Bayesian Inference for Contemporary Lattice Quantum Field Theory

Partially based on
arXiv:2302.06550
10.5281/zenodo.7612101

Julien Frison

40th International Symposium on Lattice Field Theory, Fermilab, 04.08.2023

Overview

Background

- **Probabilistic Programming** has made huge progress in the last decade [Hoffmann'2014], in large part thanks to LQCD [Duane'1987], combined with modern computing resources

- Widely applied to epidemiology, finance, pharmaceuticals, marketing, social sciences, ...

- LQCD still mostly uses statistical methods of early 20th century [Bernstein'27] [Aitken'35] [Doob'35] [Quenouille'49] [Efron'79]

Objectives

I want:

- efficient learning of physical parameters and functions
- well-defined probabilistic interpretation
- unified and consistent framework
- combining strengths of current methods
- flexible model building with arbitrary assumptions
- metrics to test any assumption

This talk

Mostly insist on unification and flexibility

Current methods

Bootstrap/Resampling

Poor support of auto-correlations

Γ method

Gaussian approximations and linearisation

χ^2 fit

- Gaussian likelihood
- Covariance needs to be known in advance and precisely
- Often unstable. No theoretical convergence toward smthing meaningful with finite data.

Akaike IC

- Requires a reliable knowledge of correlated χ^2
- Needs data parametrisable by a regular model
- Nb of models to explore quickly explode $\Rightarrow$ computing time ($\times$bootstrap)

New method

Make it Bayesian from the start to the end!

- We directly get distributions and confidence intervals
- Every assumption is packed into the model, which can be made arbitrarily complicated
- Distance from model to truth can always be evaluated, with a robust criterion (KL)
- The HMC (a second one) makes it doable in practice

Bayesian formalism

The Bayes formula

$$P(a|y,M) = \frac{P(y|a,M)P(a|M)}{P(y|M)} \tag{1}$$

Bayesian vocabulary
parameter
posterior distribution
likelihood
prior
marginal distribution

strong priors
positivity
unitarity
convergence
literature
external inputs
power counting
fit stability
intuition
loose priors

Part of the family of *generative* machine-learning models thanks to the PPD:

$$P(y'|y,M) = \int P(y'|a,M)P(a|y,M)da \tag{2}$$

Applying the HMC

We do not need a close formula for $P(a|y, M)$, we can just draw $a_1, a_2, a_3, \ldots$.
Exactly what our good old HMC does!

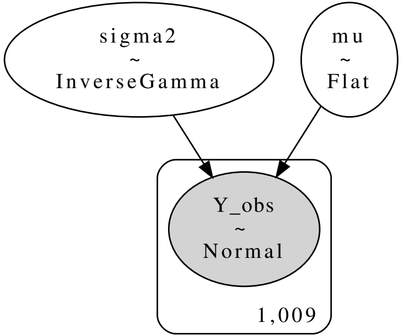
$$P(a|y, M) = \frac{P(y|a, M)P(a|M)}{P(y|M)} \tag{3}$$

Bayesian vocabulary	LQCD analogue
parameter	configuration
posterior distribution	
likelihood	e^{-S}
negative log-likelihood	action (or χ^2)
prior	
marginal distribution	partition function

Software

- I will show tests with PyMC
  In Python and simple to use, but several alternatives exist
- Vectorisation and Automatic Differentiation (HMC forces) handled by PyTensor
  Made for somewhat complex ML methods & Deep Learning
- The user only needs to write models, and it can be anything:
  - Likelihood function can be arbitrarily complicated
  - Does not have to be parametric:
    Bayesian Bootstrap, Gaussian Processes, Bayesian Neural Networks, ...
    Nor to have less parameters than data points
  - Does not need to be the *true* model:
    Models are always an approximation, to be checked a posteriori on data (IC)
  - Analytical computations only needed to speed up with marginalisation
- Runs on a laptop but scales with cluster/GPU

Bayesian networks and plate notation



- Example: unidimensional Gaussian inference
- parameters/priors upwards, observable below
  can have many layers
- Data: one number Y_{obs} $\times$ 1009 configurations
- μ uncertainty and σ are two different things
- The MLE gives us the usual point estimates:
  empirical mean and variance

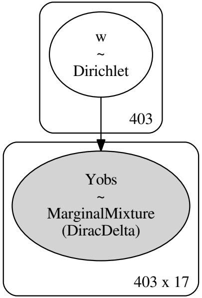
A few models

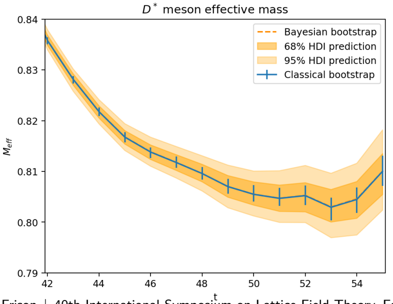
Data

- All the following models use a two-point function (CLS H101)
- Not the only possible application in LQCD
- Not an exhaustive list of models
- Show how many **different** things one can do from the **same** data within the same framework

Bayesian Bootstrap

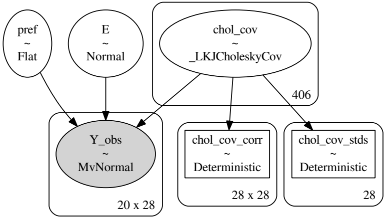
- Intuitively, Bootstrap feels “half-Bayesian”
- Actually it **is** a Bayesian model [Rubin’81]
- Simple example of non-parametric model (=∞ params)
  irrelevant params are marginalised
- Somewhat useful analogy to have a model of reference to compare with



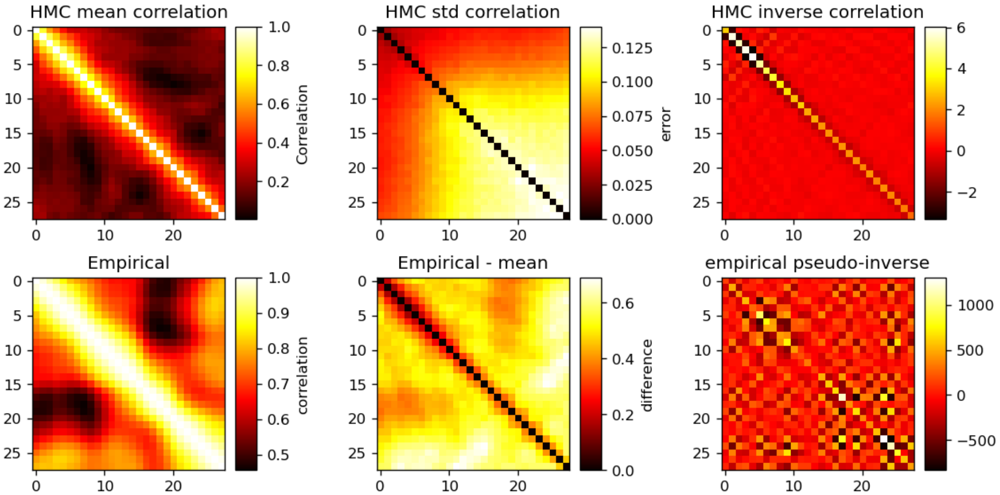


Infering a covariance (1)

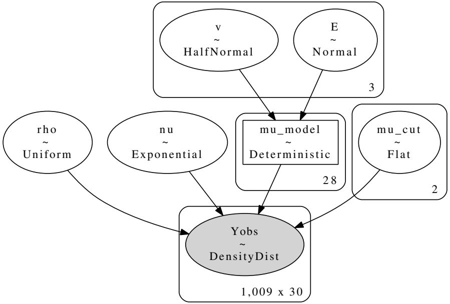
- Generalised Least Square is notoriously delicate in LQCD
- The *sample* covariance is not the *true* covariance
- We spend a lot of effort on evaluating the uncertainty on the mean but usually neglect the **uncertainty on the covariance**!
- In practice often leads to get **badly conditioned or non-positive** matrices
- Some regularisations are well-motivated but it does not propagate uncertainty, nor communicate with the model

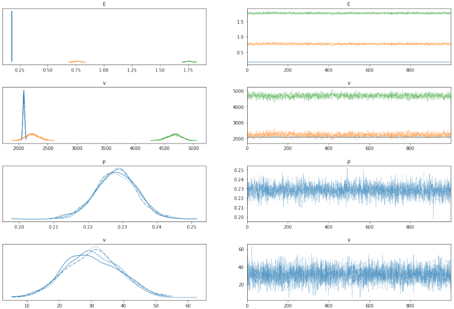


Infering a covariance (2)



Multi-exponential model with correlations



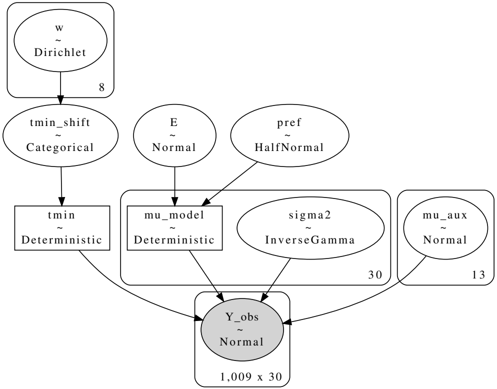


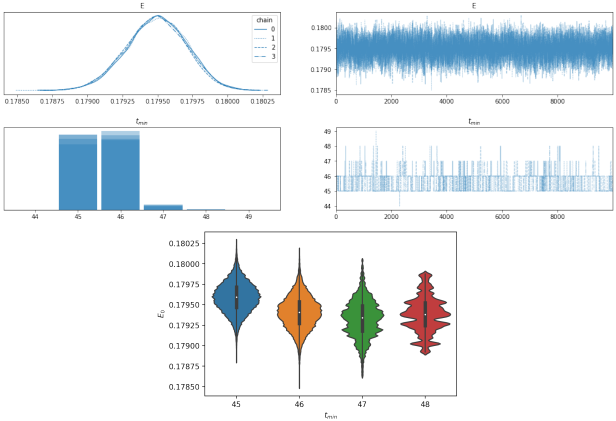
- Mixing correlated model (in Euclidian time) with marginalised Wishart prior...
- ... and auto-regressive model (auto-correlation between configurations)

$$y_\tau(t) = \rho_0 + \sum_{i=1}^{r} \rho_i y_{\tau-i}(t) + \xi_\tau(t), \qquad \langle \xi_\tau(t)\xi_\tau(t')\rangle \neq 0 \tag{4}$$

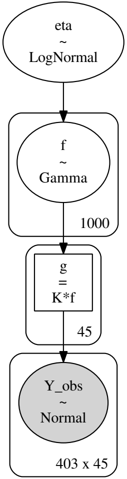
Model averaging

- There are now **many ways** to do model averaging (see [Neil] for ICs and more details)
- One is to simply put everything into a **single model**:

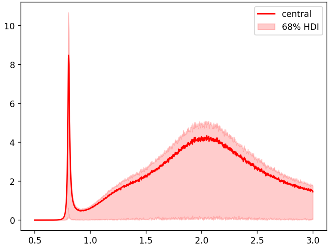


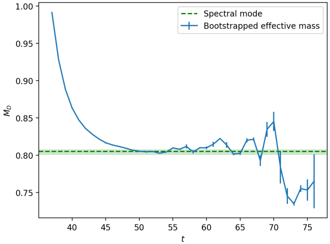


Spectral function



- Bayesian Reconstruction model from [Rothkopf'21]
- ≈equivalent to infinite multi-exponential
- does not include smoothing (Gaussian Processes being studied)
- makes heavy use of positivity
- Applied to full correlator without t_{min} cut
- Ground state determined with high precision





Conclusion

Conclusion

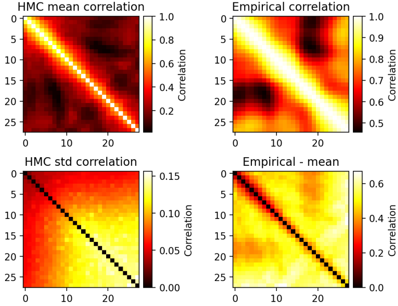
- I presented how a bayesian framework can provide **well-grounded** LQCD results with full propagation of **all uncertainties**
- I demonstrated its application with **many different models**
- Unifies things which were not obviously related
  - Bootstrap (i.e. non-gaussianity)
  - Covariance regularisation
  - auto-correlations
  - χ^2 fits
  - Model averaging
  - Inverse problem
  - ...and more that I have not covered
- Model-building is only limited by your imagination
- You can adapt this code to your needs: [10.5281/zenodo.7612101]

Thanks for your attention!

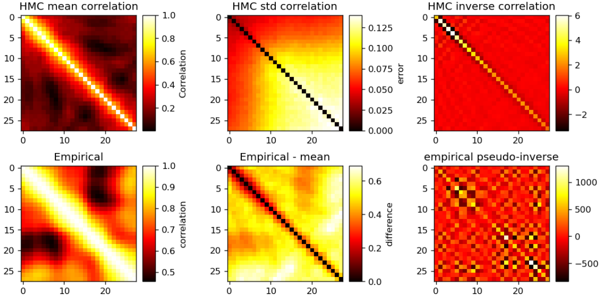
Backup slides

Trivial-vs-exp models for covariance

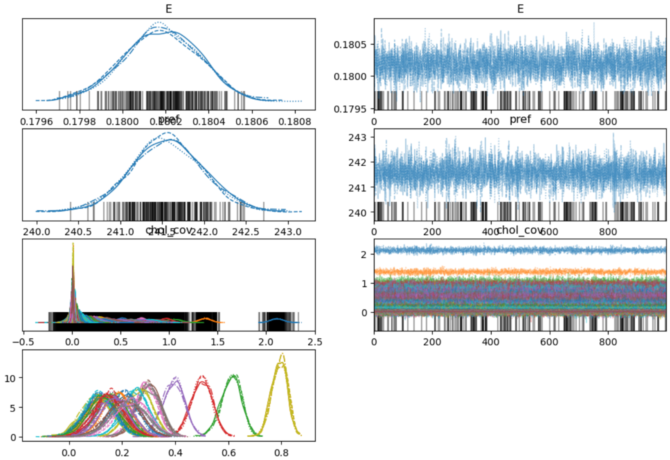
Trivial



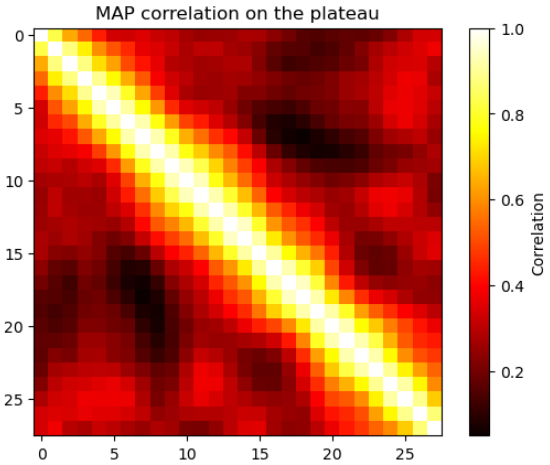
Exponential



LKJ-exp sampling



Covariance MAP



Regularisation