Correlated Particle Metropolis

Michael Pitt
King's College London
Joint work George Deligiannidis & Arnaud Doucet (Oxford)

Imperial College, May 2018

Organisation of the Talk

- Latent variable models
- The pseudo-marginal method
- The correlated pseudo-marginal method
- Illustrations

Sequential Monte Carlo Estimator

- Assume $\{X_t\}_{t\geq 1}$ is a latent Markov process, i.e. $X_1 \sim \mu_\theta(\cdot)$ and

$$X_{t+1}|\left(X_t = x\right) \sim f_\theta\left(\cdot|\, x\right), \quad Y_t|\left(X_t = x\right) \sim g_\theta\left(\cdot|\, x\right).$$

- Observations $\{Y_t\}_{t\geq 1}$ are conditionally independent given $\{X_t\}_{t\geq 0}$.
- Likelihood of $y_{1:T} = (y_1, ..., y_T)$ is

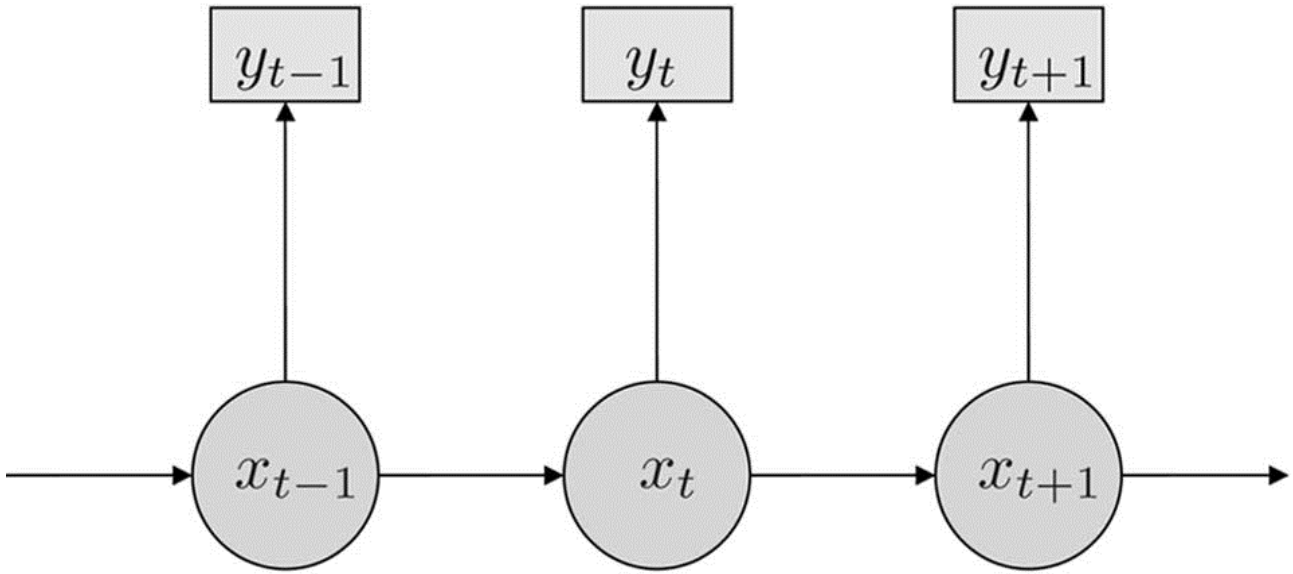
$$p(y_{1:T};\theta) = \int_{\mathbb{X}^{T+1}} p(x_{0:T}, y_{1:T};\theta)\, dx_{0:T}.$$

Sequential Monte Carlo Estimator

- SMC provides an unbiased estimator of relative variance $\mathcal{O}\left(T/N\right)$ where N is the number of particles.
- Whatever being $N \geq 1$, the pseudo-marginal MH admits $\pi\left(\theta\right)$ as invariant distribution.

A directed acyclic graph (DAG) of the problem:

$$y_t \sim g(y_t|x_t)$$



$$x_t \sim f(x_t|x_{t-1})$$

SMC: Likelihood Estimation

- Both $g(y_t|x_t)$, $f(x_t|x_{t-1})$ may be indexed by fixed parameters θ.
- Filtering density
$$p(x_t|y_{1:t};\theta).$$
- As an important byproduct we also obtain the one step predictive density
$$p(y_t|y_{1:t-1};\theta).$$
- Yields the likelihood (KF)
$$p(y_{1:T}|\theta) = p(y_1|\theta)\prod_{t=1}^{T-1} p(y_{t+1}|y_{1:t};\theta).$$

Particle Filter Estimation

- Simulation based methods to perform filtering in nonlinear/ non-Gaussian state space models.
- See Gordon, Salmond and Smith (1993) (GSS), Kitagawa (1996), Pitt and Shephard (1999) and reviewed by Doucet et al. (2000).
- We aim to have 'particles', $x_t^1, ..., x_t^N$ with associated discrete probability masses $\pi_t^1, ..., \pi_t^N$, drawn from the density $f(x_t|y_{1:t})$.

Bootrap Filter: GSS (1993, IEE)

We start at $t = 0$ *with samples from* $x_0^k \sim p(x_0)$. *For t=1,..,T:*
We have samples $x_t^k \sim p(x_t|y_{1:t})$ *for* $k = 1, ..., N$.

1. *For* $k = 1 : N$, *sample* $\widetilde{x}_{t+1}^k \sim f(x_{t+1}|x_t^k)$.
2. *For* $k = 1 : N$,
$$\pi_{t+1}^k = \frac{g(y_{t+1}|\widetilde{x}_{t+1}^k)}{\sum_{i=1}^N g(y_{t+1}|\widetilde{x}_{t+1}^i)}.$$
3. *For* $j = 1 : N$, *sample* $x_{t+1}^j \sim \sum_{k=1}^N \pi_{t+1}^k \delta(x_{t+1}^j - \widetilde{x}_{t+1}^k)$.

Step 3 multinomial (or stratified) sampling *(from the mixture).*
This will yield an *approximate sample* the desired posterior density, $f(x_t|y_{1:t})$ as t varies.

SMC Likelihood Estimation

- Parameter estimation using likelihood function, via prediction decomposition given by;

$$\log L(\theta) = \log p(y_{1,\ldots,.}y_T|\theta) = \sum_{t=1}^{T} \log p(y_{t+1}|\theta; y_{1:t}).$$

- We need to estimate the function :

$$\widehat{p}(y_{1:T}|\theta) = \widehat{p}(y_1|\theta) \prod_{t=1}^{T-1} \widehat{p}(y_{t+1}|y_{1:t};\theta),$$

$$\widehat{p}(y_{t+1}|\theta; y_{1:t}) = \frac{1}{N} \sum_{i=1}^{N} p(y_{t+1}|\widetilde{x}_{t+1}^{i}).$$

where $\widetilde{x}_{t+1}^{i} \sim f(x_{t+1}|y_{1:t};\theta)$, from step (2).

- Remarkably (just like in IS) the lik estimator $\widehat{p}(y_{1:T}|\theta)$ is unbiassed for $p(y_{1:T}|\theta)$ regardless of N, Del Moral (04).

Bayesian Inference

- Likelihood function $p(y;\theta)$ where $\theta \in \Theta \subseteq \mathbb{R}^d$.
- Prior distribution of density $p(\theta)$.
- Bayesian inference relies on the posterior

$$\pi(\theta) = p(\theta|\, y) = \frac{p(y;\theta)\, p(\theta)}{\int_\Theta p\left(y;\theta'\right) p\left(\theta'\right) d\theta'}.$$

- For non-trivial models, inference relies typically on MCMC.

Particle Metropolis: Intractable Likelihood Function

- In numerous scenarios, $p(y;\theta)$ cannot be evaluated pointwise; e.g.

$$p(y;\theta) = \int p(x, y;\theta)\, dx$$

  where the integral cannot be evaluated.
- A standard "solution" consists of using MCMC to sample from

$$p(\theta, x|\, y) = \frac{p(x, y;\theta)\, p(\theta)}{p(y)}$$

Standard MCMC Approaches

- Standard MCMC schemes target

$$p\left(\theta, x_{1:T} \middle| y_{1:T}\right) \propto p\left(\theta\right) p_\theta\left(x_{1:T}, y_{1:T}\right)$$

using Gibbs type strategy; i.e. sample alternately $X_{1:T} \sim p_\theta\left(\cdot \middle| y_{1:T}\right)$ and $\theta \sim p\left(\cdot \middle| y_{1:T}, X_{1:T}\right)$.

Standard MCMC Approaches

- Standard MCMC schemes target

$$p\left(\theta, x_{1:T} \middle| y_{1:T}\right) \propto p\left(\theta\right) p_\theta\left(x_{1:T}, y_{1:T}\right)$$

  using Gibbs type strategy; i.e. sample alternately $X_{1:T} \sim p_\theta\left(\cdot \middle| y_{1:T}\right)$ and $\theta \sim p\left(\cdot \middle| y_{1:T}, X_{1:T}\right)$.
- **Problem** 1: It can be difficult to sample $p_\theta\left(x_{1:T} \middle| y_{1:T}\right)$; e.g. non-Gaussian state-space models.

Standard MCMC Approaches

- Standard MCMC schemes target

$$p\left(\theta, x_{1:T} \middle| y_{1:T}\right) \propto p\left(\theta\right) p_\theta\left(x_{1:T}, y_{1:T}\right)$$

  using Gibbs type strategy; i.e. sample alternately $X_{1:T} \sim p_\theta\left(\cdot \middle| y_{1:T}\right)$ and $\theta \sim p\left(\cdot \middle| y_{1:T}, X_{1:T}\right)$.
- **Problem** 1: It can be difficult to sample $p_\theta\left(x_{1:T} \middle| y_{1:T}\right)$; e.g. non-Gaussian state-space models.
- **Problem** 2: Even when it is implementable, Gibbs can converge very slowly e.g. diffusions

Standard MCMC Approaches

- Standard MCMC schemes target

$$p\left(\theta, x_{1:T} \middle| y_{1:T}\right) \propto p\left(\theta\right) p_{\theta}\left(x_{1:T}, y_{1:T}\right)$$

  using Gibbs type strategy; i.e. sample alternately $X_{1:T} \sim p_{\theta}\left(\cdot \middle| y_{1:T}\right)$ and $\theta \sim p\left(\cdot \middle| y_{1:T}, X_{1:T}\right)$.
- **Problem** 1: It can be difficult to sample $p_{\theta}\left(x_{1:T} \middle| y_{1:T}\right)$; e.g. non-Gaussian state-space models.
- **Problem** 2: Even when it is implementable, Gibbs can converge very slowly e.g. diffusions
- **Problem** 3: It may only be possible to generate from $f\left(x_t \middle| x_{t-1}\right)$ not to evaluate it, e.g. DSGE models.

Standard MCMC Approaches

- Standard MCMC schemes target

$$p\left(\theta, x_{1:T} \middle| y_{1:T}\right) \propto p\left(\theta\right) p_{\theta}\left(x_{1:T}, y_{1:T}\right)$$

  using Gibbs type strategy; i.e. sample alternately $X_{1:T} \sim p_{\theta}\left(\cdot \middle| y_{1:T}\right)$ and $\theta \sim p\left(\cdot \middle| y_{1:T}, X_{1:T}\right)$.
- **Problem** 1: It can be difficult to sample $p_{\theta}\left(x_{1:T} \middle| y_{1:T}\right)$; e.g. non-Gaussian state-space models.
- **Problem** 2: Even when it is implementable, Gibbs can converge very slowly e.g. diffusions
- **Problem** 3: It may only be possible to generate from $f\left(x_t \middle| x_{t-1}\right)$ not to evaluate it, e.g. DSGE models.
- Pseudo-marginal methods mimic an algorithm targetting directly $\pi(\theta) = p\left(\theta \middle| y_{1:T}\right)$ instead of $p\left(\theta, x_{1:T} \middle| y_{1:T}\right)$.

Pseudo Metropolis (PM)

- Let $\widehat{p}(y;\theta,U)$ be an *unbiased non-negative estimator of the likelihood* where $U \sim m_\theta(\cdot)$; i.e.

$$p(y;\theta) = \int_{\mathbb{U}} \widehat{p}(y;\theta,u)\, m_\theta(u)\, du.$$

- Introduce a target distribution on $\Theta \times \mathbb{U}$ of density

$$\overline{\pi}(\theta,u) = \pi(\theta)\frac{\widehat{p}(y;\theta,u)}{p(y;\theta)} m_\theta(u) = \frac{p(\theta)\,\widehat{p}(y;\theta,u)\, m_\theta(u)}{p(y)}$$

- Then unbiasedness yields

$$\int_{\mathbb{U}} \overline{\pi}(\theta,u)\, du = \pi(\theta)$$

Any MCMC algorithm sampling from $\overline{\pi}(\theta,u)$ yields samples from $\pi(\theta)$.

Pseudo-Marginal Metropolis-Hastings algorithm

- Can form an unbiased estimator, based on N particles $\widehat{p}(y;\theta,U)$.
- Set $\left(\vartheta^{(0)},U^{(0)}\right)$ and iterate for $j=1,2,...$

Sample $\vartheta\sim q\left(\cdot|\,\vartheta^{(j-1)}\right)$, $U\sim m_{\vartheta}(\cdot)$ to obtain $\widehat{p}(y;\vartheta,U)$.

1. Compute

$$\alpha=1\wedge\frac{\widehat{p}(y;\vartheta,U)}{\widehat{p}\left(y;\vartheta^{(j-1)},U^{(j-1)}\right)}\frac{p(\vartheta)}{p\left(\vartheta^{(j-1)}\right)}\frac{q\left(\vartheta^{(j-1)}\middle|\,\vartheta\right)}{q\left(\vartheta|\,\vartheta^{(j-1)}\right)}$$

2. With proba α, set $\left(\vartheta^{(j)},U^{(j)}\right):=(\vartheta,U)$ and stay where you are otherwise.

A Nonlinear State-Space Model

- Standard non-linear model

$$X_t = \frac{1}{2}X_{t-1} + 25\frac{X_{t-1}}{1+X_{t-1}^2} + 8\cos(1.2t) + V_t, \quad V_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma_V^2\right),$$

$$Y_t = \frac{1}{20}X_t^2 + W_t, \quad W_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma_W^2\right).$$

- $T = 200$ data points with $\theta = \left(\sigma_V^2, \sigma_W^2\right) = (10, 10)$.
- Difficult to perform standard MCMC as $p\left(x_{1:T} \middle| y_{1:T}, \theta\right)$ is highly multimodal.
- We sample from $p\left(\theta \middle| y_{1:T}\right)$ using a random walk pseudo-marginal MH where $p\left(y_{1:T}; \theta\right)$ is estimated using SMC with N particles.

A Nonlinear State-Space Model

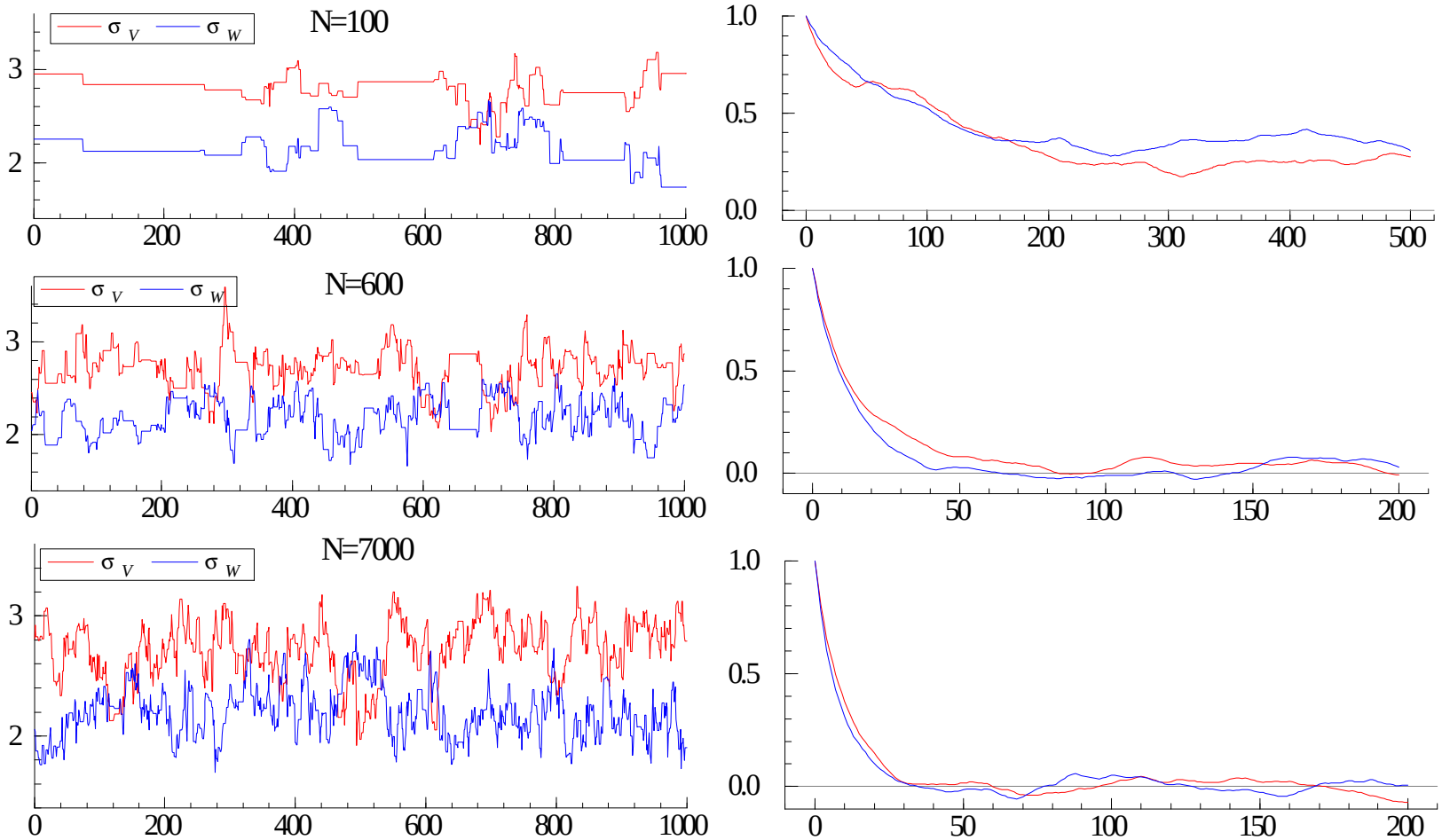


Figure: Autocorrelation of $\left\{\sigma_V^{(i)}\right\}$ and $\left\{\sigma_W^{(i)}\right\}$ of the MH sampler for various N.

How to Select the Number of Samples

- If N is too small, then the algorithm mixes poorly and will require many MCMC iterations.
- If N is too large, then each iteration is expensive due to estimating the likelihood.
- Equivalently we will examine σ^2, the variance of the estimator of the log-likelihood to trade off these two concerns.
- **Simple Guideline**: We find the optimal value for σ around 1. *JoE* (2012, Kohn, Giordani, Silva), *Biometrika* (2015, Doucet, Deligiannidis, Kohn).

Pseudo-Marginal Metropolis-Hastings algorithm

- Consider the error in the log-likelihood estimator

$$Z = \log \widehat{p}\,(y; \theta, U) - \log p(y; \theta) \sim g_\theta(\cdot)$$

- In the (θ, Z) parameterization, the target is

$$\overline{\pi}\,(\theta, u) = \pi(\theta) \frac{\widehat{p}\,(y; \theta, u)}{p\,(y; \theta)} m_\theta\,(u) \Rightarrow \overline{\pi}(\theta, z) = \pi(\theta) \exp(z) g_\theta(z).$$

How precise should the log-likelihood estimator be?

- **Aim**: Minimize the "computational time"

$$\mathrm{CT}\left(Q, h\right) = \mathrm{IACT}\left(Q, h\right) / \sigma^2$$

as $\sigma^2 \propto 1/N$ and computational efforts proportional to N

How precise should the log-likelihood estimator be?

- **Aim**: Minimize the "computational time"

$$\mathrm{CT}\left(Q, h\right) = \mathrm{IACT}\left(Q, h\right) / \sigma^2$$

  as $\sigma^2 \propto 1/N$ and computational efforts proportional to N

- where

$$\mathrm{IACT}\left(Q, h\right) \;=\; \text{Integrated Autocorrelation Time of } \left\{h\left(\vartheta_i\right)\right\}_{i \geq 1}$$

How precise should the log-likelihood estimator be?

- The IACT is

$$\mathrm{IACT}\left(Q, h\right) = 1 + 2\sum_{\tau=1}^{\infty} \mathrm{corr}_{\overline{\pi},Q}\left\{h\left(\theta_0\right), h\left(\theta_\tau\right)\right\}$$

- where Q is the pseudo-marginal kernel with acceptance criterion

$$\min\left\{1, r\left(\theta, \vartheta\right) \exp\left(w - z\right)\right\}.$$

- where $r\left(\theta, \vartheta\right)$ is exact Metropolis ratio,
- and

$$z = \log\{\widehat{p}_\theta(y_{1:T}) / p_\theta(y_{1:T})\}, \qquad w = \log\{\widehat{p}_\vartheta(y_{1:T}) / p_\vartheta(y_{1:T})\}.$$

Aim of the Analysis

- **Simplifying Assumption**: The noise Z is independent of θ and Gaussian; i.e. $g\left(z|\sigma\right) = \mathcal{N}\left(z; -\sigma^2/2; \sigma^2\right)$:

$$\overline{\pi}\left(\theta, z\right) = \pi\left(\theta\right) \underbrace{\exp\left(z\right) g\left(z\right)}_{\pi_Z\left(z|\sigma\right)} = \pi\left(\theta\right) \mathcal{N}\left(z; \sigma^2/2; \sigma^2\right).$$

- Justified empirically and theoretically (a CLT and concentration).

Pseudo-proof:

IID data, fixed θ (IS)

$$Z = \log \widehat{p}(y|\theta) - \log p(y|\theta) = \sum_{t=1}^{T} \log \left\{ 1 + \frac{\gamma_t}{\sqrt{N}} \varepsilon_t \right\},$$

where ε_t white noise, variance 1 and $N = \beta T$,

$$\simeq \sum_{t=1}^{T} \frac{\gamma_t}{\sqrt{\beta T}} \varepsilon_t - \frac{1}{2} \frac{\gamma_t^2}{\beta T} \varepsilon_t^2 \longrightarrow \mathcal{N}\left(-\frac{1}{2} \frac{\overline{\gamma}^2}{\beta}, \frac{\overline{\gamma}^2}{\beta} \right),$$

where $\overline{\gamma}^2 = T^{-1} \sum_{t=1}^{T} \gamma_t^2$.

- More detail (and more generality, PFs) in Bérard, Del Moral, Doucet (EJP, 2014).
- Need concentration (in θ as $T \longrightarrow \infty$), unproven.

Simulation:

SSF (from JoE(2012, Kohn, Giordani, Silva))

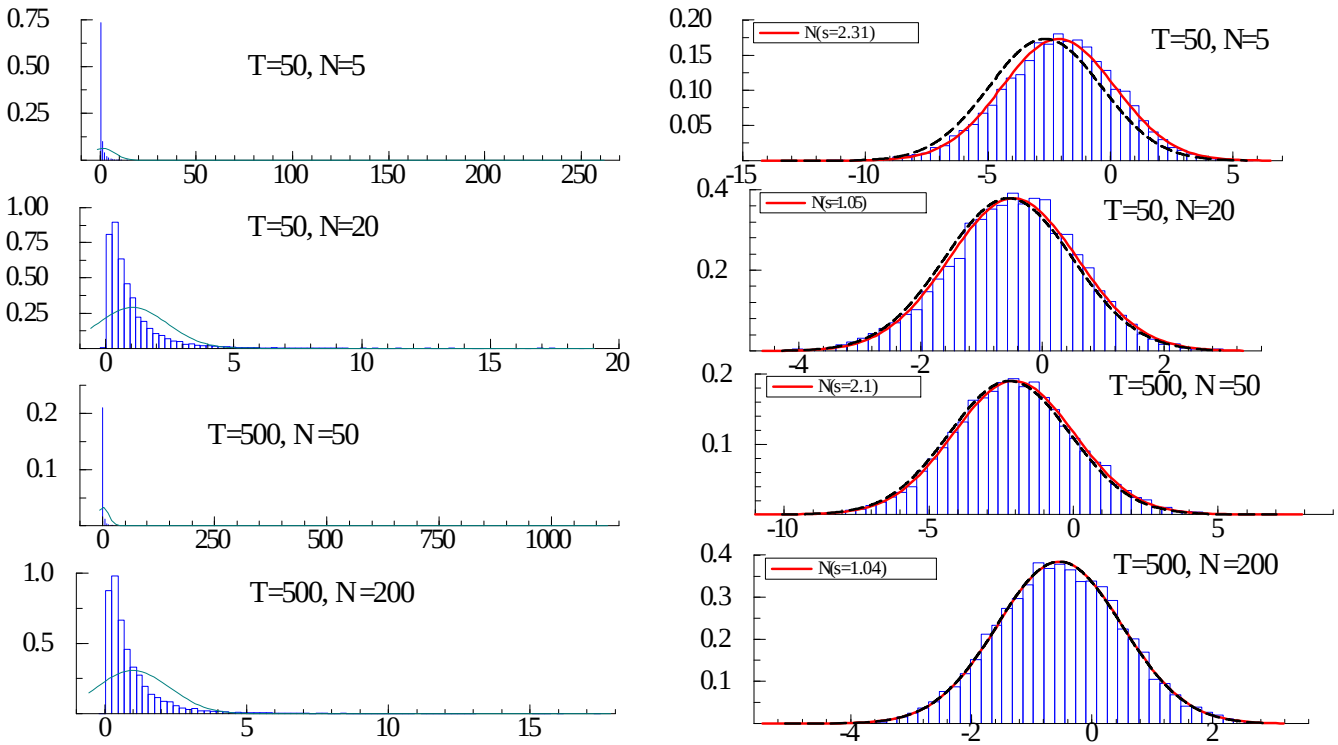


Figure: AR(1) plus noise model with fixed parameters. replications is $10,000$. SIR likelihood estimator (divided by the true likelihood) on LEFT and for the error in the log of the SIR likelihood estimator on RIGHT. Both N and T vary

Special cases

- When $q(\vartheta|\theta) = \pi(\vartheta)$, $\sigma_{\text{opt}} = 0.92$ (Pitt et al., JoE 2012). The acf simply reduces to:

$$\phi_n(\theta, Q) = \int \Pr(R \mid z; \sigma)^n \pi_Z(z|\sigma)\, dz$$

$$IACT(\sigma) = \int \frac{1 + \Pr(R \mid z; \sigma)}{1 - \Pr(R \mid z; \sigma)} \pi_Z(z|\sigma)\, dz.$$

  $\Pr(R \mid z; \sigma)$ analytically available.

- Sherlock, Thiery, Roberts and Rosenthal, (Annals 2015) consider the (joint) optimisation for a limiting target in the important case of RWM. Again anytically available and exact. Similar results.

Sketch of the Analysis

- For general proposals and targets, direct minimization of $CT_h^Q(\sigma) = IF_h^Q(\sigma)/\sigma^2$ impossible so minimize an upper bound over it.

Sketch of the Analysis

- For general proposals and targets, direct minimization of $CT_h^Q(\sigma) = IF_h^Q(\sigma)/\sigma^2$ impossible so minimize an upper bound over it.
- We introduce an auxiliary $\overline{\pi}(\theta, z)-$reversible kernel

$$Q^*\{(\theta, z), (d\vartheta, dw)\} = q(\vartheta|\theta)g(w)\alpha_{\mathrm{EX}}(\theta, \vartheta)\alpha_{\mathrm{Z}}(z, w)\, d\vartheta dw + \{1 - \varrho_{\mathrm{EX}}(\theta)\varrho_{\mathrm{Z}}(z)\}\delta_{(\theta,z)}(d\vartheta, dw),$$

where we have a product acceptance criterion:

$$\alpha_{\mathrm{EX}}(\theta, \vartheta) = \min\{1, r(\theta, \vartheta)\}, \quad \alpha_{\mathrm{Z}}(z, w) = \min\{1, \exp(w - z)\}.$$

Sketch of the Analysis

- For general proposals and targets, direct minimization of $CT_h^Q(\sigma) = IF_h^Q(\sigma)/\sigma^2$ impossible so minimize an upper bound over it.
- We introduce an auxiliary $\overline{\pi}(\theta, z)-$reversible kernel

$$Q^*\{(\theta,z),(d\vartheta,dw)\} = q(\vartheta|\theta)g(w)\alpha_{\mathrm{EX}}(\theta,\vartheta)\alpha_{\mathrm{Z}}(z,w)\,d\vartheta dw + \{1-\varrho_{\mathrm{EX}}(\theta)\varrho_{\mathrm{Z}}(z)\}\delta_{(\theta,z)}(d\vartheta,dw),$$

where we have a product acceptance criterion:

$$\alpha_{\mathrm{EX}}(\theta,\vartheta) = \min\{1, r(\theta,\vartheta)\}, \quad \alpha_{\mathrm{Z}}(z,w) = \min\{1,\exp(w-z)\}.$$

- Peskun's theorem (1973) guarantees that $IF_h^Q(\sigma) \leq IF_h^{Q^*}(\sigma)$ so that $CT_h^Q(\sigma) \leq CT_h^{Q^*}(\sigma)$.

pseudo-proof: Jump chain

IACT: $\mathbb{E}_\pi[x] = 0$, on J.C.s Douc and Robert (Annals, 2011)

$$\widetilde{\pi}(x) = \frac{\pi(x)p(x)}{P_A}$$

$$\begin{aligned}
(IACT + 1)\mathbb{E}_\pi[x_o^2] &= 2\sum_{t=0}^{\infty} \mathbb{E}_\pi[x_o x_t] \\
&= 2\sum_{\tau=0}^{\infty} \mathbb{E}_{\pi,\widetilde{\pi}}[x_o \tau \widetilde{x}_\tau] \\
&= 2\sum_{\tau=0}^{\infty} \mathbb{E}_{\pi,\widetilde{\pi}}\left[x_o \frac{\widetilde{x}_\tau}{p(\widetilde{x}_\tau)}\right] \text{ because sojourns Geometri} \\
&= 2P_A \sum_{\tau=0}^{\infty} \mathbb{E}_{\widetilde{\pi}}\left[\frac{\widetilde{x}_o}{p(\widetilde{x}_0)} \frac{\widetilde{x}_\tau}{p(\widetilde{x}_\tau)}\right] \text{ change of measure for} \\
&= 2P_A \left(\widetilde{IACT} + 1\right) \mathbb{E}_{\widetilde{\pi}}\left[\frac{\widetilde{x}_o^2}{p(\widetilde{x}_0)^2}\right].
\end{aligned}$$

Simpler Bounds on the Relative Inefficiency

- We obtain an explicit expression for $IF_h^{Q^*}(\sigma)$.
- If $IF_{h/\varrho_{EX}}^{\widetilde{Q}^{EX}} \geq 1$, e.g. $\widetilde{Q}^{EX}$ is a positive kernel, then

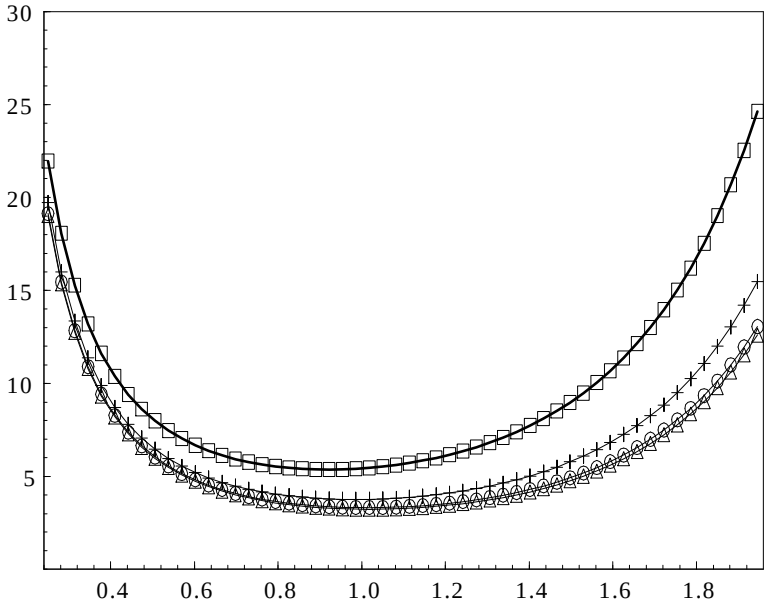
$$\frac{IF_h^{Q}(\sigma)}{IF_h^{EX}} \leq \frac{IF_h^{Q^*}(\sigma)}{IF_h^{EX}} \leq \frac{1}{2}\left(1+\frac{1}{IF_h^{EX}}\right)\pi_Z^{\sigma}\left(1/\varrho_Z^{\sigma}\right) - \frac{1}{IF_h^{EX}}$$

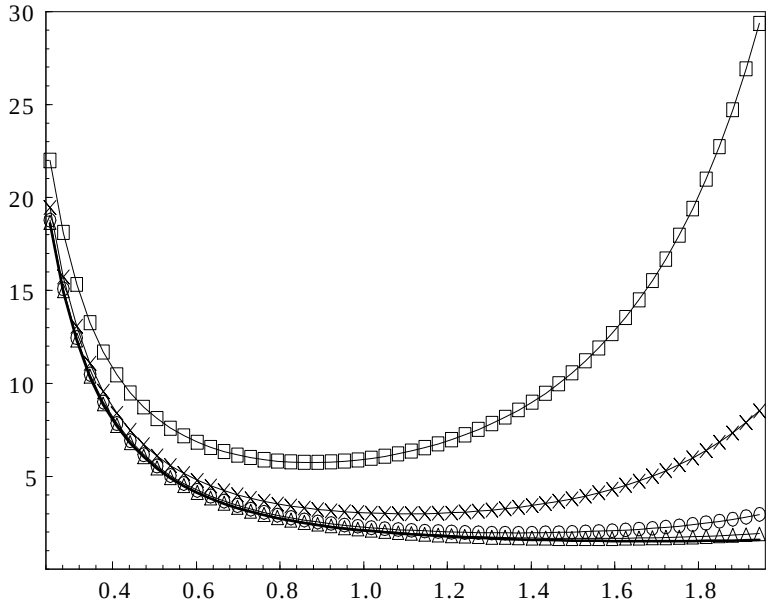
  and the bound is tight as $IF_h^{EX} \longrightarrow 1$ or $\sigma \longrightarrow 0$.
- As $IF_{J,h/\varrho_{EX}}^{EX} \longrightarrow \infty$,

$$\frac{IF_h^{Q^*}(\sigma)}{IF_h^{EX}} \longrightarrow \frac{1}{\pi_Z^{\sigma}\left(\varrho_Z^{\sigma}\right)} = \frac{1}{P_A(\sigma)}.$$

- Results used to minimize w.r.t σ upper bounds on $CT_h^{Q}(\sigma) = IF_h^{Q}(\sigma)/\sigma^2$.

Bounds on Relative Computational Time





Left: upper bound on $CT_h^{Q^*}(\sigma)$ as a function of σ for $IF_h^{\mathrm{EX}} = 1$ (square), 4 (crosses), 20 (circles), 80 (triangles). Right: upper bounds on $CT_h^{Q^*}(\sigma)$ as a function of σ for $IF_{J,h//\varrho_{\mathrm{EX}}}^{\mathrm{EX}} = 1$ for $IF_{J,h//\varrho_{\mathrm{EX}}}^{\mathrm{EX}} = 1, 4, 20, 80$ and lower bound (solid line).

Application: Stochastic Volatility Model

- Chernov et al., *J. Econometrics* (2003) and Huang & Tauchen, *J. Financial Econometrics* (2005):

$$
\begin{aligned}
dv_1(t) &= -k_1\{v_1(t) - \mu_1\}dt + \sigma_1 dW_1(t), \\
dv_2(t) &= -k_2 v_2(t) + \{1 + \beta_{12} v_2(t)\} dW_2(t), \\
d\log P(t) &= \mu_y dt + \text{s-exp}\left[\{v_1(t) + \beta_2 v_2(t)\}/2\right] dB(t),
\end{aligned}
$$

  with $\phi_1 = \text{corr}\{B(t), W_1(t)\}$ and $\phi_2 = \text{corr}\{B(t), W_2(t)\}$.
- Euler discretization of the volatilities $v_1(t)$ and $v_2(t)$ provides closed form expression for $Y(s) = \log P(\tau_{s+1}) - \log P(\tau_s)$.
- Straightforward to estimate the likelihood and simulate forward.

Application: Stochastic Volatility Model

- Daily returns $y = (y_1, ..., y_T)$ of the S&P 500 index.
- Bayesian Inference on $\theta = \left(k_1, \mu_1, \sigma_1, k_2, \beta_{12}, \beta_2, \mu_y, \phi_1, \phi_2\right)$.
- Performance of the pseudo-marginal for RW proposal w.r.t σ, standard deviation of $\log \widehat{p}_\theta(y)$ at posterior mean $\overline{\theta}$.

Empirical vs Assumed Distributions of Z for SV model:

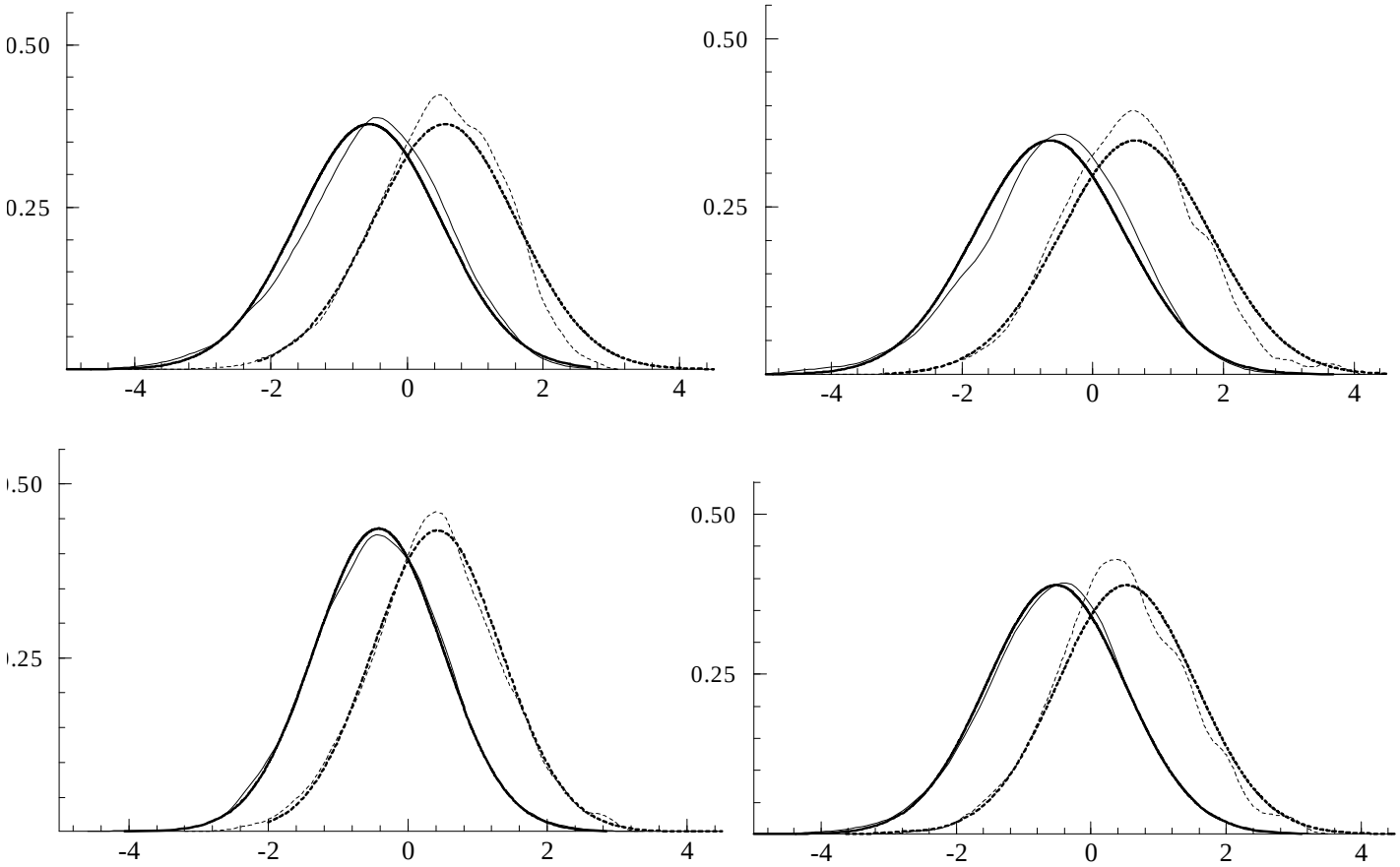


Figure: Empirical distributions (dashed) vs assumed Gaussians (solid) of Z . $T = 300$ and $T = 2700$ at $\overline{\theta}$ (left) and marginalized over $\pi(\theta)$.

Integrated Autocorrelation Time of Pseudo-Marginal MH

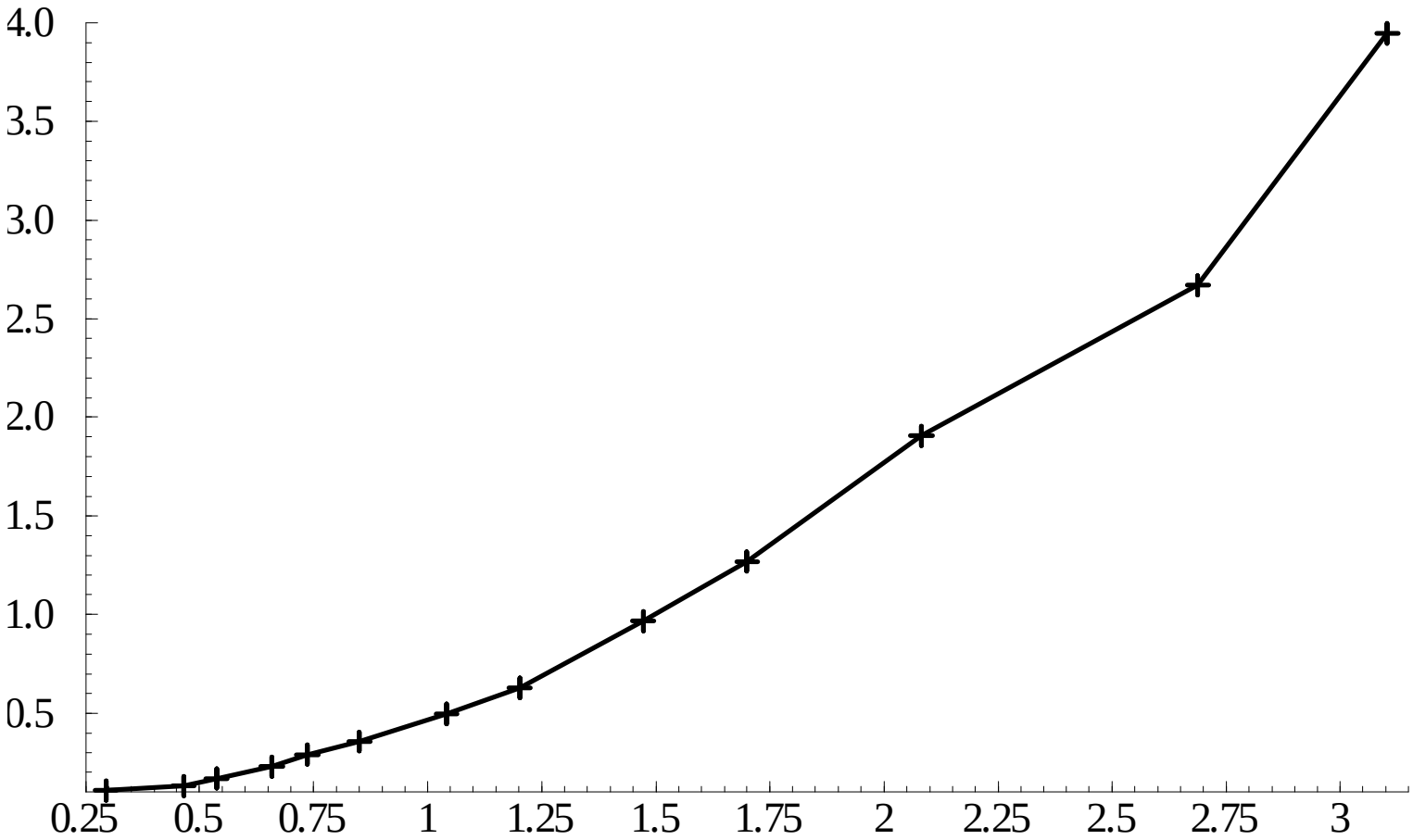


Figure: Average over the 9 parameter components of the log-integrated autocorrelation time of pseudo-marginal chain as a function of σ for $T = 300$.

Computational time for the SV model

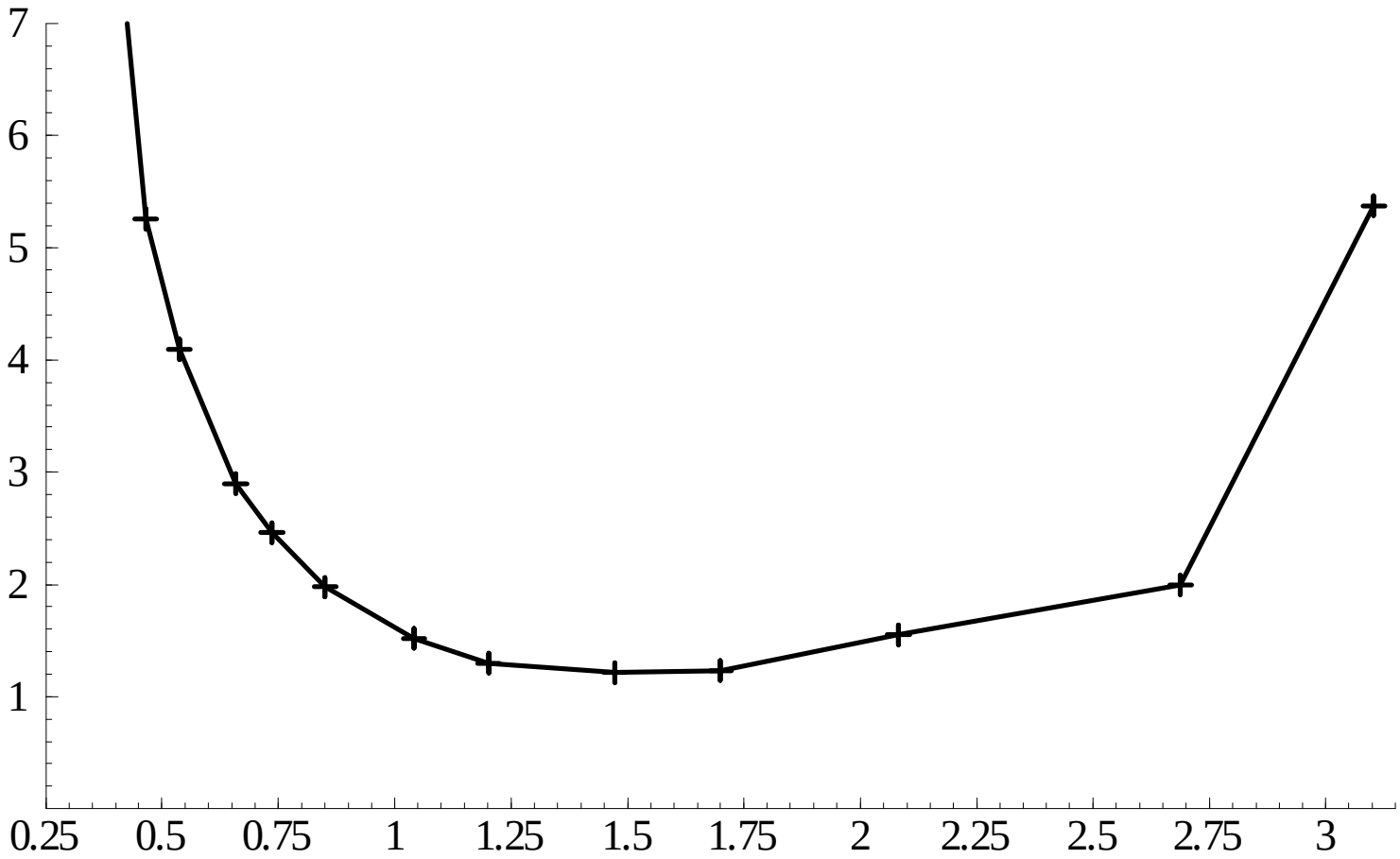


Figure: Computational time as a function of σ

Guideline and Discussion

- **Guideline:** Optimal σ depends on efficiency of the ideal MH algorithm but $\sigma \approx 1.2 - 1.3$ is a sweet spot.

Guideline and Discussion

- **Guideline:** Optimal σ depends on efficiency of the ideal MH algorithm but $\sigma \approx 1.2 - 1.3$ is a sweet spot.
- Pseudo-marginal scales in $\mathcal{O}\left(T^2\right)$ at each iteration as we require $N \propto T$.

Guideline and Discussion

- **Guideline:** Optimal σ depends on efficiency of the ideal MH algorithm but $\sigma \approx 1.2 - 1.3$ is a sweet spot.
- Pseudo-marginal scales in $\mathcal{O}\left(T^2\right)$ at each iteration as we require $N \propto T$.
- For i.i.d. data, simulated ML estimator is efficient as long as N increases at a rate faster than $\sqrt{T}$; e.g. Lee, *Econometric Theory*, 1999.

Guideline and Discussion

- **Guideline:** Optimal σ depends on efficiency of the ideal MH algorithm but $\sigma \approx 1.2 - 1.3$ is a sweet spot.
- Pseudo-marginal scales in $\mathcal{O}\left(T^2\right)$ at each iteration as we require $N \propto T$.
- For i.i.d. data, simulated ML estimator is efficient as long as N increases at a rate faster than $\sqrt{T}$; e.g. Lee, *Econometric Theory*, 1999.
- **Problem**: the ratio $p_\vartheta\left(y_{1:T}\right) / p_\theta\left(y_{1:T}\right)$ is estimated by estimating independently the numerator and denominator in pseudo-marginal.

Correlated Pseudo-Marginal Algorithm

- Previously, we consider the likelihood estimator $\widehat{p}_\theta\left(y_{1:T}; U\right)$ where $U \sim m_\theta\left(\cdot\right)$.
- Reparameterize the likelihood estimator so that $U \sim \mathcal{N}\left(0, I\right)$.
- Correlate estimators of $p_\theta\left(y_{1:T}\right)$ and $p_\vartheta\left(y_{1:T}\right)$ by setting

$$\widehat{p}_\vartheta\left(y_{1:T}\right) = \widehat{p}_\vartheta\left(y_{1:T}; V\right)$$

where

$$V = \rho U + \sqrt{1-\rho^2}\varepsilon,\ \varepsilon \sim \mathcal{N}\left(0, I\right).$$

- In practice, ρ will be selected close to 1.
- The invariance of the PMH unaffected.
- Can be seen as a solution of a standard O-U over $[0, \delta]$, $\rho = \exp(-\delta)$.

Correlated Pseudo-Marginal Algorithm

- Correlated pseudo-marginal simulates a Markov chain $\{\vartheta_i, U_i\}_{i \geq 1}$ of limiting distribution $\overline{\pi}(\theta, u)$.

Correlated Pseudo-Marginal Algorithm

- Correlated pseudo-marginal simulates a Markov chain $\{\vartheta_i, U_i\}_{i \geq 1}$ of limiting distribution $\overline{\pi}(\theta, u)$.

Correlated Pseudo-Marginal Algorithm

- Correlated pseudo-marginal simulates a Markov chain $\{\vartheta_i, U_i\}_{i\geq 1}$ of limiting distribution $\overline{\pi}(\theta, u)$.

At iteration i

- Sample $\vartheta \sim q\left(\cdot \middle| \vartheta_{i-1}\right)$ and $U = \rho U_{i-1} + \sqrt{1-\rho^2}\varepsilon$, $\varepsilon \sim \mathcal{N}\left(0, I\right)$.

Correlated Pseudo-Marginal Algorithm

- Correlated pseudo-marginal simulates a Markov chain $\{\vartheta_i, U_i\}_{i \geq 1}$ of limiting distribution $\overline{\pi}(\theta, u)$.

At iteration i

- Sample $\vartheta \sim q\left(\cdot \middle| \vartheta_{i-1}\right)$ and $U = \rho U_{i-1} + \sqrt{1-\rho^2}\varepsilon$, $\varepsilon \sim \mathcal{N}\left(0, I\right)$.
- Compute the estimate $\widehat{p}_{\vartheta}\left(y_{1:T}; U\right)$ of $p_{\vartheta}\left(y_{1:T}\right)$.

Correlated Pseudo-Marginal Algorithm

- Correlated pseudo-marginal simulates a Markov chain $\{\vartheta_i, U_i\}_{i\geq 1}$ of limiting distribution $\overline{\pi}(\theta, u)$.

At iteration i

- Sample $\vartheta \sim q\left(\cdot \middle| \vartheta_{i-1}\right)$ and $U = \rho U_{i-1} + \sqrt{1-\rho^2}\varepsilon$, $\varepsilon \sim \mathcal{N}\left(0, I\right)$.
- Compute the estimate $\widehat{p}_{\vartheta}\left(y_{1:T}; U\right)$ of $p_{\vartheta}\left(y_{1:T}\right)$.
- With probability

$$\min\{1, \frac{\widehat{p}_{\vartheta}\left(y_{1:T}; U\right)}{\widehat{p}_{\vartheta_{i-1}}\left(y_{1:T}; U_{i-1}\right)} \frac{p\left(\vartheta\right)}{p\left(\vartheta_{i-1}\right)} \frac{q\left(\vartheta_{i-1} \middle| \vartheta\right)}{q\left(\vartheta \middle| \vartheta_{i-1}\right)}\}$$

set $\vartheta_i = \vartheta$, $U_i = U$, otherwise set $\vartheta_i = \vartheta_{i-1}$, $U_i = U_{i-1}$.

Likelihood estimation for state-space models

- If the likelihood is computed using standard SMC, $(\theta, u) \longmapsto \widehat{p}_\theta(y_{:}u)$ is not continuous and displays large variations for moderate N.

Likelihood estimation for state-space models

- If the likelihood is computed using standard SMC, $(\theta, u) \longmapsto \widehat{p}_\theta(y_{:}, u)$ is not continuous and displays large variations for moderate N.
- Discontinuities arise from the resampling step: you can end up picking resampling particles very far from each other even with small variations in (θ, u).

Likelihood estimation for state-space models

- If the likelihood is computed using standard SMC, $(\theta, u) \longmapsto \widehat{p}_\theta(y_{:}u)$ is not continuous and displays large variations for moderate N.
- Discontinuities arise from the resampling step: you can end up picking resampling particles very far from each other even with small variations in (θ, u).
- For $n = 1$, (Pitt et al, 2012) propose sorting particles $X_t^{\sigma_t(1)} \leq \ldots \leq X_t^{\sigma_t(N)}$.

Likelihood estimation for state-space models

- If the likelihood is computed using standard SMC, $(\theta, u) \longmapsto \widehat{p}_\theta(y_{:}u)$ is not continuous and displays large variations for moderate N.
- Discontinuities arise from the resampling step: you can end up picking resampling particles very far from each other even with small variations in (θ, u).
- For $n = 1$, (Pitt et al, 2012) propose sorting particles $X_t^{\sigma_t(1)} \leq \ldots \leq X_t^{\sigma_t(N)}$.
- For $n \geq 2$, (Gerber & Chopin, 2015) use Hilbert curve; e.g. map particles to $[0,1]^n$ using $\psi : \mathbb{R}^n \longrightarrow [0,1]^n$ (e.g. logistic) and $h : [0,1]^n \longrightarrow [0,1]$ (pseudo-inverse Hilbert curve) and sort projected particles on $[0,1]$.

Likelihood estimation for state-space models

- If the likelihood is computed using standard SMC, $(\theta, u) \longmapsto \widehat{p}_\theta(y_{:}, u)$ is not continuous and displays large variations for moderate N.
- Discontinuities arise from the resampling step: you can end up picking resampling particles very far from each other even with small variations in (θ, u).
- For $n = 1$, (Pitt et al, 2012) propose sorting particles $X_t^{\sigma_t(1)} \leq \ldots \leq X_t^{\sigma_t(N)}$.
- For $n \geq 2$, (Gerber & Chopin, 2015) use Hilbert curve; e.g. map particles to $[0,1]^n$ using $\psi : \mathbb{R}^n \to [0,1]^n$ (e.g. logistic) and $h : [0,1]^n \to [0,1]$ (pseudo-inverse Hilbert curve) and sort projected particles on $[0,1]$.
- Alternative coupling ideas have been used to mitigate these fluctuations (Jacob et al., arXiv 2016).

Large sample analysis of the correlated pseudo-marginal

- **Assumption 1** - *Asymptotic Normality*: $\exists\ \widehat{\theta}^T \xrightarrow{P} \overline{\theta}$ and Σ a p.d. matrix s.t.

$$\int \left| p\left(\theta \middle| Y_{1:T}\right) - \phi(\theta; \widehat{\theta}^T, \Sigma / T) \right| \mathrm{d}\theta \xrightarrow{P} 0.$$

Large sample analysis of the correlated pseudo-marginal

- **Assumption 1** - *Asymptotic Normality*: $\exists\ \widehat{\theta}^T \xrightarrow{P} \overline{\theta}$ and Σ a p.d. matrix s.t.

$$\int \left| p\left(\theta \middle| Y_{1:T}\right) - \phi(\theta; \widehat{\theta}^T, \Sigma / T) \right| \mathrm{d}\theta \xrightarrow{P} 0.$$

- **Assumption 2** - *Proposal*: $\vartheta = \theta + \varepsilon / \sqrt{T}$ where $\varepsilon \sim v\left(\cdot\right)$ with $v\left(\varepsilon\right) = v\left(-\varepsilon\right)$.

Large sample analysis of the correlated PM - i.i.d. case

Proposition. Let $N \to \infty$ as $T \to \infty$ with $N = o(T)$. When $U \sim \overline{\pi}_T(\cdot|\theta)$ and $U' = \rho U + \sqrt{1-\rho^2}\varepsilon$ with $\rho = \exp\left(-\psi\frac{N}{T}\right)$ then as $T \to \infty$

$$\log\left\{\frac{\widehat{p}_{\theta+\xi/\sqrt{T}}\left(Y_{1:T};U'\right)}{\widehat{p}_{\theta}\left(Y_{1:T};U\right)} \Big/ \frac{p_{\theta+\xi/\sqrt{T}}\left(Y_{1:T}\right)}{p_{\theta}\left(Y_{1:T}\right)}\right\} \Bigg| \mathcal{Y}^T, \mathcal{U}^T \Rightarrow \mathcal{N}\left(-\frac{\kappa^2(\theta)}{2}, \kappa^2(\theta)\right)$$

where

$$\kappa^2(\theta) = 2\psi\mathbb{E}\left[\left\|\partial_u \varpi\left(U, Y; \theta\right)\right\|^2\right]$$

- This CLT is conditional on the observation sequence and the current auxiliary variables.

Large sample analysis of the correlated PM - i.i.d. case

Proposition. Let $N \to \infty$ as $T \to \infty$ with $N = o(T)$. When $U \sim \overline{\pi}_T(\cdot|\theta)$ and $U' = \rho U + \sqrt{1-\rho^2}\varepsilon$ with $\rho = \exp\left(-\psi \frac{N}{T}\right)$ then as $T \to \infty$

$$\log\left\{\frac{\widehat{p}_{\theta+\xi/\sqrt{T}}\left(Y_{1:T}; U'\right)}{\widehat{p}_{\theta}\left(Y_{1:T}; U\right)} \Big/ \frac{p_{\theta+\xi/\sqrt{T}}\left(Y_{1:T}\right)}{p_{\theta}\left(Y_{1:T}\right)}\right\} \Big| \mathcal{Y}^T, \mathcal{U}^T \Rightarrow \mathcal{N}\left(-\frac{\kappa^2(\theta)}{2}, \kappa^2(\theta)\right)$$

where

$$\kappa^2(\theta) = 2\psi \mathbb{E}\left[\left\|\partial_u \varpi\left(U, Y; \theta\right)\right\|^2\right]$$

- This CLT is conditional on the observation sequence and the current auxiliary variables.
- Asymptotically the distribution of the log-ratio decouples from the current location of the Markov chain.

Large sample analysis of the correlated PM - i.i.d. case

Proposition. Let $N \to \infty$ as $T \to \infty$ with $N = o(T)$. When $U \sim \overline{\pi}_T(\cdot|\theta)$ and $U' = \rho U + \sqrt{1-\rho^2}\varepsilon$ with $\rho = \exp\left(-\psi \frac{N}{T}\right)$ then as $T \to \infty$

$$\log\left\{\frac{\widehat{p}_{\theta+\xi/\sqrt{T}}\left(Y_{1:T}; U'\right)}{\widehat{p}_{\theta}\left(Y_{1:T}; U\right)} \Big/ \frac{p_{\theta+\xi/\sqrt{T}}\left(Y_{1:T}\right)}{p_{\theta}\left(Y_{1:T}\right)}\right\} \Big| \mathcal{Y}^T, \mathcal{U}^T \Rightarrow \mathcal{N}\left(-\frac{\kappa^2(\theta)}{2}, \kappa^2(\theta)\right)$$

where

$$\kappa^2(\theta) = 2\psi \mathbb{E}\left[\left\|\partial_u \varpi\left(U, Y; \theta\right)\right\|^2\right]$$

- This CLT is conditional on the observation sequence and the current auxiliary variables.
- Asymptotically the distribution of the log-ratio decouples from the current location of the Markov chain.
- The asymptotic variance is $O(1)$ even for $N \sim \log(T)$.

Large sample analysis of the correlated PM - i.i.d. case

- Let $\Theta_T := \{\vartheta_i^T\}_{i \geq 0}$ be the stationary *non-Markovian* sequence of the correlated PM targetting $p\left(\theta \middle| Y_{1:T}\right)$.

Large sample analysis of the correlated PM - i.i.d. case

- Let $\Theta_T := \{\vartheta_i^T\}_{i \geq 0}$ be the stationary *non-Markovian* sequence of the correlated PM targetting $p\left(\theta \middle| Y_{1:T}\right)$.
- **Proposition** (Deligiannidis et al., 2016): The sequences $\{\Theta_T\}_{T \geq 1}$ converge weakly as $T \to \infty$ to a stationary Markov chain of invariant density $\phi\left(\widetilde{\theta}; 0, \Sigma\right)$ and kernel given for $\widetilde{\theta} \neq \widetilde{\vartheta}$ by

$$\widetilde{Q}(\widetilde{\theta}, \mathrm{d}\widetilde{\theta}') = v(\widetilde{\theta}' - \widetilde{\theta}) \mathbb{E}_{R \sim \mathcal{N}(-\kappa^2/2, \kappa^2)} \left[\min \left\{ 1, \frac{\phi(\widetilde{\theta}'; 0, \Sigma)}{\phi(\widetilde{\theta}; 0, \Sigma)} \exp R \right\} \right] \mathrm{d}\widetilde{\theta}'$$

where $\kappa := \kappa\left(\overline{\theta}\right)$.

Large sample analysis of the correlated PM - i.i.d. case

- Let $\Theta_T := \{\vartheta_i^T\}_{i \geq 0}$ be the stationary *non-Markovian* sequence of the correlated PM targetting $p(\theta | Y_{1:T})$.
- **Proposition** (Deligiannidis et al., 2016): The sequences $\{\Theta_T\}_{T \geq 1}$ converge weakly as $T \to \infty$ to a stationary Markov chain of invariant density $\phi\left(\widetilde{\theta}; 0, \Sigma\right)$ and kernel given for $\widetilde{\theta} \neq \widetilde{\vartheta}$ by

$$\widetilde{Q}(\widetilde{\theta}, \mathrm{d}\widetilde{\theta}') = v(\widetilde{\theta}' - \widetilde{\theta}) \mathbb{E}_{R \sim \mathcal{N}(-\kappa^2/2, \kappa^2)} \left[\min \left\{ 1, \frac{\phi(\widetilde{\theta}'; 0, \Sigma)}{\phi(\widetilde{\theta}; 0, \Sigma)} \exp R \right\} \right] \mathrm{d}\widetilde{\theta}'$$

  where $\kappa := \kappa\left(\overline{\theta}\right)$.
- It is tempting to use this result to provide guidelines on the optimization of CPM... but one has to be careful.

Decomposition in fast and slow components

- For a stationary CPM chain (ϑ_i, U_i), decompose

$$h(\vartheta_i) = \underbrace{\mathbb{E}[h(\vartheta) | U_i]}_{\text{slow}} + \underbrace{h(\vartheta_i) - \mathbb{E}[h(\vartheta) | U_i]}_{\text{fast}}.$$

Decomposition in fast and slow components

- For a stationary CPM chain (ϑ_i, U_i), decompose

$$h(\vartheta_i) = \underbrace{\mathbb{E}[h(\vartheta)|\, U_i]}_{\text{slow}} + \underbrace{h(\vartheta_i) - \mathbb{E}[h(\vartheta)|\, U_i]}_{\text{fast}}.$$

- U_i is proposed according to AR scheme with persistency $\approx 1 - \psi N/T$: when N grows to slowly with T, IACT of $h(\vartheta_i)$ is dominated by IACT of $\mathbb{E}[h(\vartheta)|\, U_i]$ where for scalar $h(\vartheta) = \vartheta$

$$\mathbb{E}[\vartheta|\, U_i] = \underbrace{\widehat{\theta}_T + \frac{\Sigma}{T}\nabla_\theta \log \frac{\widehat{p}\left(Y_{1:T}|\, \widehat{\theta}_T, U_i\right)}{p\left(Y_{1:T}|\, \widehat{\theta}_T\right)}}_{\Psi(U_i)} + O_P\left(T^{-2}\right).$$

Decomposition in fast and slow components

- For a stationary CPM chain (ϑ_i, U_i), decompose

$$h(\vartheta_i) = \underbrace{\mathbb{E}[h(\vartheta)|\, U_i]}_{\text{slow}} + \underbrace{h(\vartheta_i) - \mathbb{E}[h(\vartheta)|\, U_i]}_{\text{fast}}.$$

- U_i is proposed according to AR scheme with persistency $\approx 1 - \psi N/T$: when N grows to slowly with T, IACT of $h(\vartheta_i)$ is dominated by IACT of $\mathbb{E}[h(\vartheta)|\, U_i]$ where for scalar $h(\vartheta) = \vartheta$

$$\mathbb{E}[\vartheta|\, U_i] = \underbrace{\widehat{\theta}_T + \frac{\Sigma}{T}\nabla_\theta \log \frac{\widehat{p}\left(Y_{1:T}|\, \widehat{\theta}_T, U_i\right)}{p\left(Y_{1:T}|\, \widehat{\theta}_T\right)}}_{\Psi(U_i)} + O_P\left(T^{-2}\right).$$

- **Proposition**. Let $N \propto T^\alpha$ for $0 < \alpha < 1$ then $\mathrm{IACT}(Q, \Psi) \gtrsim T^{1-2\alpha}$.

Decomposition in fast and slow components

- For a stationary CPM chain (ϑ_i, U_i), decompose

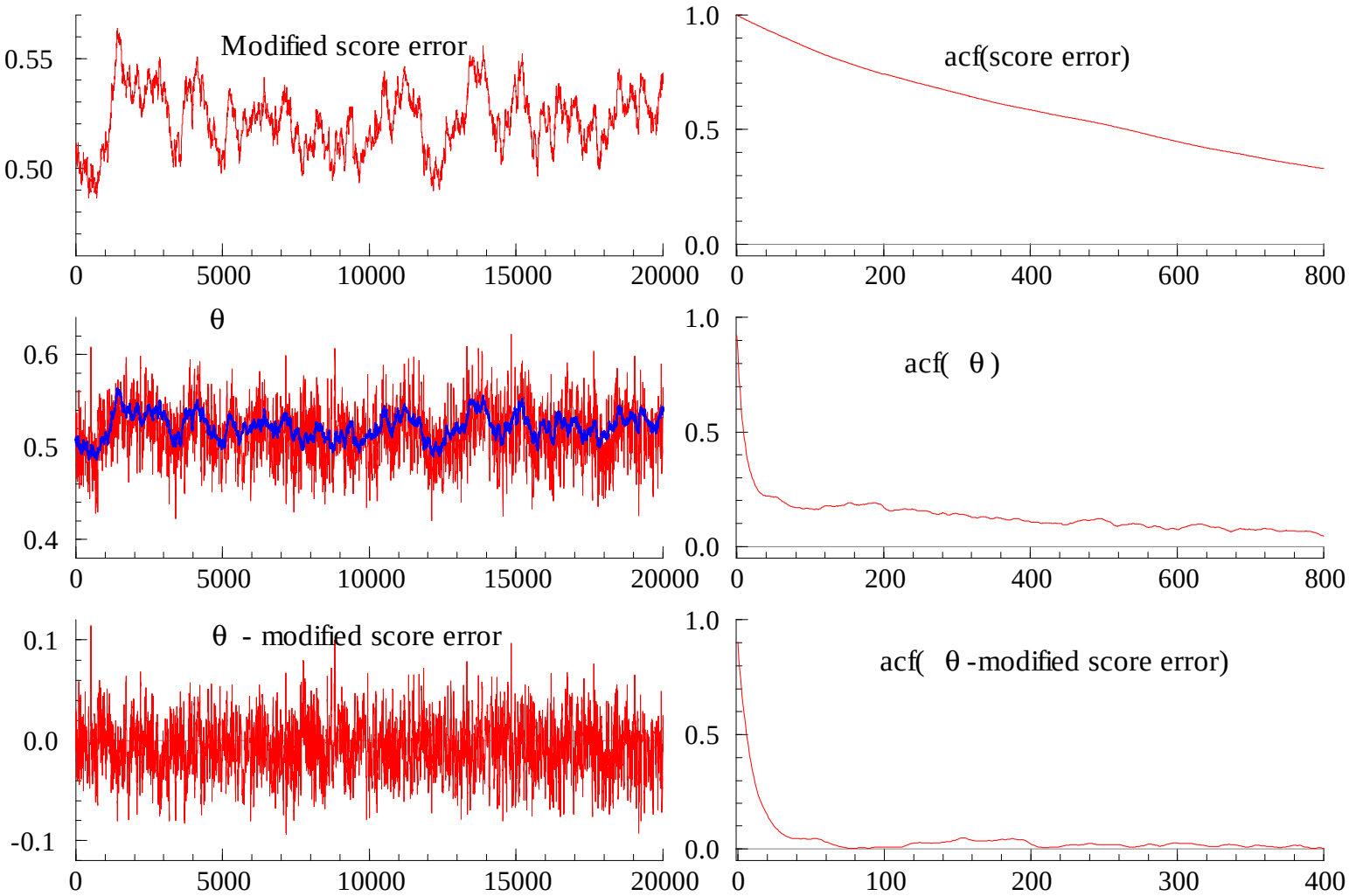
$$h(\vartheta_i) = \underbrace{\mathbb{E}[h(\vartheta)|\, U_i]}_{\text{slow}} + \underbrace{h(\vartheta_i) - \mathbb{E}[h(\vartheta)|\, U_i]}_{\text{fast}}.$$

- U_i is proposed according to AR scheme with persistency $\approx 1 - \psi N/T$: when N grows to slowly with T, IACT of $h(\vartheta_i)$ is dominated by IACT of $\mathbb{E}[h(\vartheta)|\, U_i]$ where for scalar $h(\vartheta) = \vartheta$

$$\mathbb{E}[\vartheta|\, U_i] = \underbrace{\widehat{\theta}_T + \frac{\Sigma}{T}\nabla_\theta \log \frac{\widehat{p}\left(Y_{1:T}|\, \widehat{\theta}_T, U_i\right)}{p\left(Y_{1:T}|\, \widehat{\theta}_T\right)}}_{\Psi(U_i)} + O_P\left(T^{-2}\right).$$

- **Proposition**. Let $N \propto T^\alpha$ for $0 < \alpha < 1$ then $\mathrm{IACT}(Q, \Psi) \gtrsim T^{1-2\alpha}$.
- This result suggests we need at least $\sqrt{T}/N = O(1)$.

Graphical Illustration



Example: Gaussian Latent Variable Model

MH ($T = 8192$)		IACT(θ)	
		15.6	
PM ($\rho = 0.0$)			
N		RIACT(θ)	RCT(θ)
5000		2.2	11210
CPM ($\rho = 0.9963$)			
N	κ	RIACT(θ)	RCT(θ)
10	3.1	14.0	126.2
20	2.2	4.7	93.3
25	2.0	2.8	69.3
35	1.7	1.7	61.1
56	1.3	1.6	87.0

Here $\mathrm{RIACT} = \mathrm{IACT}/\mathrm{IACT}_{MH}$ and $\mathrm{RCT} = N \times \mathrm{RIACT}$. Improvement by 180 fold.

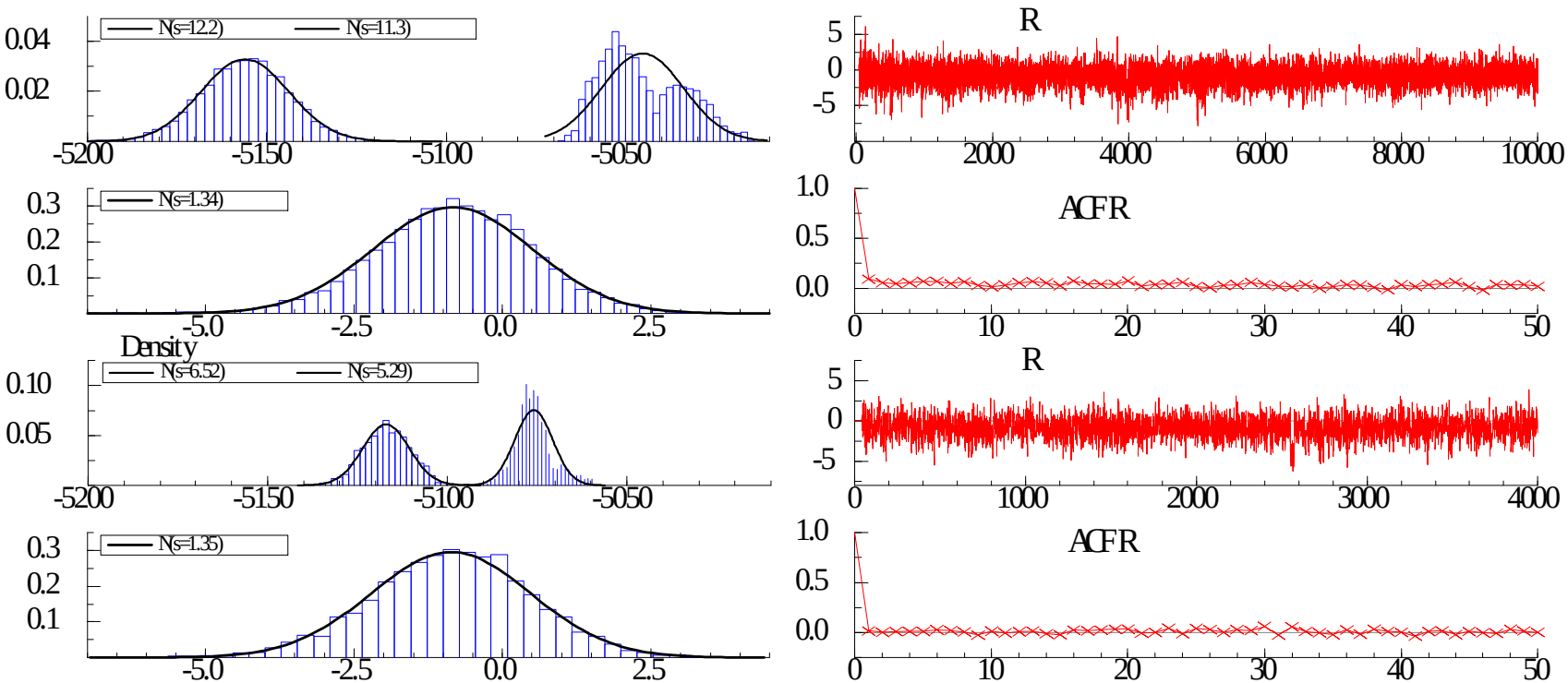
Example: Noisy Autoregressive Model

MH ($T = 16,000$)		IACT(θ)	
		5.8	
PM ($\rho = 0.0$)			
N		RIACT(θ)	RCT(θ)
2500		3.1	8427.0
CPM ($\rho = 0.9965$)			
N	κ	RIACT(θ)	RCT(θ)
6	6.7	43.8	262.8
10	3.3	8.7	86.7
16	1.9	6.0	85.8
22	1.3	3.9	85.6
35	0.8	2.4	85.0
40	0.7	2.4	94.8

Improvement by 100 fold.

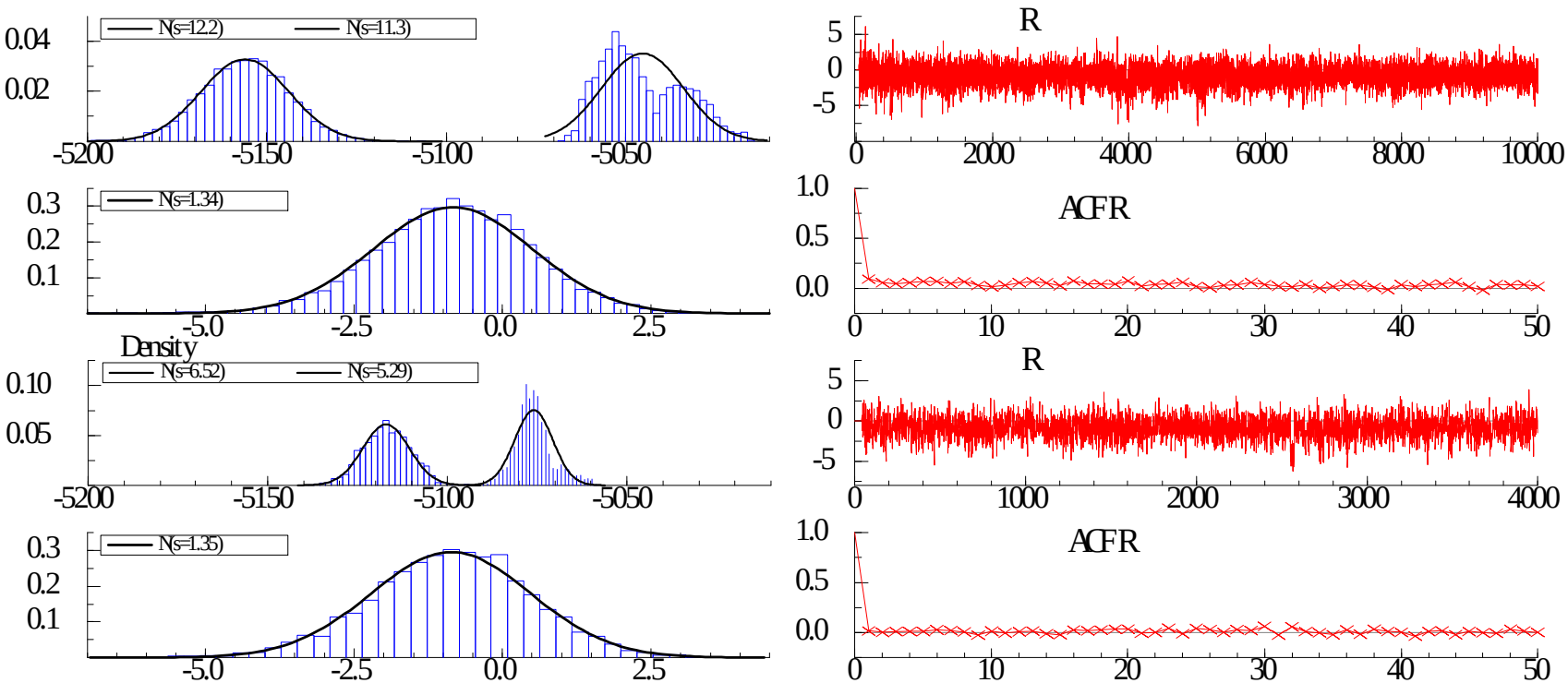
Heston Stochastic Volatility Model

- Inference for a discretized one-dimensional SDE, 40,000 latent variables. 100-fold gain compared to PM.



Heston Stochastic Volatility Model

- Inference for a discretized one-dimensional SDE, 40,000 latent variables. 100-fold gain compared to PM.
- Real data: 4,000 returns from the S&P 500 index from 15/08/1990 to 03/07/2006.



Higher-dimensional SSM

- CLT appears to hold only for N at least of order $T^{n/n+1}$ where n is state-dimension.

Higher-dimensional SSM

- CLT appears to hold only for N at least of order $T^{n/n+1}$ where n is state-dimension.
- Long-range effect is consequently limited.

Higher-dimensional SSM

- CLT appears to hold only for N at least of order $T^{n/n+1}$ where n is state-dimension.
- Long-range effect is consequently limited.
- Still significant gains over PM: over 50 fold for 2-d complex SV model, over 70-fold for 4-d model in (Jacob et al., 2016).

Discussion

- Large sample analysis of pseudo-marginal algorithm provides useful guidelines, overall complexity is $O\left(T^2\right)$.

Discussion

- Large sample analysis of pseudo-marginal algorithm provides useful guidelines, overall complexity is $O\left(T^2\right)$.
- Correlated pseudo-marginal can achieve very substantial improvement.

Discussion

- Large sample analysis of pseudo-marginal algorithm provides useful guidelines, overall complexity is $O\left(T^2\right)$.
- Correlated pseudo-marginal can achieve very substantial improvement.
- In i.i.d. case, analysis shows $\sqrt{T}/N = O(1)$ is necessary and we conjecture it is sufficient leading to complexity $O(T^{3/2})$ vs $O\left(T^2\right)$.

Discussion

- Large sample analysis of pseudo-marginal algorithm provides useful guidelines, overall complexity is $O\left(T^2\right)$.
- Correlated pseudo-marginal can achieve very substantial improvement.
- In i.i.d. case, analysis shows $\sqrt{T}/N = O(1)$ is necessary and we conjecture it is sufficient leading to complexity $O(T^{3/2})$ vs $O\left(T^2\right)$.
- Implementation for state-space models in state dimension $n > 1$ relies on non-standard particle scheme (e.g., Gerber & Chopin, 2015): our analysis does not capture these cases, experimental results suggest $O(T^{1+\frac{n}{n+1}})$.

Some References

- C. Andrieu, A.D. & R. Holenstein, "Particle Markov chain Monte Carlo methods", *JRSS* B, 2000.
- J. Berard, P. Del Moral & A.D., "A Lognormal CLT for Particle Approximations of Normalizing Constants", *Elec. J. Proba.*, 2014.
- G. Deligiannidis, A.D. and M.K. Pitt, "The Correlated Pseudo-marginal Method", arXiv:1511.04992, 2015.
- A.D., M.K. Pitt, G. Deligiannidis and R. Kohn, "Efficient Implementation of Markov Chain Monte Carlo when Using an Unbiased Likelihood Estimator", *Biometrika*, 2015.
- L. Lin, K. Lin & J. Sloan, "A Noisy Monte Carlo Algorithm", *Phys. Rev.* D, 2000.