Multi Level MCMC methods

Andrea Boselli
Carlo Ghiglione
Eleonora Spizzi
Erica Manfrin
Randeep Singh

November 11^{th} 2021

Politecnico di Milano,
joint project Bayesian Statistics - Computational Statistics
Tutors:
Prof. Alessandra Guglielmi - Department of Mathematics, Politecnico di Milano
Prof. Andrea Manzoni - Department of Mathematics, Politecnico di Milano

This is us

What are Multi Level MCMC methods?

Uncertainty Quantification and Bayesian Statistics

Mathematical models can be defined as input-output relationships:

$$Y = f(\theta_1, ..., \theta_k)$$

The **parameters** (or **inputs**) $\{\theta_1, ..., \theta_k\} =: \underline{\theta}$:

- influence the model output
- are usually unknown

Stochastic Modelling (Bayesian Approach)

1. Assign a probability $\pi(\underline{\theta})$, based on prior (expert) knowledge
2. With observations of the model output F_{obs}, update the knowledge $\pi(\underline{\theta}|F_{obs})$ on model parameters

Research Topic

In this framework, computing the likelihood may require the solution of complicated integrals, or differential problems (e.g. ODEs, PDEs).

- **no exact solution available**, but only approximations
- very **time consuming** computations

Multilevel Monte Carlo (MLMC)
$+$
Monte Carlo Markov Chain (MCMC)
$\rightarrow$ **Multilevel MCMC (MLMCMC)**

Multilevel Monte Carlo (MLMC)

Suppose we want to compute $\mathbb{E}[A(\underline{\theta})]$, where $A(\underline{\theta})$ is a quantity of interest, arising from the model $Y = f(\underline{\theta})$.

Suppose we can sample iid samples $\{\underline{\theta}^i\}_{i=1}^N$ from a desired distribution.

Usually, $A(\underline{\theta})$ cannot be computed exactly, but we have only a hierarchy of approximations $\{A_\ell\}_{\ell\in\mathbb{N}}$.

- Increasing ($\uparrow$) **computational cost**
- Decreasing ($\downarrow$) **approximation error**

Multilevel Monte Carlo (MLMC)

Standard Monte Carlo method

$$\mathbb{E}[A] \approx \mathbb{E}[A_L] \approx \frac{1}{N} \sum_{i=1}^{N} A_L(\underline{\theta}^i)$$

Alternatively, we can use a hierarchy of Control Variates:

$$\mathbb{E}[A_L] = \mathbb{E}[A_0] + \sum_{\ell=1}^{L} \mathbb{E}[A_\ell - A_{\ell-1}]$$

and estimate each term of the RHS independently.

Multilevel Monte Carlo (MLMC)

Multilevel Monte Carlo method

$$\mathbb{E}[A] \approx \frac{1}{N_0}\sum_{i=1}^{N_0} A_0(\underline{\theta}^{(i,0)}) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}[A_\ell(\underline{\theta}^{(i,\ell)}) - A_{\ell-1}(\underline{\theta}^{(i,\ell)})]$$

- Many **cheap** computations of **coarse** solutions
- Few **expensive** computations of **fine** solutions
- We <u>can't</u> sample directly from the posterior distribution

Multilevel Monte Carlo Markov Chain (MLMCMC)

Consider again a **hierarchy** $\{Q_\ell\}_{\ell=1}^{L}$ of L approximations $Q_\ell := Q_{M_\ell,R_\ell}$ of a given quantity of interest Q, where:

- M_ℓ: **model complexity**
- R_ℓ: **number of input parameters** $\underline{\theta}_\ell \in \mathbb{R}^{R_\ell}$, $\underline{\theta}_\ell | F_{obs} \sim \nu^\ell$

Important!

$\underline{\theta}_\ell | F_{obs} \sim \nu^\ell$ is our **posterior** at level ℓ.

Multilevel Monte Carlo Markov Chain (MLMCMC)

Like in MLMC methods, we stop at level L and exploit the identity:

$$\mathbb{E}[Q] \approx \mathbb{E}_{\nu^L}[Q_L] = \mathbb{E}_{\nu^0}[Q_0] + \sum_{\ell=1}^{L} (\mathbb{E}_{\nu^\ell}[Q_\ell] - \mathbb{E}_{\nu^{\ell-1}}[Q_{\ell-1}])$$

We estimate:

- $\mathbb{E}_{\nu^0}[Q_0]$ through MCMC;
- each $Y_\ell := (\mathbb{E}_{\nu^\ell}[Q_\ell] - \mathbb{E}_{\nu^{\ell-1}}[Q_{\ell-1}])$ from 2 Markov chains: $\{\underline{\theta}_\ell^n\}_{n\in\mathbb{N}}$ and $\{\underline{\Theta}_{\ell-1}^n\}_{n\in\mathbb{N}}$.

$$\hat{Y}_{\ell,N_\ell} = \frac{1}{N_\ell} \sum_{n=1}^{N_\ell} [Q_\ell(\underline{\theta}_\ell^n) - Q_{\ell-1}(\underline{\Theta}_{\ell-1}^n)]$$

Multilevel Monte Carlo Markov Chain (MLMCMC)

$$\hat{Y}_{\ell,N_\ell} = \frac{1}{N_\ell}\sum_{n=1}^{N_\ell}[Q_\ell(\underline{\theta}_\ell^n) - Q_{\ell-1}(\underline{\Theta}_{\ell-1}^n)]$$

Required assumptions

- To achieve an unbiased estimate of $\mathbb{E}_{\nu^L}[Q_L]$, $\{\underline{\theta}_\ell^n\}$ and $\{\underline{\Theta}_\ell^n\}$ both need to be drawn from $\nu^\ell\ \forall\ell$.
- To ensure that $Var[Y_\ell] \to 0$, $\{\underline{\theta}_\ell^n\}$ and $\{\underline{\Theta}_{\ell-1}^n\}$ need to be correlated.

$\{\underline{\theta}_\ell^n\}$ and $\{\underline{\Theta}_{\ell-1}^n\}$ are sampled through a MH-like algorithm (see final appendix), satisfying all the requirements above.

Our goals

Our goals

- **Understand and apply** recently proposed **MLMCMC techniques** for large-scale applications with high-dimensional parameter spaces, e.g., in uncertainty quantification related with systems described in terms of (partial) differential equations.
- Explore the **impact** of the **choice of the prior and proposal distributions** for the models at each level, investigating alternative options to build models at different levels or fidelities.
- Ultimately, apply the MLMCMC technique to a (slightly) **more advanced case of interest** modeled through **PDEs**.

Our progress so far

COMET: a first case study

To practice with MLMCMC implementation, we decided to start from a simple problem, where each parameter has a clear physical meaning.

Comet equation

$$\begin{cases} -\mu\Delta u + 10(\cos\theta, \sin\theta) \cdot \nabla u = 10e^{-100\left\|\underline{x}-\underline{x}_0\right\|_2} & \underline{x} \in \Omega = [0,1]^2 \\ u = 0 & \underline{x} \in \partial\Omega \end{cases}$$

- $\mu \in (0, \infty)$: diffusion parameter
- $\theta \in (0, 2\pi)$: angle of advection term
- $\underline{x}_0 = (0.5, 0.5)$: centre of the forcing bump

COMET: a first case study

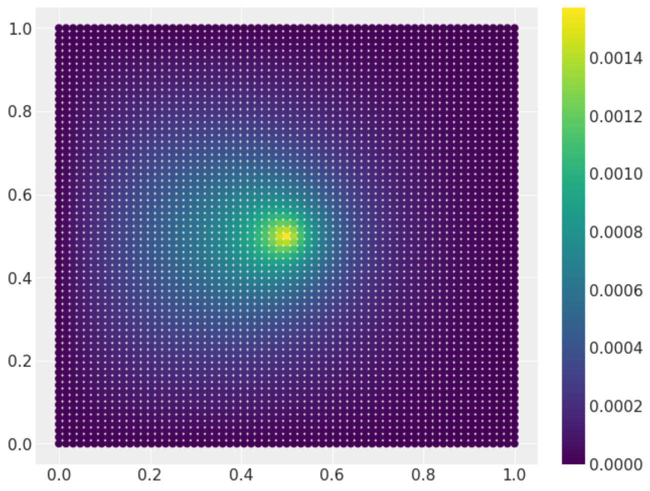

Figure 1: Plot of the solution for $\mu = 2$ and $\theta = \pi$ in a 64x64 grid

COMET - Likelihood

We fix a value for $\{\mu^*, \theta^*\}$ as a ground truth: the corresponding solution u_{obs} (plus a gaussian noise) constitute the **observed data**.

Models at different levels differ only for the **mesh refinement**, not for the set of parameters.

At level ℓ, the **likelihood** is chosen as:

$$u_{obs}|\mu, \theta \sim \mathcal{N}(u_\ell(\mu, \theta), \sigma_\ell \mathcal{I}_{M_\ell})$$

where $u_\ell(\mu, \theta)$ is the solution of the PDE for μ, θ, evaluated on the vertices of u_{obs} (through a proper interpolation).

Ideally, the likelihood should be concentrated on $\{\mu^*, \theta^*\}$.

COMET - Priors setting

For the beginning we set **flat priors**:

$$\mu \sim \mathcal{U}(0.5, 5)$$
$$\theta \sim \mathcal{U}(0, 2\pi)$$

We expect that the method is able to estimate a **posterior distribution** concentrated on the exact values $\{\mu^*, \theta^*\}$.

COMET - Some results

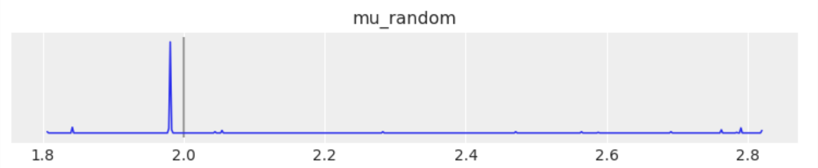

Figure 2: Posterior distribution for μ

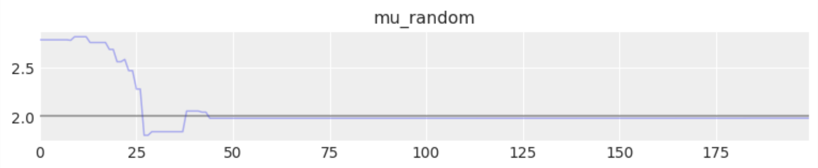

Figure 3: Trace plot for μ

Acceptance rate: 0.125

Normalized ESS: 0.0140

COMET - Some results

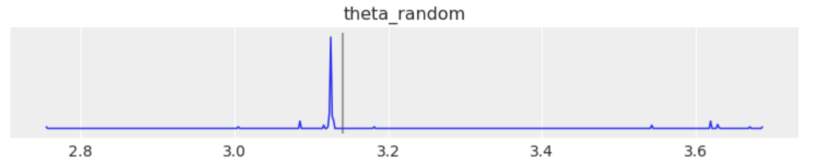

Figure 4: Posterior distribution for θ

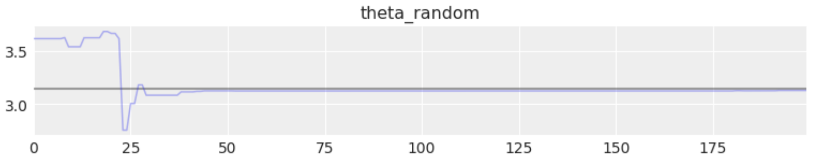

Figure 5: Trace plot for θ

Acceptance rate: 0.125

Normalized ESS: 0.0105

COMET - Interpretations

These results are due to:

- the method is used to estimate the distribution of μ and θ that are **fixed parameters** and not random fields;
- very **high number** of available observations;
- available observations have very **small noise**.

$\rightarrow$ It's not a real "stochastic" context!

Low acceptance rate and ESS $\leftarrow$ The chain converges to the correct parameters and it is stucked by the **very precise** and **numerous** observations that confirm the reached state.

Roadmap

Our next steps is to recover a more "stochastic" and applicative scenario. To do it, we propose to:

- **Increase the noise** of the available observations of the exact solution;
- **Decrease** the number of available observations of the exact solution;
- Try to apply the method to a **more complex problem**.

Bibliography

- T. J. Dodwell, C. Ketelsen, R. Scheichl, and A. L. Teckentrup, Multilevel Markov Chain Monte Carlo. SIAM Review 61(3), pp. 509–545, 2019.
- T. J. Dodwell, C. Ketelsen, R. Scheichl, andA. L. Teckentrup, A Hierarchical Multilevel Markov Chain Monte Carlo Algorithm with Applications to Uncertainty Quantification in Subsurface Flow. SIAM/ASA J. Uncertainty Quantification 3, pp. 1075–1108, 2015.
- Python PyMC3 package (providing an implementation for MLMCMC and a few examples)
- FEniCS (providing an easy implementation for the discretization and solution of PDEs through Finite Element Methods (FEM))

Appendix: M-H for MCMC for $\mathbf{Q}_\ell - Q_{\ell-1}$

Suppose we can generate i.i.d samples from $\nu^{\ell-1}$ (using recursive subsampling algorithm). Set as initial states $\Theta^0_{\ell-1} \sim \nu^{\ell-1}$ and $\theta^0_\ell := [\Theta^0_{\ell-1}, \theta^0_{\ell,F}]$.
For $n > 0$:

- **On level** $\ell - 1$: Generate an independent sample $\Theta^{n+1}_{\ell-1}$ from the distribution $\nu^{\ell-1}$
- **On level** ℓ: Given θ^n_ℓ and $\Theta^{n+1}_{\ell-1}$, generate θ^{n+1}_ℓ using Metropolis-Hastings MCMC algorithm with the specific proposal distribution $q^\ell_{ML}(\theta'_\ell|\theta^n_\ell)$ induced by taking $\theta'_{\ell,C} := \Theta^{n+1}_{\ell-1}$ and by generating a proposal for $\theta'_{\ell,F}$ from some proposal distribution $q^{\ell,F}_{ML}(\theta'_{\ell,F}|\theta^n_{\ell,F})$ that is independent of the coarse modes. The acceptance probability is:

$$\alpha^\ell_{ML}(\theta'_\ell|\theta^n_\ell) = min\left\{1, \frac{\pi^\ell(\theta'_\ell)q^\ell_{ML}(\theta^n_\ell|\theta'_\ell)}{\pi^\ell(\theta^n_\ell)q^\ell_{ML}(\theta'_\ell|\theta^n_\ell)}\right\}$$