Bayesian static parameter estimation using Multilevel Monte Carlo

Kody Law

joint with A. Jasra (NUS), K. Kamatani (Osaka), & Y. Zhou

MANCHESTER 1824
The University of Manchester

MCQMC 2018
Rennes, FR

July 4, 2018

Outline

Outline

Orientation

Aim: Approximate posterior expectations of the state path and static parameters associated to a discretely observed diffusion, which must be finitely approximated.

Solution: Apply an approximate coupling strategy so that the multilevel Monte Carlo (MLMC) method can be used within a particle marginal Metropolis-Hastings (PMMH) algorithm [B02, AR08, ADH10].

- MLMC methods *reduce cost to* error$= \mathcal{O}(\varepsilon)$ [H00, G08];
- Recently this methodology has been applied to *inference*, mostly in cases where target can be evaluated up to a normalizing constant [HSS13, DKST15, HTL16, BJLTZ17].
- Here it cannot, but using PMMH we are able to sample consistently from an approximate coupling of successive targets [JKLZ18].

Outline

Example: expectation for SDE [G08]

Estimation of expectation of solution of intractable stochastic differential equation (SDE).

$$dX = f(X)dt + \sigma(X)dW, \quad X_0 = x_0 \ .$$

Aim: estimate $\mathbb{E}(g(X_T))$.
We need to

(1) Approximate, e.g. by Euler-Maruyama method with resolution h:

$$X_{n+1} = X_n + hf(X_n) + \sqrt{h}\sigma(X_n)\xi_n, \quad \xi_n \sim N(0,1).$$

(2) Sample $\{X_{N_T}^{(i)}\}_{i=1}^{N}$, $N_T = T/h$.

Multilevel Monte Carlo [H00, G08, CGST11]

Assume $h_l = 2^{-l}$, X_l approximates X_T with stepsize h_l, and there are $\alpha, \beta, \zeta > 0$ such that

(i) weak error $|\mathbb{E}[g(X_l) - g(X)]| = \mathcal{O}(h_l^\alpha)$.

(ii) strong error $\mathbb{E}|g(X_l) - g(X)|^2 = \mathcal{O}(h_l^\beta) \Rightarrow V_l = \mathcal{O}(h_l^\beta)$,

(iii) computational cost for a realization of $g(X_l) - g(X_{l-1})$, $C_l \propto h_l^{-\zeta}$.

Then, it is possible to choose L and $\{N_l\}_{l=0}^L$ so as to achieve a mean squared error of $\mathcal{O}(\varepsilon^2)$ at a cost

$$\mathrm{COST} \leq C \begin{cases} \varepsilon^{-2}, & \text{if} \quad \beta > \gamma, \\ \varepsilon^{-2}|\log(\epsilon)|^2, & \text{if} \quad \beta = \gamma, \\ \varepsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)}, & \text{if} \quad \beta < \gamma. \end{cases}$$

Compared with $\mathcal{O}(\varepsilon^{-2-\gamma/\alpha})$ for single level.

Outline

Parameter inference

Estimate the posterior expectation of a function φ of the joint path $X_{1:T}$ and *parameters* θ, of an intractable SDE

$$dX = f_\theta(X)dt + \sigma_\theta(X)dW, \quad X_0 \sim \mu_\theta ,$$

given noisy partial observations

$$Y_n \sim g_\theta(X_n, \cdot) , \quad n = 1, \ldots, T .$$

Aim: estimate $\mathbb{E}[\varphi(\theta, X_{0:T})|y_{1:T}]$, where $y_{1:T} := \{y_1, \ldots, y_T\}$.

The hidden process $\{X_n\}$ is a Markov chain.

Discretize with resolution h and denote the transition kernel $F_{\theta,h}\left(x_{p-1}, \mathrm{d}x_p\right)$ – this can be *simulated from*, but its density cannot be *evaluated*.

Return to ML

The joint measure is

$$\Pi_h(\mathrm{d}\theta, \mathrm{d}x_{0:n}) \propto \Pi(\mathrm{d}\theta)\mu_\theta(\mathrm{d}x_0)\prod_{p=1}^{n} g_\theta(x_p, y_p)F_{\theta,h}(x_{p-1}, \mathrm{d}x_p) \ ,$$

For $+\infty > h_0 > \cdots > h_L > 0$, we would like to compute

$$\mathbb{E}_{\Pi_{h_L}}[\varphi(\theta, X_{0:n})] = \sum_{l=0}^{L}\left\{\mathbb{E}_{\Pi_{h_l}}[\varphi(\theta, X_{0:n})] - \mathbb{E}_{\Pi_{h_{l-1}}}[\varphi(\theta, X_{0:n})]\right\}$$

where $\mathbb{E}_{\Pi_{h_{-1}}}[\cdot] := 0$.

Outline

Approximate coupling

Consider a single pair $\mathbb{E}_{\Pi_h}[\varphi(\theta, X_{0:n})] - \mathbb{E}_{\Pi_{h'}}[\varphi(\theta, X_{0:n})]$, $h < h'$.

Let $z = (x, x')$ and let $Q_{\theta,h,h'}(z, \mathrm{d}\bar{z})$ be a coupling of $(F_{\theta,h}(x, \mathrm{d}\bar{x}), F_{\theta,h'}(x', \mathrm{d}\bar{x}'))$.

Let $G_{p,\theta}(z) = \max\{g_\theta(x, y_p), g_\theta(x', y_p)\}$.

We will sample from the joint coarse/fine filter

$$\Pi_{h,h'}(\mathrm{d}\theta, \mathrm{d}z_{0:n}) \propto \Pi(\mathrm{d}\theta)\nu_\theta(\mathrm{d}z_0) \prod_{p=1}^{n} G_{p,\theta}(z_p) Q_{\theta,h,h'}(z_{p-1}, \mathrm{d}z_p),$$

where ν_θ is the initial coupling

$$\nu_\theta(\mathrm{d}(x, x')) = \mu_\theta(\mathrm{d}x)\delta_x(\mathrm{d}x').$$

Change of measure

We have

$$\mathbb{E}_{\Pi_h}[\varphi(\theta, X_{0:n})] - \mathbb{E}_{\Pi_{h'}}[\varphi(\theta, X'_{0:n})] =$$

$$\frac{\mathbb{E}_{\Pi_{h,h'}}[\varphi(\theta, X_{0:n})H_{1,\theta}(\theta, Z_{0:n})]}{\mathbb{E}_{\Pi_{h,h'}}[H_{1,\theta}(\theta, Z_{0:n})]} - \frac{\mathbb{E}_{\Pi_{h,h'}}[\varphi(\theta, X'_{0:n})H_{2,\theta}(\theta, Z_{0:n})]}{\mathbb{E}_{\Pi_{h,h'}}[H_{2,\theta}(\theta, Z_{0:n})]}$$

where

$$\begin{aligned} H_{1,\theta}(\theta, z_{0:n}) &= \prod_{p=1}^{n} \frac{g_\theta(x_p, y_p)}{G_{p,\theta}(z_p)} \\ H_{2,\theta}(\theta, z_{0:n}) &= \prod_{p=1}^{n} \frac{g_\theta(x'_p, y_p)}{G_{p,\theta}(z_p)} . \end{aligned}$$

Outline

Particle filter, for fixed θ

Let $M \geq 1$ and θ be fixed, and introduce $a_{0:n-1} \in \{1, \ldots, M\}^n$. The bootstrap particle filter [Del04] approximates

$$\Pi_{h,h'}(\mathrm{d}z_{0:n}|\theta) \propto \nu_\theta(\mathrm{d}z_0) \prod_{p=1}^{n} G_{p,\theta}(z_p) Q_{\theta,h,h'}(z_{p-1}, \mathrm{d}z_p)$$

by sampling from

$$P(a_{0:n-1}^{1:M}, \mathrm{d}z_{0:n}^{1:M}|\theta) = \left(\prod_{i=1}^{M} \nu_\theta(\mathrm{d}z_0^i)\right) \prod_{p=1}^{n} \prod_{i=1}^{M} \left(\frac{G_{p-1,\theta}(z_{p-1}^{a_{p-1}^i})}{\sum_{j=1}^{M} G_{p-1,\theta}(z_{p-1}^j)} Q_{\theta,h,h'}(z_{p-1}^{a_{p-1}^i}, \mathrm{d}z_p^i) \right),$$

where $G_{0,\theta} := 1$,

i.e. $z_{p-1}^{a_{p-1}^i}$ is resampled with the probability $\frac{G_{p-1,\theta}(z_{p-1}^{a_{p-1}^i})}{\sum_{j=1}^{M} G_{p-1,\theta}(z_{p-1}^j)}$.

Unbiased estimator, for fixed θ

Draw J with probability proportional to $G_{n,\theta}(z_n^i)$ for $i = 1, \ldots, M$. Let $\widehat{z}_n = z_n^J$, and trace its ancestral lineage

$$\widehat{z}_{n-1} = z_{n-1}^{a_{n-1}^J}, \quad \widehat{z}_{n-2} = z_{n-2}^{a_{n-2}^{a_{n-1}^J}}, \quad \text{and so on}.$$

Define $p^M_{h,h'}(y_{0:n}|\theta) = \prod_{p=1}^n \left(\frac{1}{M} \sum_{j=1}^M G_{p,\theta}(z_p^j) \right)$, let $\mathbb{E}$ denote expectation w.r.t. $(A^{1:M}_{0:n-1}, Z^{1:M}_{0:n}, J)$, and note that [Del04]

$$\mathbb{E}\left[\varphi(\widehat{z}_{0:n}) p^M_{h,h'}(y_{0:n}|\theta)\right] = \int_{Z^{n+1}} \varphi(z_{0:n}) \nu_\theta(\mathrm{d}z_0) \prod_{p=1}^n G_{p,\theta}(z_p) Q_{\theta,h,h'}(z_{p-1}, \mathrm{d}z_p) .$$

Therefore, one can run a MH chain $\{\theta^i, \widehat{z}_{0:n}(\theta^i)\}$ targeting $\propto p^M_{h,h'}(y_{0:n}|\theta)\Pi(\mathrm{d}\theta)$, and it is consistent with respect to $\Pi_{h,h'}(\mathrm{d}z_{0:n}, \mathrm{d}\theta)$ [B03, AR08, ADH10].

Particle marginal MH (PMMH) [ADH10]

1. Sample $\theta^0 \sim \pi(\mathrm{d}\theta)$ and $(a_{0:n-1}^{1:M}, z_{0:n}^{1:M})$ from particle filter $P(\mathrm{d}a_{0:n-1}^{1:M}, \mathrm{d}z_{0:n}^{1:M}|\theta^0)$, and store $p_{h,h'}^M(y_{0:n}|\theta^0)$.
2. Select a path $\widehat{z}_{0:n}^0$: draw z_n^j with probability proportional to $G_{n,\theta^0}(z_n^j)$, let $\widehat{z}_n^0 = z_n^j$, and trace back its ancestral lineage
$$\widehat{z}_{n-1}^0 = z_{n-1}^{a_{n-1}^j}, \quad \widehat{z}_{n-2}^0 = z_{n-2}^{a_{n-2}^{a_{n-1}^j}}, \quad \text{and so on} ; \quad \text{Set } i = 1 .$$
3. Sample $\theta^*|\theta^{i-1}$ according to $R(\mathrm{d}\theta^*|\theta^{i-1}) = r(\theta^*|\theta^{i-1})\mathrm{d}\theta^*$, then sample from particle filter $P(\mathrm{d}a_{0:n-1}^{1:M}, \mathrm{d}z_{0:n}^{1:M}|\theta^*)$. Select one path $\widehat{z}_{0:n}^*$ as above.
4. Set $\theta^i = \theta^*$, $\widehat{z}_{0:n}^i = \widehat{z}_{0:n}^*$ with probability:
$$\min\left\{1, \frac{p_{h,h'}^M(y_{0:n}|\theta^*)}{p_{h,h'}^M(y_{0:n}|\theta^{i-1})} \frac{\pi(\theta^*)r(\theta^{i-1}|\theta^*)}{\pi(\theta^{i-1})r(\theta^*|\theta^{i-1})}\right\}$$
otherwise $\theta^i = \theta^{i-1}$, $\widehat{z}_{0:n}^i = \widehat{z}_{0:n}^{i-1}$.
5. Set $i = i + 1$ and return to the start of 3.

PMMH increment estimator

$$\frac{\frac{1}{N}\sum_{i=1}^{N}\varphi(\theta^i,\widehat{x}^i_{0:n})H_{1,\theta}(\theta^i,\widehat{z}^i_{0:n})}{\frac{1}{N}\sum_{i=1}^{N}H_{1,\theta}(\theta^i,\widehat{z}^i_{0:n})} - \frac{\frac{1}{N}\sum_{i=1}^{N}\varphi(\theta^i,\widehat{x}'^i_{0:n})H_{2,\theta}(\theta^i,\widehat{z}^i_{0:n})}{\frac{1}{N}\sum_{i=1}^{N}H_{2,\theta}(\theta^i,\widehat{z}^i_{0:n})}.$$

$$\overrightarrow{N\to\infty}$$

$$\frac{\mathbb{E}_{\Pi_{h,h'}}[\varphi(\theta,X_{0:n})H_{1,\theta}(\theta,Z_{0:n})]}{\mathbb{E}_{\Pi_{h,h'}}[H_{1,\theta}(\theta,Z_{0:n})]} - \frac{\mathbb{E}_{\Pi_{h,h'}}[\varphi(\theta,X'_{0:n})H_{2,\theta}(\theta,Z_{0:n})]}{\mathbb{E}_{\Pi_{h,h'}}[H_{2,\theta}(\theta,Z_{0:n})]}$$

$$=$$

$$\mathbb{E}_{\Pi_h}[\varphi(\theta,X_{0:n})] - \mathbb{E}_{\Pi_{h'}}[\varphi(\theta,X'_{0:n})]$$

Outline

Multilevel estimator

Consider

$$\sum_{l=0}^{L} \bar{E}_l^{N_l}(\varphi) \,, \qquad \bar{E}_l^{N_l}(\varphi) = E_l^{N_l}(\varphi) - E_l(\varphi) \,, \tag{1}$$

where

$$E_l^{N_l}(\varphi) = \frac{\frac{1}{N_l}\sum_{i=1}^{N_l} \varphi(\theta^i, , \widehat{x}_{0:n}^i) H_{1,\theta}(\theta^i, \widehat{z}_{0:n}^i)}{\frac{1}{N_l}\sum_{i=1}^{N_l} H_{1,\theta}(\theta^i, \widehat{z}_{0:n}^i)} - \frac{\frac{1}{N_l}\sum_{i=1}^{N_l} \varphi(\theta^i, \widehat{x}'^i_{0:n}) H_{2,\theta}(\theta^i, \widehat{z}_{0:n}^i)}{\frac{1}{N_l}\sum_{i=1}^{N_l} H_{2,\theta}(\theta^i, \widehat{z}_{0:n}^i)}$$

is a consistent estimator of

$$E_l(\varphi) := \mathbb{E}_{\Pi_{h_l}}[\varphi(\theta, X_{0:n})] - \mathbb{E}_{\Pi_{h_{l-1}}}[\varphi(\theta, X_{0:n})] \,.$$

$\Rightarrow$ (1) is a consistent estimator of $\mathbb{E}_{\Pi_{h_L}}[\varphi(\theta, X_{0:n})]$.

One must bound

$$\mathbb{E}[(\sum_{l=0}^{L} \bar{E}_l^{N_l}(\varphi))^2] = \sum_{l=0}^{L} \mathbb{E}[\bar{E}_l^{N_l}(\varphi)^2] \,.$$

Assumptions

(A1) $\forall\, y \in \mathsf{T}$, $\exists\, C > 0$ such that $\forall\, x \in \mathsf{S}$, $\theta \in \Theta$,

$$C \leq g_\theta(x, y) \leq C^{-1}.$$

And $\forall\, y \in \mathsf{T}$, $g_\theta(x, y)$ is globally Lipschitz on $\mathsf{S} \times \Theta$.

(A2) $\forall\, 0 \leq k \leq n$, $\exists\, \beta > 0$ such that $\forall$
$\varphi \in \mathcal{B}_b(\Theta \times \mathsf{S}^{k+1}) \cap \mathrm{Lip}(\Theta \times \mathsf{S}^{k+1})$ $\exists\, C > 0$

$$\int_{\Theta \times \mathsf{S}^{2k+2}} |\varphi(\theta, x_{0:k}) - \varphi(\theta, x'_{0:k})|^2 \Pi(\mathrm{d}\theta)\nu_\theta(\mathrm{d}z_0) \prod_{p=1}^{k} Q_{\theta,h,h'}(z_{p-1}, \mathrm{d}z_p) \leq C(h')^\beta.$$

(A3) Suppose that $\forall\, n > 0$, $\exists\, \xi \in (0, 1)$ and $\nu \in \mathcal{P}(\mathsf{W})$ such that for each $w \in \mathsf{W}$, $\varphi \in \mathcal{B}_b(\mathsf{W}) \cap \mathrm{Lip}(\mathsf{W})$, h, h':

$$\int_{\mathsf{W}} \varphi(w')K(w, dw') \geq \xi \int_{\mathsf{W}} \varphi(v)\nu(dv).$$

K is η-reversible, where η is the joint on the extended space.

Main result

Theorem (JKLZ18)

Assume (A1-3). Then $\forall\ n > 0$, $\exists\ \beta > 0$ *such that* $\forall$ $\varphi \in \mathcal{B}_b(\Theta \times \mathsf{S}^{n+1}) \cap \mathrm{Lip}(\Theta \times \mathsf{S}^{n+1})$ $\exists\ C > 0$ *such that*

$$\mathbb{E}[\bar{E}_l^{N_l}(\varphi)^2] \leq \frac{C h_l^{\beta}}{N_l},$$

and β *is from (A2).*

(A4) $\exists\, \gamma, \alpha, C > 0$ such that the cost to simulate $E_l^{N_l}$ is controlled by $\mathrm{C}(E_l^{N_l}) \leq C N_l h_l^{-\gamma}$, and the bias is controlled by

$$|\mathbb{E}_{\Pi_{h_L}}(\varphi(\theta, X_{0:n})) - \mathbb{E}_{\Pi_0}(\varphi(\theta, X_{0:n}))| \leq C h_L^{\alpha} \, .$$

Corollary

Assume (A1-4). $\forall\, n > 0$ *and* $\varphi \in \mathcal{B}_b(\Theta \times \mathsf{S}^{n+1}) \cap \mathrm{Lip}(\Theta \times \mathsf{S}^{n+1})$ $\exists$ $C > 0$ *such that* $\forall\, \epsilon > 0$ *one can choose* $(L, \{N_l\}_{l=0}^{L})$ *so*

$$\mathbb{E}\left[\left| \sum_{l=0}^{L} E_l^{N_l}(\varphi) - \mathbb{E}_{\Pi_0}(\varphi(\theta, X_{0:n})) \right|^2 \right] \leq C\epsilon^2 \, ,$$

with a total cost (per time step)

$$\mathrm{COST} \leq C \begin{cases} \epsilon^{-2}, & \text{if} \quad \beta > \gamma, \\ \epsilon^{-2}|\log(\epsilon)|^2, & \text{if} \quad \beta = \gamma, \\ \epsilon^{-\left(2 + \frac{\gamma - \beta}{\alpha}\right)}, & \text{if} \quad \beta < \gamma. \end{cases}$$

Outline

Ornstein-Uhlenbeck process

$$
\begin{aligned}
\mathrm{d}X_t &= \theta(\mu - X_t)\mathrm{d}t + \sigma \mathrm{d}W_t\,, \qquad X_0 = x_0\,,\\
Y_k | X_{\delta k} &\sim \mathcal{N}(X_{\delta k}, \tau^2)\,,\\
\theta &\sim \mathcal{G}(1,1)\,,\\
\sigma &\sim \mathcal{G}(1,0.5)\,.
\end{aligned}
$$

- $\mathcal{N}(m, \tau^2)$ denotes the Normal with mean m and variance τ^2.
- $\mathcal{G}(a, b)$ denotes the Gamma with shape a and scale b.
- $x_0 = 0$, $\mu = 0$, $\delta = 0.5$, and $\tau^2 = 0.2$.
- 100 observations simulated with $\theta = 1$ and $\sigma = 0.5$.

Langevin SDE

$$\mathrm{d}X_t = \frac{1}{2}\nabla \log \pi(X_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \qquad X_0 = x_0$$
$$Y_k | X_k \sim \mathcal{N}(0, \tau^2 \exp X_k)\,,$$
$$\theta \sim \mathcal{G}(1, 1)\,,$$
$$\sigma \sim \mathcal{G}(1, 0.5)\,.$$

- $\pi(x)$ denotes the probability density function of a Student's t-distribution with θ degrees of freedom.
- $x_0 = 0$.
- 1,000 observations simulated with $\theta = 10$, $\sigma = 1$, and $\tau^2 = 1$.

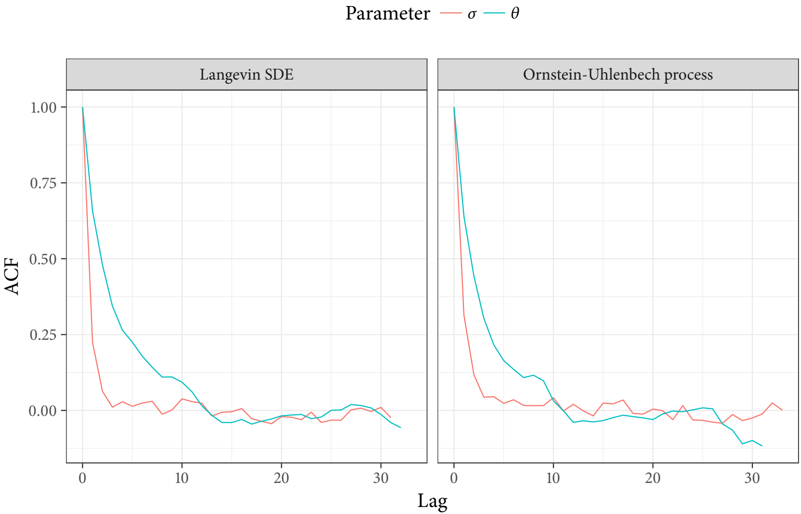

Figure: Autocorrelation of a typical PMCMC chain.

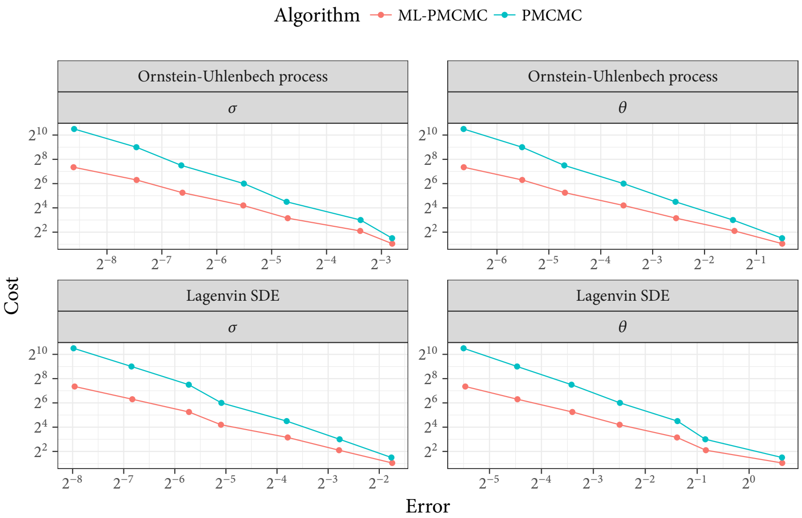

Figure: Cost vs. MSE for the 2 parameters for each of the 2 SDEs.

Model	Parameter	ML-PMCMC	PMCMC
OU	θ	-1.022	-1.463
	σ	-1.065	-1.522
Langevin	θ	-1.060	-1.508
	σ	-1.023	-1.481

Table: Estimated rates of convergence of MSE with respect to cost for various parameters, fitted to the curves.

Outline

Summary

- New approximate coupling strategy can be used to apply MLMC to PMCMC for static parameter estimation [JKLZ18.i].
- Same strategy can be employed for multi-index MCMC [JKLZ18.ii].
- In progress: $MISMC^2$ [JLX18.s]
- PhD and postdocs wanted – please inquire !

References

- **[G08]**: Giles. "Multilevel Monte Carlo path simulation." Op. Res., 56, 607-617 (2008).
- **[H00]** Heinrich. "Multilevel Monte Carlo methods." LSSC proceedings (2001).
- **[CGST11]** Cliffe, Giles, Scheichl, & Teckentrup. "MLMC and applications to elliptic PDEs." Computing and Visualization in Science, 14(1), 3 (2011).
- **[D04]**: Del Moral. "Feynman-Kac Formulae." Springer: New York (2004).
- **[B02]** Beaumont. "Estimation of population growth..." Genetics 164(3) 1139-1160 (2003).
- **[AR08]** Andrieu & Roberts. "The pseudo-marginal approach." Annals of Stat. 37(2) 697-725 (2009).
- **[ADH10]** Andrieu, Doucet, and Holenstein. "Particle MCMC methods." JRSSB 72(3) 269-342 (2010).

References

- **[JKLZ18.i]**: Jasra, Kamatani, Law, Zhou. "MLMC for static Bayesian parameter estimation." SISC 40, A887-A902 (2018).
- **[JKLZ18.ii]**: Jasra, Kamatani, Law, Zhou. "A Multi-Index Markov Chain Monte Carlo Method." IJUQ 8(1), 61-73 (2018).
- **[JLX18.s]**: Jasra, Law, Xu. "Multi-index SMC^2." Submitted.
- **[BJLTZ15]**: Beskos, Jasra, Law, Tempone, Zhou. "Multilevel Sequential Monte Carlo samplers." SPA 127:5, 1417–1440 (2017).
- **[HSS13]**: Hoang, Schwab, Stuart. Inverse Prob., 29, 085010 (2013).
- **[DKST13]**: Dodwell, Ketelsen, Scheichl, Teckentrup. "A hierarchical MLMCMC algorithm." SIAM/ASA JUQ 3(1) 1075-1108 (2015).

References

- **[CHLNT17]** Chernov, Hoel, Law, Nobile, Tempone. "Multilevel ensemble Kalman filtering for spatio-temporal processes." arXiv:1608.08558 (2017).
- **[JLS17]** Jasra, Law, Suciu. "Advanced Multilevel Monte Carlo Methods." arXiv:1704.07272 (2017).
- **[DJLZ16]**: Del Moral, Jasra, Law, Zhou. "Multilevel Sequential Monte Carlo samplers for normalizing constants." ToMACS 27(3), 20 (2017).
- **[JKLZ15]**: Jasra, Kamatani, Law, Zhou. "Multilevel particle filter." SINUM 55(6), 3068–3096 (2017).
- **[JLZ16]**: Jasra, Law, and Zhou. "Forward and Inverse UQ MLMC Algorithms for an Elliptic Nonlocal Equation." IJUQ 6(6), 501–514 (2016).
- **[HLT15]**: Hoel, Law, Tempone. "Multilevel ensemble Kalman filter." SINUM 54(3), 1813–1839 (2016).

Thank you