Unbiased Estimation and Exact Simulation for Bayesian Inverse Problems

Sergios Agapiou

Department of Mathematics and Statistics, University of Cyprus

Joint work with G. Roberts, A. Stuart and S. Vollmer

Applied Inverse Problems 2017
29 May - 2 June 2017, Hangzhou

Outline

Outline

Bayesian Inverse Problems in Function Space

AIM: Estimate expectation of function of interest f wrt intractable measure μ, $\mathbb{E}_\mu[f(\cdot)]$.

- e.g. μ is the posterior arising in BIP.

$$y = \mathcal{G}(u) + \eta$$

- $u \in \mathcal{X}$ unknown, $\mathcal{X}$ function space.
- $y \in \mathbb{R}^J$ observation, $\mathcal{G} : \mathcal{X} \to \mathbb{R}^J$ forward operator, $\eta \sim N(0, \Gamma)$ observational noise.

Bayesian Formulation:

- Prior $u \sim \mu_0$
- Posterior $u|y \sim \mu^y$

$$\frac{d\mu^y}{d\mu_0}(u; y) \propto \exp\left(-\frac{1}{2}\|\Gamma^{-\frac{1}{2}}(y - \mathcal{G}(u))\|^2\right).$$

Problem Overview

- Approximations introduce bias:
  - construct Markov chain targeting μ^y, use finite-time distributions $\mu^{y,k} \Rightarrow$ burn-in time issues.
  - Markov chain targets approximation μ_i^y in $\mathbb{R}^i$, use finite-time distributions $\mu_i^{y,k} \Rightarrow$ discretization bias and burn-in time issues.
- Need to distribute resources between sources of approximation $\Rightarrow$ ε-error cost is suboptimal.

Unbiased Estimation and Exact Simulation

- Unbiased Estimation of $\mathbb{E}_\mu[f]$ using biased samples

  
  S. Agapiou, G. O. Roberts and S. J. Vollmer, *Unbiased Monte Carlo: posterior estimation for intractable/infinite dimensional models*, to appear in Bernoulli arXiv:1411.7713

  
  C. H. Rhee, *Unbiased estimation with biased samples*, PhD thesis, Stanford University, 2013 (supervisor P. W. Glynn).

- Exact Simulation from μ: ongoing work with G. Roberts and A. Stuart. Ideas from exact simulation of diffusions see

  
  K. Latuszynski, I. Kosmidis, O. Papaspiliopoulos, G. Roberts, *Simulating Events of Unknown Probabilities via Reverse Time Martingales*, Random Structures & Algorithms, 2011.

Outline

Unbiased Estimation Using Biased Samples

- We study unbiased estimation of $\mathbb{E}_\mu[f]$ using biased samples, $u_i \sim \mu_i$.
- Assume $\mathbb{E}_{\mu_i}[f] \xrightarrow{i} \mathbb{E}_\mu[f]$.
- Define $\Delta_i := f(u_i) - f(u_{i-1})$.
- If Fubini applies

$$\mathbb{E}_\mu[f] = \sum_{i=1}^{\infty}(\mathbb{E}_{\mu_i}[f] - \mathbb{E}_{\mu_{i-1}}[f]) = \sum_{i=1}^{\infty}\mathbb{E}\Delta_i \overset{?}{=} \mathbb{E}\sum_{i=1}^{\infty}\Delta_i.$$

- $\sum_{i=1}^{\infty}\Delta_i$ is unbiased but requires infinite computing time.

Debiasing Idea - John von Neumann, Stanislaw Ulam

$$Z := \sum_{i=0}^{N} \frac{\Delta_i}{\mathbb{P}(N \geq i)},$$

N integer-valued r.v. independent of Δ_i, s.t. $\mathbb{P}(N \geq i) > 0, \forall i$.

- If Fubini applies

$$\mathbb{E}[Z] = \mathbb{E}\left[\sum_{i=0}^{\infty} \frac{\mathbb{1}_{\{N \geq i\}} \Delta_i}{\mathbb{P}(N \geq i)}\right] \stackrel{?}{=} \sum_{i=0}^{\infty} \frac{\mathbb{E}[\mathbb{1}_{\{N \geq i\}} \Delta_i]}{\mathbb{P}(N \geq i)} = \sum_{i=0}^{\infty} \mathbb{E}\Delta_i = \mathbb{E}_\mu[f].$$

- Z unbiased and requires finite but random computing time.
- To be practical, Z needs to have finite variance and finite expected computing time.
- For $Z^{(m)} \overset{iid}{\sim} Z$, define $Z_M = \frac{1}{M}\sum_{m=1}^{M} Z^{(m)}$ achieving optimal ε-error cost.

Unbiasing Theory of Glynn and Rhee

Proposition (Glynn and Rhee 13)

Assume

$$\sum_{i \le \ell} \frac{\|\Delta_i\|_2 \|\Delta_\ell\|_2}{\mathbb{P}(N \ge i)} < \infty, \qquad \text{where } \|\Delta_i\|_2^2 = \mathbb{E}(|\Delta_i|^2).$$

Then Z is an UE for $\mathbb{E}_\mu[f]$ with finite variance. Can use mutually independent Δ_i.

Unbiasing Theory of Glynn and Rhee

Proposition (Glynn and Rhee 13)

Assume

$$\sum_{i\leq \ell} \frac{\|\Delta_i\|_2 \|\Delta_\ell\|_2}{\mathbb{P}(N \geq i)} < \infty, \qquad \text{where } \|\Delta_i\|_2^2 = \mathbb{E}(|\Delta_i|^2).$$

Then Z is an UE for $\mathbb{E}_\mu[f]$ with finite variance. Can use mutually independent Δ_i.

- t_i expected cost of generating Δ_i. Expected computing time of Z

$$\mathbb{E}(\tau) = \mathbb{E}\sum_{i=0}^{N} t_i = \mathbb{E}\sum_{i=1}^{\infty} t_i \mathbb{1}_{\{N\geq i\}} = \sum_{i=0}^{\infty} t_i \mathbb{P}(N \geq i).$$

- To be possible to choose $\mathbb{P}(N \geq i)$ s.t. Z practical, suffices to generate Δ_i's with correct expectation s.t. $\|\Delta_i\|_2^2$ decays sufficiently faster than t_i blows-up.
- Can optimize choice of N by minimizing $\mathbb{E}[\tau].\mathrm{Var}(Z)$.

Removing Function Space Discretization Bias

- $\mathcal{X} = L^2[0,1]$, $\{\varphi_\ell\}$ complete orthonormal basis.
- μ Gaussian measure in $\mathcal{X}$ given via the Karhunen-Loeve expansion:

$$\mu = \mathcal{L}\left(\sum_{\ell=1}^{\infty} \ell^{-a}\xi_\ell\varphi_\ell\right), \qquad \xi_\ell \overset{iid}{\sim} N(0,1), \quad a > \frac{1}{2}.$$

- To estimate $\mathbb{E}_\mu[f]$, need to truncate introducing discretization bias.

AIM: unbiasedly estimate $\mathbb{E}_\mu[f]$ in finite time for $f : \mathcal{X} \to \mathbb{R}$ Lipschitz.

- Approximations $\mu_i = \mathcal{L}\left(\sum_{\ell=1}^{j_i} \ell^{-a}\xi_\ell\varphi_\ell\right)$, j_i increasing.
- $\Delta_i = f(u_i) - f(u_{i-1})$, where $u_i \sim \mu_i$, $u_{i-1} \sim \mu_{i-1}$ with same randomness.

Removing Function Space Discretization Bias

Theorem 1 (A., Roberts and Vollmer '14)

Assume $a > 1$. Then $\exists$ choices j_i and $\mathbb{P}(N \geq i)$, s.t. $Z = \sum_{i=1}^{N} \frac{\Delta_i}{\mathbb{P}(N \geq i)}$ is unbiased estimator of $\mathbb{E}_\mu[f]$ with finite variance and finite expected computing time.

Proof.

- Consider $j_i = 2^i$. Use Proposition from GR13.
- $\|\Delta_i\|_2^2 = \mathbb{E}(|f(u_i) - f(u_{i-1})|^2) \leq \|f'\|_\infty^2 \, \mathbb{E}(\|u_i - u_{i-1}\|^2) = \mathcal{O}(j_{i-1}^{1-2a} - j_i^{1-2a}) = \mathcal{O}(2^{i(1-2a)})$.
- Cost of Δ_i, $t_i = \mathcal{O}(j_i) = \mathcal{O}(2^i)$ $\quad$ ($\#$ $N(0,1)$ draws).
- For $a > 1$, $\|\Delta_i\|_2^2$ decays sufficiently faster than t_i blows-up.
- Can choose $\mathbb{P}(N \geq i)$ s.t. $\mathbb{E}(\tau), \mathrm{Var}(Z) < \infty$.

□

Removing Burn-in Time Bias

- $\mathcal{X}$ finite-dim state space.
- μ cannot be sampled directly but can construct Markov chain $\mathbb{X} = (X_n)_{n\in\mathbb{N}}$ with transition kernel P and stationary distribution μ.
- e.g. μ posterior in BIP considered at a fixed discretization level.
- a_i increasing positive integers.
- To estimate $\mathbb{E}_\mu[f]$, use finite-time distributions $\mu_i = \mathcal{L}(X_{a_i})$ introducing burn-in issues.

AIM: unbiasedly estimate $\mathbb{E}_\mu[f]$ in finite time for $f : \mathcal{X} \to \mathbb{R}$ Lipschitz.

Removing burn-in time bias

- Weak convergence of μ_i not enough to get convergence of Δ_i in L_2.
- Contracting coupling assumption: we can simultaneously generate chains started at different states s.t. they come together geometrically quickly.
- Use top level chain $\mathcal{T}_\cdot^i$ running for a_i steps and bottom level chain $\mathcal{B}_\cdot^i$ running for a_{i-1} steps, coupled as follows:

$$\begin{array}{rccccccc} x_0 = & \mathcal{B}^i_{-a_{i-1}} & \dots & \mathcal{B}^i_{-a_0} & \dots & \mathcal{B}^i_0 & & \\ & | & & | & & | & \} \Delta_i = f(\mathcal{T}_0^i) - f(\mathcal{B}_0^i) & \\ x_0 = \mathcal{T}^i_{-a_i} & \dots & \mathcal{T}^i_{-a_{i-1}} & \dots & & \dots & \mathcal{T}^i_0 & \end{array}$$

Removing Burn-in Time Bias

- Can estimate

$$\|\Delta_i\|_2^2 \leq c\, r^{a_{i-1}}.$$

- Cost of Δ_i, $t_i = \mathcal{O}(a_i)$ ($\#$ steps).
- $\|\Delta_i\|_2^2$ decays exponentially in a_i while t_i increases linearly, hence can again show

Theorem 2 (A., Roberts and Vollmer '14)

$\exists$ choices a_i and $\mathbb{P}(N \geq i)$, s.t. $Z = \sum_{i=1}^{N} \frac{\Delta_i}{\mathbb{P}(R \geq i)}$ is UE of $\mathbb{E}_\mu[f]$ with finite variance and finite expected computing time.

UE for BIP in function space

- Combining can perform UE of $\mathbb{E}_\mu[f]$ for μ both ∞-dim and only accessible in the limit of a Markov chain.
- Approximation using finite-time distributions and discretizing space: top chain $\mathcal{T}_\cdot^i$ more steps and higher discretization level than bottom chain $\mathcal{B}_\cdot^i$

$$\begin{array}{rccccccccc} j_{i-1}: & & & x_0 = & \mathcal{B}^i_{-a_{i-1}} & \dots & \mathcal{B}^i_{-a_0} & \dots & \mathcal{B}^i_0 & \\ & & & & | & & | & & | & \} \Delta_i = f(\mathcal{T}_0^i) - f(\mathcal{B}_0^i) \\ j_i: & x_0 = & \mathcal{T}^i_{-a_i} & \dots & \mathcal{T}^i_{-a_{i-1}} & \dots & & \dots & \mathcal{T}^i_0 & \end{array}$$

- In ARV14, achieve this:
  1. in non-linear BIPs with uniform priors, using independence sampler;
  2. in non-linear BIPs with Gaussian priors, using pCN algorithm.

(MH with proposal $X_{k+1} = \lambda X_k + \sqrt{1-\lambda^2}\xi$)

Outline

Sampling the Posterior in Bayesian Inverse Problems

- Posterior in BIP

$$\frac{d\mu^y}{d\mu_0}(u;y) \propto \exp(-\frac{1}{2}\|\Gamma^{-\frac{1}{2}}(y-\mathcal{G}(u))\|^2),$$

  $\mathcal{G}$ often involves solving a PDE/ODE.

- Assume u is finite-dim, want to sample the finite-dim posterior μ^y.

- Need to construct Markov chain with μ^y as invariant distribution.

Metropolis-Hastings Algorithm

- Metropolis-Hastings: at state u propose $u' \sim Q(u'|u)$ and accept/reject based on carefully chosen acceptance probability $\alpha(u, u')$.
- e.g. for prior-reversible proposals

$$\alpha(u, u') = 1 \wedge \exp\left(\frac{1}{2}\|y - \mathcal{G}(u)\|_\Gamma^2 - \frac{1}{2}\|y - \mathcal{G}(u')\|_\Gamma^2\right)$$

- At each step of MH we evaluate acceptance probability, involving the solution of the PDE/ODE $\rightarrow$ discretize with mesh-size J to integrate.
- Limit of Markov chain $\mu_J^y \approx \mu^y$.

Standard Metropolis-Hastings Algorithm

- Fix discretization level J and build Markov chain with μ_J^y as limiting distribution.

Metropolis-Hastings algorithm

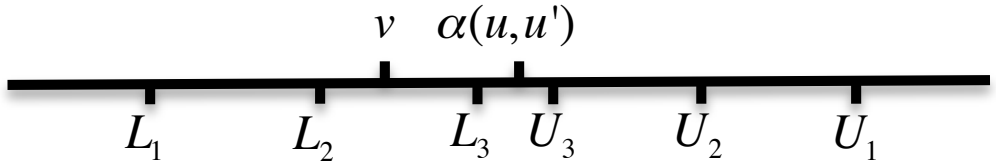
Pick u_0. For $k = 0, \ldots, K$ repeat:

1. Propose new state $u' \sim Q(u'|u_k)$;
2. Compute acceptance probability $\alpha_J(u_k, u')$;
3. Draw $v \sim U[0, 1]$.
   - If $v \leq \alpha_J(u_k, u')$ accept move, set $u_{k+1} = u'$;
   - Else reject, set $u_{k+1} = u_k$.

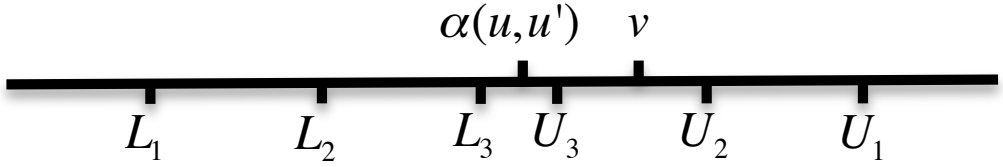
Exact Simulation Idea - Luc Devroye's Series Sampling

- Assume that for each (u, u') we can compute upper and lower bounds on $\alpha(u, u')$ depending on the discretization level J.
- Assume $J = 2^i$. We have two sequences $L_i \uparrow \alpha(u, u')$ and $U_i \downarrow \alpha(u, u')$.
- To accept/reject a proposed move, draw $v \sim U[0, 1]$ and determine if
  - $v \leq \alpha(u, u')$ hence accept;
  - $v > \alpha(u, u')$ hence reject.
- Can find this out by finding which sequence crosses v:

Accept



Reject



Exact Metropolis-Hastings Algorithm

- Can build Markov chain with μ^y (not μ_J^y) as limiting distribution.

Exact Metropolis-Hastings algorithm

Pick u_0. For $k = 0, \ldots, K$ repeat:

1. Propose new state $u' \sim Q(u'|u_k)$;
2. Draw $v \sim U[0,1]$;
3. Set $i = 0$, $L_0 = 0$ and $U_0 = 1$.
   - While $v \in [L_i, U_i]$ set $i = i + 1$ and compute bounds on $\alpha(u_k, u')$, L_i, U_i.
   - If $L_i > v$ accept move, set $u_{k+1} = u'$;
   - Else reject, set $u_{k+1} = u_k$.

Computing Time

- Computing time T until making a decision is random:

$$\mathbb{E}[T] = \sum_{i=1}^{\infty} t_i \mathbb{P}(T \geq t_i),$$

  where t_i cost of computing L_i, U_i and $\mathbb{P}(T \geq t_i) = |U_i - L_i|$.

- For algorithm to be practical need $\mathbb{E}[T] < \infty$.

- Computation becomes nontrivial if bounds depend on u whose randomness also needs to be taken into account.

- e.g. $\mathbb{E}[T] < \infty$ in 1D groundwater flow ip with uniform prior and FEM with hat functions.

Outline

Unbiased Estimation and Exact Simulation Often Feasible in BIP's.

Unbiased Estimation

- Fairly developed theory.
- In toy simulations competitive, need comparisons in problems of higher complexity.
- Optimization wrt parameters is crucial especially in function space setting.
- Easily parallelizable: a) use independent copies of Z, b) Δ_i's independent.

Exact Simulation

- Ongoing development of theory (conditions for finite computing time, user impatience bias, ...).
- Some FEM libraries come with error estimates which could be used in exact simulation in complex problems.
- Extend to simulation with no error due to discretization of the unknown.

http://www.sergiosagapiou.com/

- S. Agapiou, G. O. Roberts and S. J. Vollmer, *Unbiased Monte Carlo: posterior estimation for intractable/infinite dimensional models*, to appear in Bernoulli arXiv:1411.7713
- C. H. Rhee, *Unbiased estimation with biased samples*, PhD thesis, Stanford University, 2013, (supervisor P. W. Glynn).
- M. Pollock, *Some Monte Carlo Methods for Jump Diffucsions*, PhD thesis, University of Warwick, 2013 (supervisors A. Johansen and G. Roberts).
- J. G. Propp and D. B. Wilson, *Exact sampling with coupled Markov chains and applications to statistical mechanics*, Random Structures and Algorithms, 1996.
- M. Dashti and A. M. Stuart, *The Bayesian approach to inverse problems*, arXiv:1302.6989.
- M. Hairer, A. M. Stuart and S. J. Vollmer *Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions*, The Annals of Applied Probability, 2014.