Introduction to the Bayesian Approach to Inverse Problems - Part 2: Algorithms

Lecture 4

Claudia Schillings, Aretha Teckentrup[*,†]

[*]School of Mathematics, University of Edinburgh
[†]Alan Turing Institute, London

LMS Short Course - May 11, 2017

THE UNIVERSITY *of* EDINBURGH
School of Mathematics

The Alan Turing Institute

Outline of second lecture

1. Metropolis Hastings Markov chain Monte Carlo (MH-MCMC)
2. Convergence of MH-MCMC
3. Multilevel MH-MCMC

Metropolis Hastings Markov chain Monte Carlo (MH-MCMC)

Sampling methods based on non-i.i.d. samples

- In the last lecture, we used Bayes' Theorem to write

$$\mathbb{E}_{\mu^y}[\phi] = \frac{\mathbb{E}_{\mu_0}[\phi \exp[-\Phi]]}{\mathbb{E}_{\mu_0}[\exp[-\Phi]]},$$

  and then used Monte Carlo to estimate the two prior expectations.

- In this lecture, we want to estimate the posterior expectation directly by a sampling method:

$$\mathbb{E}_{\mu^y}[\phi] = \int_{\mathbb{R}^n} \phi(u)\, d\mu^y(u) \approx \sum_{i=1}^{N} w_i \phi(u^{(i)}) \approx \sum_{i=1}^{N} w_i \phi_h(u^{(i)}),$$

  where $u^{(i)} \sim \mu^y$, for all $1 \leq i \leq N$, and ϕ_h is a numerical approximation to ϕ with step length h.

- Since μ^y is not known in closed form, it is generally not possible to generate i.i.d. samples distributed according to μ^y.

Metropolis Hastings Markov chain Monte Carlo (MH-MCMC)

Markov chain Monte Carlo (MCMC) [Robert, Casella '99]

A Markov chain Monte Carlo (MCMC) estimator of $\mathbb{E}_{\mu^y}[\phi]$ is of the form

$$\widehat{E}_{h,N}^{\text{MCMC}} := \frac{1}{N} \sum_{i=1}^{N} \phi_h(u^{(i)}),$$

where $\{u^{(i)}\}_{i=1}^{\infty}$ is a Markov chain.

Definition (Markov chain)

The family of random variables $\{u^{(i)}\}_{i=1}^{\infty}$ is called a *Markov chain* if

$$\Pr\big[\, u^{(i)} = x_i \,\big|\, u^{(1)} = x_1 \,, \ldots ,\, u^{(i-1)} = x_{i-1} \,\big] = \Pr\big[\, u^{(i)} = x_i \,\big|\, u^{(i-1)} = x_{i-1} \,\big],$$

for all $i \geq 2$ and $x_1, \ldots, x_i \in \mathbb{R}^n$.

We want the distribution of each $u^{(i)}$ to be (close to) μ^y.

Metropolis Hastings Markov chain Monte Carlo (MH-MCMC)

Standard Algorithm [Robert, Casella '99]

A particular example is the Metropolis Hastings (MH-MCMC) estimator, which uses the following algorithm to construct $\{u^{(i)}\}_{i=1}^{\infty}$:

ALGORITHM 1. (Standard MH-MCMC)

- Choose $u^{(1)}$ with $\rho^y(u^{(1)}) > 0$.
- At state $u^{(i)}$, sample a proposal u' from density $q(u' \,|\, u^{(i)})$.
- Accept sample u' with probability

$$\alpha(u' \,|\, u^{(i)}) = \min\left(1, \frac{\rho^y(u')\, q(u^{(i)} \,|\, u')}{\rho^y(u^{(i)})\, q(u' \,|\, u^{(i)})}\right),$$

i.e. $u^{(i+1)} = u'$ with probability $\alpha(u' \,|\, u^{(i)})$; otherwise stay at $u^{(i+1)} = u^{(i)}$. Here ρ^y denotes the density of μ^y.

- The proposal density q is chosen to be easy to sample from.
- The accept/reject step is added in order to obtain samples from μ^y.
- Knowledge of the normalising constant Z of μ^y is not required.

Metropolis Hastings Markov chain Monte Carlo (MH-MCMC)

Computable Algorithm

- To evaluate $\alpha(u' \mid u^{(i)})$, we need to compute $\rho^y(u')$ and $\rho^y(u^{(i)})$.
- Since $\rho^y(u) = \frac{1}{Z}\exp[-\Phi(u)]\rho_0(u)$, this requires the computation of $\Phi(u')$ and $\Phi(u^{(i)})$, which usually cannot be done exactly in practice.
- With Φ_h a numerical approximation to Φ, we define the approximate posterior distribution μ_h^y with density $\rho_h^y(u) = \frac{1}{Z}\exp[-\Phi_h(u)]\rho_0(u)$.

ALGORITHM 2. (Computable MH-MCMC)

- Choose $u^{(1)}$ with $\rho_h^y(u^{(1)}) > 0$.
- At state $u^{(i)}$, sample a proposal u' from density $q(u' \mid u^{(i)})$.
- Accept sample u' with probability

$$\alpha_h(u' \mid u^{(i)}) = \min\left(1, \frac{{\rho_h}^y(u')\, q(u^{(i)} \mid u')}{{\rho_h}^y(u^{(i)})\, q(u' \mid u^{(i)})}\right),$$

i.e. $u^{(i+1)} = u'$ with probability $\alpha_h(u' \mid u^{(i)})$; otherwise stay at $u^{(i+1)} = u^{(i)}$. Here ρ_h^y denotes the density of μ_h^y.

Metropolis Hastings Markov chain Monte Carlo (MH-MCMC)

Choice of proposal density q [Cotter, Dashti, Stuart '12]

- It is desirable to choose a proposal q that is prior reversible, i.e.

$$\rho_0(u^{(i)})\, q(u' \mid u^{(i)}) = \rho_0(u')\, q(u^{(i)} \mid u').$$

- The acceptance probability then becomes

$$\alpha_h(u'|u^{(i)}) = \min\left(1, \frac{\exp[-\Phi_h(u')]}{\exp[-\Phi_h(u^{(i)})]}\right),$$

  which depends on u' only through its likelihood $\exp[-\Phi_h(u')]$.

- With prior reversible q, the MH-MCMC algorithm is well-defined in the infinite dimensional setting, i.e. in the limit as the dimension n of $u \in \mathbb{R}^n$ tends to infinity. In particular, the performance of MH-MCMC is robust with respect to n.

Metropolis Hastings Markov chain Monte Carlo (MH-MCMC)

Choice of proposal density q [Cotter, Dashti, Stuart '12]

- A choice of q that is prior reversible is the preconditioned Crank-Nicholson (pCN) proposal in [Cotter, Dashti, Stuart '12].

- The specific form of the pCN proposal depends on the prior. If μ_0 is $\mathcal{N}(0, C_0)$, then $q(u' \,|\, u^{(i)})$ is defined by

$$u' = \sqrt{1 - \beta^2}\, u^{(i)} + \beta \xi_i, \qquad \text{where } \xi_i \sim \mathcal{N}(0, C_0), \quad \beta > 0.$$

  β is a step size parameter that needs to be tuned.

- Using the additive properties of Gaussian random variables, we have

$$\xi_i, u^{(i)} \sim \mathcal{N}(0, C_0) \quad \Rightarrow \quad u' \sim \mathcal{N}(0, C_0).$$

Convergence of MH-MCMC

Distribution of the Markov chain [Robert, Casella '99]

What is the distribution of the samples $\{u^{(i)}\}_{i=1}^{\infty}$ produced by Algorithm 2?

Definition (Transition density and kernel)

The *transition density* of the Markov chain $\{u^{(i)}\}_{i=1}^{\infty}$ is

$$\Pr\big[u^{(i)} = u' \,|\, u^{(i-1)} = u\big] =: K(u'|u) =$$

$$\underbrace{\alpha_h(u'|u)q(u'|u)}_{\text{accept } u'} + \underbrace{\left(1 - \int_{\mathbb{R}^n} \alpha_h(u''|u)q(u''|u)\mathrm{d}u''\right)}_{\text{reject any other proposal}}\delta(u - u').$$

The corresponding *transition kernel* is $K(\cdot\,|\,u)$:

$$\Pr\big[u^{(i)} \in A \,|\, u^{(i-1)} = u\big] =: K(A\,|\,u) = \int_A K(u'\,|\,u)\mathrm{d}u', \qquad \forall A \in \mathcal{B}(\mathbb{R}^n).$$

Convergence of MH-MCMC

Existence of Stationary Distribution [Robert, Casella '99]

Definition (Stationary distribution)

A probability distribution π is a *stationary distribution* of the chain $\{u^{(i)}\}_{i=1}^{\infty}$ with transition kernel K if

$$\pi(A) = \int_{\mathbb{R}^n} K(A \,|\, u)\pi(\mathrm{d}u).$$

If π is a stationary distribution of $\{u^{(i)}\}_{i=1}^{\infty}$, then $u^{(i)} \sim \pi \Rightarrow u^{(i+1)} \sim \pi$.

Define the sets $\mathcal{E}$ and $\mathcal{D}$ by

$$\mathcal{E} := \{u \mid \rho_h^y(u) > 0\},$$
$$\mathcal{D} := \{u \mid q(u \,|\, u^*) > 0 \text{ for some } u^* \in \mathcal{E}\}.$$

Theorem (Stationary distribution of MH-MCMC)

If $\mathcal{E} \subseteq \mathcal{D}$, then μ_h^y is a stationary distribution of the chain $\{u^{(i)}\}_{i=1}^{\infty}$.

$\mathcal{E} \subseteq \mathcal{D}$ is sufficient for detailed balance of K with respect to μ_h^y.

Convergence of MH-MCMC

Convergence in total variation [Robert, Casella '99]

Definition (m-step transition kernel)

The m-step transition kernel K^m of $\{u^{(i)}\}_{i=1}^{\infty}$ is defined by $K^1 = K$ and

$$\Pr\left[u^{(m+1)} \in A \,|\, u^{(1)} = u\right] =: K^m(A\,|u) = \int_{\mathbb{R}^n} K^{m-1}(A\,|u'')K(\mathrm{d}y\,|u)\mathrm{d}u''.$$

Theorem (Convergence to stationary distribution)

Suppose $\Pr[\alpha_h = 1] < 1$ *and*

$$q(u\,|\,u^*) > 0 \quad \textit{for all} \quad u, u^* \in \mathcal{E} \times \mathcal{E}.$$

Then, as $m \to \infty$,

$$K^m(\cdot\,|\,u^{(1)}) \xrightarrow{TV} \mu_h^y. \qquad (u^m \sim \mu_h^y)$$

Here, $\xrightarrow{TV}$ denotes convergence in total variation:

$$\sup_{A \in \mathcal{B}(\mathbb{R}^n)} |K^m(A\,|\,u^{(1)}) - \mu_h^y(A)| \to 0.$$

$\Pr[\alpha_h = 1] < 1$ is sufficient for aperiodicity, $q(u\,|\,u^*) > 0$ is sufficient for irreducibility.

Convergence of MH-MCMC

Strong Law of Large Numbers [Robert, Casella '99]

Theorem (Strong Law of Large Numbers)

Suppose $\mathbb{E}_{\mu_h^y}[|\phi_h|] < \infty$, *and*

$$q(u \,|\, u^*) > 0 \text{ for all } u, u^* \in \mathcal{E} \times \mathcal{E}.$$

Then as $N \to \infty$, *with* $\{u^{(i)}\}_{i=1}^{\infty}$ *from Algorithm 2,*

$$\widehat{E}_{h,N}^{\mathrm{MCMC}} \xrightarrow{a.s.} \mathbb{E}_{\mu_h^y}[\phi_h].$$

Convergence of MH-MCMC

Central Limit Theorem [Robert, Casella '99]

Define an auxiliary chain $\{\tilde{u}^{(i)}\}_{i=1}^{\infty}$, generated by Algorithm 2 with $\tilde{u}^{(1)} \sim \mu_h^y$. The covariance structure is implicitly defined by Algorithm 2. Define the *asymptotic variance*

$$\sigma_\phi^2 := \mathbb{V}[\phi_h(\tilde{u}^{(1)})] + 2\sum_{i=2}^{\infty} \mathbb{C}\mathrm{ov}[\phi_h(\tilde{u}^{(1)}), \phi_h(\tilde{u}^{(i)})].$$

Theorem (Central Limit Theorem)

Suppose $\sigma_\phi^2 < \infty$, $\Pr[\alpha_h = 1] < 1$ *and*

$$q(u \,|\, u^*) > 0 \quad \textit{for all} \quad u, u^* \in \mathcal{E} \times \mathcal{E}.$$

Then, as $N \to \infty$, *we have*

$$\widehat{E}_{h,N}^{\mathrm{MCMC}} \xrightarrow{D} \mathcal{N}(\mathbb{E}_{\mu_h^y}[\phi_h], \frac{\sigma_\phi^2}{N}).$$

Note that the asymptotic variance σ_ϕ^2 is larger than in the i.i.d. case.

Convergence of MH-MCMC

Mean Square Error [Dodwell et al '15]

The mean square error of the MH-MCMC estimator $\widehat{E}_{h,N}^{\mathrm{MCMC}}$ is

$$e(\widehat{E}_{h,N}^{\mathrm{MCMC}})^2 = \mathbb{E}\Big[\big(\widehat{E}_{h,N}^{\mathrm{MCMC}} - \mathbb{E}_{\mu^y}(\phi)\big)^2\Big].$$

Theorem (Mean Square Error)

$$e(\widehat{E}_{h,N}^{\mathrm{MCMC}})^2 \leq$$

$$\underbrace{\mathbb{V}[\widehat{E}_{h,N}^{\mathrm{MCMC}}] + 2\big(\mathbb{E}[\widehat{E}_{h,N}^{\mathrm{MCMC}}] - \mathbb{E}_{\mu_h^y}[\widehat{E}_{h,N}^{\mathrm{MCMC}}]\big)^2}_{\textit{sampling error}} + \underbrace{2\big(\mathbb{E}_{\mu_h^y}[\phi_h] - \mathbb{E}_{\mu^y}[\phi]\big)^2}_{\textit{numerical error}}.$$

Proof: This follows directly from the linearity of expectation and the triangle inequality.

Convergence of MH-MCMC

Sampling Error [Rudolf '11]

For standard Monte Carlo estimators based on i.i.d. sampling, the sampling error is

$$\mathbb{V}[\widehat{E}^{\mathrm{MC}}_{h,N}] = \frac{\mathbb{V}[f_h]}{N}.$$

What can we say about the sampling error for MH-MCMC estimators?

Theorem (Non-asymptotic bound on sampling error)

Under certain regularity assumptions given in [Rudolf, '11], *we have*

$$\mathbb{V}[\widehat{E}^{\mathrm{MCMC}}_{h,N}] + 2\big(\mathbb{E}\big[\widehat{E}^{\mathrm{MCMC}}_{h,N}\big] - \mathbb{E}_{\mu_h^y}[\widehat{E}^{\mathrm{MCMC}}_{h,N}]\big)^2 \;\leq\; C\,\frac{\mathbb{E}[\phi_h^2]}{N},$$

for a constant C independent of N and h.

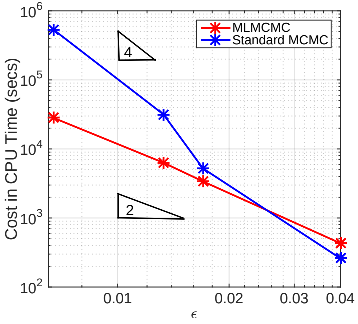
Multilevel MH-MCMC

Motivation and challenges

- The cost of standard MH-MCMC can be prohibitively large in practical applications.
- It is possible to extend the idea of multilevel Monte Carlo to the setting of MH-MCMC, see e.g. [Hoang, Schwab, Stuart '13], [Dodwell et al '15] and [Efendiev et al '15].
- Since multilevel methods use a sequence of step lengths $\{h_\ell\}_{\ell=0}^{L}$, and the step length changes the posterior distribution μ_h^y, the multilevel estimator has to be defined carefully.

Numerical Comparison: Mean Square Error

- We compute $\mathbb{E}_{\mu^y}[\phi]$ for a typical model problem in groundwater flow, using standard MH-MCMC and multilevel MH-MCMC estimators.
- Computational Cost is measured by CPU time.



[Dodwell, Ketelsen, Scheichl, Teckentrup '15]

Conclusions

- In the second part of this mini-course, we have looked at sampling methods to approximate expectations with respect to the posterior distribution μ^y.
- In the first lecture, we showed how we can use samples from the prior distribution to do this. In this context, it is usually possible to use sampling methods based on i.i.d. samples.
- In the second lecture, we considered using samples from the posterior distribution. In this case, it is generally not possible to generate i.i.d. samples, so we introduced the Metropolis Hastings algorithm.

References I

- P. Billingsley, *Probability and measure*, John Wiley & Sons, 1995.
- N. Collier, A.-L. Haji-Ali, F. Nobile, E. von Schwerin, and R. Tempone, *A continuation multilevel Monte Carlo algorithm*, BIT Numerical Mathematics, 55 (2015), pp. 399–432.
- S. Cotter, M. Dashti, and A. Stuart, *Variational data assimilation using targetted random walks*, International Journal for Numerical Methods in Fluids, 68 (2012), pp. 403–421.
- T. Dodwell, C. Ketelsen, R. Scheichl, and A. Teckentrup, *A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow*, SIAM/ASA Journal on Uncertainty Quantification, 3 (2015), pp. 1075–1108.

References II

Y. Efendiev, B. Jin, M. Presho, and X. Tan, *Multilevel Markov Chain Monte Carlo Method for High-Contrast Single-Phase Flow Problems*, Communications in Computational Physics, 17 (2015), pp. 259–286.

M. Giles, *Multilevel Monte Carlo Path Simulation*, Operations Research, 56 (2008), pp. 607–617.

S. Heinrich, *Multilevel Monte Carlo Methods*, in International Conference on Large-Scale Scientific Computing, Springer, 2001, pp. 58–67.

V. Hoang, C. Schwab, and A. Stuart, *Complexity analysis of accelerated MCMC methods for Bayesian inversion*, Inverse Problems, 29 (2013), p. 085010.

J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*, Springer, 2004.

References III

- P. L'Ecuyer, *Random number generation*, in Handbook of Computational Statistics, Springer, 2011, pp. 35–71.
- C. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, 1999.
- D. Rudolf, *Explicit error bounds for Markov chain Monte Carlo*, arXiv preprint arXiv:1108.3201, (2011).
- R. Scheichl, A. Stuart, and A. Teckentrup, *Quasi-Monte Carlo and Multilevel Monte Carlo Methods for Computing Posterior Expectations in Elliptic Inverse Problems*, arXiv preprint arXiv:1602.04704, (2016).
- A. Stuart, *Inverse Problems: A Bayesian Perspective*, Acta Numerica, 19 (2010), pp. 451–559.