Bayesian Approach to Inverse Problems

L Del Debbio

Higgs Centre for Theoretical Physics
University of Edinburgh

work in collaboration with

M Wilson, A Candido, T Giani, G Petrillo, A Lupo, M Panero, N Tantalo

inverse problems

ubiquitous in physics, geosciences, engineering...

$$\boxed{y_I = \int dx\, C_I(x)\, f(x)}$$

... are known to be ill-defined problems

$\hookrightarrow$ simple parametrization could lead to a biased result for f

examples in particle physics:

- PDFs from DIS/lattice: y_I structure function data, $f(x)$ PDFs
- spectral densities: y_t Euclidean correlators, $f(x)$ spectral function

multiple approaches: fits to fixed functional forms, NN, Backus-Gilbert

bayesian approach

- f is promoted to be a *stochastic process*
- $f(x)$ for $x \in \mathcal{I}$ is a set of stochastic variables
- for any given $\mathbf{f}$, where $f_i = f(x_i)$, we have a prior $p(\mathbf{f})$
- all a priori knowledge about f is encoded in p (more later)
- posterior distribution obtained from Bayes theorem

$$\tilde{p}(\mathbf{f}) = p(\mathbf{f}|y) = \frac{p(y|\mathbf{f})p(\mathbf{f})}{p(y)}$$

- knowledge about the solution is encoded in the posterior, eg

$$\begin{aligned}&\text{central value} : E_{\tilde{p}}[\mathbf{f}]\\&\text{covariance} : \mathrm{Cov}_{\tilde{p}}[\mathbf{f}, \mathbf{f}']\end{aligned}$$

gaussian process

GPs are a specific kind of stochastic process

$$f \sim \mathcal{GP}\left(m, k\right) ,$$

where

$$m : \mathcal{I} \to \mathbb{R}\,, \quad k : \mathcal{I} \times \mathcal{I} \to \mathbb{R}$$

for a GP, the vector of stochastic variables $\mathbf{f}$

$$\mathbf{x} = \{x_i; i = 1, \ldots, N\}\,, \quad \mathbf{f} = f(\mathbf{x}) = \begin{pmatrix} f_1 \\ \vdots \\ f_N \end{pmatrix} \in \mathbb{R}^N\,, \quad f_i = f(x_i)$$

is distributed as a multidimensional Gaussian

$$\mathbf{f} \sim \mathcal{N}(\mathbf{m}, K)\,,$$

prior distribution

mean & covariance

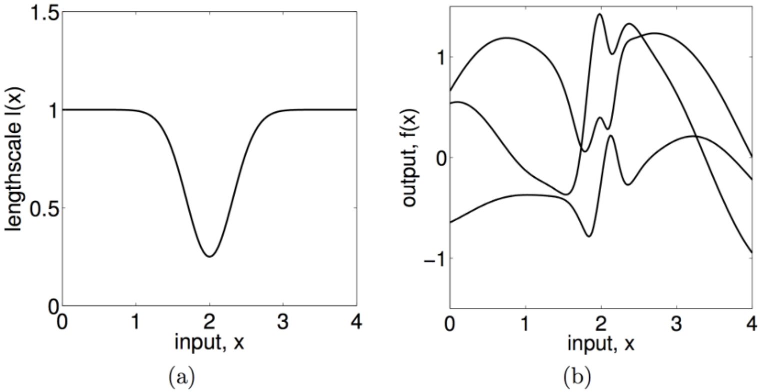
$$\mathbf{m} = m(\mathbf{x})\,, \quad K = k(\mathbf{x}, \mathbf{x}^T)\,,$$

$$\begin{aligned} E[f_i] &= m_i = m(x_i)\,, \\ \mathrm{Cov}[f_i, f_j] &= K_{ij} = k(x_i, x_j)\,. \end{aligned}$$

specific choices for this work: zero mean and Gibbs kernel

$$\begin{aligned} m(x) &= 0 \\ k(x, x') &= \sigma^2 \sqrt{\frac{2l\,(x)\,l\,(y)}{l^2\,(x) + l^2\,(y)}} \exp\left[-\frac{(x-y)^2}{l^2\,(x) + l^2\,(y)}\right] \end{aligned}$$

Gibbs kernel interpretation



we use in this work

$$l\,(x) = l_0 \times (x + \delta)$$

$$\text{hyperparameters} : \theta = (\sigma, l_0)$$

setting the problem

sampling f at points $\mathbf{x} = \{x_i; i = 1, \ldots, N\}$ and $\mathbf{x}^* = \{x_i^*; i = 1, \ldots, M\}$

$$\mathbf{f} \in \mathbb{R}^N , \quad \mathbf{f}^* \in \mathbb{R}^M ,$$

the prior probability distribution is

$$p(\mathbf{f}, \mathbf{f}^* | \theta) = \frac{1}{\sqrt{\det(2\pi K)}} \times \exp\left\{ -\frac{1}{2} \left((\mathbf{f} - \mathbf{m})^T, (\mathbf{f}^* - \mathbf{m}^*)^T \right) K^{-1} \begin{pmatrix} \mathbf{f} - \mathbf{m} \\ \mathbf{f}^* - \mathbf{m}^* \end{pmatrix} \right\} ,$$

K is now an $(N + M) \times (N + M)$ matrix

$$K = \begin{pmatrix} k(\mathbf{x}, \mathbf{x}^T) & k(\mathbf{x}, \mathbf{x}^{*T}) \\ k(\mathbf{x}^*, \mathbf{x}^T) & k(\mathbf{x}^*, \mathbf{x}^{*T}) \end{pmatrix} = \begin{pmatrix} K_{\mathbf{xx}} & K_{\mathbf{xx}^*} \\ K_{\mathbf{x}^*\mathbf{x}} & K_{\mathbf{x}^*\mathbf{x}^*} \end{pmatrix} .$$

data and theory predictions

dataset central values: $\mathbf{y} = \{y_I, I = 1, \ldots, N_{\mathrm{dat}}\}$

dataset fluctuations: $\epsilon \sim \mathcal{N}(0, C_Y)$

linear dependence on f:

$$T_I = \int_{\mathcal{I}} dx\, C_I(x) f(x) \approx \sum_{i=1}^{N} (\mathrm{FK})_{Ii} f_i$$

NB: applies to both quasi/pseudo-PDFs and spectral densities

$$E[T_I] = (\mathrm{FK})_{Ij} m_j$$
$$\mathrm{Cov}[T_I, T_J] = (\mathrm{FK})_{Ii} \left(K_{\mathbf{xx}}\right)_{ij} (\mathrm{FK})^T_{jJ}$$

posterior distribution

we want to determine

$$\tilde{p}(\mathbf{f}, \mathbf{f}^*) = p(\mathbf{f}, \mathbf{f}^*|y) = \int d\theta\, p\left(\mathbf{f}, \mathbf{f}^*, \theta|y\right)$$

$$p\left(\mathbf{f}, \mathbf{f}^*, \theta|y\right) = p\left(\mathbf{f}, \mathbf{f}^*|\theta, y\right) p\left(\theta|y\right)$$

compute each factor independently

$$p(\mathbf{f}, \mathbf{f}^*|\theta, y) \propto \exp\left\{-\frac{1}{2}\left((\mathbf{f}-\mathbf{m})^T, (\mathbf{f}^*-\mathbf{m}^*)^T\right) K^{-1} \begin{pmatrix} \mathbf{f}-\mathbf{m} \\ \mathbf{f}^*-\mathbf{m}^* \end{pmatrix}\right\}$$
$$\times\ \exp\left\{-\frac{1}{2}((\mathrm{FK})\mathbf{f}-y)^T C_Y^{-1}((\mathrm{FK})\mathbf{f}-y)\right\}\,.$$

posterior distribution

integrating over $\mathbf{f}^*$ yields

$$\int d\mathbf{f}^* \, p(\mathbf{f}, \mathbf{f}^*|\theta, y) \propto \exp\left\{-\frac{1}{2}(\mathbf{f}-\mathbf{m})^T K_{\mathbf{xx}}^{-1}(\mathbf{f}-\mathbf{m})\right\}$$
$$\times \; \exp\left\{-\frac{1}{2}((\mathrm{FK})\mathbf{f}-y)^T C_Y^{-1}((\mathrm{FK})\mathbf{f}-y)\right\}$$

posterior distribution is Gaussian

$$p\left(\mathbf{f}|\theta, y\right) = \mathcal{N}\left(\mathbf{f}; \tilde{\mathbf{m}}, \tilde{K}_{\mathbf{xx}}\right)$$

$$\tilde{\mathbf{m}} = \mathbf{m} + K_{\mathbf{xx}}(\mathrm{FK})^T \, C_{YT}^{-1} \left(\mathbf{y} - (\mathrm{FK})\mathbf{m}\right)$$
$$\tilde{K}_{\mathbf{xx}} = K_{\mathbf{xx}} - K_{\mathbf{xx}}(\mathrm{FK})^T C_{YT}^{-1} \, (\mathrm{FK}) K_{\mathbf{xx}}$$
$$C_{YT} = (\mathrm{FK}) K_{\mathbf{xx}} (\mathrm{FK})^T + C_Y$$

posterior distribution

integrating over $\mathbf{f}$

$$p\left(\mathbf{f}^*|\theta, y\right) = \mathcal{N}\left(\tilde{\mathbf{m}}^*, \tilde{K}^*_{\mathbf{xx}}\right)$$

$$\tilde{\mathbf{m}}^* = \mathbf{m}^* + K_{\mathbf{x}^*\mathbf{x}}(\mathrm{FK})^T C_{YT}^{-1}\left(\mathbf{y} - (\mathrm{FK})\mathbf{m}\right)\ ,$$

$$\tilde{K}_{\mathbf{x}^*\mathbf{x}^*} = K_{\mathbf{x}^*\mathbf{x}^*} - K_{\mathbf{x}^*\mathbf{x}}(\mathrm{FK})^T\, C_{YT}^{-1}\,(\mathrm{FK})K_{\mathbf{x}\mathbf{x}^*}\ .$$

- correction to the mean proportional to $(\mathbf{y} - (\mathrm{FK})\mathbf{m})$
- correlations in the prior allow to make predictions for $\mathbf{f}^*$

inference for hyperparameters

using Bayes theorem

$$p(\theta|y) = \frac{p(y|\theta)\, p_\theta(\theta)}{\int d\theta\, p(y|\theta)\, p_\theta(\theta)}\,,$$

on the RHS

$$p(y|\theta) = \frac{e^{-\frac{1}{2}\,(y-(\mathrm{FK})\mathbf{m})^T C_{YT}^{-1} (y-(\mathrm{FK})\mathbf{m})}}{\sqrt{\det\left[2\pi C_{YT}\right]}}\,.$$

$p(\theta|y)$ can be sampled by MCMC

PDF from DIS - closure test

study the triplet PDF T_3

$$T_3 = (u + \bar{u}) - \left(d + \bar{d}\right)$$

using DIS structure function from BCDMS

$$y = F_2^p - F_2^d = C_{T_3} \otimes T_3$$

$$C_Y = \mathrm{Cov}\left[F_2^p, F_2^p\right] + \mathrm{Cov}\left[F_2^d, F_2^d\right] - 2\mathrm{Cov}\left[F_2^p, F_2^d\right]$$

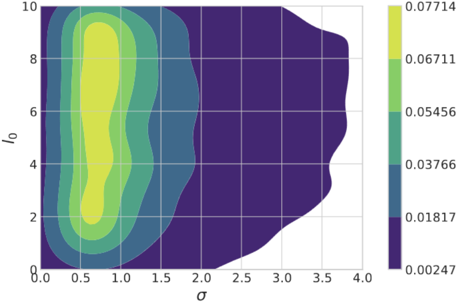
test the methodology using synthetic data

$$y = (\mathrm{FK})\,\mathbf{f}_0 + \eta\,, \quad \text{with} \quad \eta \sim \mathcal{N}\left(0, C_Y\right)$$

where $\mathbf{f}_0$ is taken from a known PDF set (NNPDF4.0)

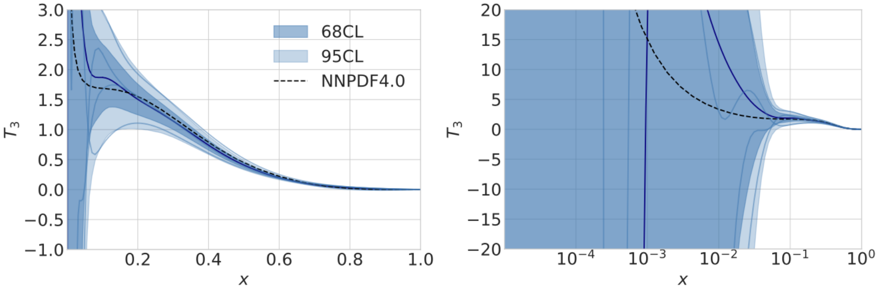
inference for the hyperparameters

starting from **flat** priors for the hyperparameters, we get for $p(\theta|y)$



and $p(\mathbf{f}^*|\theta, y)$ is known analytically

inference for the PDF



interpretation of the results - closure test

vanishing exp errors

$$y = y_0 = (\mathrm{FK})\mathbf{f}_0\,, \quad C_Y = 0$$

yields

$$\tilde{\mathbf{m}} = R^{(0)}_{\mathbf{xx}}\,\mathbf{f}_0\,, \quad \tilde{\mathbf{m}}^* = R^{(0)}_{\mathbf{x}^*\mathbf{x}}\,\mathbf{f}_0$$

where we introduced the smearing kernel

$$R^{(0)}_{\mathbf{xx}} = K_{\mathbf{xx}}\,(\mathrm{FK})^T\left[(\mathrm{FK})K_{\mathbf{xx}}(\mathrm{FK})^T\right]^{-1}(\mathrm{FK})$$

the result of Bayesian inference is a smeared version of the 'true' answer

$$\tilde{\mathbf{m}} - \mathbf{f}_0 = \left[R^{(0)}_{\mathbf{xx}} - \mathbb{1}\right]\mathbf{f}_0\,, \quad \tilde{K}_{\mathbf{xx}} = \left(\mathbb{1} - R^{(0)}_{\mathbf{xx}}\right)K_{\mathbf{xx}}$$

in data space, consider *bias* and *variance*

$$\mathcal{B} = (\mathrm{FK})\,(\tilde{\mathbf{m}} - \mathbf{f}_0) = (\mathrm{FK})\left(R^{(0)}_{\mathbf{xx}} - \mathbb{1}\right)\mathbf{f}_0 = 0\,,$$
$$\mathcal{V} = (\mathrm{FK})\,\tilde{K}\,(\mathrm{FK})^T = (\mathrm{FK})\left(\mathbb{1} - R^{(0)}_{\mathbf{xx}}\right)K_{\mathbf{xx}}(\mathrm{FK})^T = 0\,.$$

- the GP methodology reconstructs the input data exactly, independently of the specific values of the hyperparameters
- the model function is not in general reconstructed exactly, *i.e.* $\tilde{\mathbf{m}} \neq \mathbf{f}_0$
- in the functional space, a residual reconstruction error is still present

adding experimental errors

reconstruction kernel

$$R_{\mathbf{xx}} = K_{\mathbf{xx}}\,(\mathrm{FK})^T \left[(\mathrm{FK})K_{\mathbf{xx}}(\mathrm{FK})^T + C_Y\right]^{-1}(\mathrm{FK})$$

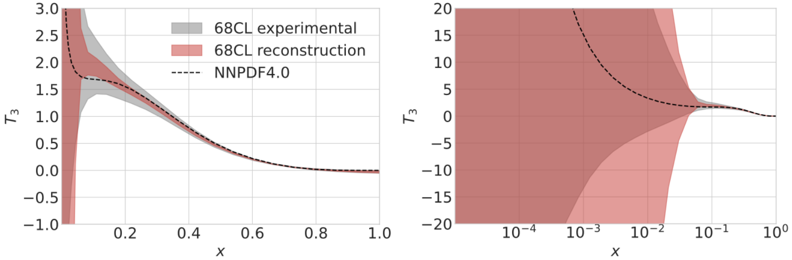
comparison with the 'true' input

$$\begin{aligned}\tilde{\mathbf{m}} - \mathbf{f}_0 &= \left[R_{\mathbf{xx}} - \mathbb{1}\right]\mathbf{f}_0 + a_{\mathbf{xx}}^T\eta \\ \tilde{K}_{\mathbf{xx}} &= \left(\mathbb{1} - R_{\mathbf{xx}}\right)K_{\mathbf{xx}}\left(\mathbb{1} - R_{\mathbf{xx}}\right)^T + a_{\mathbf{xx}}^T C_Y a_{\mathbf{xx}}\end{aligned}$$

in data space

$$\begin{aligned}\mathcal{B} &= (\mathrm{FK})\left[R_{\mathbf{xx}} - \mathbb{1}\right]\mathbf{f}_0 + (\mathrm{FK})a_{\mathbf{xx}}^T\eta \\ \mathcal{V} &= (\mathrm{FK})\left(\mathbb{1} - R_{\mathbf{xx}}\right)K_{\mathbf{xx}}\left(\mathbb{1} - R_{\mathbf{xx}}\right)^T(\mathrm{FK})^T + (\mathrm{FK})a_{\mathbf{xx}}^T C_Y a_{\mathbf{xx}}(\mathrm{FK})^T\end{aligned}$$

- limited reconstruction due to smearing, functional uncertainty
- functional uncertainty is not cured by more precise data
- the term proportional to η is the propagation of the experimental error in the reconstructed function, experimental uncertainty



conclusions

- bayesian analysis offers an independent tool to look at inverse problems
- all hypotheses are explicitly spelled out in the prior
- for linear data, we get analytical results useful to build intuition
- Backus-Gilbert methods can be rephrased as GP processes with specific prior
- interesting connection with NN parametrizations (no time today!)
- being used for PDFs and spectral densities