Bayesian Model Mixing

A Nuclear Physics Perspective

Vojtech Kejzlar
Department of Statistics and Probability, MSU

Collaborators: Léo Neufcourt, Frederi Viens, Taps Maiti, Witold Nazarewicz
(*STT/FRIB at MSU*)

ISNET-6
Oct 8-12, 2018, TU Darmstadt

Outline

Introduction

Common scenario in nuclear physics: Use experimental measurements $(y_i)_{i=1}^n$ at known values $(x_i)_{i=1}^n$ of a physical process $x \mapsto y(x)$ to predict its values $y^* = y(x^*)$ at new input points x^* using a *computer model*.

General Bayesian Solution:

1. Consider a quantity of interest interest Δ (unknown value of a observable y^* for given nuclear configuration x^*).
2. Approximate $p(\Delta|y, \mathcal{M})$ given the data $y = (y_1, \ldots, y_n)$ and model $\mathcal{M}$ obtained from Bayes formula.

Introduction

Common scenario in nuclear physics: Use experimental measurements $(y_i)_{i=1}^n$ at known values $(x_i)_{i=1}^n$ of a physical process $x \mapsto y(x)$ to predict its values $y^* = y(x^*)$ at new input points x^* using a *computer model*.

General Bayesian Solution:

1. Consider a quantity of interest interest Δ (unknown value of a observable y^* for given nuclear configuration x^*).
2. Approximate $p(\Delta|y, \mathcal{M})$ given the data $y = (y_1, \ldots, y_n)$ and model $\mathcal{M}$ obtained from Bayes formula.

Situation with K competing models $\mathcal{M}_1, ..., \mathcal{M}_K$:

- $p(\Delta|y, \mathcal{M}) \Rightarrow p(\Delta|y, \mathcal{M}_k)$.
- BMA posterior distribution $p(\Delta|y)$ accounts for modeling uncertainty and can lead to better predictions (Witek's talk).

Practical Considerations and Challenges

- **Calibration:** Models under consideration rely on calibration parameters θ_k.

Practical Considerations and Challenges

- **Calibration:** Models under consideration rely on calibration parameters θ_k.

- **Varying domains:** $\mathcal{M}_1, ..., \mathcal{M}_K$ can be defined on different input domains.
  - Models constrained by observables of different nature.
  - Models are appropriate for different parts of nuclear domain.
  - One model has wider parameter space for a specific observable (models mostly fitted on binding energies or charged radii).

Practical Considerations and Challenges

- **Calibration:** Models under consideration rely on calibration parameters θ_k.
- **Varying domains:** $\mathcal{M}_1, ..., \mathcal{M}_K$ can be defined on different input domains.
  - Models constrained by observables of different nature.
  - Models are appropriate for different parts of nuclear domain.
  - One model has wider parameter space for a specific observable (models mostly fitted on binding energies or charged radii).
- **Numerical integration:** Posteriors on models $p(\mathcal{M}_k|y)$ need to be approximated.

Bayesian Calibration of Computer Models

Statistical formulation for a single model (Kennedy, O'Hagan, 2001):

$$y_i^m = \Psi_m(f(x_i, \theta)) + \delta(x_i) + \sigma_m(x_i)\epsilon_i \tag{1}$$

Note: Ψ_m is here to emphasize that the same model typically produces and is fitted to m types of observables.

- $\Psi_m(f(x_i, \theta))$ is the computer model.
- $\delta(x)$ is the systematic error of the model, $\delta(x) \sim \mathcal{GP}(m_\delta(x), k_\delta(x, x'))$.
- ϵ_i are scaled measurement errors, $\epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$.

Bayesian Calibration of Computer Models

For expensive models $f(x, \theta)$ that we cannot evaluate at wish, it is common practice to model them as a Gaussian process:

$$f(x, \theta) \sim \mathcal{GP}(m_f(x, \theta), k_f((x, \theta), (x', \theta')))$$

Computationally efficient alternative (linearization):

$$f(x, \theta) \approx f(x, \theta_0) + \nabla_\theta f(x, \theta_0)^T \cdot (\theta - \theta_0) \tag{2}$$

Pros	**Cons**
Intuitive	Requires previous optimization
Computationally inexpensive	

Bayesian Calibration of Computer Models - UNEDF0

Averaging Models With Different Domains

Recurrent scenario: K competing models $\mathcal{M}_1, ..., \mathcal{M}_K$ defined on different input domains (Energy Density Functionals (EDF) vs many-body computations, or even between different EDFs)

$\Rightarrow$ Each model $\mathcal{M}_k$ is calibrated on a subset $y^{(k)}$ of the data y.

The generic BMA allows us to compute the posteriors $p(\Delta|y^{(k)})$ and $p(\Delta|y^{(\infty)})$, where $y^{(\infty)} := \bigcap_k y^{(k)}$.

$\Rightarrow$ Additional argument needed, because we want $p(\Delta|y)$ for $y = \bigcup_k y^{(k)}$.

Averaging Models With Different Domains

Fortunately

$$p(\theta_k|y, \mathcal{M}_k) \propto p(\theta_k|y^{(k)}, \mathcal{M}_k),$$

and

$$p(\mathcal{M}_k|y) \propto p(y^{(-k)}|y^{(k)})p(\mathcal{M}_k|y^{(k)}).$$

The BMA posterior for a general quantity Δ of interest:

$$p(\Delta|y) \propto \sum_{k=1}^{K} p(\Delta|y^{(k)}, \mathcal{M}_k)p(y^{(-k)}|y^{(k)})p(\mathcal{M}_k|y^{(k)}). \quad (3)$$

Averaging Models With Different Domains

Fortunately

$$p(\theta_k|y, \mathcal{M}_k) \propto p(\theta_k|y^{(k)}, \mathcal{M}_k),$$

and

$$p(\mathcal{M}_k|y) \propto p(y^{(-k)}|y^{(k)})p(\mathcal{M}_k|y^{(k)}).$$

The BMA posterior for a general quantity Δ of interest:

$$p(\Delta|y) \propto \sum_{k=1}^{K} p(\Delta|y^{(k)}, \mathcal{M}_k)p(y^{(-k)}|y^{(k)})p(\mathcal{M}_k|y^{(k)}). \quad (3)$$

No major complication, we just need to compute an additional multiplicative corrective factor $p(y^{(-k)}|y^{(k)})$:)

Computing Evidence Integral

The evidence integral of the model $\mathcal{M}_k$

$$p(y^{(k)}|\mathcal{M}_k) = \int_{\theta_k} p(y^{(k)}|\theta_k, \mathcal{M}_k)\pi(\theta_k|\mathcal{M}_k)d\theta_k \tag{4}$$

is the key quantity to the posterior probability of the different models, which are given according to Bayes formula.

Popular methods:

- Monte Carlo integration
- Laplace approximation

Fast alternative (recycling samples from Bayesian calibration):

$$\widehat{p(y^{(k)}|\mathcal{M}_k)} = \frac{n_C}{\sum_i p(y^{(k)}|\theta_k^{(i)}, \mathcal{M}_k)\pi(\theta_k^{(i)}|\mathcal{M}_k)}.$$

Illustrative Example - Toy Model

Scenario:

- Synthetic dataset with observations generated independently from $N(0, 10^{-3})$ at input points
  $x = (\pm 1, \pm 2, \pm 3, \pm 4, \pm 5, \pm 6, \pm 7, \pm 8, \pm 9)$.

- Two competing computer models $\mathcal{M}_1$ and $\mathcal{M}_2$ defined as follows:

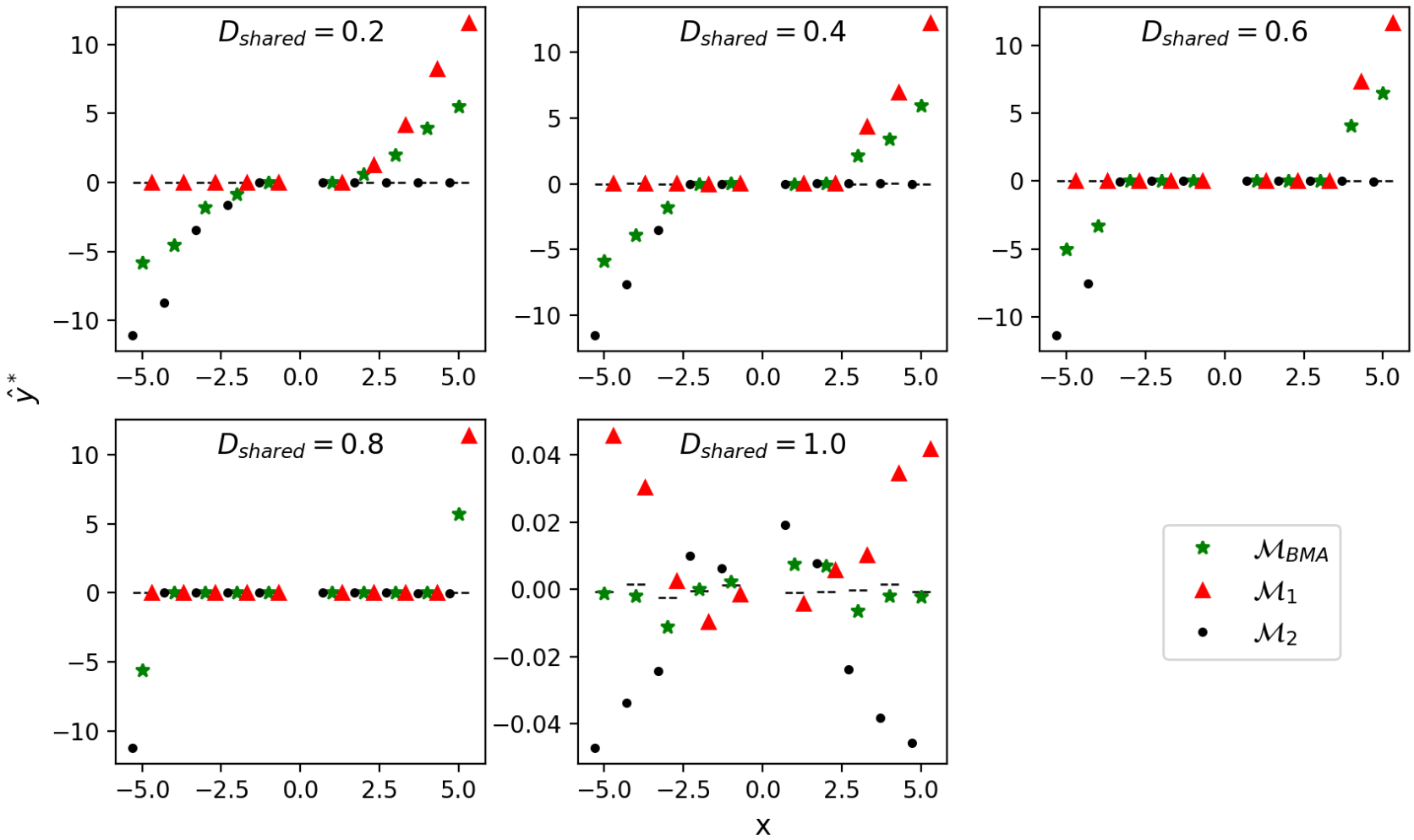
$$\begin{aligned} f_1(x, \theta_1) &= 0.5x^2 + \theta_1, \\ f_2(x, \theta_2) &= -0.5x^2 + \theta_2, \end{aligned} \tag{5}$$

Illustrative Example - Toy Model

Each of the models $\mathcal{M}_1$ and $\mathcal{M}_2$ was assigned different training dataset $y^{(k)}$ for various proportion of shared observations (D_{shared}):

		Training dataset $y^{(k)}$																
D_{shared}	Model	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9
0.2	$\mathcal{M}_1$	x	x	x	x	x	x	x	x	x	x								
	$\mathcal{M}_2$									x	x	x	x	x	x	x	x	x	x
0.4	$\mathcal{M}_1$		x	x	x	x	x	x	x	x	x	x							
	$\mathcal{M}_2$								x	x	x	x	x	x	x	x	x	x	
0.6	$\mathcal{M}_1$			x	x	x	x	x	x	x	x	x	x						
	$\mathcal{M}_2$							x	x	x	x	x	x	x	x	x	x		
0.8	$\mathcal{M}_1$				x	x	x	x	x	x	x	x	x	x					
	$\mathcal{M}_2$						x	x	x	x	x	x	x	x	x	x			
1	$\mathcal{M}_1$					x	x	x	x	x	x	x	x	x	x				
	$\mathcal{M}_2$					x	x	x	x	x	x	x	x	x	x				

Illustrative Example - Toy Model



Illustrative Example - Toy Model

D_{shared}	Model	$\widehat{PMSE}$	$p(y^{(k)}\|\mathcal{M}_k)$	$p(y^{(-k)}\|y^{(k)})$	$\hat{r}_{BMA}$	$P_{1,2}$
0.2	$\mathcal{M}_1$	22.0	$1.17 \cdot 10^{-24}$	$9.88 \cdot 10^{-23}$	0.504	
	$\mathcal{M}_2$	21.3	$1.18 \cdot 10^{-24}$	$9.07 \cdot 10^{-23}$	0.488	0.91
	$\mathcal{M}_{BMA}$	10.9	-	-	-	
0.4	$\mathcal{M}_1$	21.7	$5.52 \cdot 10^{-23}$	$2.13 \cdot 10^{-19}$	0.516	
	$\mathcal{M}_2$	20.4	$5.84 \cdot 10^{-23}$	$2.13 \cdot 10^{-19}$	0.485	0.95
	$\mathcal{M}_{BMA}$	10.5	-	-	-	
0.6	$\mathcal{M}_1$	19.0	$3.98 \cdot 10^{-21}$	$1.03 \cdot 10^{-14}$	0.498	
	$\mathcal{M}_2$	18.6	$3.99 \cdot 10^{-21}$	$1.30 \cdot 10^{-14}$	0.487	1.26
	$\mathcal{M}_{BMA}$	9.54	-	-	-	
0.8	$\mathcal{M}_1$	13.0	$2.65 \cdot 10^{-19}$	$5.52 \cdot 10^{-9}$	0.508	
	$\mathcal{M}_2$	12.7	$2.67 \cdot 10^{-19}$	$5.61 \cdot 10^{-9}$	0.497	1.01
	$\mathcal{M}_{BMA}$	6.39	-	-	-	
1	$\mathcal{M}_1$	$6.16 \cdot 10^{-4}$	$2.14 \cdot 10^{-18}$	-	0.956	
	$\mathcal{M}_2$	$8.68 \cdot 10^{-4}$	$2.17 \cdot 10^{-18}$	-	0.969	0.99
	$\mathcal{M}_{BMA}$	$2.71 \cdot 10^{-5}$	-	-	-	

Figure: BMA results for model scenario with different domains. Here, $P_{1,2} = p(y^{(-1)}|y^{(1)})p(\mathcal{M}_1|y^{(1)})/[p(y^{(-2)}|y^{(2)})p(\mathcal{M}_2|y^{(2)})]$

Thank you!

kejzlarv@stt.msu.edu