frac{p(\text{theory})}{p(\text{data})} \]

• 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 distribution: \( p(x) = \frac{a}{2} \sin(ax) \)
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:

\[ \langle x \rangle = \frac{\pi}{2a} \]
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:

\[- \log p, a = - (X - \langle X \rangle)^T \langle \langle XX^T \rangle \rangle^{-1} \langle X \rangle, a.\]

- Here, \(X = (x, y, ..., x^2, xy, y^2, ..., x^3, x^2 y, xy^2, y^3, ...)^T\)
But which moments?

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

\[- \log p,_{\alpha} = -(X - \langle X \rangle)^T \langle \langle XX^T \rangle \rangle^{-1} \langle X \rangle,_{\alpha}.\]

- 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...

\[ \langle x \rangle = \frac{\pi}{2a} \]

\[ \langle x^2 \rangle = \frac{\pi^2}{2a^2} - \frac{2}{a^2} \]
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 (pyDELFII)
Max-likelihood-value comparisons

![Graph showing comparisons between different models with correlation coefficients: momento vs. C-SimLow: 0.93, momento vs. pydelfi: 0.93, C-SimLow vs. pydelfi: 0.99.]
\( \tau \) constraints

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Likelihood</th>
<th>( \tau (EE) )</th>
<th>( \tau (TTTEEE) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planck 2018</td>
<td>C-SimLow</td>
<td>0.0530 ( \pm ) 0.0071</td>
<td>...</td>
</tr>
<tr>
<td>momento</td>
<td></td>
<td>0.0507 ( \pm ) 0.0063</td>
<td>0.0527 ( \pm ) 0.0058</td>
</tr>
<tr>
<td>pydelfi</td>
<td></td>
<td>0.0517 ( \pm ) 0.0070</td>
<td>0.0513 ( \pm ) 0.0078</td>
</tr>
<tr>
<td>SRoll12</td>
<td>C-SimLow</td>
<td>0.0582 ( \pm ) 0.0057</td>
<td>...</td>
</tr>
<tr>
<td>momento</td>
<td></td>
<td>0.0581 ( \pm ) 0.0055</td>
<td>0.0604 ( \pm ) 0.0052</td>
</tr>
<tr>
<td>pydelfi</td>
<td></td>
<td>0.0588 ( \pm ) 0.0054</td>
<td>0.0580 ( \pm ) 0.0064</td>
</tr>
</tbody>
</table>
Since gone on to constrain $r$ (arXiv:2207.04903)

\begin{figure}[h]
\centering
\includegraphics[width=\columnwidth]{figure3}
\caption{Posters for $r$ (marginalised over $\tau$) from two-dimensional low-$\ell$ 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.}
\end{figure}
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