Monte Carlo methods and proper weighting

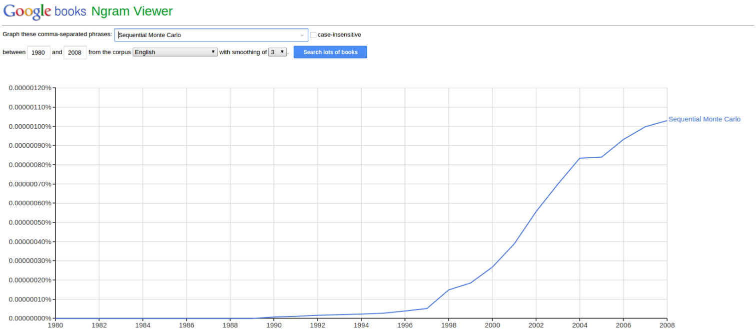
Christian A. Naesseth

Department of Electrical Engineering, Linköping University

LINKÖPING UNIVERSITY

Importance sampling & sequential Monte Carlo

- Introduce importance sampling-based algorithms and concept of *proper weighting*.
- Look at some of our recent work in this area: *Nested SMC* and *Hamiltonian IS*



LINKÖPING UNIVERSITY

LINKÖPING UNIVERSITY

Bayesian inference

Averaging over uncertainty for the latent variable x

$$\mathbb{E}[h(x)] = \int_{\mathsf{X}} h(x)\bar{\pi}(x)\mathrm{d}x =: \bar{\pi}(h),$$

for some function h with $\bar{\pi}(x) = Z_{\pi}^{-1}\pi(x)$.

Bayesian inference – examples

- **Learning:**

$$\bar{\pi}(\theta) = \frac{f(y|\theta)p(\theta)}{p(y)}$$

- **Smoothing & Filtering:**

$$\bar{\pi}(x_{1:n}) = \frac{\prod_{t=1}^{n} f(x_t|x_{t-1})g(y_t|x_t)}{p(y_{1:n})} \quad \text{(SSM)}$$

$$\bar{\pi}(x_{1:n}) = \frac{\prod_{t=1}^{n} f(x_t|x_{1:t-1})g(y_t|x_{1:t})}{p(y_{1:n})} \quad \text{(non-Markovian)}$$

Monte Carlo basics

The goal is to compute $\mathbb{E}[h(x)] = \bar{\pi}(h)$

- Sample $X^i \overset{\text{i.i.d.}}{\sim} \bar{\pi}(x),\ i = 1, \ldots, N$
- Estimator

$$\bar{\pi}^N_{\text{MC}}(h) = \frac{1}{N} \sum_{i=1}^{N} h(X^i) \longrightarrow \bar{\pi}(h), \quad a.s. \text{ (LLN)}$$

Fact: We can derive a CLT for the estimator $\bar{\pi}^N$ with standard rate $1/\sqrt{N}$.
Issue: This requires us to be able to sample *exactly* from $\bar{\pi}$.

LINKÖPING UNIVERSITY

A first importance sampler - I/II

Assume we can evaluate $\bar{\pi}$ and that we have access to some proposal $\bar{q}(x)$ such that $\bar{q}(x) \neq 0$ whenever $\bar{\pi}(x)h(x) \neq 0$.

$$\int_{\mathsf{X}} h(x)\bar{\pi}(x)\mathrm{d}x = \int_{\mathsf{X}} h(x)\frac{\bar{\pi}(x)}{\bar{q}(x)}\bar{q}(x)\mathrm{d}x$$

- Sample $X^i \overset{\text{i.i.d.}}{\sim} \bar{q}(x),\ i = 1, \ldots, N$
- Estimator

$$\bar{\pi}^N_{\text{bIS}}(h) = \frac{1}{N}\sum_{i=1}^{N} w(X^i)h(X^i),\ w(x) = \frac{\bar{\pi}(x)}{\bar{q}(x)}$$

A first importance sampler - II/II

This basic importance sampler

- is *unbiased*: $\mathbb{E}[\bar{\pi}^N_{\text{bIS}}(h)] = \bar{\pi}(h)$,
- is *consistent*: $\bar{\pi}^N_{\text{bIS}}(h) \longrightarrow \bar{\pi}(h)$, a.s., (again due to LLN)
- satisfies a *CLT* with rate $N^{-1/2}$.

Issue: requires normalised $\bar{\pi}$ (and $\bar{q}$) to evalute weights.

Self-normalised importance sampler - I/II

Assume we can evaluate π and that we have access to some proposal $\bar{q}(x)$ such that $\bar{q}(x) \neq 0$ whenever $\bar{\pi}(x) \neq 0$.

$$\int_{\mathsf{X}} h(x)\bar{\pi}(x)\mathrm{d}x = \frac{\int_{\mathsf{X}} h(x)\frac{\pi(x)}{q(x)}\bar{q}(x)\mathrm{d}x}{\int_{\mathsf{X}} \frac{\pi(x)}{q(x)}\bar{q}(x)\mathrm{d}x}$$

- Sample $X^i \overset{\text{i.i.d.}}{\sim} \bar{q}(x),\ i = 1, \ldots, N$
- Estimator

$$\bar{\pi}^N_{\text{IS}}(h) = \sum_{i=1}^{N} \frac{w(X^i)}{\sum_\ell w(X^\ell)} h(X^i),\ w(x) = \frac{\pi(x)}{q(x)}$$

Self-normalised importance sampler - II/II

The self-normalised importance sampler

- *is not* unbiased: $\mathbb{E}[\bar{\pi}^N_{\text{IS}}(h)] \neq \bar{\pi}(h)$,
- *is* consistent: $\bar{\pi}^N_{\text{IS}}(h) \longrightarrow \bar{\pi}(h)$, a.s.,
- *does* satisfy a CLT, given that $\bar{q}(w^2 h^2) < \infty$,

$$\sqrt{N}\left\{\bar{\pi}^N_{\text{IS}}(h) - \bar{\pi}(h)\right\} \xrightarrow{\text{d}} \mathcal{N}(0, \sigma^2_{\text{IS}}),$$

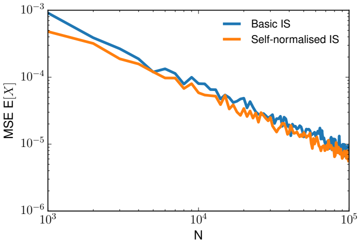
  with $\sigma^2_{\text{IS}} = \int \frac{\bar{\pi}(x)^2}{\bar{q}(x)}(h(x) - \bar{\pi}(h))^2 \text{d}x$.

- gives unbiased estimate of $\frac{Z_\pi}{Z_q}$, i.e. $N^{-1}\sum_i w(X^i)$,
- can have lower MSE than bIS, trade-off bias-variance.

Simple example

We consider a simple Gaussian example:

$$\pi(x) = e^{-\frac{1}{2}x^2}, Z_\pi = \sqrt{2\pi}, \quad \bar{q}(x) = \mathcal{N}(0, 1.2^2).$$



LINKÖPING UNIVERSITY

A more general class of algorithms?

- Learning/inference when evaluating π is intractable (use "random weights")
  - Discretely observed diffusions [Beskos et al., 2006, Fearnhead et al., 2008]
  - Intractable likelihood $f(y|\theta)$ [Tran et al., 2013, Everitt et al., 2015]
- Algorithms where we can generate samples and weights that "work" without being able to explicitly state $q(x)$ [Naesseth et al., 2015, Stern, 2015]
- New population/sequential Monte Carlo variants (see IS with Hamiltonian dynamics later)

Proper weighting

Definition (Properly weighted sample)**.** *A (random) pair* $(W, X) \in \mathbb{R} \times \mathsf{X}$ *is properly weighted for* $\bar{\pi}$ *if* $\mathbb{E}[Wh(X)] = c \cdot \bar{\pi}(h)$ *for all measurable functions* h*, where* $c \in \mathbb{R}_+$ *is a constant.*

Note that this is not new, it is a slight modification of the conditions in Liu [2001].

Importance sampling with proper weights

Assume we have some distribution $\bar{Q}$ that can generate samples $(W^i, X^i) \overset{\text{i.i.d.}}{\sim} \bar{Q}$, that are properly weighted with respect to $\bar{\pi}$.

- Sample $(W^i, X^i) \overset{\text{i.i.d.}}{\sim} \bar{Q},\ i = 1, \dots, N$
- Estimator

$$\bar{\pi}^N_{\text{pwIS}}(h) = \sum_{i=1}^{N} \frac{W^i}{\sum_\ell W^\ell} h(X^i),$$

All the above importance samplers follow as a special case when $\bar{Q}$ is a *degenerate* distribution, i.e. W^i is deterministic given $X^i \sim \bar{q}(x)$.

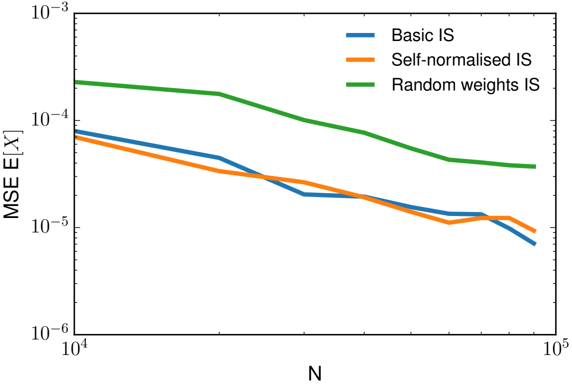
Properties of the properly weighted importance sampler

The completely general pwIS (and its extensions) is still a work in progress, however it

- *is not* unbiased: $\mathbb{E}[\bar{\pi}^N_{\text{pwIS}}(h)] \neq \bar{\pi}(h)$,
- *is* consistent: $\bar{\pi}^N_{\text{pwIS}}(h) \longrightarrow \bar{\pi}(h)$, a.s.,
- if the common $c = Z_\pi$, then $N^{-1}\sum_i W^i$ is an unbiased estimate of Z_π,
- should satisfy a CLT with rate $N^{-1/2}$ under the right conditions, (WIP)
- in general has higher variance (“random weights” version) than deterministic/degenerate weights.

Properly weighted importance sampler – Example

Same example as before, however we add $\mathcal{N}(0, 1)$ to the importance weights.



Rejection sampling – I/II

If we would like to use rejection sampling to estimate expectations. (assume $\frac{\pi(x)}{\bar{q}(x)} \leq C$ for some constant)

- Sample $X^i \overset{\text{i.i.d.}}{\sim} \bar{q}(x),\ i = 1, \ldots, N$
- Draw $U^i \overset{\text{i.i.d.}}{\sim} \mathcal{U}[0, 1]$ and set $W^i = 1$ if $U^i < \frac{\pi(x)}{C\bar{q}(x)}$ (*accept*) and $W^i = 0$ otherwise (*reject*)
- Estimator

$$\bar{\pi}^N_{\text{RS}}(h) = \sum_{i=1}^{N} \frac{W^i}{\sum_\ell W^\ell} h(X^i),$$

Rejection sampling – II/II

Now, given this process to generate (W^i, X^i) we can show that these are properly weighted:

$$\mathbb{E}[W^i h(X^i)] = \mathbb{E}\left[h(X^i)\int_0^1 w^i(u)\mathrm{d}u\right] = \mathbb{E}\left[h(X^i)\int_0^{\frac{\pi(X^i)}{C\bar{q}(X^i)}} 1\mathrm{d}u\right]$$

$$= \mathbb{E}\left[h(X^i)\frac{\pi(X^i)}{C\bar{q}(X^i)}\right] = \int h(x)\frac{\pi(x)}{C\bar{q}(x)}\bar{q}(x)\mathrm{d}x = \frac{Z_\pi}{C}\bar{\pi}(h)$$

Thus we can view rejection sampling as a special case of pwIS!

Sampling importance resampling – I/II

Rubin [1987] suggested to make use of *resampling* in importance sampling to generate unweighted samples.

- Sample $(W^i, X^i) \overset{\text{i.i.d.}}{\sim} \bar{Q},\ i = 1, \ldots, N$
- Resample according to W^i to get $(N^{-1}, \tilde{X}^i)$
- Estimator

$$\bar{\pi}_{\text{SIR}}^N(h) = \frac{1}{N} \sum_{i=1}^{N} h(\tilde{X}^i),$$

This means we have $\tilde{X}^i \overset{\text{approx.}}{\sim} \bar{\pi}$.

Sampling importance resampling – II/II

Let ν^i be integer-valued random variables corresponding to the number of times sample i was resampled. ($\mathbb{E}[\nu^i] = NW^i$)

$$\bar{\pi}^N_{\text{SIR}}(h) = \frac{1}{N}\sum_{i=1}^N h(\tilde{X}^i) = \sum_{i=1}^N \frac{\nu^i}{\sum_\ell \nu^\ell} h(X^i),$$

and (ν^i, X^i) is properly weighted with respect to $\bar{\pi}$!

- Yet another special case of pwIS.
- This idea together with sequential importance sampling (later if there is time) are the key components of the particle filter.

Nested IS – I/II

Here we are interested in making an approximation of the *proposal* $\bar{q}(x)$. We will do this by running IS, with proposal $\bar{r}(x)$, within an outer IS (hence the word *nested*):

- Simulate $\tilde{X}^{i,j} \overset{\text{i.i.d.}}{\sim} \bar{r}(x),\ i = 1, \ldots, N,\ j = 1, \ldots, M$
- Set $\tilde{W}^{i,j} = \frac{q(\tilde{X}^{i,j})}{\bar{r}(\tilde{X}^{i,j})}$
- Simulate $A^i \sim \text{Cat}\left(\left\{\tilde{W}^{i,j}(\sum_\ell \tilde{W}^{i,\ell})^{-1}\right\}_{j=1}^{M}\right)$ and set $X^i = \tilde{X}^{i,A^i}$
- Set $W^i = \frac{\pi(X^i)}{q(X^i)} \cdot \frac{1}{M}\sum_{j=1}^{M} \tilde{W}^{i,j}$
- Estimator

$$\bar{\pi}^N_{\text{NIS}}(h) = \sum_{i=1}^{N} \frac{W^i}{\sum_\ell W^\ell} h(X^i),$$

Nested IS – II/II

Now why would we want to approximate the proposal?

- Well, typically a reasonable proposal for any test function h is $\bar{q}(x) = \bar{\pi}(x)$. This means we can use a nested IS with M samples to approximate an exact draw from $\bar{\pi}$!
- Extra useful in a sequential IS with resampling algorithm, see Naesseth et al. [2015].
- This also opens up for novel parallelization options. (cloud computing etc.)

We can show that NIS is in fact a special case of pwIS as well! The distribution/proposal $\bar{Q}$ that generates (W^i, X^i) is given by the procedure on the previous slide.

LINKÖPING UNIVERSITY

Iterated importance sampling - I/II

One key problem in IS is how to choose a good proposal $\bar{Q}$. The iterated importance sampler (or population Monte Carlo) does this by *adaptation*. Assume we have a set $\{\bar{Q}_j\}$ of proposals.

- Assume we have $\{W_{t-1}^i, X_{t-1}^i\}_{i=1}^N$ (here $W_{t-1} \geq 0$) that are properly weighted w.r.t. $\bar{\pi}$
- Simulate $A_t^i \sim \text{Cat}\left(\left\{W_{t-1}^j (\sum_\ell W_{t-1}^\ell)^{-1}\right\}_{j=1}^N\right)$
- Simulate $(X_t^i, W_t^i) \sim \bar{Q}_{A_t^i}$
- Estimator

$$\bar{\pi}_{\text{pwIIS}}^N(h) = \sum_{i=1}^N \frac{W_t^i}{\sum_\ell W_t^\ell} h(X_t^i),$$

Iterated importance sampling - II/II

Note a few things:

- The proposals $\bar{Q}_j, j = 1, \ldots, N$ can depend on the whole previous particle population. Example:

$$X_t^i \sim \mathcal{N}\left(X_{t-1}^{A_t^i}, C_t\right),$$

$$W_t^i = \frac{\pi(X_t^i)}{\mathcal{N}\left(X_t^i; X_{t-1}^{A_t^i}, C_t\right)},$$

  where C_t is empirical covariance estimate.

- The general framework is still very much a work in progress, we can show that we get proper weights.
- However, due to correlation between samples consistency needs some work and there will be some restrictions on $\bar{Q}_j$.

Recent examples

- Adaptive SMC sampler [Fearnhead et al., 2013]
- Neural Adaptive SMC [Gu et al., 2015] (and more stuff being done here!)
- Gradient Importance Sampling [Schuster, 2015]
- Kernel Adaptive SMC sampler [Schuster et al., 2015]
- Importance sampling with Hamiltonian dynamics [Naesseth and Lindsten, 2015]

Hamiltonian dynamics – I/II

We extend the state-space with a momentum variable $p \in \mathbb{R}$.

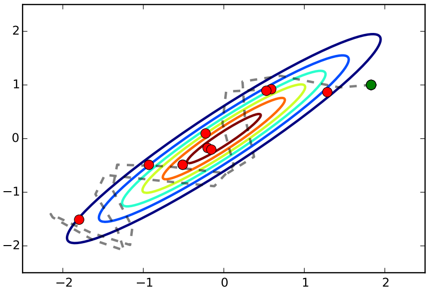
- Hamiltonian: $H(x, p) = -\log \pi(x) - \frac{p^\top M^{-1} p}{2}$
- Hamiltonian dynamics, with initial conditions x_0 and $p_0 \sim \mathcal{N}(0, M)$, is given by solving

$$\dot{p} = -\frac{\partial H}{\partial x},$$
$$\dot{x} = \frac{\partial H}{\partial p}.$$

- Solution by numerical integration $(x_t, p_t) = \Phi_\tau(x_0, p_0)$

Hamiltonian dynamics – II/II

Let's define our target distribution $\tilde{\pi}(x,p) = e^{-H(x,p)}$, thus by construction $\tilde{\pi}(x) = \pi(x)$.



A key property of Φ_τ is that its Jacobian satisfies $|J_{\Phi_\tau}| = 1$, i.e. it is volume preserving.

Importance sampling with Hamiltonian dynamics – I/III

- Assume we have $\{W_{t-1}^i, X_{t-1}^i\}_{i=1}^N$ that are properly weighted w.r.t. $\bar{\pi}$
- Simulate $p_0^i \sim \mathcal{N}(0, M)$ and set $x_0^i = X_{t-1}^i$
- Numerical integration: $(X_t^i, P_t^i) = \Phi_\tau(x_0^i, p_0^i)$
- Set $W_t^i = W_{t-1}^i \exp\left\{-H(X_t^i, P_t^i) + H(x_0^i, p_0^i)\right\}$

Resampling step can be added if needed.

Importance sampling with Hamiltonian dynamics – II/III

$$
\begin{aligned}
\mathbb{E}[W_t^i h(X_t^i)] &= \mathbb{E}\left[W_{t-1}^i \exp\left\{-H(X_t^i, P_t^i) + H(x_0^i, p_0^i)\right\} h(X_t^i)\right] \\
&= \mathbb{E}\Big[W_{t-1}^i \underbrace{\int e^{-H(x_t^i, p_t^i) + H(x_{t-1}^i, p_0^i)} h(x_t^i) \mathcal{N}(p_0^i; 0, M) \mathrm{d}p_0^i}_{g(X_{t-1}^i)}\Big] \\
&= c \int \bar{\pi}(x_{t-1}^i) e^{-H(x_t^i, p_t^i) + H(x_{t-1}^i, p_0^i)} h(x_t^i) \mathcal{N}(p_0^i; 0, M) \mathrm{d}p_0^i \mathrm{d}x_{t-1}^i \\
&= \frac{c}{Z_\pi \sqrt{(2\pi)^d |M|}} \int e^{-H(x_t^i, p_t^i)} h(x_t^i) \mathrm{d}p_0^i \mathrm{d}x_{t-1}^i \\
&= \frac{c}{Z_\pi \sqrt{(2\pi)^d |M|}} \int e^{-H(x_t^i, p_t^i)} h(x_t^i) \underbrace{|J_{\Phi_\tau}|}_{=1} \mathrm{d}p_t^i \mathrm{d}x_t^i = c \cdot \bar{\pi}(h)
\end{aligned}
$$

Importance sampling with Hamiltonian dynamics - III/III

- Possible also to consider sequential targets $\pi_t(x)$ (annealing) or $\pi_t(x_{1:t})$ (smoothing)
- No rejection, only *proper weighting*
- Various sorts of adaptation possible (that doesn't work for MCMC with Hamiltonian dynamics)
- Still a work in progress so many unanswered questions, e.g.
    - How does it perform compared to other methods or SMC with Hamiltonian MCMC kernel?
    - Exact requirements on convergence?

LINKÖPING UNIVERSITY

Sequential Monte Carlo

In SMC we consider a sequence of targets: $\bar{\pi}(x_{1:t}), t = 1, \ldots, T$.
The key idea is to combine sequential IS:

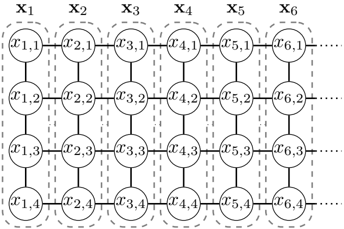
- $X_t^i \sim \bar{q}_t(x_t | X_{1:t-1}^i)$, $X_{1:t}^i = (X_{1:t-1}^i, X_t^i)$
- $W_t^i = W_{t-1}^i \frac{\pi_t(X_{1:t}^i)}{\pi_{t-1}(X_{1:t-1}^i)\bar{q}_t(X_t^i | X_{1:t-1}^i)}$

with the resampling procedure that we discussed earlier and get:

- $\mathbb{P}(A_t^i = j) = \frac{W_{t-1}^j}{\sum_\ell W_{t-1}^\ell}$
- $X_t^i \sim \bar{q}_t(x_t | X_{1:t-1}^{A_t^i})$, $X_{1:t}^i = (X_{1:t-1}^{A_t^i}, X_t^i)$
- $W_t^i = \frac{\pi_t(X_{1:t}^i)}{\pi_{t-1}(X_{1:t-1}^{A_t^i})\bar{q}_t(X_t^i | X_{1:t-1}^{A_t^i})}$

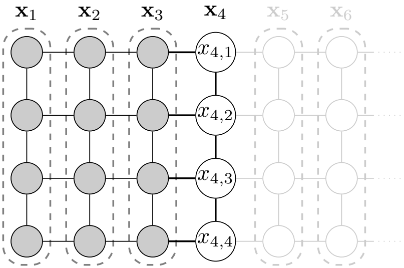
Nested sequential Monte Carlo – I/III

In Nested SMC we use the ideas from Nested IS and generate properly weighted samples w.r.t. $\bar{q}_t$.
Ex: 2D MRF with 1 spatial + 1 temporal dimension



$$\bar{\pi}_t(\mathbf{x}_{1:t}) = Z_{\pi_t}^{-1} \boldsymbol{\phi}_1(\mathbf{x}_1) \prod_{s=2}^{t} \left\{ \boldsymbol{\phi}_s(\mathbf{x}_s) \boldsymbol{\psi}_s(\mathbf{x}_{s-1}, \mathbf{x}_s) \right\}.$$

Nested sequential Monte Carlo – II/III



$$q_t(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \boldsymbol{\phi}_t(\mathbf{x}_t)\boldsymbol{\psi}_t(\mathbf{x}_{t-1}, \mathbf{x}_t)$$

$$= \left\{ \prod_{k=1}^{d} G_{t,k}(x_{t,k}) \prod_{k=2}^{d} m(x_{t,k-1}, x_{t,k}) \right\} \left\{ \prod_{k=1}^{d} \psi(x_{t-1,k}, x_{t,k}) \right\}$$

Nested sequential Monte Carlo – III/III

Gaussian spatio-temporal model in the form of a 2D MRF, $d \times t$, i.e. $\dim x_t = d$.

$$p(\mathbf{x}_{1:t}, y_{1:t}) \propto \prod_{s=1}^{t} \underbrace{\mathcal{N}(\mathbf{y}_s; \mathbf{x}_s, \tau^{-1} I)}_{G} \underbrace{\mathcal{N}(\mathbf{x}_s; a\mathbf{x}_{s-1}, I)}_{\psi} \underbrace{\mathcal{N}(\mathbf{x}_s; 0, \Sigma)}_{m}$$

where Σ^{-1} is a banded matrix (reflecting local dependencies).

$d = 50$ $d = 100$ $d = 200$

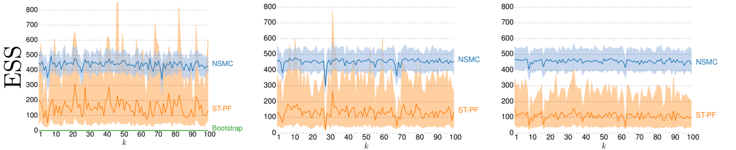


Figure: Median (over dimension) effective sample size (ESS) and 15–85% percentiles. $N = 500$ and $M = 2d$. (Results for 100 independent runs.)

ST-PF as proposed by Beskos et al. [2014].

Summary & conclusions

- Importance sampling-based algorithms rely on the concept of *proper weighting*.
- These kind of sample population methods can better handle multi-modality and parallelization.
- The proposal is perhaps the most important design choice in a importance sampler.
- Adaptation is emerging as a important way of choosing the proposal.
- Exciting applications in signal processing, statistical mechanics, probabilistic programming, machine learning, and many many more areas!

Open tenure track position in ML

Tenure track position in Machine Learning at Uppsala University.

UPPSALA
UNIVERSITET

If you are interested, send an e-mail to Thomas Schön thomas.schon@it.uu.se

If you know someone else who might be interested feel free to help us spread this information as much as you can and want.

Deadline for applications: November 30, 2015.

All details are availble here
www.uu.se/en/about-uu/join-us/details/?positionId=75817

There is also a post-doc opening, send Thomas an e-mail if you are interested.

References

A. Beskos, D. Crisan, A. Jasra, K. Kamatani, and Y. Zhou. A stable particle filter in high-dimensions. arXiv.org, arXiv:1412.3501, December 2014.

Alexandros Beskos, Omiros Papaspiliopoulos, Gareth O Roberts, and Paul Fearnhead. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):333–382, 2006.

R. G. Everitt, A. M. Johansen, E. Rowing, and M. Evdemon-Hogan. Bayesian model comparison with intractable likelihoods. *ArXiv e-prints*, April 2015.

P. Fearnhead, O. Papaspiliopoulos, and G. O. Roberts. Particle filters for partially observed diffusions. *Journal of the Royal Statistical Society: Series B*, 70(4):755–777, 2008.

Paul Fearnhead, Benjamin M Taylor, et al. An adaptive sequential monte carlo sampler. *Bayesian Analysis*, 8(2):411–438, 2013.

S. Gu, Z. Ghahramani, and R. E. Turner. Neural Adaptive Sequential Monte Carlo. *ArXiv e-prints*, June 2015.

J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, 2001.

C. A. Naesseth, F. Lindsten, and T. B. Schön. Nested sequential Monte Carlo methods. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, Lille, France, July 2015.

Christian A. Naesseth and Fredrik Lindsten. Importance sampling with hamiltonian dynamics, 2015. forthcoming.

D. B. Rubin. A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR algorithm. *Journal of the American Statistical Association*, 82(398):543–546, June 1987. Comment to Tanner and Wong: The Calculation of Posterior Distributions by Data Augmentation.

I. Schuster. Gradient Importance Sampling. *ArXiv e-prints*, July 2015.

I. Schuster, H. Strathmann, B. Paige, and D. Sejdinovic. Kernel Adaptive Sequential Monte Carlo. *ArXiv e-prints*, October 2015.

Rafael B. Stern. *A statistical contribution to historical linguistics*. PhD thesis, Carnegie Mellon University, May 2015.

Thank you!

www.liu.se

LINKÖPING UNIVERSITY