Inferring Parton Distributions with Gaussian Processes

The 43rd International Symposium on Lattice Field Theory
July 29th, 2026

HadStruct Collaboration: J. Karpie, H. Dutrieux, K. Orginos, S. Zafeiropoulos

Speaker: Yamil Cahuana Medrano

Based on: JHEP 04 (2026) 182, arXiv:2510.21041

CENTER for NUCLEAR FEMTOGRAPHY

WILLIAM & MARY
CHARTERED 1693

FULBRIGHT

Jefferson Lab

Outline

Particularities of fitting and solving the inverse problem…

- ✓ Motivation
- ✓ PDFs on Euclidean Lattice
- ✓ Gaussian processes
  - ➢ Bayesian approach
  - ➢ Levels of inference (3 ways to write Bayes)
- ✓ Results
- ✓ Conclusion

Joe's talk about GPs for GPDs

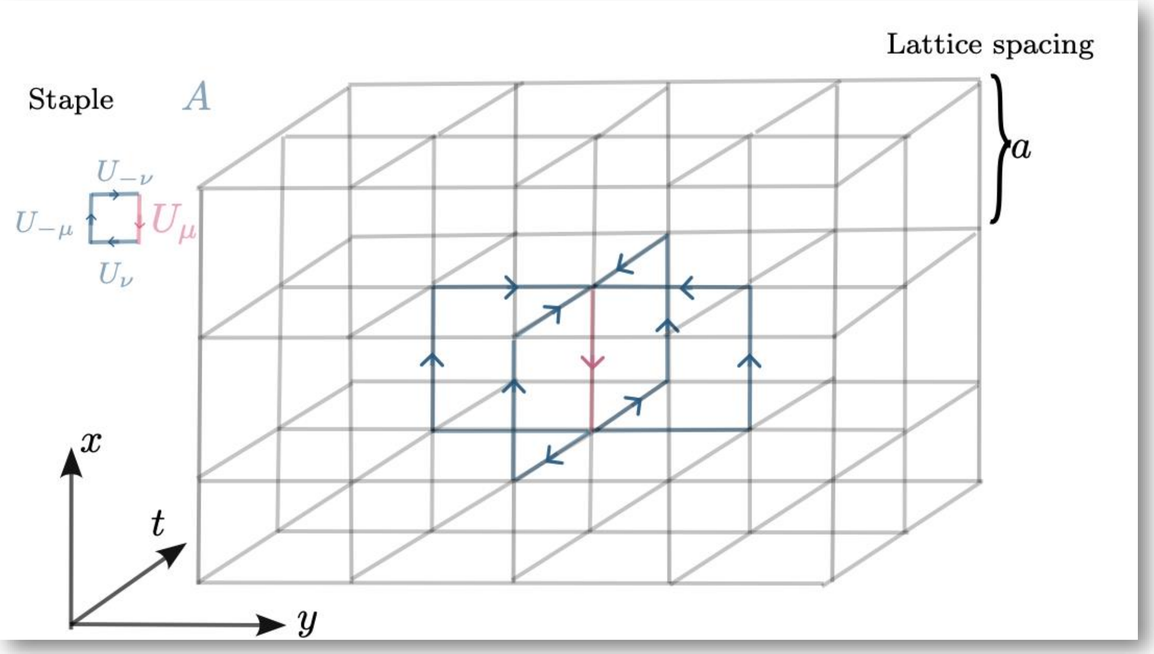
Motivation

PDF ⟷ LQCD

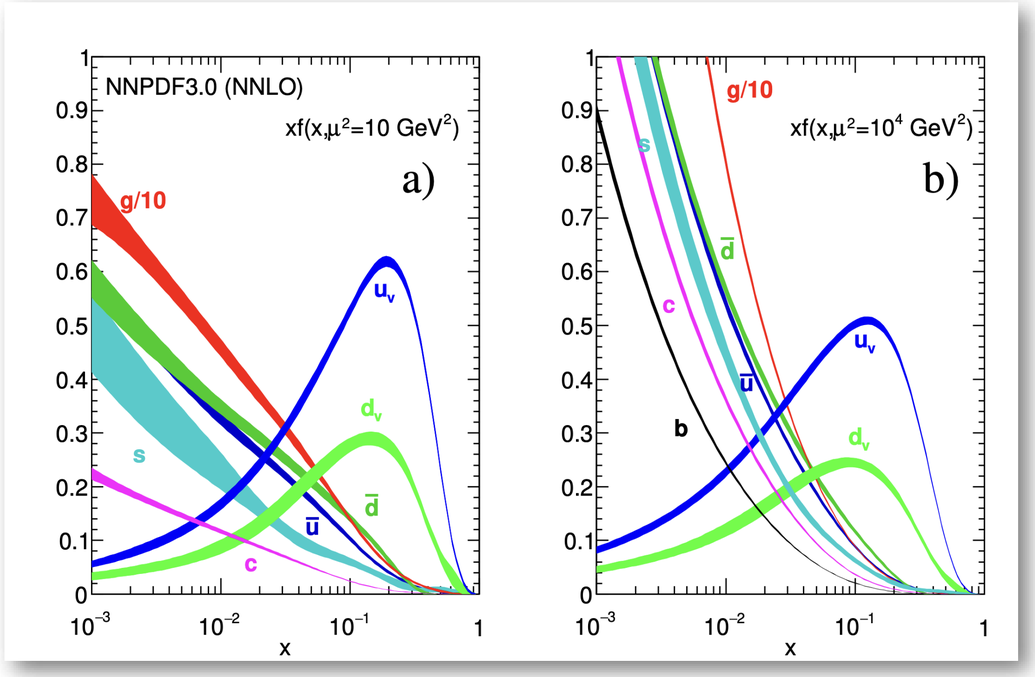
Introduction to LQCD and PDFs

$$\mathcal{L}_{QCD} = -\frac{1}{4}F^{\alpha}_{\mu\nu}F^{\mu\nu}_{\alpha} + \overline{\psi}(i\not{D} - m_i)\psi_i$$

$$\langle \mathcal{O}(fields) \rangle = \frac{1}{Z}\int D[fields]\mathcal{O}(fields)e^{-S(fields)}$$

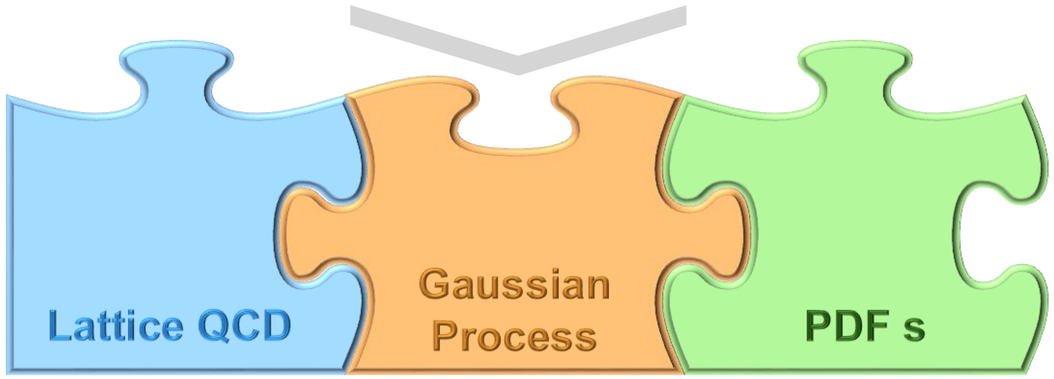


`Parton distributions and lattice QCD calculations
arXiv:1711.07916v3



Parton distributions for the LHC run II
10.1007/jhep04(2015)040

$$f_q(x,\mu) = \int_{-\infty}^{\infty}\frac{dz^-}{2\pi}e^{-xP^+z^-}\,\langle P|\,\bar{\psi}(z^-)U(0;z^-)\psi(0)\,|P\rangle$$



PDFs on a Euclidean Lattice

J. Collins, PARTON DISTRIBUTION AND DECAY FUNCTIONS*

X. Ji, Phys. Rev. Lett. 110, 262002 (2013).

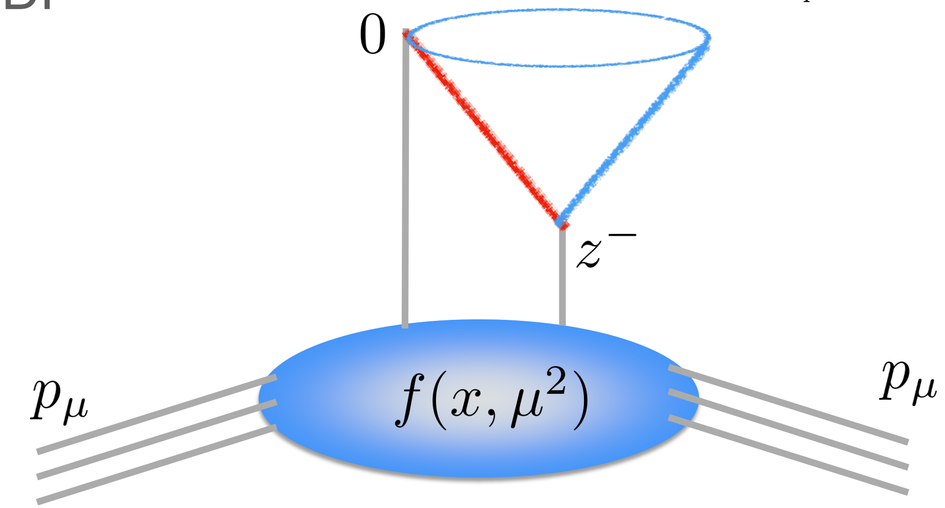
A.Radyushkin, Phys. Rev. D 96, 034025 (2017)

Pseudo-PDFs

$$f(x) = \int_{-\infty}^{\infty} \frac{dz^-}{2\pi} e^{-xP^+z^-} \langle P| \bar{\psi}(z^-)\gamma^+ U(0;z^-)\psi(0) |P\rangle$$

$$M^\alpha(p,z) =< p|\bar{\psi}(z)\gamma^\alpha U(z;0)\psi(0)|p >= p^\alpha \mathcal{M}(\nu, z^2) + z^\alpha \mathcal{N}(\nu, z^2)$$

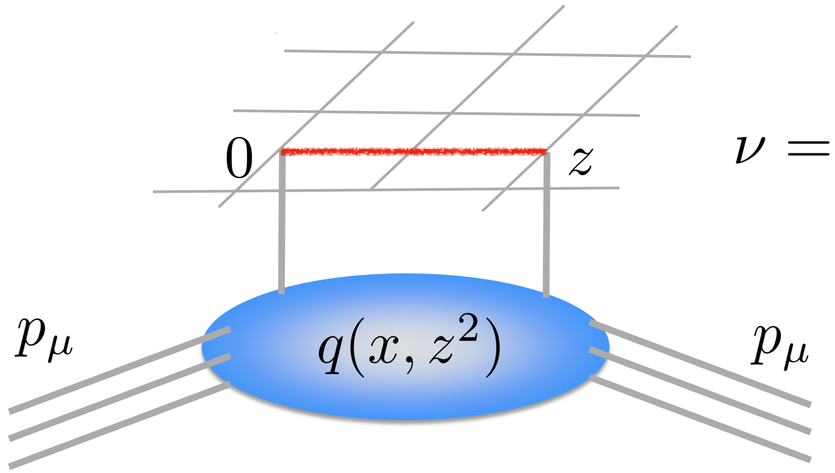
PDF $\alpha = +$ $z_\alpha = (0, z^-, 0_T)$ $p_\alpha = (p^+, \frac{m_p^2}{2p^+}, 0_T)$



$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\nu \mathcal{M}(-\nu, 0) e^{-ix\nu}$$

$$\mathcal{M}(-p_+ z_-, 0) = \int_{-1}^{1} dx f(x) e^{-ixp_+z_-}$$

Pseudo-PDF $\alpha = 0$ $z_\alpha = (0, 0, 0, z_3)$ $p_\alpha = (p_0, 0, 0, p_3)$



$$\nu = p \cdot z$$

$$q(x, z^2) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\nu \mathcal{M}(\nu, z^2) e^{ix\nu}$$

$$\mathfrak{M}(\nu, z^2) = \frac{\mathcal{M}(\nu, z^2)}{\mathcal{M}(0, z^2)} = \int_{-1}^{1} dx q(x, z^2) e^{-ix\nu}$$

Inverse problem

Fourier transform or Inverse Fourier transform?

$$q(x, z^2) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\nu e^{-ix\nu} \mathfrak{M}(\nu, z^2)$$

We may need to make assumptions about the behavior of ITD at large v

or

A.Radyushkin, Phys. Rev. D 96, 034025 (2017)

$$\mathfrak{M}(\nu, z^2) = \int_{-1}^{1} dx e^{ix\nu} q(x, z^2)$$

Reconstructing parton distribution functions from Ioffe time data:
from Bayesian methods to Neural Networks, JHEP 04 (2019) 057

There is plenty of discussion and details:

- **LaMET's Asymptotic Extrapolation vs. Inverse Problem**
  J.-W. Chen et al., *Phys. Rev. D 113*, 014509 (2026),
- **Inverse Problem in the LaMET Framework**
  H. Dutrieux, et al., *Phys.Rev.D* 113 (2026) 7, 074524

$$\mathcal{L} \equiv \int_{-1}^{1} dx e^{ix\nu}$$

We must solve an inverse problem…

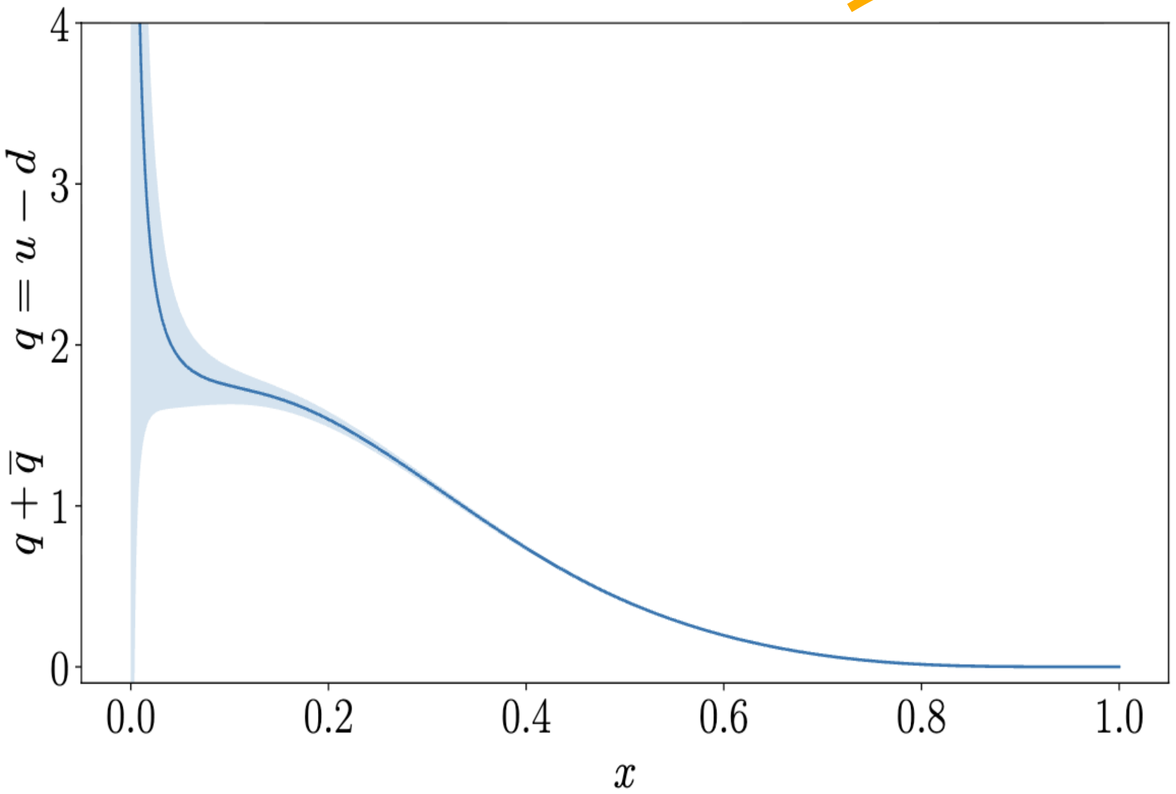
Inverse problem (Closure test)

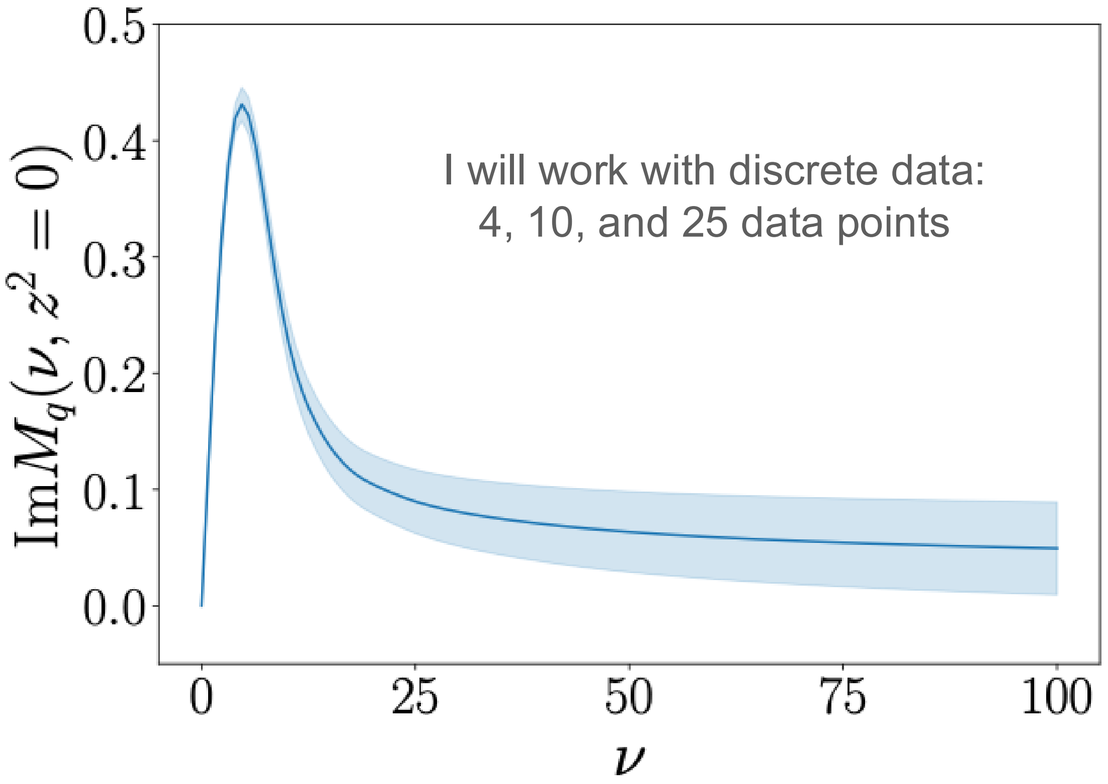
NNPDF 4.0 (imaginary component)

NNPDF Collaboration, The Path to Proton Structure at One-Percent Accuracy

Forward map = Easy

$\mathcal{L}$





I will work with discrete data:
4, 10, and 25 data points

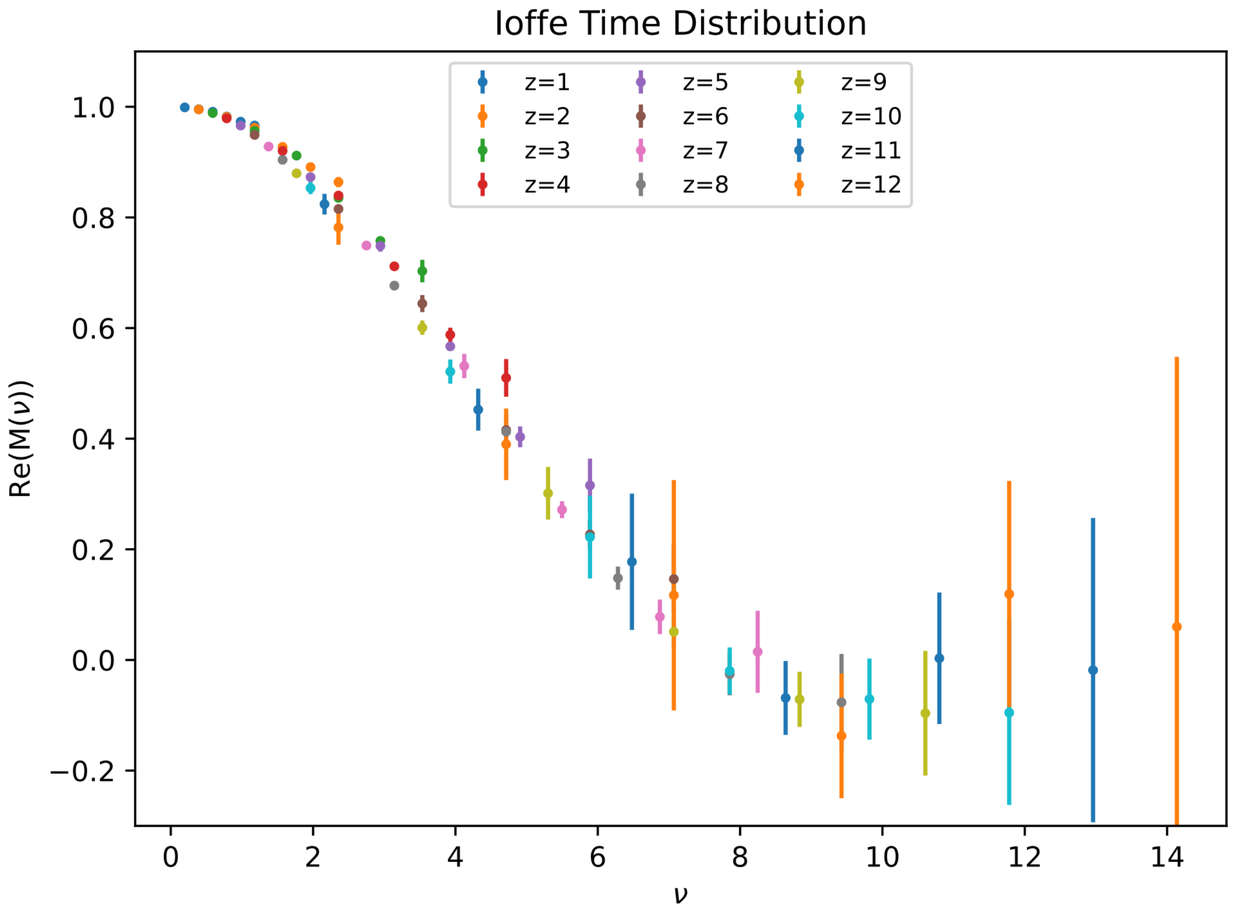
$\mathcal{L}^{-1}$

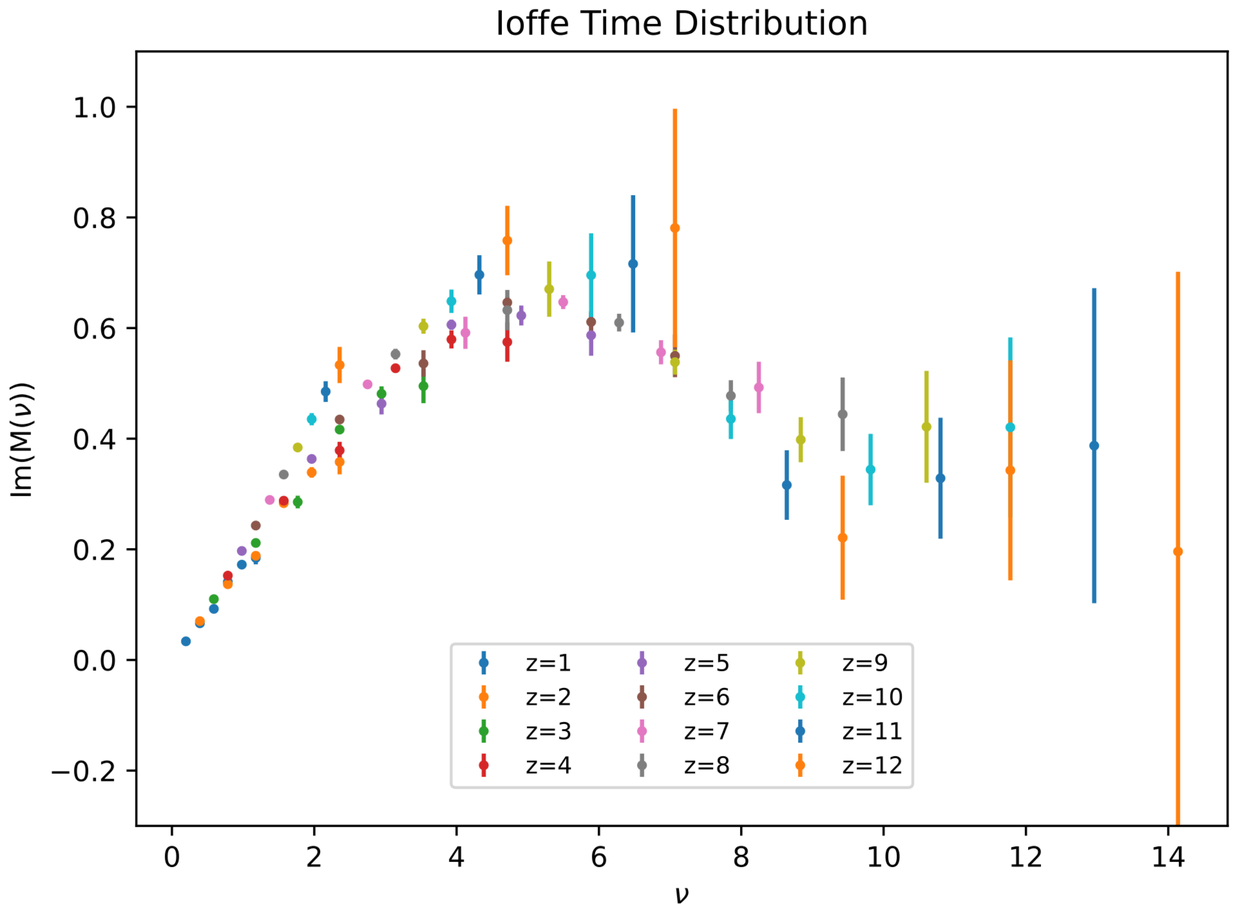
$$\mathfrak{M}(\nu) = \int_{-1}^{1} dx e^{ix\nu} q(x)$$

Lattice data

Unpolarized isovector ITD of the nucleon

Lattice details: 2+1 flavors of clover improved Wilson quarks with a lattice spacing a = 0.094(1) fm and a pion mass of 358(3) MeV



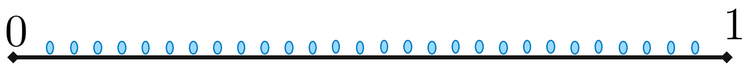


Colin Egerer, Forward and Off-Forward Parton Distributions from Lattice QCD

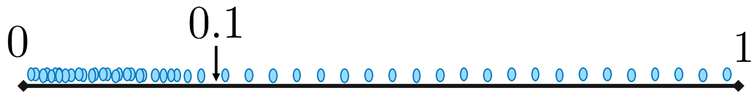
Finite elements

Grids (from continuous to discrete)

uniform

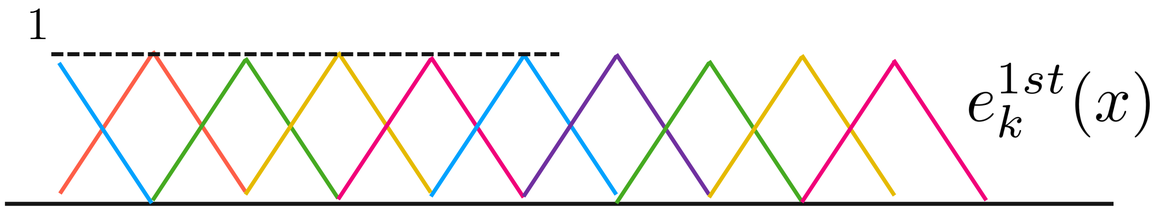


Log-uniform & uniform

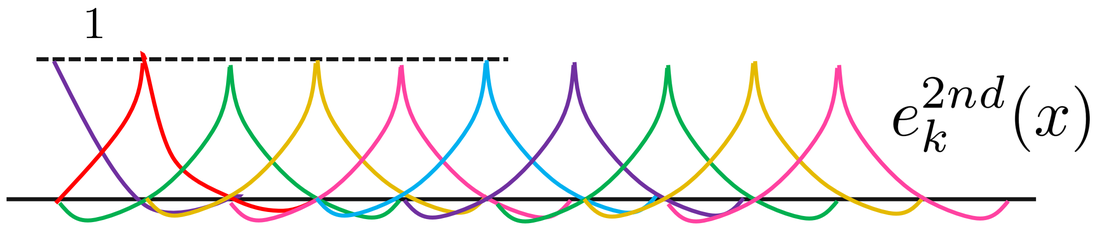


We can represent any function in x in this basis

$$q(x) = \sum_{k \in N_x} q_k e_k(x)$$



First degree finite elements (FE)

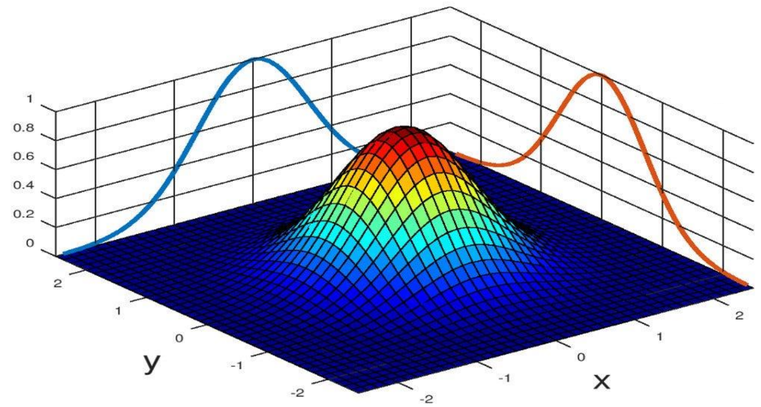


My grid in x uses second degree FE

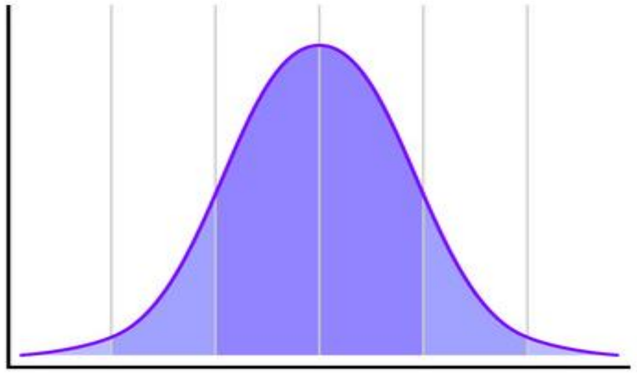
Gaussian processes!!!

= Stochastic Process…

$\vec{\mu}_i, \sigma_{ij}$



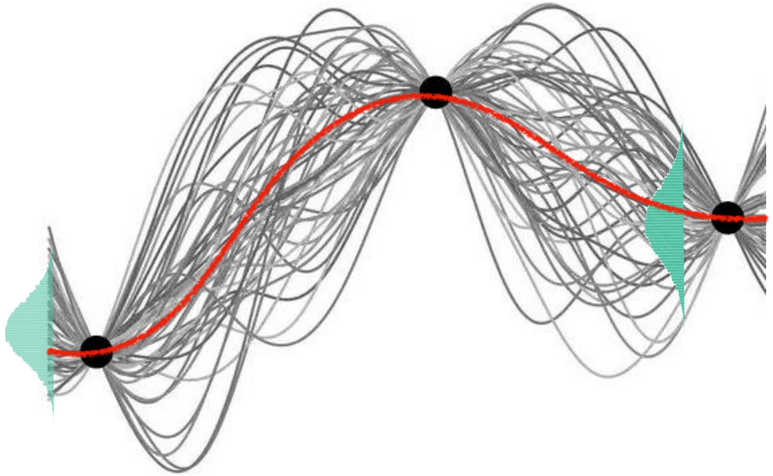
μ, σ



Normal Distribution

$$p(x) \approx e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$\mu(x), \sigma(x, x')$ or $K(x, x')$



Try to imagine an infinite dimensional normal distribution

Machine Learning of Nonlinear Partial Differential Equations. arXiv: 1708.00588
Gaussian Processes for Machine Learning, E. Rasmussen and C. K. I. Williams

Gaussian processes

Parametric vs Non-Parametric

Some Results on Tchebycheffian Spline Functions, George K., Grace Wahba, 1971

Gaussian Processes for Machine Learning, E. Rasmussen and C. K. I. Williams

Parametric

1) $$\hat{\theta} = \min_{\theta} \left(\chi^2(q(x;\theta))\right)$$

$$\chi^2(q(x;\theta)) = \frac{1}{2}(M_i - \mathcal{L}_{\nu_i} q(x;\theta)) C_{ij}^{-1} (M_j - \mathcal{L}_{\nu_j} q(x;\theta))$$

2) $$q(x;\theta)_{PDF} = N x^{\alpha}(1-x)^{\beta}$$

3) $$P(\theta|M^i) = \frac{e^{-\chi^2(q(x;\theta))} P(\theta)}{P(M^i)}$$

Bayesian, but still parametric

If we do not parametrize q, what we do?

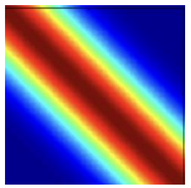
Non-Parametric

1) $$\hat{q}(x) = \min_{q \in H} \left(\chi^2(q(x)) + \|\mathcal{P}q\|_H^2\right)$$

Imposes additionally conditions on the Hilbert space $H = \{q(x) \mid$ continuous? smooth? square-integrable $\}$

$$\|\mathcal{P}q\|_H^2 \rightarrow q(x) K^{-1}(x,x') q(x')$$

Remarkable result!!! Parametrization of the inner product of the H space (RKHS)

2) 

$$K(x,x') = \sigma e^{\frac{|x-x'|^2}{2l^2}}$$

We can recover a parametric feature

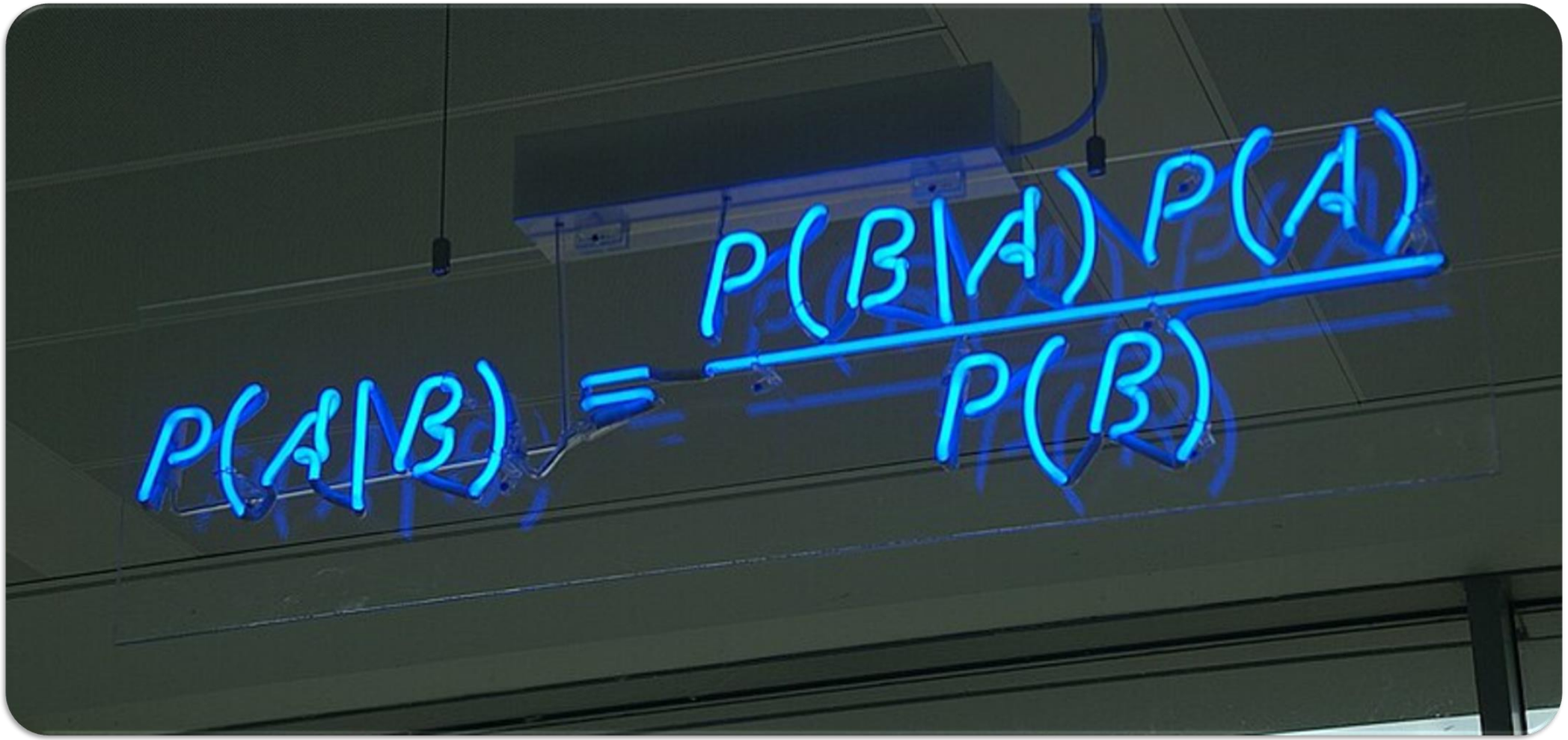
$$q(x) \rightarrow q(x) - q_{PDF}(x)$$

3) $$P(q(x)|M^i) = \frac{e^{-\chi^2(q(x))} e^{-\|\mathcal{P}q\|_H^2}}{P(M^i)}$$

Bayes theorem

$$\text{Posterior} = \frac{\text{Likelihood}\quad\text{Prior}}{\text{Evidence}}$$

A given B...

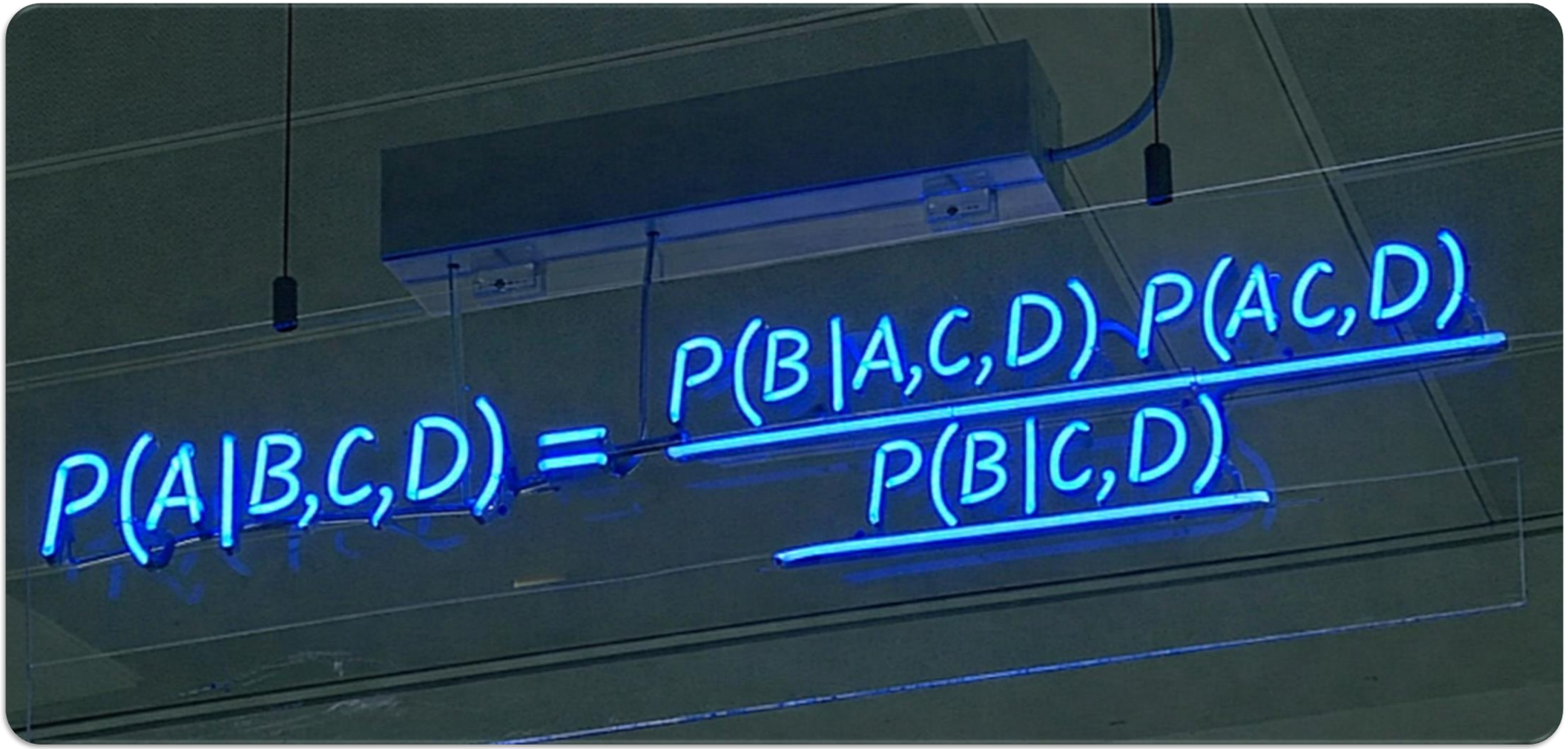


Bayes theorem (update)

$$\text{Posterior} = \frac{\text{Likelihood}\quad\text{Prior}}{\text{Evidence}}$$

for non-parametric methods

$$\{functions, Parameters, Hypothesis, Data\} \equiv \{A, C, D, B\}$$



Non-parametric Bayesian Approach

Levels of inference

$$\{functions, Parameters, Hypothesis, Data\} \equiv \{q(x), \theta, \mathcal{H}, M^l\}$$

Parametric Models

3rd

$$P(\mathcal{H}_i|M^l) = \frac{P(M^l|\mathcal{H}_i)P(\mathcal{H}_i)}{P(M^l)}$$

Average/Selection of Models

$$\mathfrak{q}(x) \quad \mathfrak{q}(x)\mathfrak{q}(x)$$

2nd

$$P(\theta|M^l, \mathcal{H}) = \frac{P(M^l|\theta, \mathcal{H})P(\theta|\mathcal{H})}{P(M^l|\mathcal{H})}$$

Solved Numerically (sample parameters)

$$< q(x) > \quad < q(x)q(x) >$$

1st

$$P(q(x)|M^l, \theta, \mathcal{H}) = \frac{P(M^l|q(x), \theta, \mathcal{H})P(q(x)|\theta, \mathcal{H})}{P(M^l|\theta, \mathcal{H})}$$

Solved analytically (path integrals)

$$\bar{q}(x) \quad \overline{q(x)q(x)}$$

Marginal Likelihood or Evidence = Likelihood of the next level of inference

Non-parametric Bayesian Approach

Mean and Covariance

$$\{functions, Parameters, Hypothesis, Data\} \equiv \{q(x), \theta, \mathcal{H}, M^l\}$$

Parametric Models

3rd

Average/Selection of Models

$$\mathfrak{q}(x) = \sum_i P(\mathcal{H}_i \mid M^l) \langle q(x) \rangle_i$$

$$\mathfrak{q}(x)\mathfrak{q}(x') = \sum_i P(\mathcal{H}_i \mid M^l) \left[\langle q(x)q(x') \rangle_i + (\langle q(x) \rangle_i - \mathfrak{q}(x))(\langle q(x') \rangle_i - \mathfrak{q}(x')) \right]$$

2nd

Solved Numerically (sample parameters)

$$\langle q(x) \rangle = \int d\theta P(\theta \mid M^l, \mathcal{H}_i) \bar{q}(x)$$

$$\langle q(x)q(x') \rangle = \int d\theta P(\theta \mid M^l, \mathcal{H}_i) \left[\overline{q(x)q(x')} + (\bar{q}(x) - \langle q(x) \rangle)(\bar{q}(x') - \langle q(x') \rangle) \right]$$

1st

Solved analytically (path integrals)

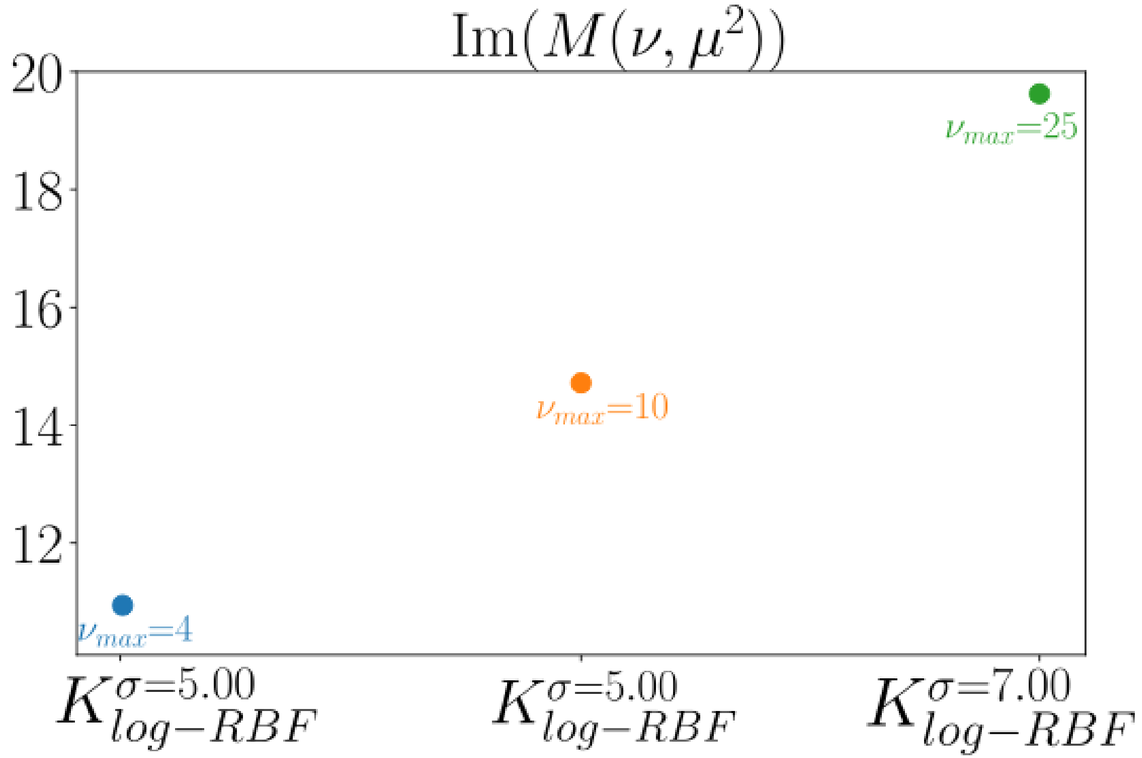
$$\bar{q}(x) = \int D[q(x)] P(q(x) \mid M^l, \theta, \mathcal{H}_i) q(x) \qquad \overline{q(x)q(x')} = \int D[q(x)] P(q(x) \mid M^l, \theta, \mathcal{H}_i) q(x) q(x')$$

Once you have the posterior, integrate over it…

KL Divergence

Information gained globally

$$D_{KL}(P[q(x)|M,\theta,\mathcal{H}]||P[q(x)|\theta,\mathcal{H}]) = \int D[q(x)]\ P[q(x)|M,\theta,\mathcal{H}] \log\left(\frac{P[q(x)|M,\theta,\mathcal{H}]}{P[q(x)|\theta,\mathcal{H}]}\right)$$

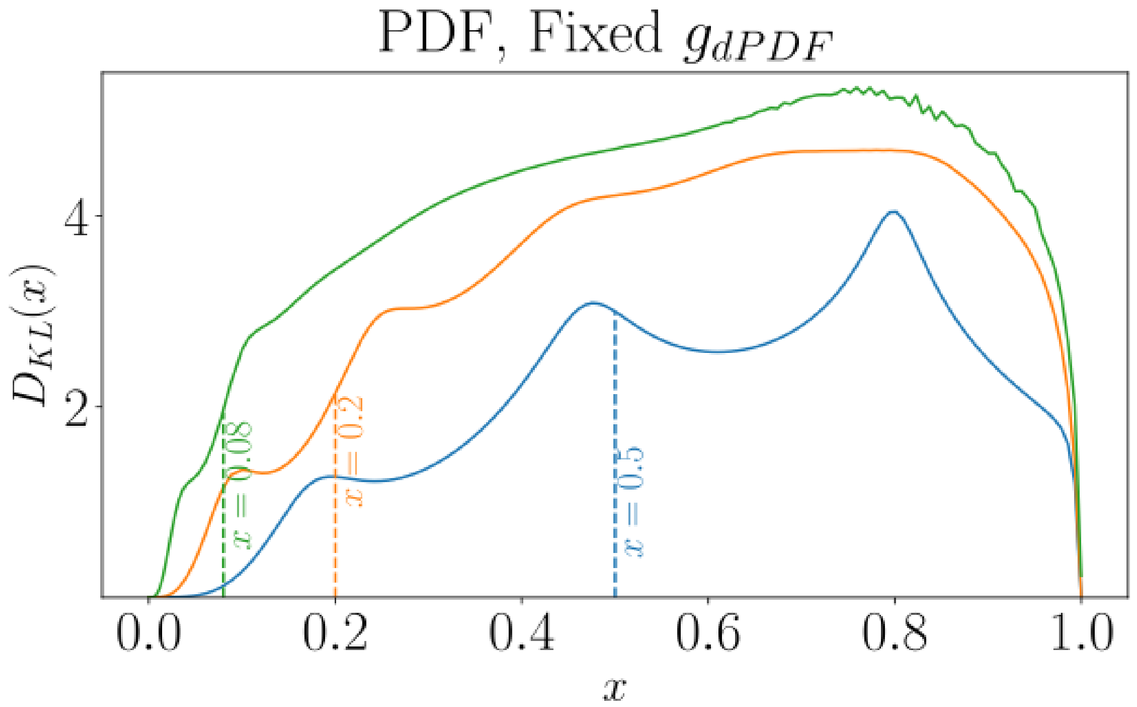


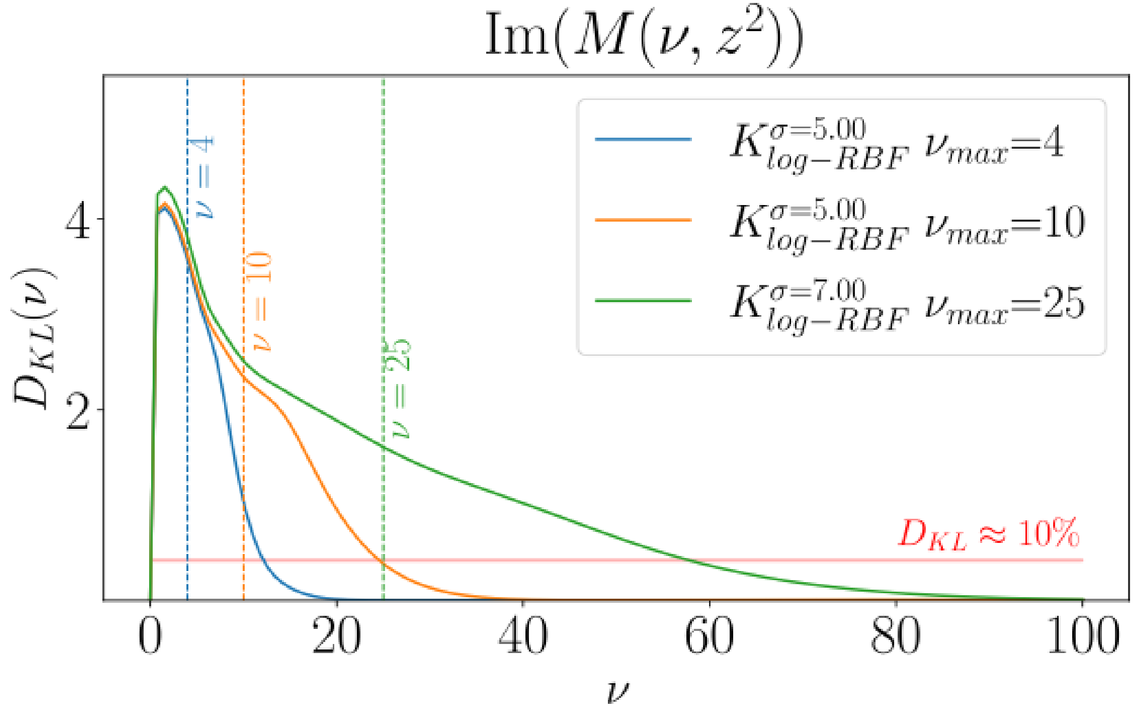
KL Divergence

Information gained locally

Valentine & Sambridge, GJI 220, 1632 (2020)

$$D_{KL}(x_i) \equiv D_{KL}(P[q_i|M,\theta,I]||P[q_i|\theta,I]) = \int D[q(x_i)]\, P[q_i|M,\theta,I] \log\left(\frac{P[q_i|M,\theta,I]}{P[q_i|\theta,I]}\right)$$





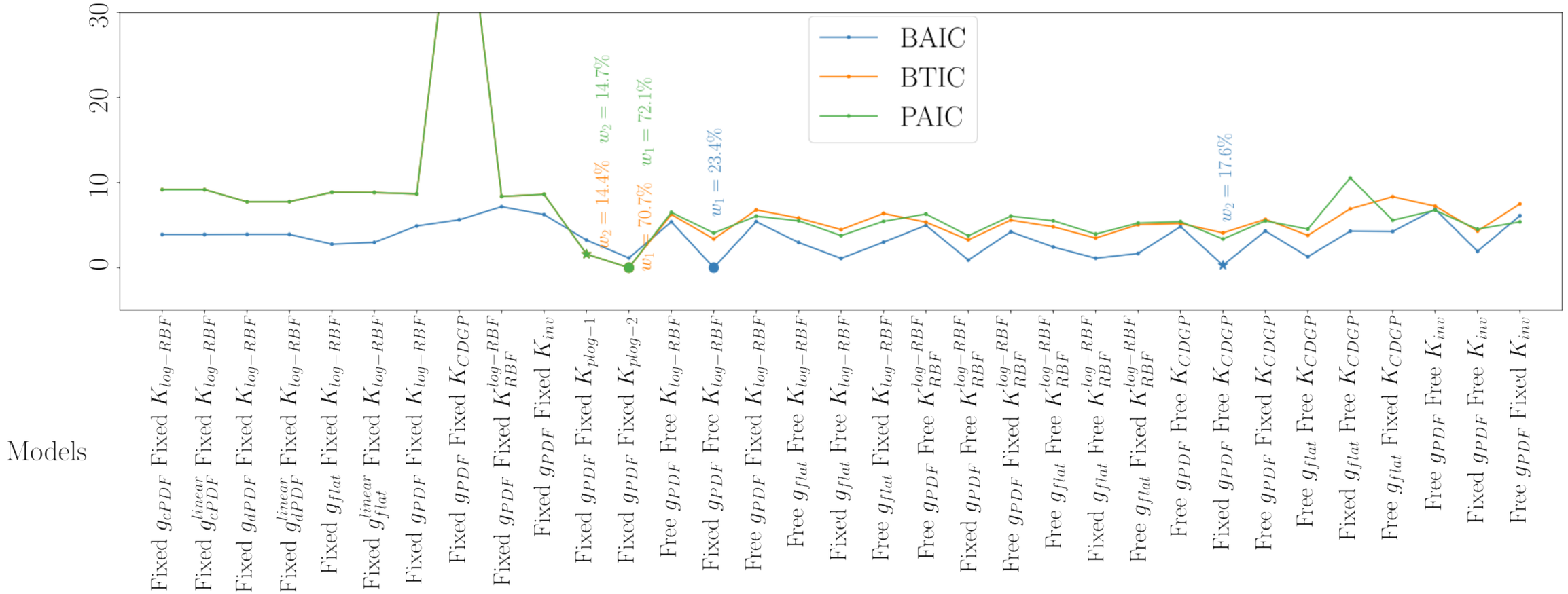
$$P_{const} = e^{-\frac{1}{2\lambda}\left(\int_0^1 dx q(x) - 1\right)^2 - \frac{1}{2\lambda_c}\left(\int_0^1 dx q(x)\delta(1-x)\right)^2}$$

3rd Level of inference

Selection ~ Averaging

$$P(\mathcal{H}_i|M^l) = \frac{P(M^l|\mathcal{H}_i)P(\mathcal{H}_i)}{P(M^l)}$$

Bayesian model averaging for analysis of lattice field theory results, William I. Jay, Ethan T. Neil

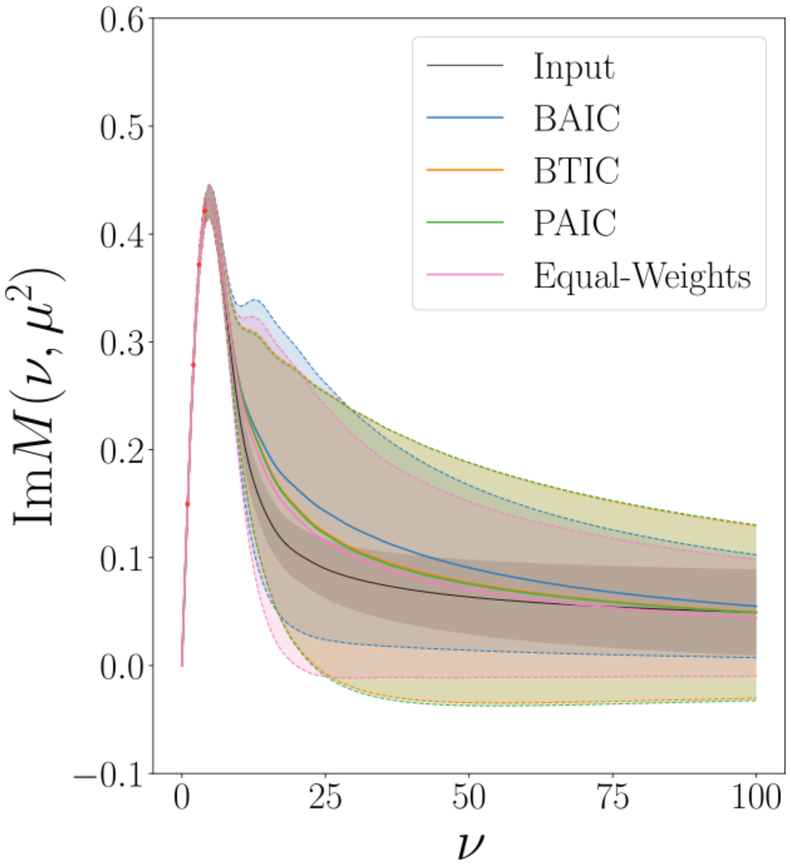


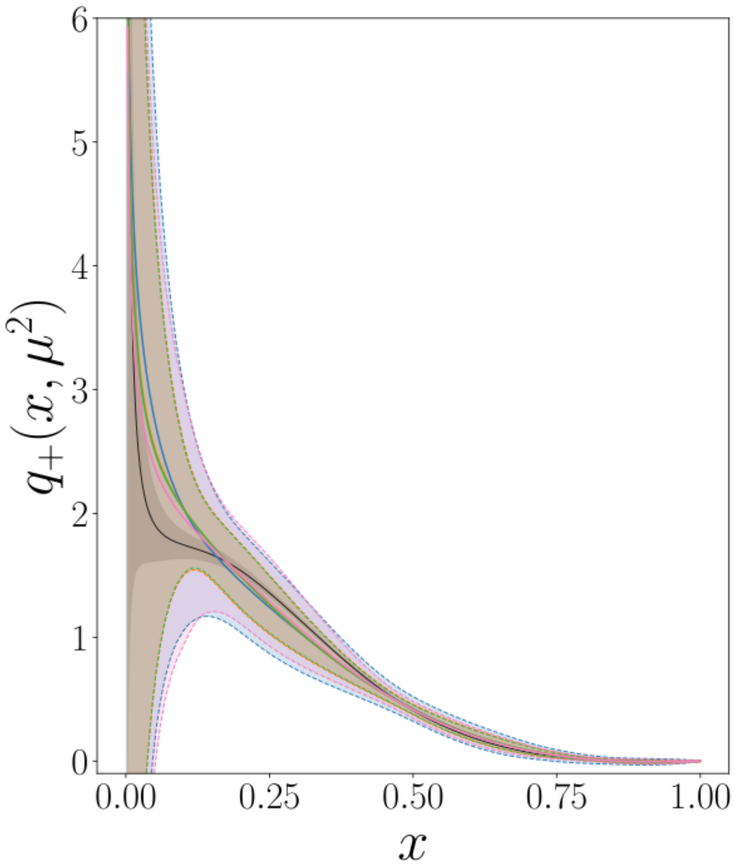
3rd Level of Inference

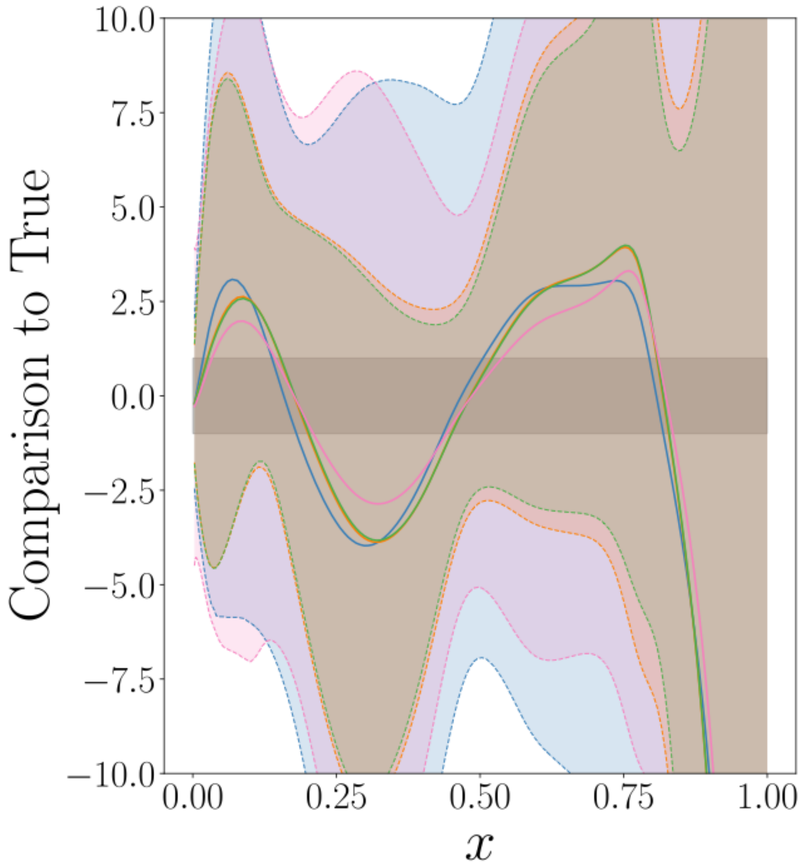
Results for Closure test (4 points)

$$P(\mathcal{H}_i|M^l) = \frac{P(M^l|\mathcal{H}_i)P(\mathcal{H}_i)}{P(M^l)}$$

4 data points





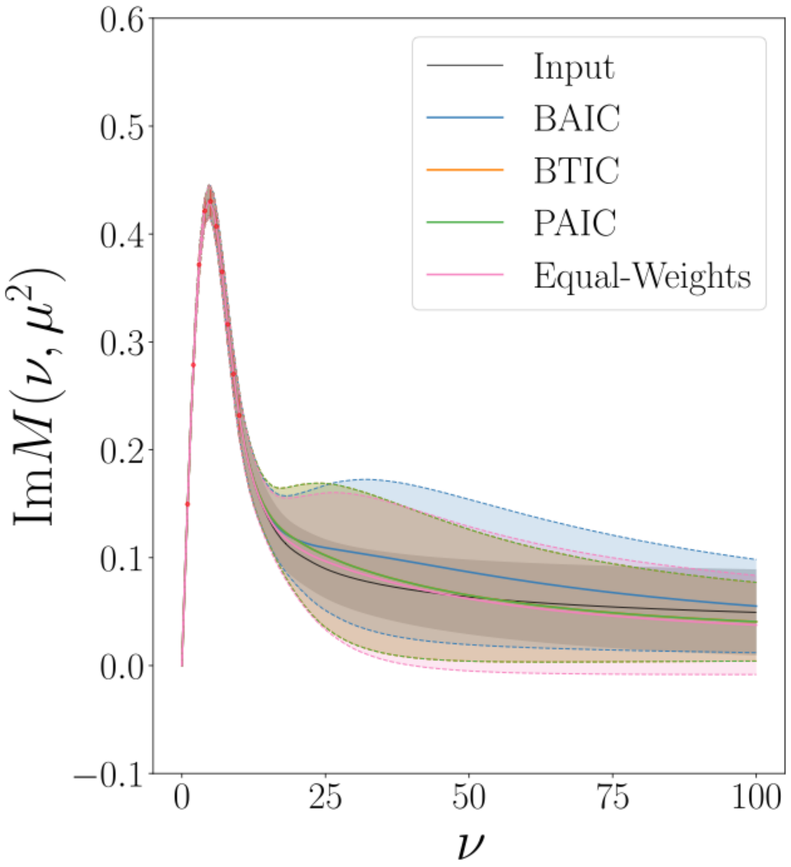


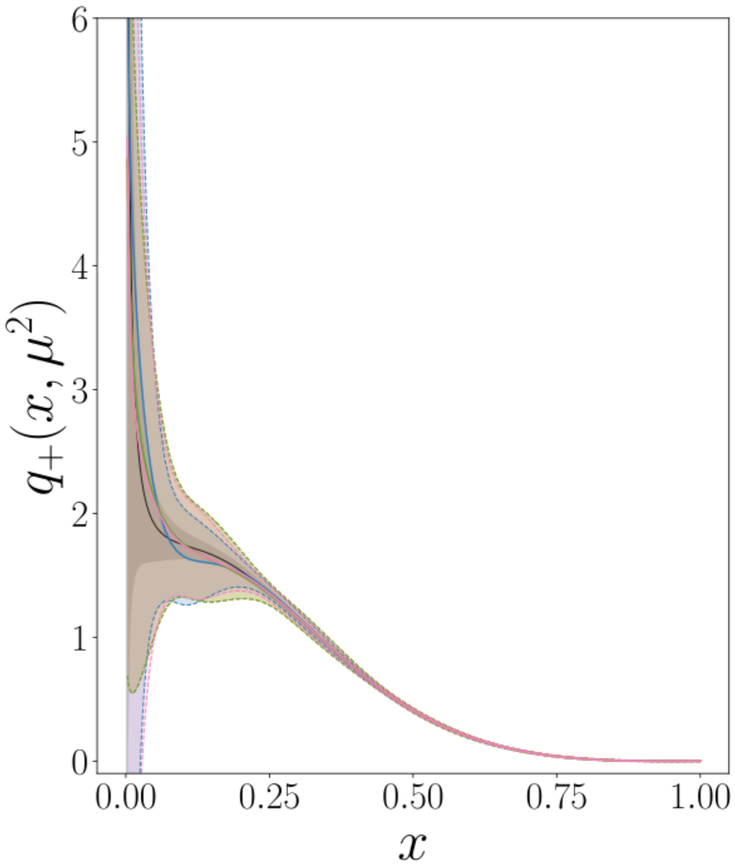
3rd Level of Inference

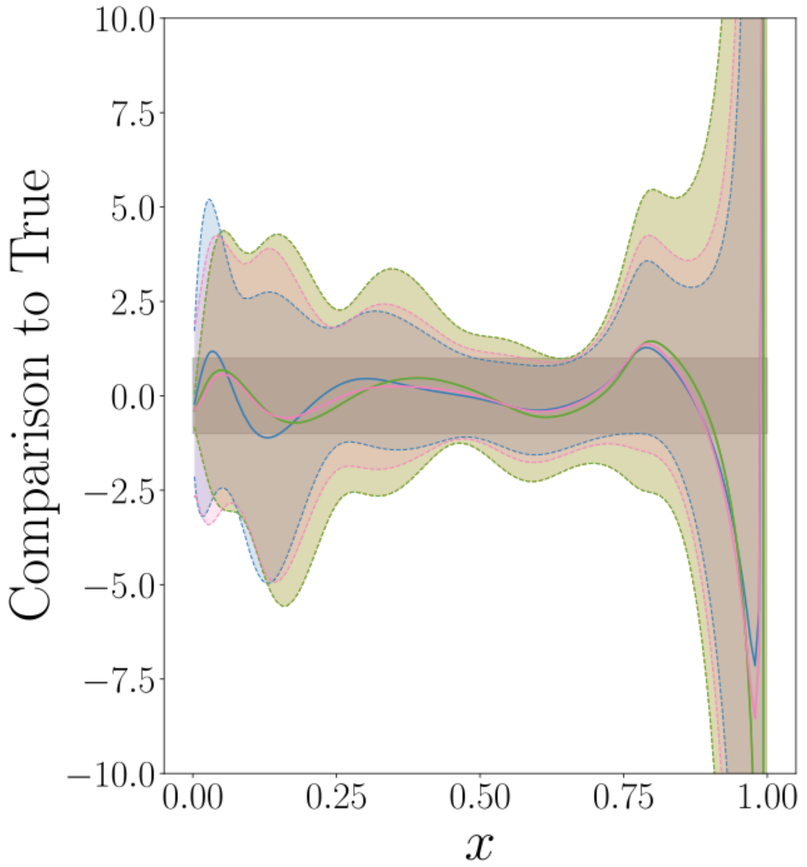
Results for Closure test (10 points)

$$P(\mathcal{H}_i|M^l) = \frac{P(M^l|\mathcal{H}_i)P(\mathcal{H}_i)}{P(M^l)}$$

10 data points





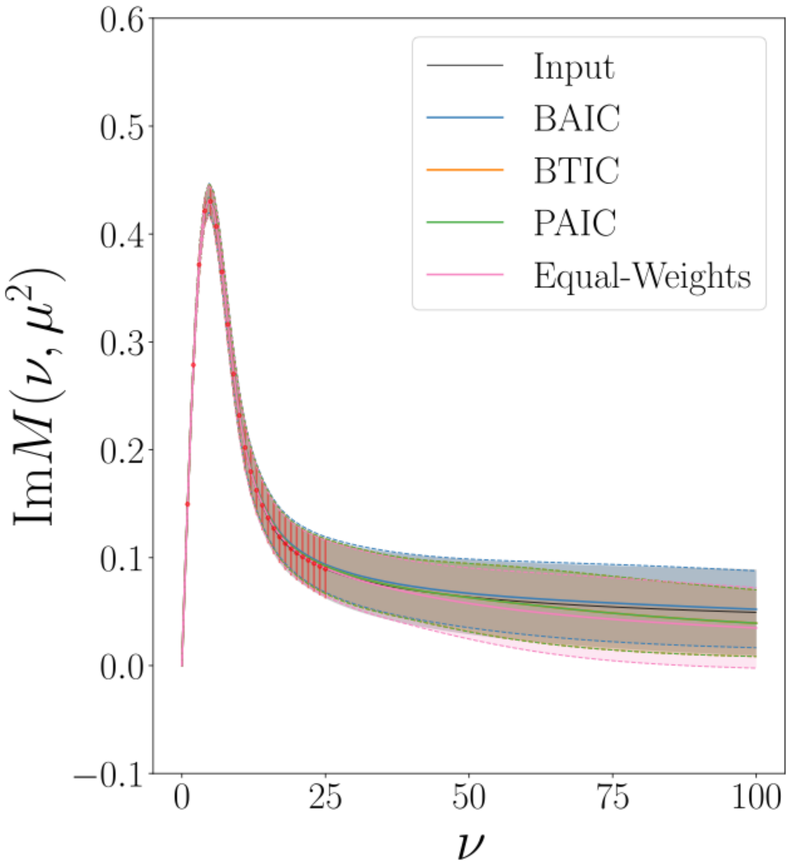


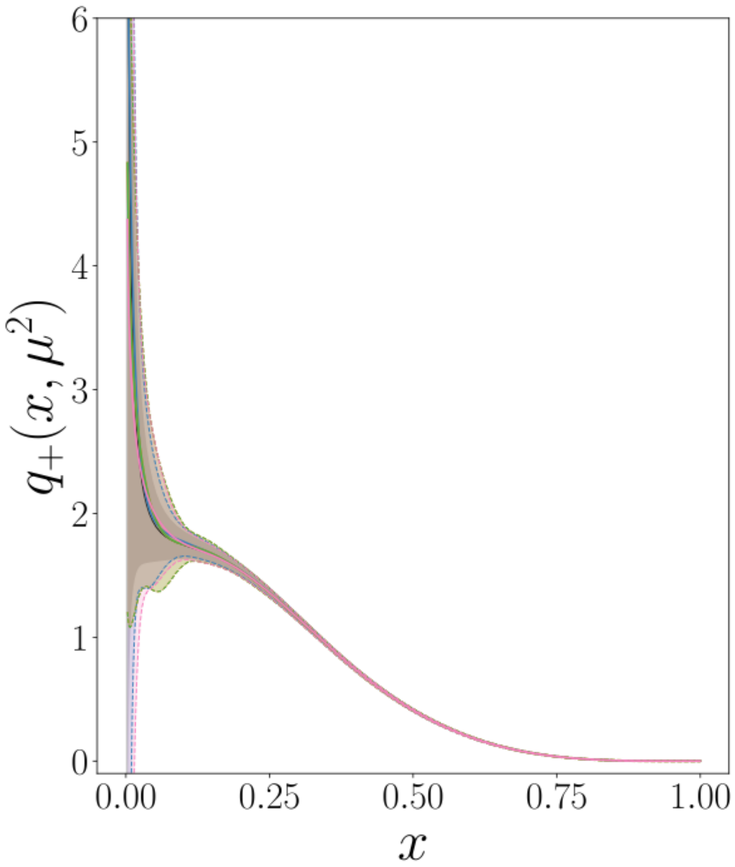
3rd Level of inference

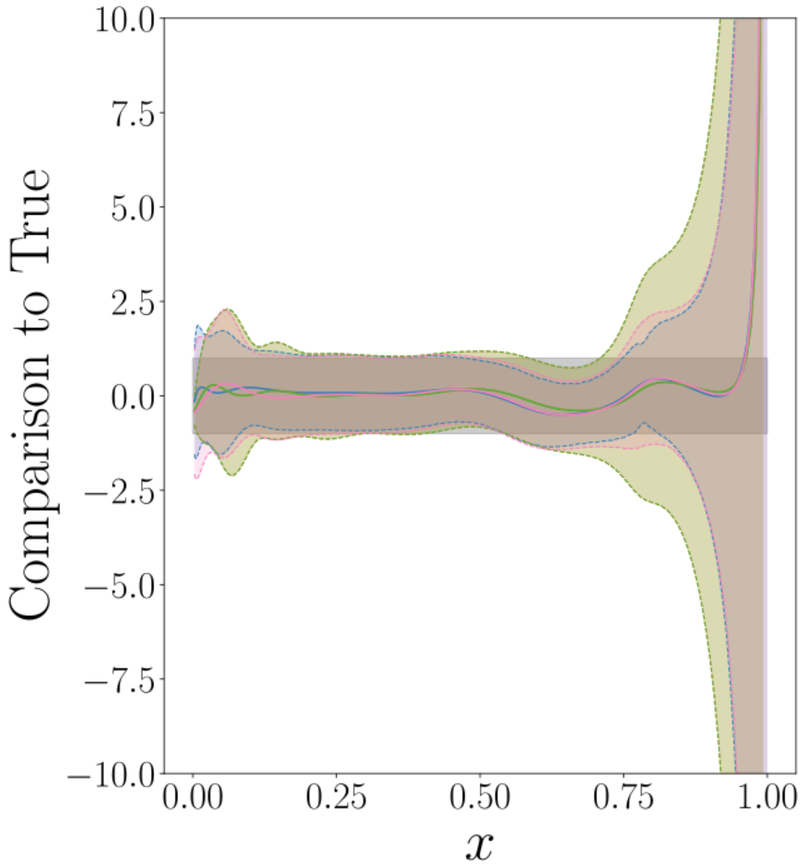
Results for Closure test (25 points)

$$P(\mathcal{H}_i|M^l) = \frac{P(M^l|\mathcal{H}_i)P(\mathcal{H}_i)}{P(M^l)}$$

25 data points



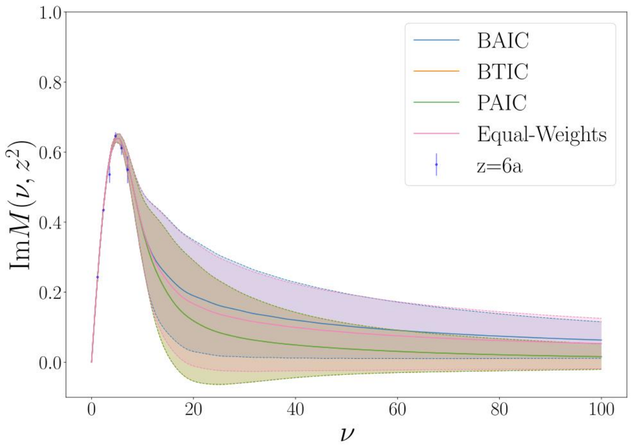


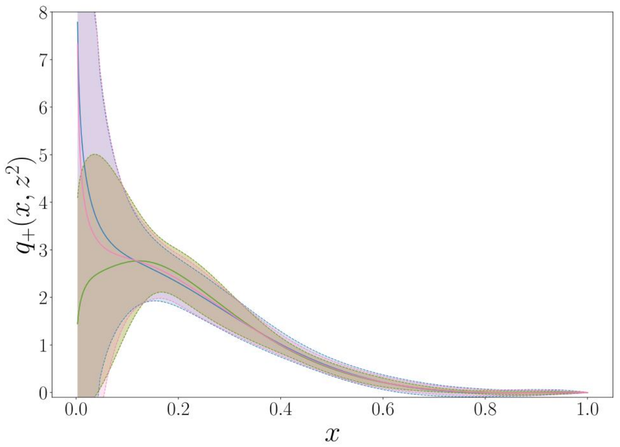


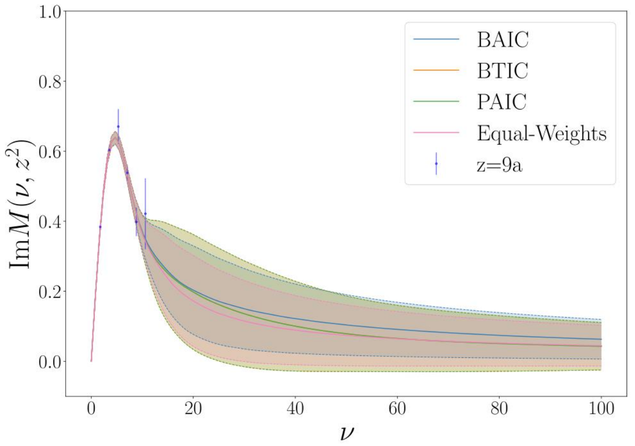
Lattice data

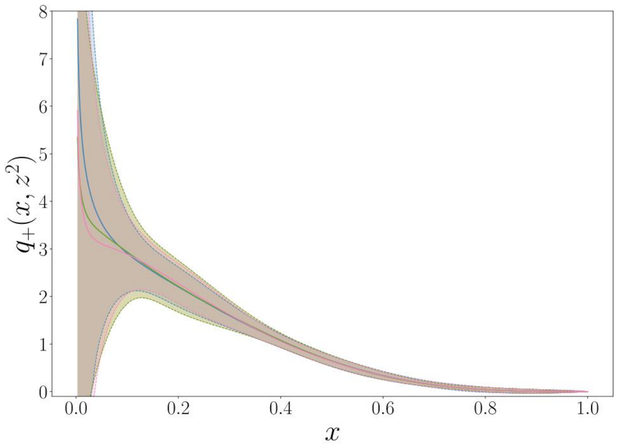
6 data points

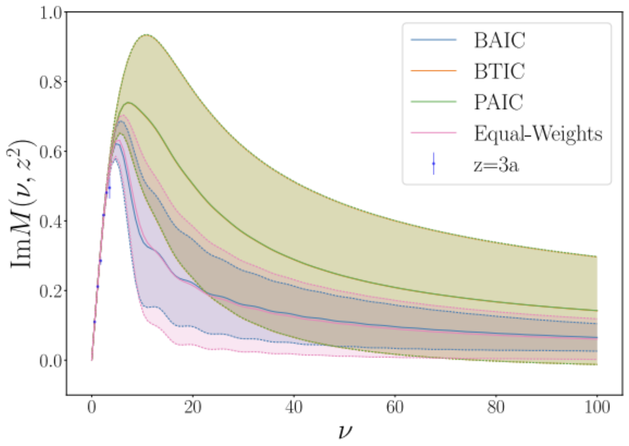
$$P(\mathcal{H}_i|M^l) = \frac{P(M^l|\mathcal{H}_i)P(\mathcal{H}_i)}{P(M^l)}$$

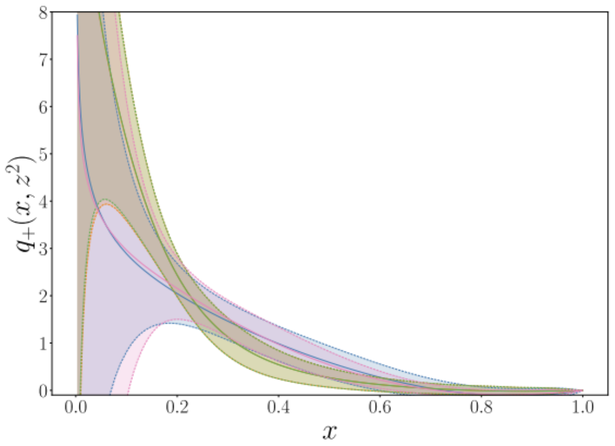




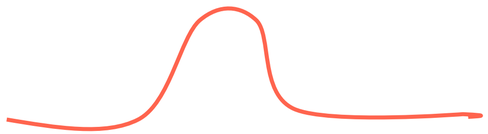








Conclusions



Key points

- GP provides a systematic and flexible way to regulate the inverse problem.
- Synthetic data evaluate the reconstructed uncertainty bands.
- Results remain stable under changes in kernels, mean functions, and hyperparameters through model averaging.
- Global and local KL divergence help us to measure the information gained.
- We learned how beautiful it is to work with just normal distributions.

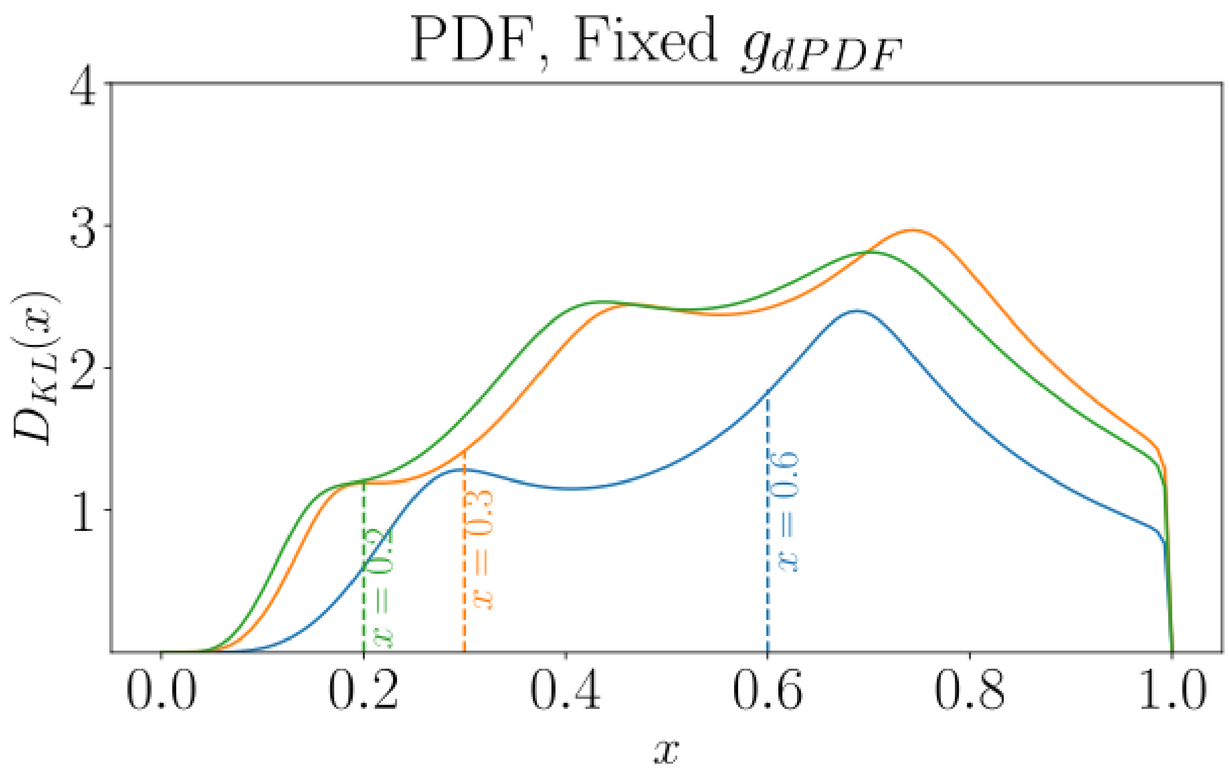
Thank you!

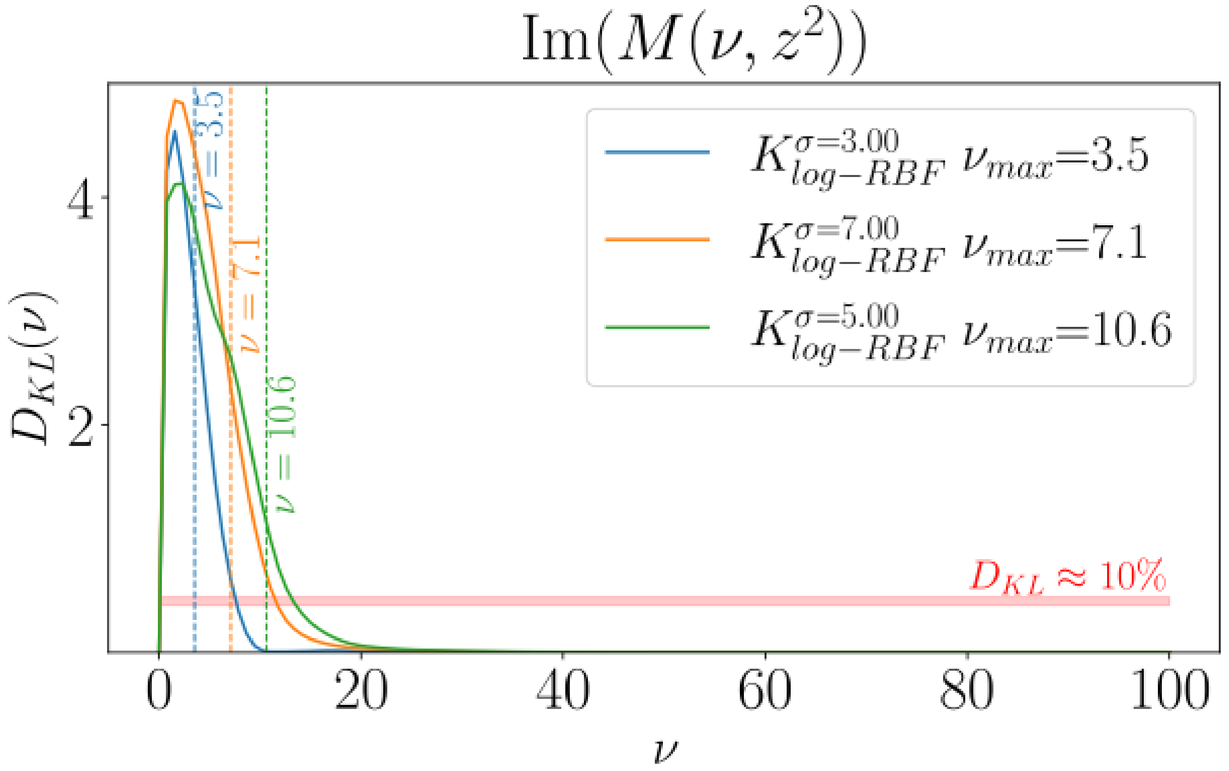
Email: yacahuanamedra@wm.edu

Back-up slides

Lattice data

KL divergence (local definition)





1st level of inference

Gaussian processes à la Feynman

$$P(q(x)|M^l,\theta,\mathcal{H}) = \frac{P(M^l|q(x),\theta,\mathcal{H})P(q(x)|\theta,\mathcal{H})}{P(M^l|\theta,\mathcal{H})}$$

Posterior = Likelihood Prior / Evidence

Everything is "gaussian" in this level of inference(It's like solving a free field theory).

My prior and likelihood have an analytic expression:

Likelihood

$$P(M^l|q(x),\theta,\mathcal{H}) = N_{likelihood} e^{-\frac{1}{2}(M_i - \mathcal{L}_{\nu_i} q(x)) C^{-1}_{ij} (M_j - \mathcal{L}_{\nu_j} q(x))}$$

Prior

$$P(q(x)|\theta,\mathcal{H}) = N_{prior} P_{const} e^{-\frac{1}{2}(\int dx dx' (q(x) - q_{PDF}(x)) K^{-1}(x,x')(q(x') - q_{PDF}(x'))}$$

$$P_{const} = e^{-\frac{1}{2\lambda}(\int_0^1 dx q(x) - 1)^2 - \frac{1}{2\lambda_c}(\int_0^1 dx q(x)\delta(1-x))^2}$$

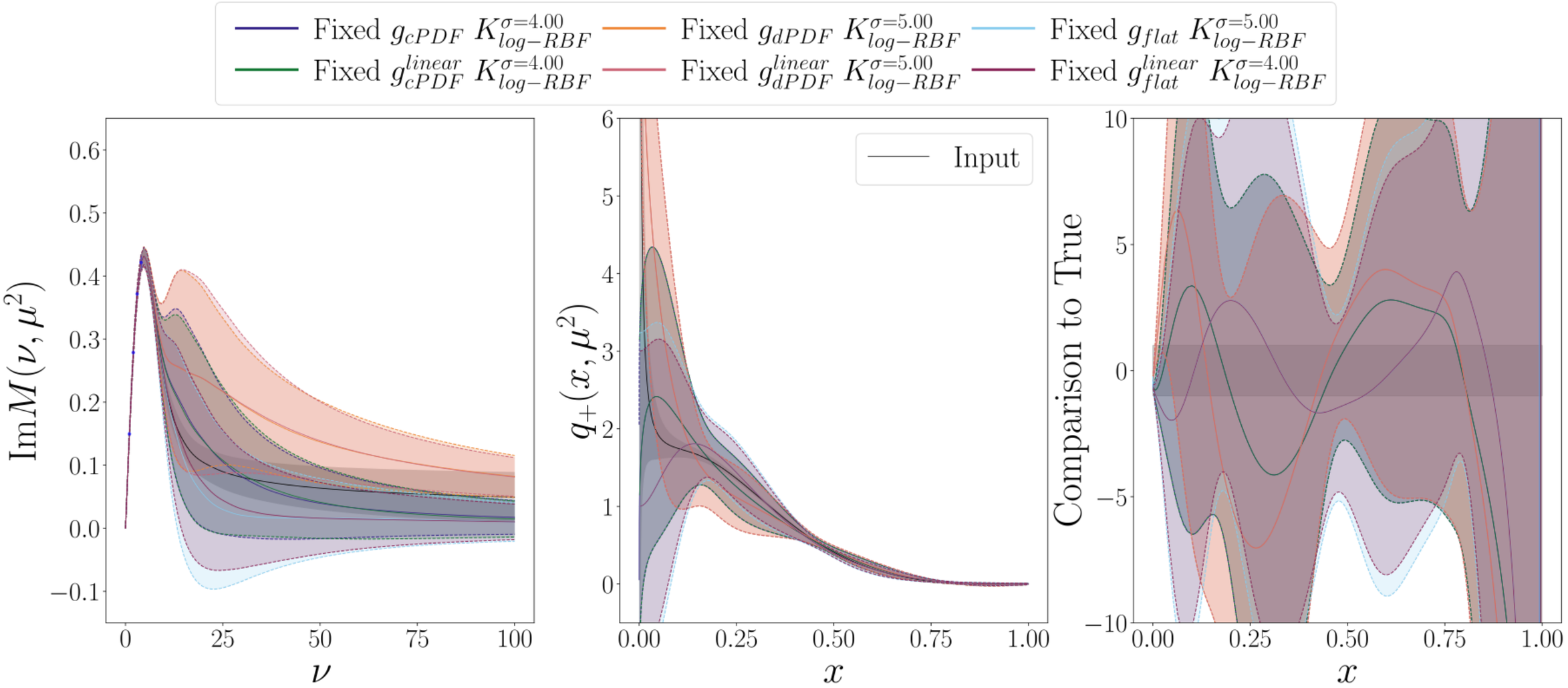
Normalization and q(x=1)=0

Evidence

$$P(M^l|\theta,\mathcal{H}) = \int D[q(x)] P(M^l|q(x),\theta,\mathcal{H}) P(q(x)|\theta,\mathcal{H})$$

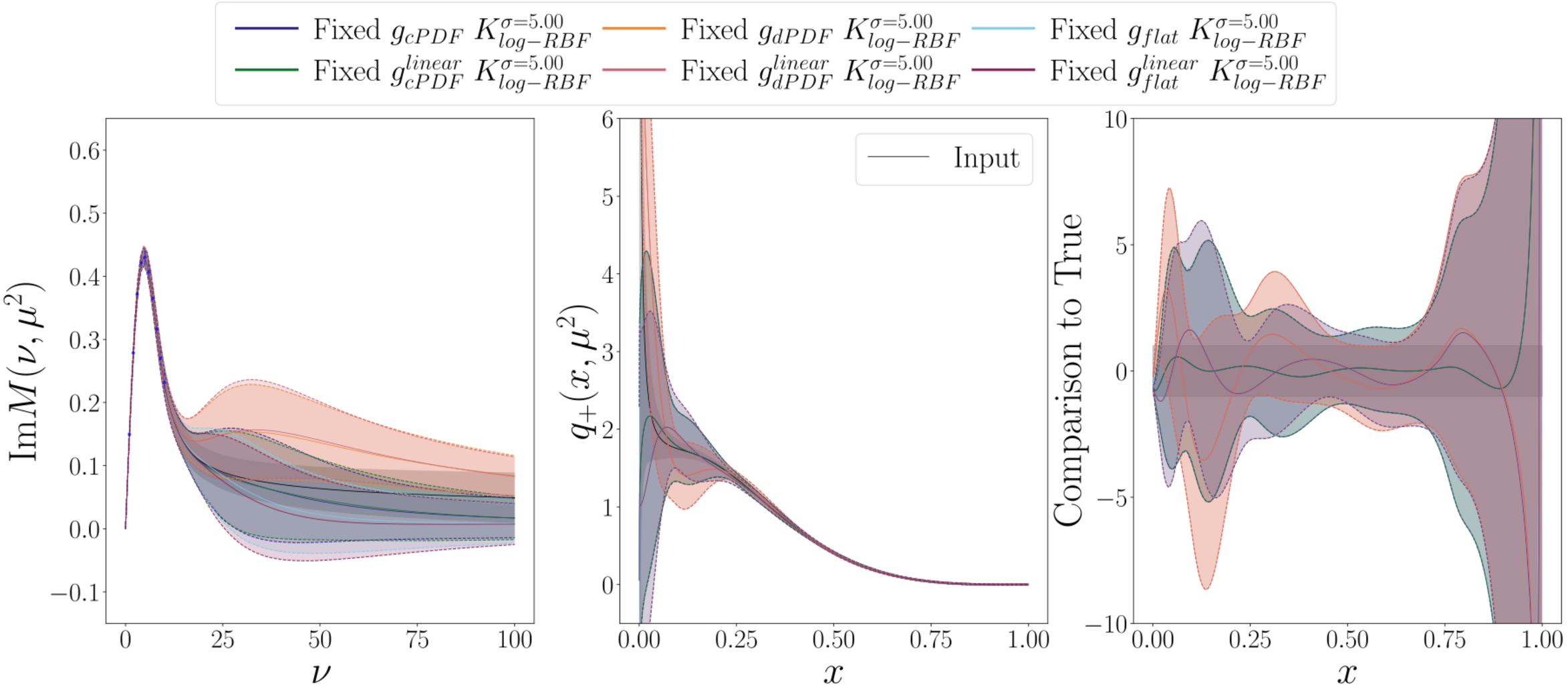
1st Level of inference

Fix parameters (4 data points)



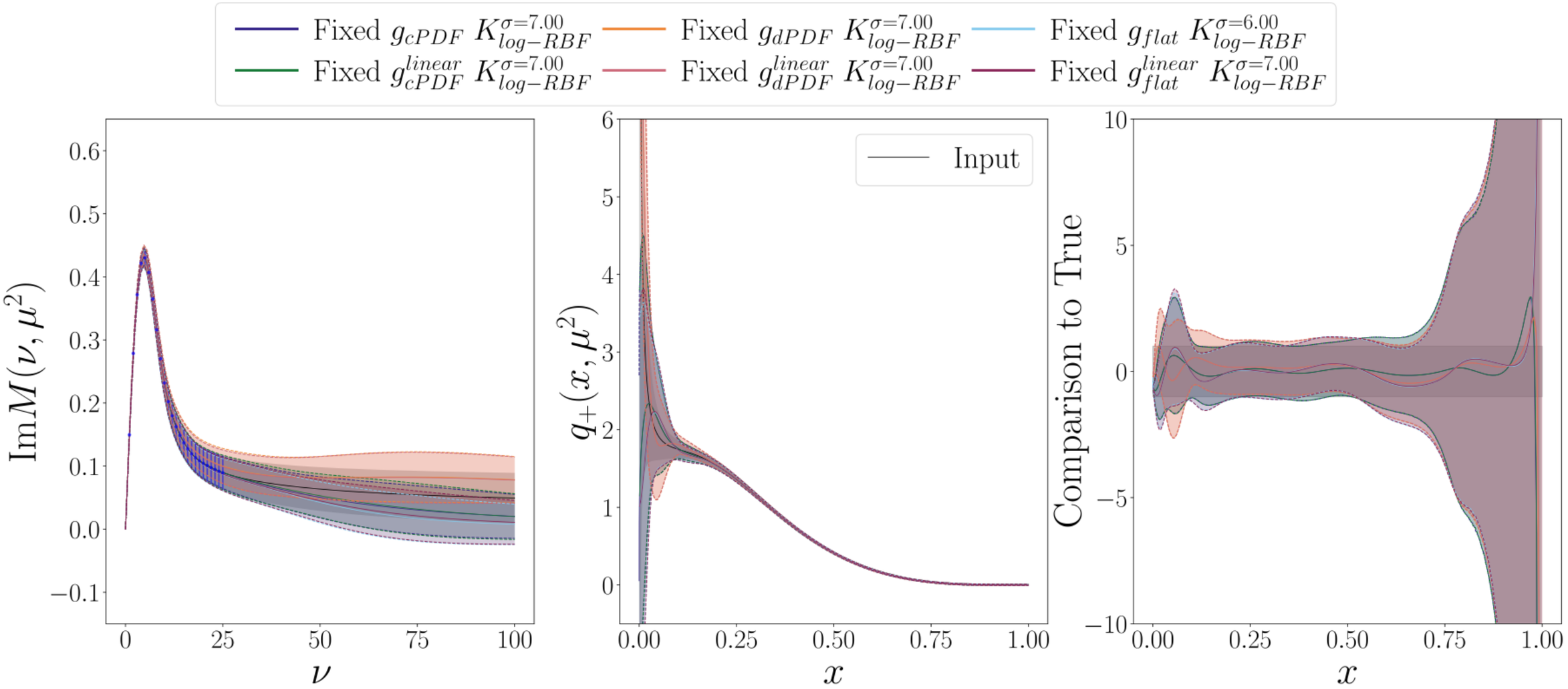
1st Level of inference

Fix parameters (10 data points)



1st Level of inference

Fix parameters (25 data points)



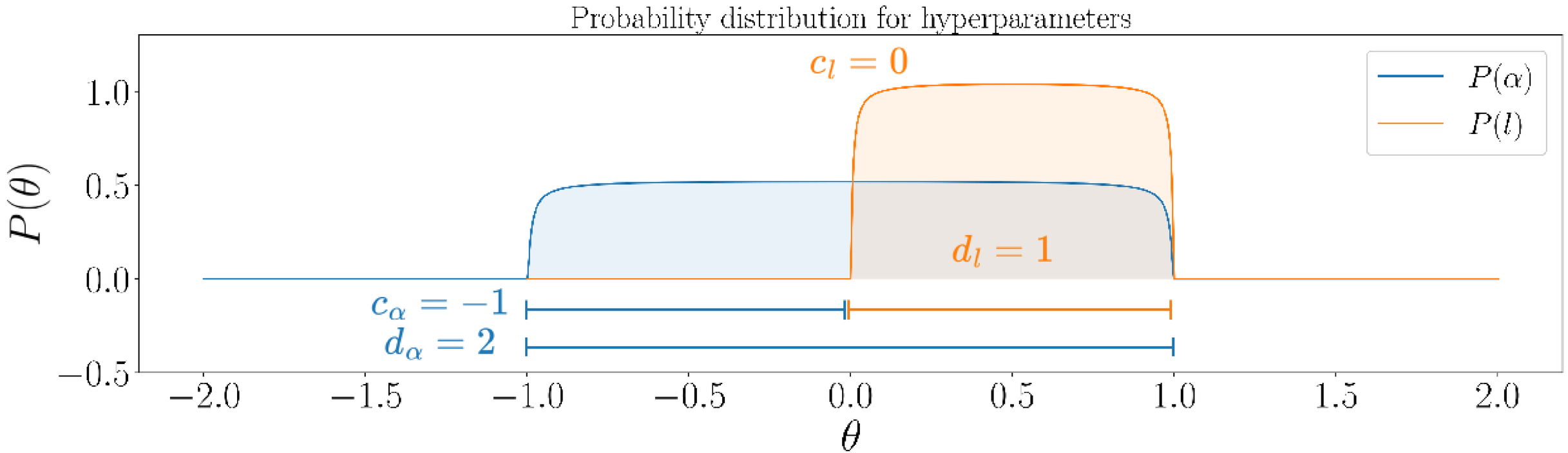
2nd Level of inference

Prior = exponential beta

$$P(\theta|M^l,\mathcal{H}) = \frac{P(M^l|\theta,\mathcal{H})P(\theta|\mathcal{H})}{P(M^l|\mathcal{H})}$$

$$P(\theta|\mathcal{H}) = ne^{-\frac{\hat{\theta}^a(1-\hat{\theta})^b}{2\cdot B(a+1,b+1)}}$$

$$\hat{\theta} = \frac{\theta - c}{d}$$

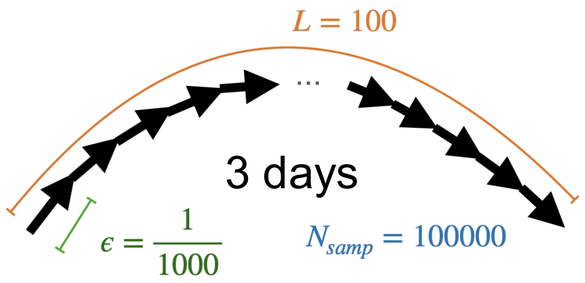


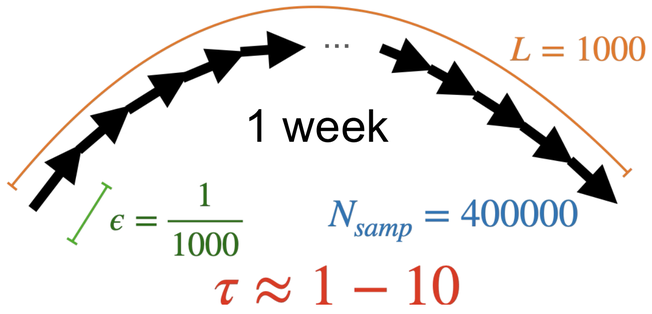
2nd Level of inference

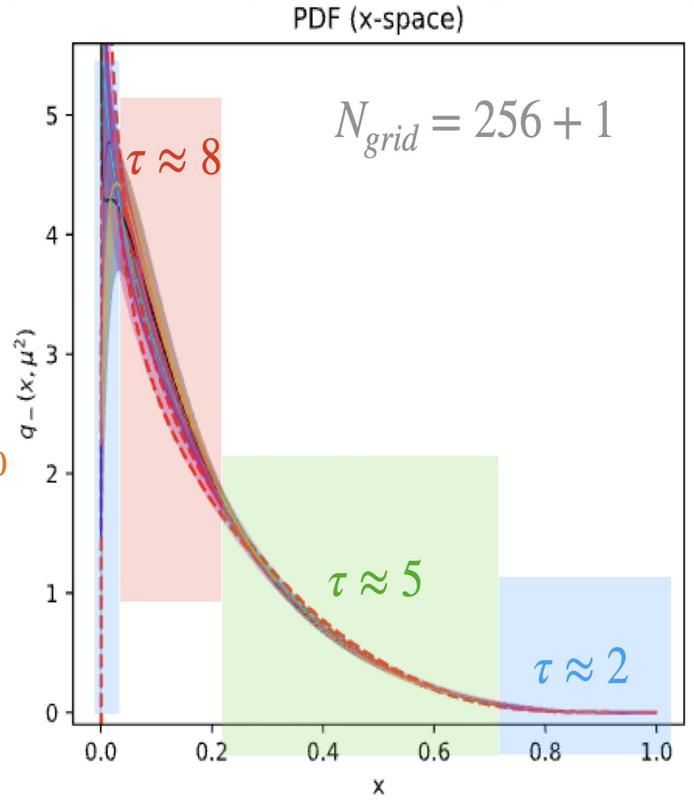
HMC -> Importance Sampling

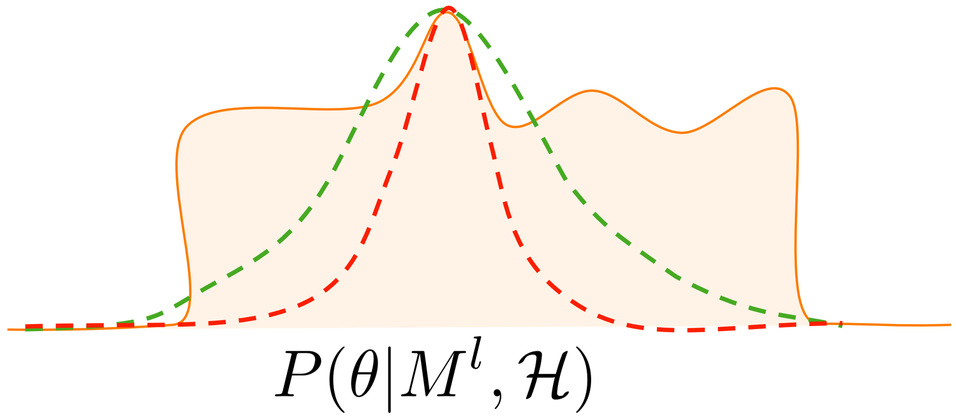
$$P(\theta|M^l,\mathcal{H}) = \frac{P(M^l|\theta,\mathcal{H})P(\theta|\mathcal{H})}{P(M^l|\mathcal{H})}$$

$$< q(x) >= \int P(\theta|M^l,\mathcal{H})\bar{q}(x;\theta)d\theta \qquad < q(x)q(x) >\equiv \int (\bar{q}(x;\theta) - < q(x) >)^2 P(\theta,|M^l,\mathcal{H})d\theta$$







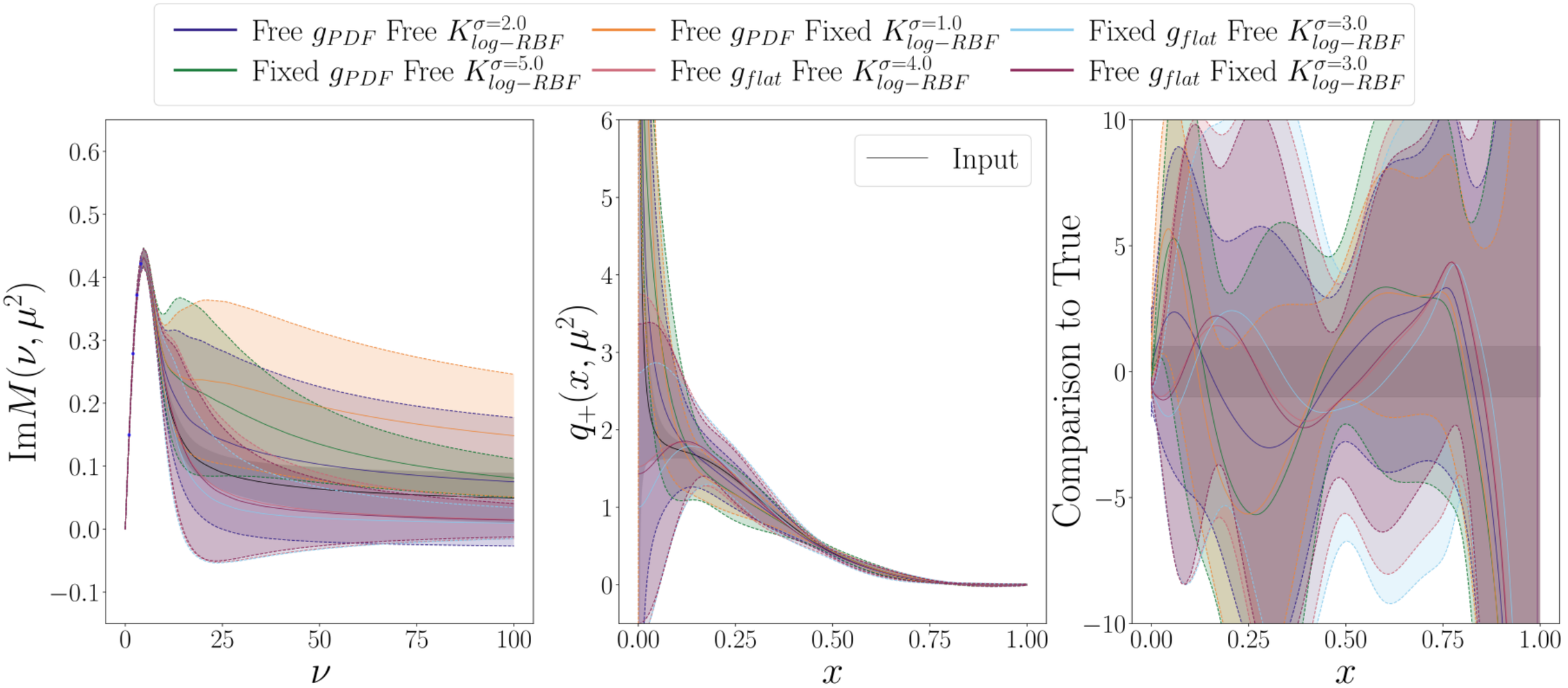


- HMC is effective but it can take a lot of computational resources and time to run.
- IS reduces the sampling process to 50 min per model

2nd Level of inference

Sampled results

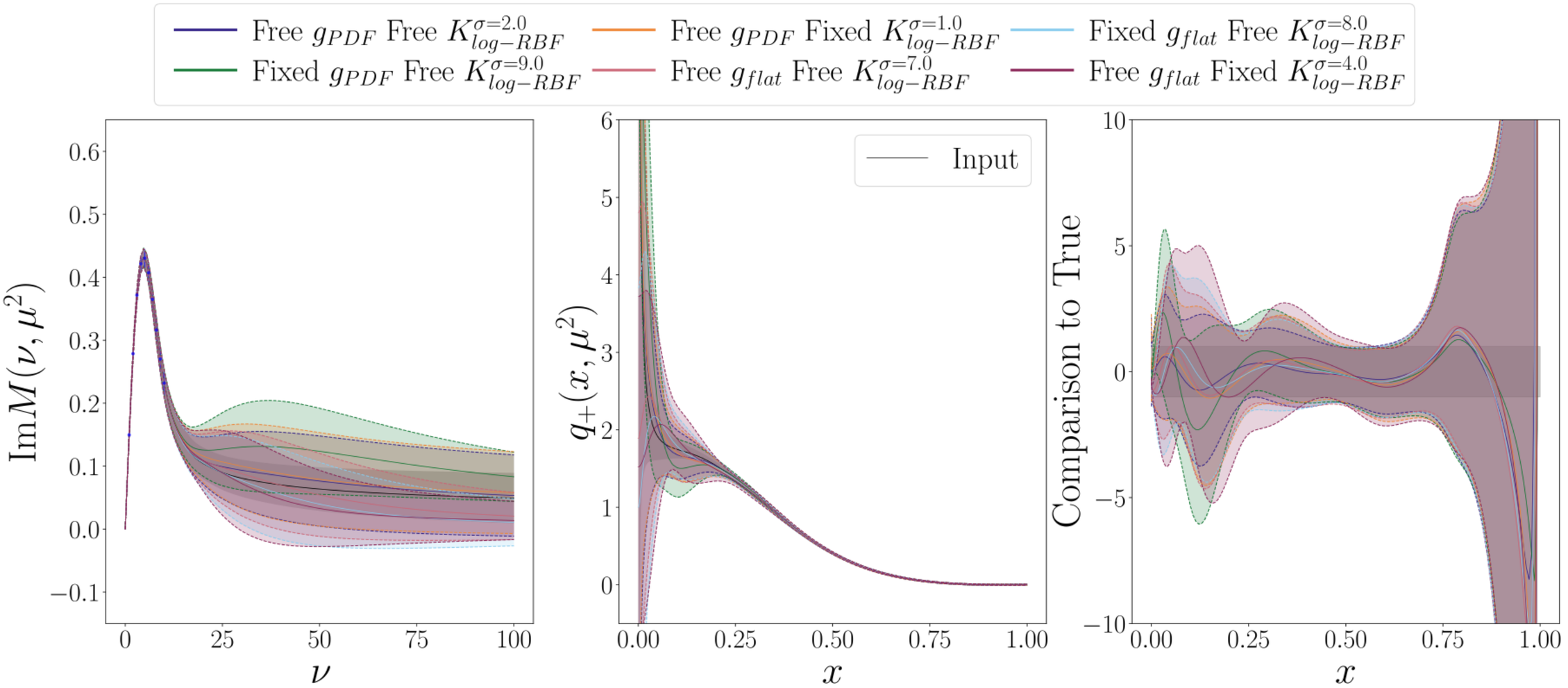
$$P(\theta|M^l,\mathcal{H}) = \frac{P(M^l|\theta,\mathcal{H})P(\theta|\mathcal{H})}{P(M^l|\mathcal{H})}$$



2nd Level of inference

Sampled results

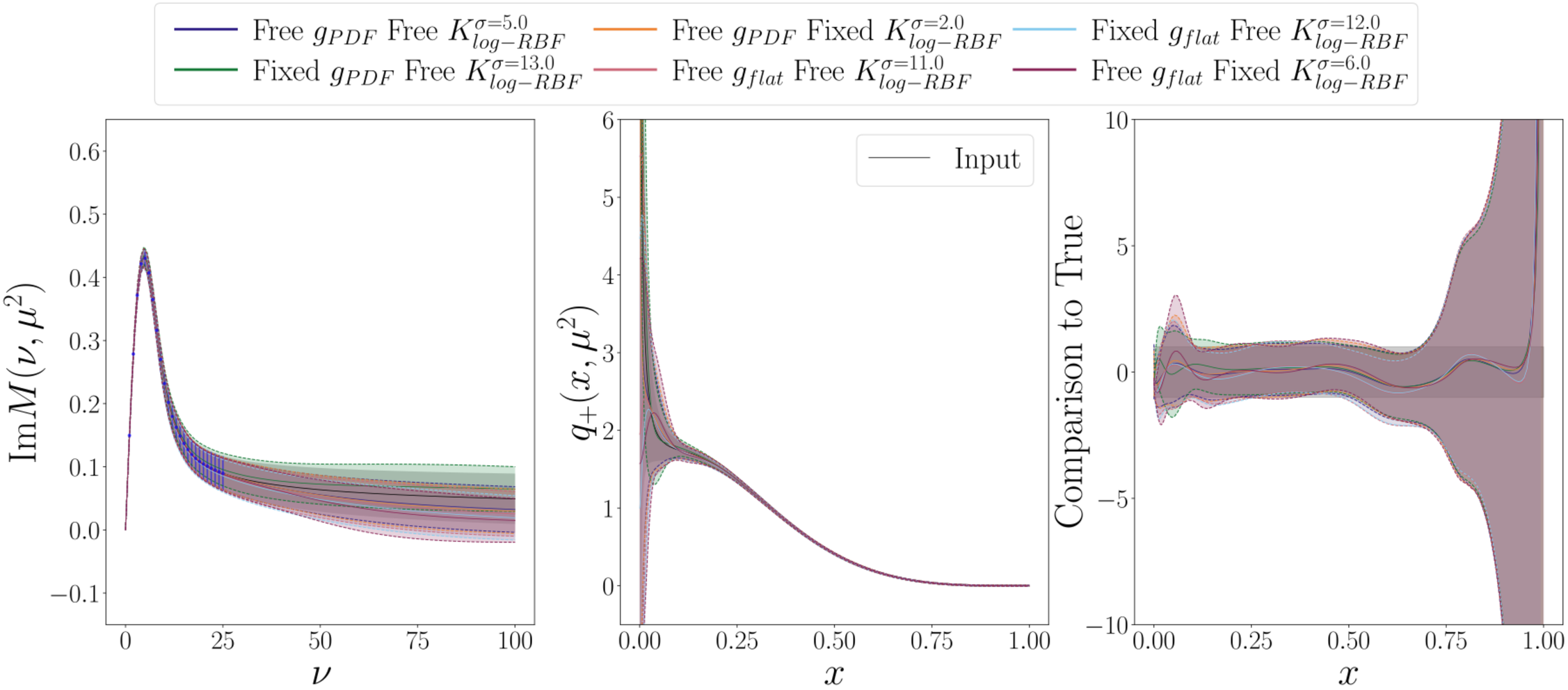
$$P(\theta|M^l,\mathcal{H}) = \frac{P(M^l|\theta,\mathcal{H})P(\theta|\mathcal{H})}{P(M^l|\mathcal{H})}$$



2nd Level of inference

Results…(we explore 30-ish models)

$$P(\theta|M^l,\mathcal{H}) = \frac{P(M^l|\theta,\mathcal{H})P(\theta|\mathcal{H})}{P(M^l|\mathcal{H})}$$



3rd Level of inference

Information criteria

$$P(\mathcal{H}_i|M^l) = \frac{P(M^l|\mathcal{H}_i)P(\mathcal{H}_i)}{P(M^l)}$$

$$P(\mathcal{H}_i) = \frac{1}{N_{models}}$$

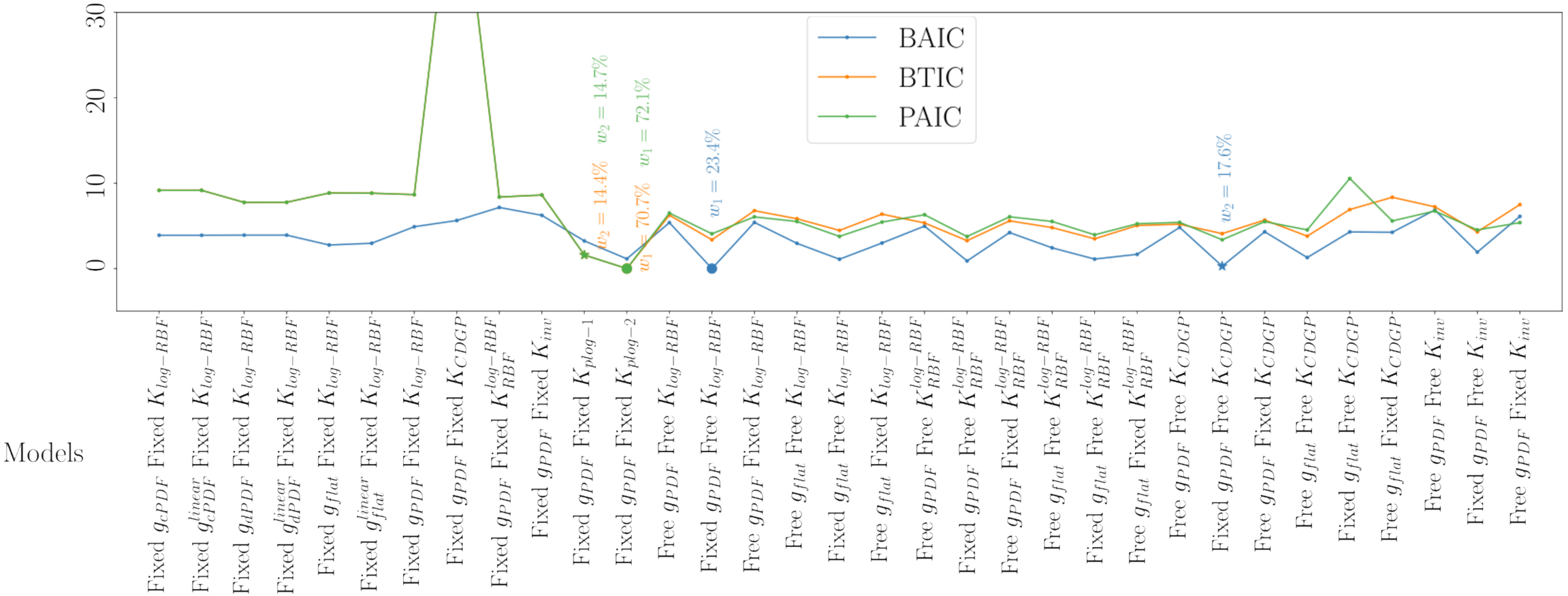
$$-2\log(P(M^l|\mathcal{H}_i)) \approx \begin{cases} BAIC = -log(P(M^l|\theta_{min},\mathcal{H})P(\theta_{min}|\mathcal{H})) + 2k \\ BTIC = -log(P(M^l|\theta_{min},\mathcal{H})P(\theta_{min}|\mathcal{H})) + 2Tr(J^{-1}(\theta_{min})I(\theta_{min})) \\ PAIC = - < log(P(M^l|\theta,\mathcal{H})P(\theta|\mathcal{H})) > + 2Tr(J^{-1}(\theta_{min})I(\theta_{min})) \end{cases}$$

$$\mathfrak{q}(x) \equiv \sum_i^{models} P(\mathcal{H}_i|M^l) < q(x) >_i \qquad \mathfrak{q}(x)\mathfrak{q}(x) \equiv \sum P(\mathcal{H}_i|M^l)[< q(x) >_i - \mathfrak{q}(x)]^2$$

3rd Level of inference

Selection ~ Averaging

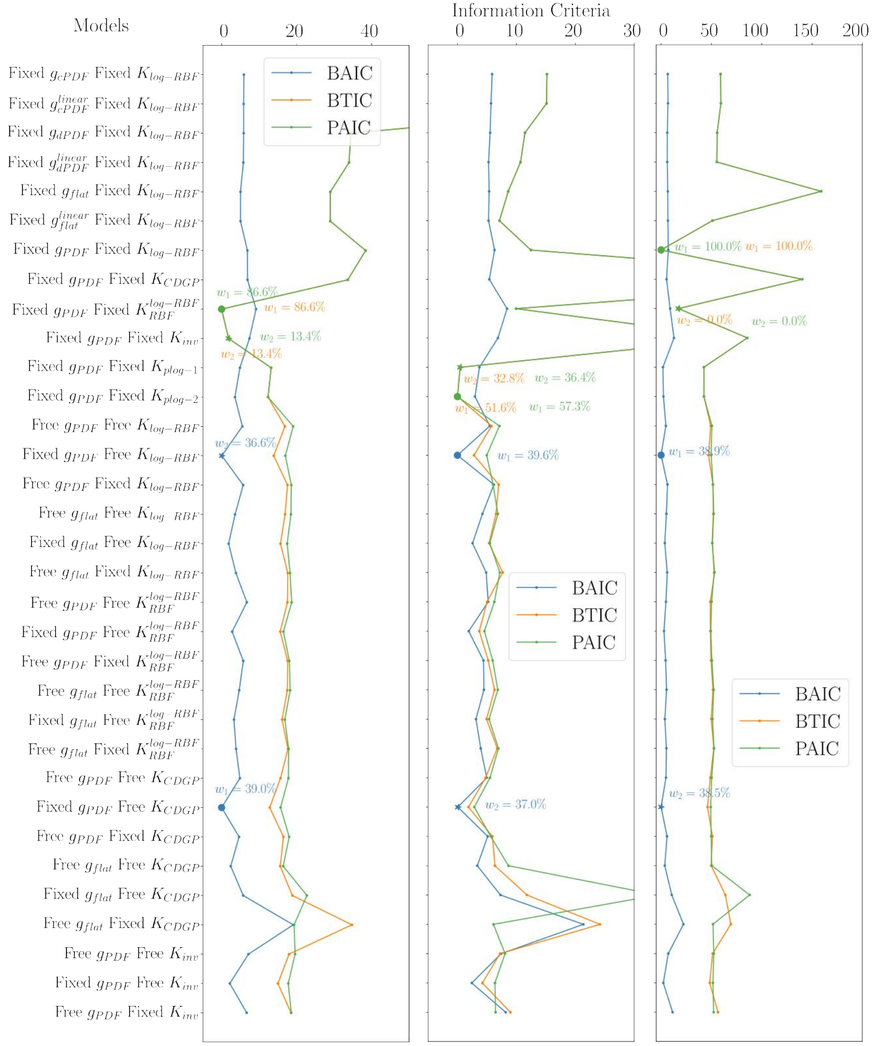
$$P(\mathcal{H}_i|M^l) = \frac{P(M^l|\mathcal{H}_i)P(\mathcal{H}_i)}{P(M^l)}$$

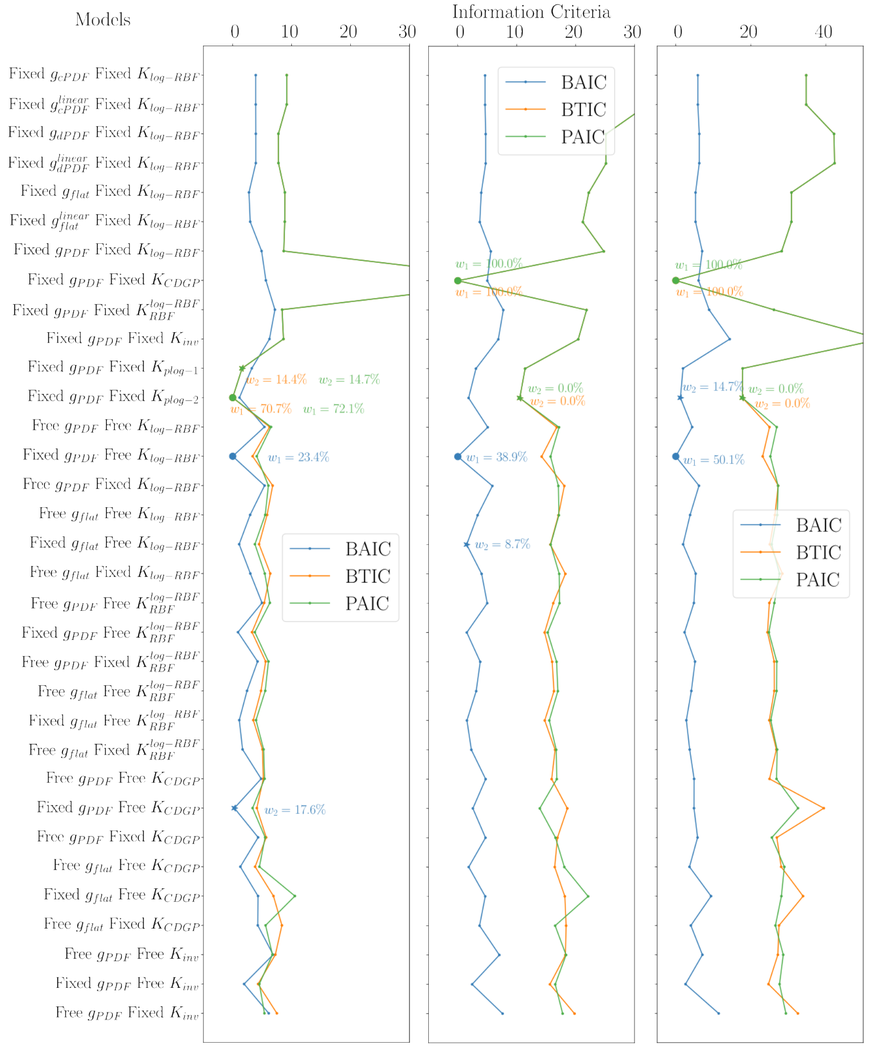


3rd Level of inference

Information criteria

$$P(\mathcal{H}_i|M^l) = \frac{P(M^l|\mathcal{H}_i)P(\mathcal{H}_i)}{P(M^l)}$$





Inverse problem

Complex = Real + i Imaginary

$$M(\nu) = \int_{-1}^{1} dx e^{ix(p\cdot z)} \cdot q(x) = \int_{-1}^{0} dx e^{ix\nu} q(x) + \int_{0}^{1} dx e^{ix\nu} q(x)$$

$$= \int_{0}^{1} dx [e^{ix\nu} q(x) + e^{-ix\nu} q(-x)] = \int_{0}^{1} dx [e^{ix\nu} q(x) - e^{-ix\nu} \bar{q}(x)]$$

$$Re(M(\nu)) = \int_{0}^{1} dx \cos(x\nu) \cdot (q(x) - \bar{q}(x)) \qquad Im(M(\nu)) = \int_{0}^{1} dx \sin(x\nu) \cdot (q(x) + \bar{q}(x))$$

$$q_-(x) = (q(x) - \bar{q}(x)) \qquad q_+(x) = (q(x) + \bar{q}(x))$$

Kernels and Models

Some equations…

$$K_{f-\mathrm{RBF}}(x,x';\theta=\{\sigma,l^2\}) = \sigma \exp\left(-\frac{||f(x)-f(x')||^2}{2l^2}\right),$$

$$K_{plog-1}(x,x') = \sigma\sum_{n=1}^{\infty}(1-x)^n(1-x')^n = \sigma\frac{(1-x)(1-x')}{1-(1-x)(1-x')}.$$

$$K_{plog-2}(x,x') = \sigma\sum_{n=1}^{\infty}\frac{(1-x)^n(1-x')^n}{n} = -\sigma\log\big(1-(1-x)(1-x')\big).$$

$$K_{\mathrm{combined}}\Big(x,x';\theta=\{\theta_1,\theta_2,s,x_0\}\Big) = \sigma(x;s,x_0)K_1\Big(x,x';\theta_1\Big)\sigma(x';s,x_0)+$$
$$(1-\sigma(x;s,x_0))K_2\Big(x,x';\theta_2\Big)(1-\sigma(x';s,x_0))$$

$$g_{\mathrm{flat}}(x;\theta=\{N\}) = N \quad ; \qquad g_{\mathrm{PDF}}(x;\theta=\{N,\alpha,\beta\}) = Nx^{\alpha}(1-x)^{\beta}$$

Introduce Mellin moments

Constraints in the 1st level

$$P_{const} = e^{-\frac{1}{2\lambda}(\int_0^1 dx q(x) - 1)^2 - \frac{1}{2\lambda_c}(\int_0^1 dx q(x)\delta(1-x))^2}$$

$$P_{Mellin} = P_{const} e^{-\frac{1}{2\lambda_n}(\int_0^1 dx q(x) x^{n-1} - b_n)^2}$$

In this case… $b_n^{lattice} = b_n \pm \delta b_n$

$$\lambda_n = \frac{1}{\delta b_n}$$

The equivalence of data points in the x space has been already implemented before, we just have to generalize it.

1st level of inference

Important results

$$P(q(x)|M^l,\theta,\mathcal{H}) = \frac{P(M^l|q(x),\theta,\mathcal{H})P(q(x)|\theta,\mathcal{H})}{P(M^l|\theta,\mathcal{H})}$$

$$\overline{q}(x;\theta) = \int q(x)P(q(x)|M^l,\theta,\mathcal{H})D[q(x)]$$ Mean of the Posterior

$$\overline{q}(x;\theta) = q_{PDF}(x) + K \cdot B^{\perp}[C + BKB^{\perp}](M^l - B^l \cdot q_{PDF})$$

$$\overline{q(x)q(x)} = \int q(x)q(x)P(q(x)|M^l,\theta,\mathcal{H})D[q(x)]$$ Covariance of the Posterior

$$H(\theta) \equiv \overline{q(x)q(x')} = K(x,x') - KB^{\perp}[C + BKB^{\perp}]^{-1}BK$$

Effective evidence

$$E(\theta) \equiv -\log(P(M^l|\theta,\mathcal{H})) = \frac{1}{2}(M_i - Bq_{PDF})^{\perp}(C + BKB^{\perp})^{-1}(M_i - Bq_{PDF}) + \frac{1}{2}\log[det(C + BKB^{\perp})] + \frac{N_x}{2}\log(2\pi)$$

2nd level of inference

Kernel and Mean functions

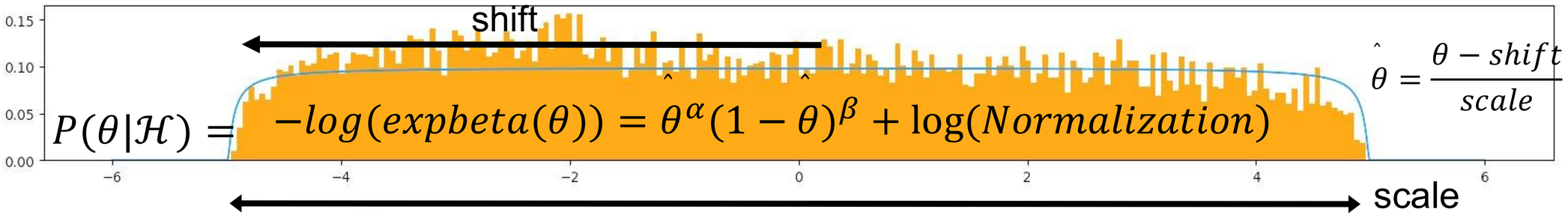
$$P(\theta|M^l,\mathcal{H}) = \frac{P(M^l|\theta,\mathcal{H})P(\theta|\mathcal{H})}{P(M^l|\mathcal{H})}$$

Where does this hyperparameter come from?

$$q_{PDF}(x) = Nx^{\alpha}(1-x)^{\beta} \quad + \quad K_{rbf}(x,x') = \sigma^2 e^{-\frac{|x-x'|^2}{2l^2}} \rightarrow \theta = \{N,\alpha,\beta,\sigma,l\}$$

How should we define the prior?

$$P(\theta|\mathcal{H}) = \{Normal(\theta), lognormal(\theta), expbeta(\theta)\}$$



We use HMC/Important Sampling to obtain the posterior

$$P(M^l|\mathcal{H}) = \int d\theta P(M^l|\theta,\mathcal{H})P(\theta|\mathcal{H})$$

No functional variables anymore, we have to integrate over hyperparameters.

2nd level of inference

Important results

$$P(\theta|M^l,\mathcal{H}) = \frac{P(M^l|\theta,\mathcal{H})P(\theta|\mathcal{H})}{P(M^l|\mathcal{H})}$$

Mean of the Posterior

$$< q(x) >= \int P(\theta|M^l,\mathcal{H})\overline{q}(x;\theta)d\theta \qquad < q(x) >= \frac{\int e^{-E(\theta)}\overline{q}(x;\theta)d\theta}{\int e^{-E(\theta)}d\theta}$$

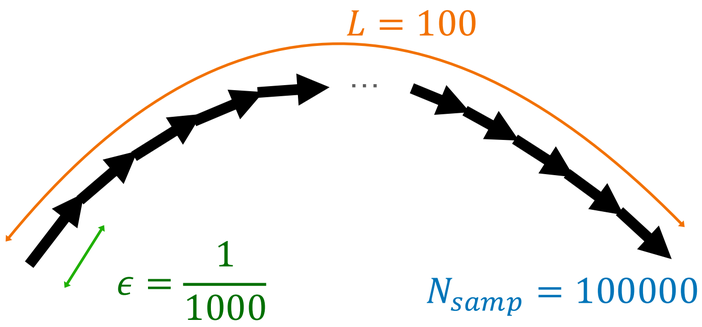
Covariance of the Posterior

$$< q(x)q(x) >\equiv \int (\overline{q}(x;\theta) - < q(x) >)^2 P(\theta,|M^l,\mathcal{H})d\theta$$

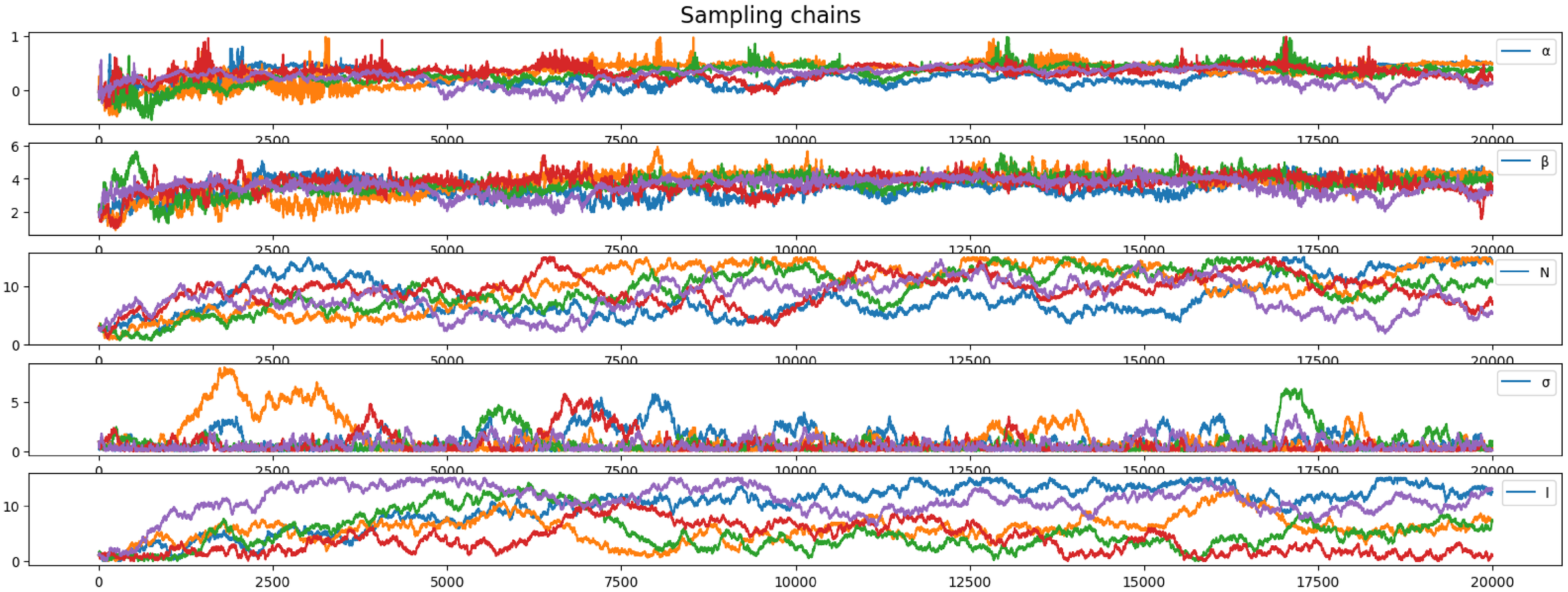
$$< q(x)q(x) >= \frac{\int e^{-E(\theta)}[H(\theta) + (\overline{q}(x;\theta) - < q(x) >)^2]P(\theta|\mathcal{H})d\theta}{\int e^{-E(\theta)}P(\theta|\mathcal{H})d\theta}$$

Hybrid Monte Carlo

Kernel and Mean functions



How does hyper parameters sampling looks like?



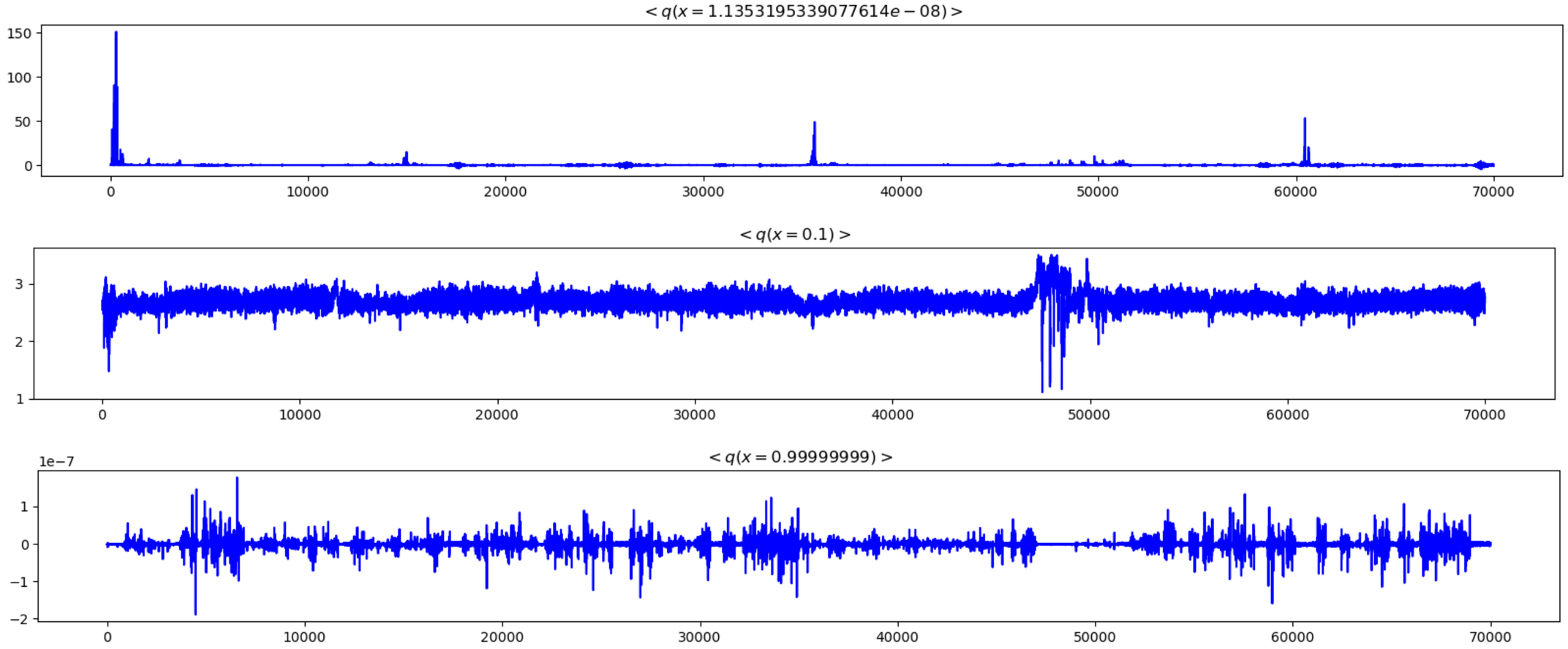
HMC observables

Kernel and Mean functions

$$P(\theta|M^l,\mathcal{H}) = \frac{P(M^l|\theta,\mathcal{H})P(\theta|\mathcal{H})}{P(M^l|\mathcal{H})}$$

How does <q(x)>, <q(x)q(x')> sampling looks like

$$< q(x) > \int d\theta P(\theta|M^l,\mathcal{H})\overline{q}(x)$$

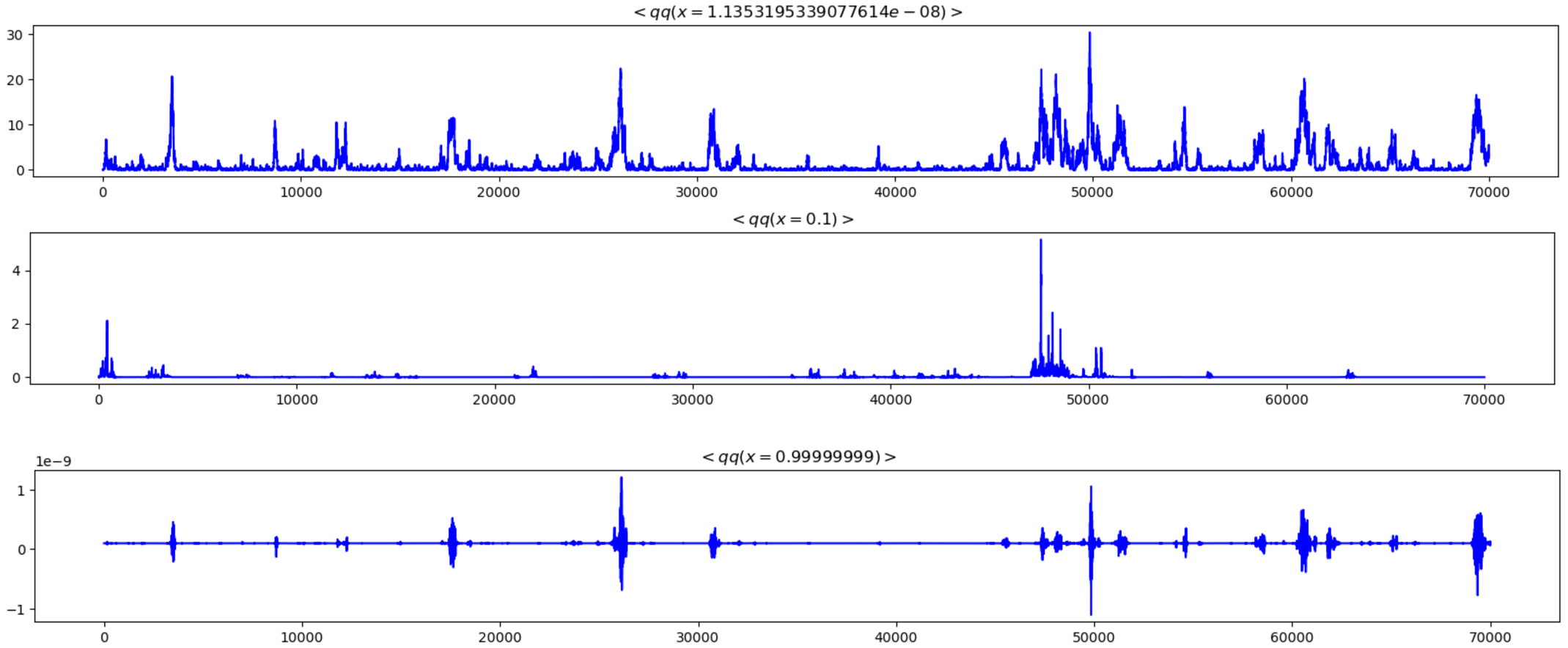


HMC observables

Kernel and Mean functions

$$P(\theta|M^l,\mathcal{H}) = \frac{P(M^l|\theta,\mathcal{H})P(\theta|\mathcal{H})}{P(M^l|\mathcal{H})}$$

How does <q(x)>, <q(x)q(x')> sampling looks like?

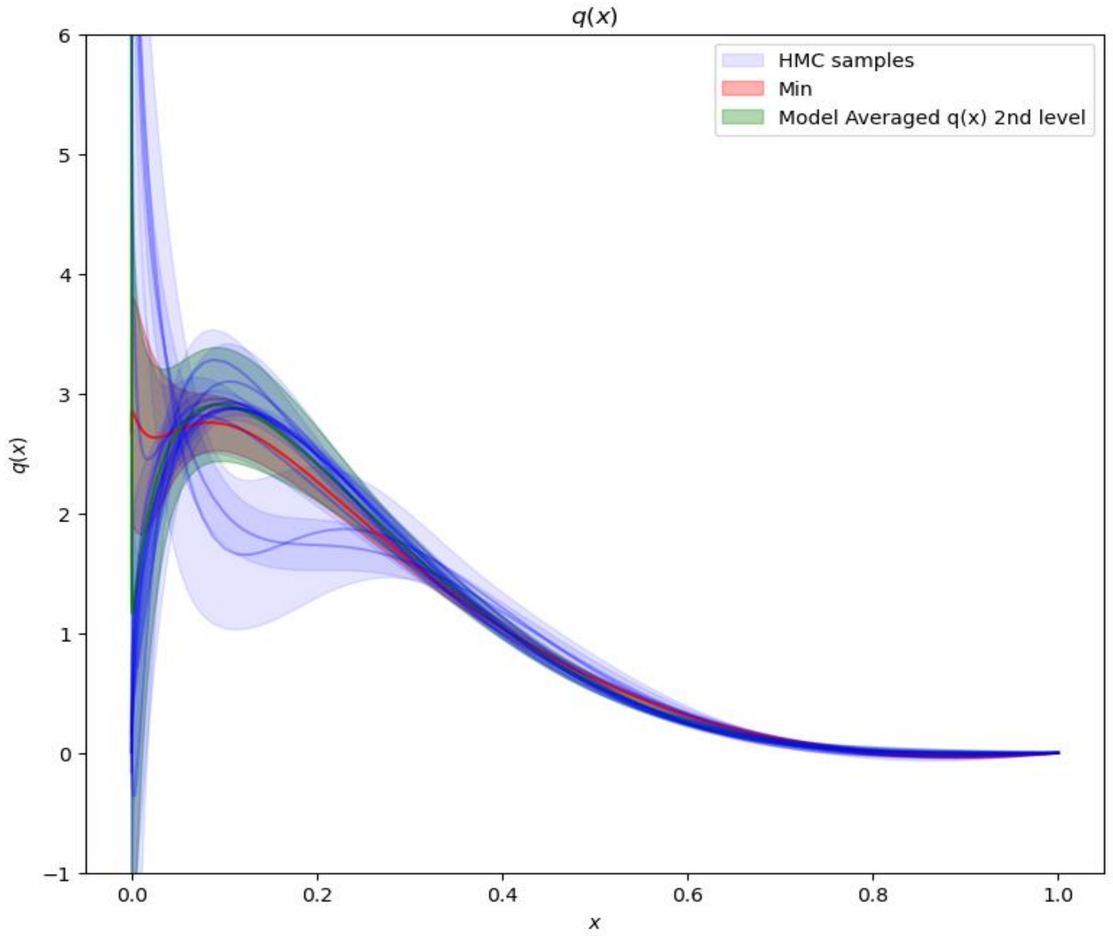


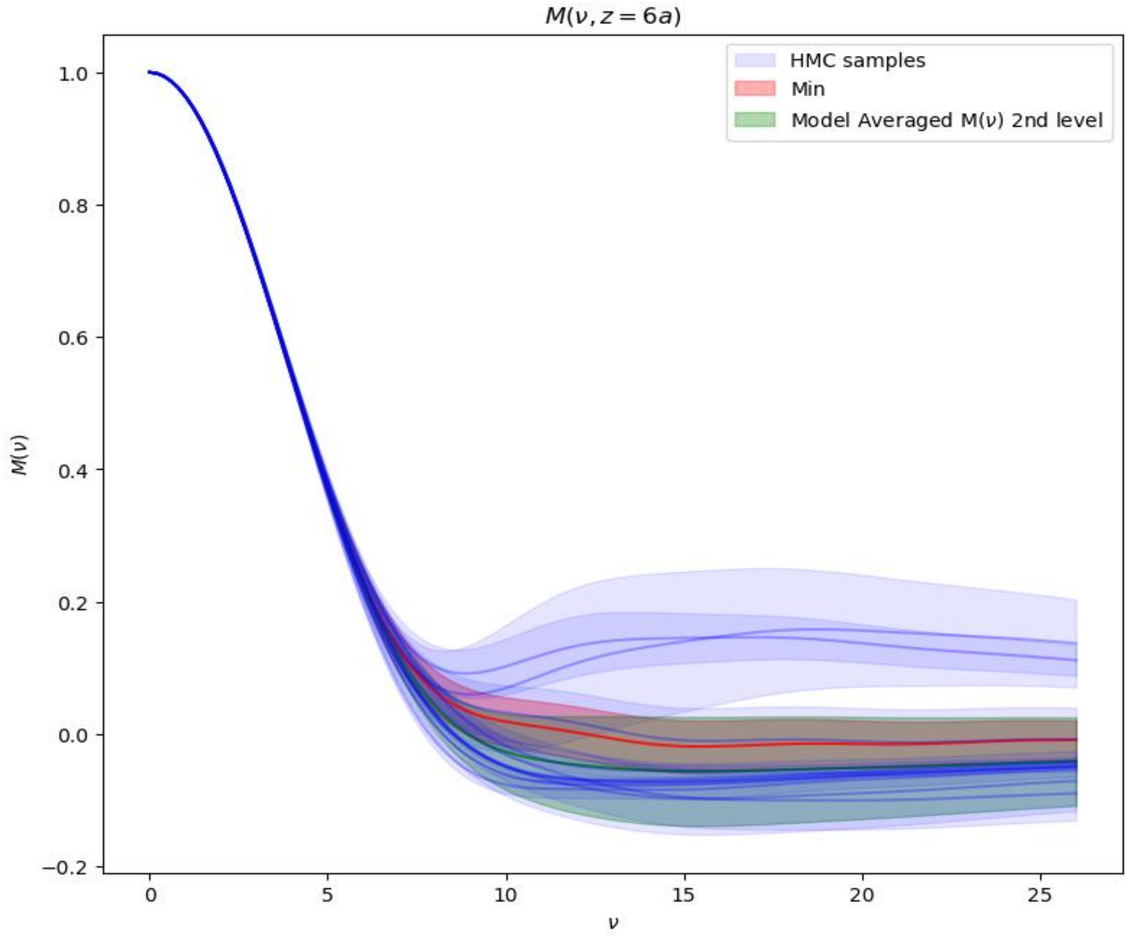
2nd level of inference

Model averaged results

$$P(\theta|M^l,\mathcal{H}) = \frac{P(M^l|\theta,\mathcal{H})P(\theta|\mathcal{H})}{P(M^l|\mathcal{H})}$$

How does <q(x)>, <q(x)q(x')> sampling looks like?





3rd level of inference

Important results

$$P(\mathcal{H}_i|M^l) = \frac{P(M^l|\mathcal{H}_i)P(\mathcal{H}_i)}{P(M^l)}$$

Mean of the Posterior

$$\mathfrak{q}(x) \equiv \sum_i^{models} P(\mathcal{H}_i|M^l) < q(x) >_i \qquad \mathfrak{q}(x) = \frac{\sum_i^{Models} e^{\frac{\Delta IC}{2}} < q(x) >_i}{\sum_i^{Models} e^{\frac{\Delta IC}{2}}}$$

Covariance of the Posterior

$$\mathfrak{q}(x)\mathfrak{q}(x) \equiv \sum P(\mathcal{H}_i|M^l)[< q(x) >_i - \mathfrak{q}(x)]^2$$

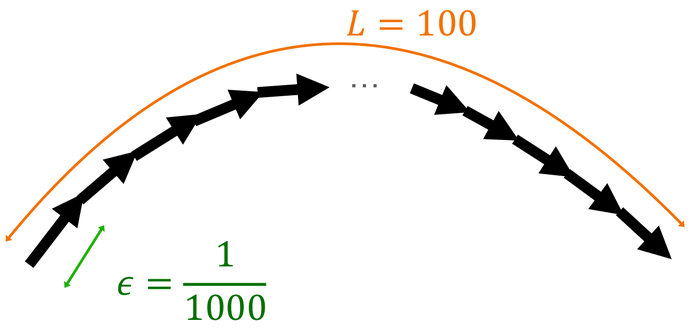
$$\mathfrak{q}(x)\mathfrak{q}(x) \equiv \frac{\sum_i^{Models} e^{\frac{\Delta IC}{2}} [< q(x)q(x) >_i + (< q(x) >_i - \mathfrak{q}(x))^2]}{\sum_i^{Models} e^{\frac{\Delta IC}{2}}}$$

Hybrid Monte Carlo

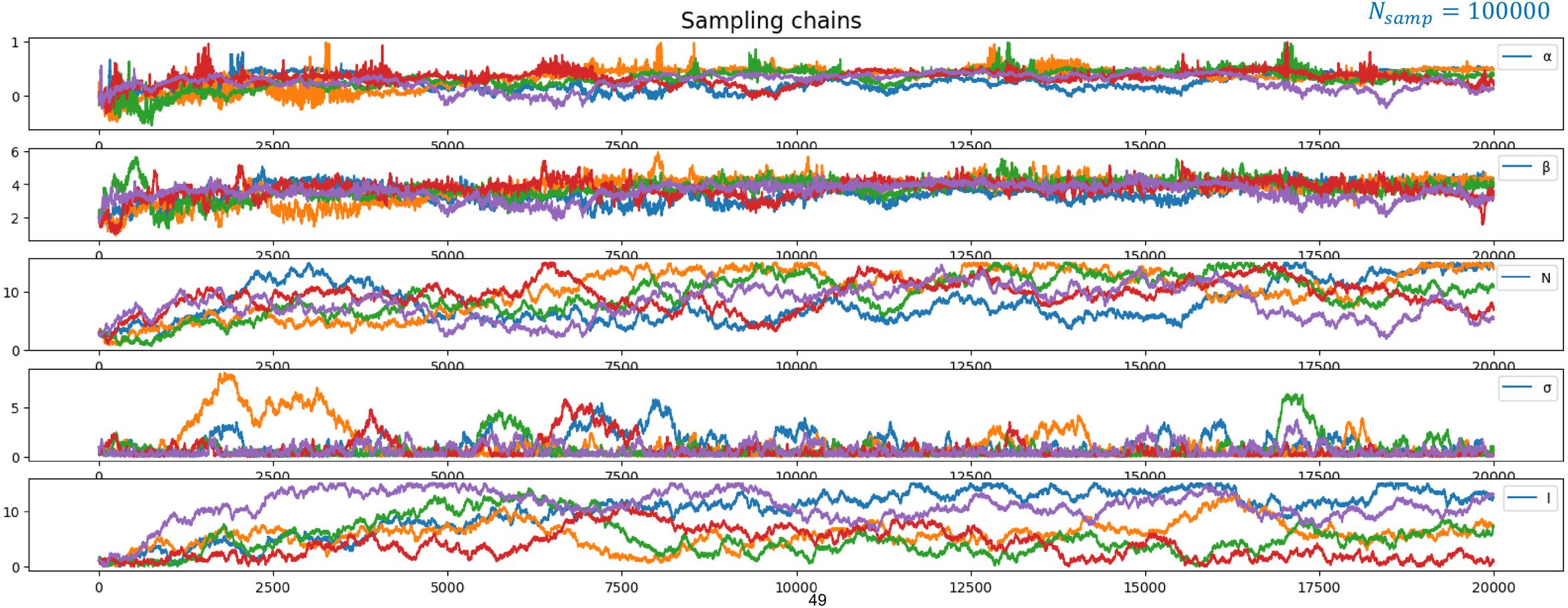
Sampling procedure

Skip 20 trajectories, 12 hours
(5 independent chains) —>

$\tau \approx 100 - 500$



$N_{samp} = 100000$



Bayesian approach

Bayes' Theorem (not that easy)

- How can we work with multiple conditions?

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(A|B,C,D) = ? \qquad P(A|B,C) = ?$$

- For now we can work with 2 conditions instead of 3

$$P(A,B|C) = P(A|B,C)P(B|C)$$

Textbook property

$$P(A,B) = P(B,A)$$

$$P(A|B,C) = \frac{P(A,B|C)}{P(B|C)} = \frac{P(B,A|C)}{P(B|C)} = \frac{P(B|A,C)P(A|C)}{P(B|C)}$$

$C \rightarrow C, D$

$$P(A|B,C,D) = \frac{P(B|A,C,D)P(A|C,D)}{P(B|C,D)}$$