Layered Adaptive Importance Sampling

Luca Martino

L. Martino, V. Elvira, D. Luengo, J. Corander. "Layered Adaptive Importance Sampling", (to appear) Statistics and Computing, 2016.

QMUL, 2016

Outline

1. Introduction and motivation
2. Layered Adaptive Importance Sampling (LAIS)
3. Consistency of the estimators (in LAIS)
4. Theoretical motivation of the proposed Markov adaptation
5. Numerical simulations

- INTRODUCTION (framework) AND MOTIVATION

Introduction and notation

- Bayesian inference:
  - $g(\mathbf{x})$: prior pdf.
  - $\ell(\mathbf{y}|\mathbf{x})$: likelihood function.
  - $\mathbf{x}$: variable of interest.
  - $\mathbf{y}$: observed data - measurements.
  - Posterior pdf and marginal likelihood (evidence)

$$\bar{\pi}(\mathbf{x}) = p(\mathbf{x}|\mathbf{y}) = \frac{\ell(\mathbf{y}|\mathbf{x})g(\mathbf{x})}{Z(\mathbf{y})},$$

$$Z(\mathbf{y}) = \int_{\mathcal{X}} \ell(\mathbf{y}|\mathbf{x})g(\mathbf{x})d\mathbf{x}.$$

- In general, $Z(\mathbf{y})$ is unknown, we can evaluate $\pi(\mathbf{x}) \propto \bar{\pi}(\mathbf{x})$:

$$\pi(\mathbf{x}) = \ell(\mathbf{y}|\mathbf{x})g(\mathbf{x}).$$

In the following, we denote $Z(\mathbf{y})$ simply as Z.

Goal

- Our goal is computing efficiently an integral w.r.t. the target pdf,

$$I = E_\pi[f(\mathbf{x})] = \frac{1}{Z}\int_{\mathcal{X}} f(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}, \qquad (1)$$

where f is a square-integrable function, for instance,

$$\widehat{\mathbf{x}}_{MMSE} = \frac{1}{Z}\int_{\mathcal{X}} \mathbf{x}\pi(\mathbf{x})d\mathbf{x},$$

and the normalizing constant,

$$Z = \int_{\mathcal{X}} \pi(\mathbf{x})d\mathbf{x}, \qquad (2)$$

via Monte Carlo.

Monte Carlo approximation

- (Monte Carlo) IDEAL CASE: Draw $\mathbf{x}^{(m)} \sim \bar{\pi}(\mathbf{x})$, $m = 1, \ldots, M$, and

$$\widehat{I} = \frac{1}{M} \sum_{m=1}^{M} f(\mathbf{x}^{(m)}) \approx I.$$

- However, in general:
  - it is not possible to draw from $\bar{\pi}(\mathbf{x})$.
  - Even in this "ideal" case it is not trivial to approximate Z, i.e., to find $\widehat{Z} \approx Z$.

Monte Carlo - Sampling methods

- Since it is impossible to draw directly from $\bar{\pi}(\mathbf{x})$:
  - Importance Sampling $\Longrightarrow$ weighted samples.
  - Markov Chain Monte Carlo (MCMC) $\Longrightarrow$ correlated samples.
- MC sampling techniques use a simpler proposal density $q(\mathbf{x})$ for generating random candidates, and then "filtering" them according to some suitable rule.

Importance sampling (IS)

- Draw $\mathbf{x}^{(m)} \sim q(\mathbf{x})$, $m = 1, \ldots, M$.
- Assign to each sample the unnormalized weights

$$w_m = \frac{\pi(\mathbf{x}^{(m)})}{q(\mathbf{x}^{(m)})}, \qquad m = 1, \ldots, M.$$

- Compute

$$\widetilde{I} = \frac{1}{Z}\frac{1}{M}\sum_{m=1}^{M} w_m f(\mathbf{x}^{(m)}).$$

or (if Z is unkonwn)

$$\widehat{I} = \sum_{m=1}^{M} \bar{w}_m f(\mathbf{x}^{(m)}) = \frac{1}{\sum_{m=1}^{M} w_m}\sum_{m=1}^{M} w_m f(\mathbf{x}^{(m)}).$$

and

$$\widehat{Z} = \frac{1}{M}\sum_{m=1}^{M} w_m \approx Z.$$

Importance sampling (IS)

- The IS approach is valid (i.e., $\widetilde{I}$ unbiased) since

$$E_\pi[f(\mathbf{x})] = E_q[w(\mathbf{x})f(\mathbf{x})],$$

$$\begin{aligned} \frac{1}{Z}\int_{\mathcal{X}} f(\mathbf{x})\pi(\mathbf{x})d\mathbf{x} &= \frac{1}{Z}\int_{\mathcal{X}} f(\mathbf{x})\frac{\pi(\mathbf{x})}{q(\mathbf{x})}q(\mathbf{x})d\mathbf{x}, \\ &= \frac{1}{Z}\int_{\mathcal{X}} f(\mathbf{x})w(\mathbf{x})q(\mathbf{x})d\mathbf{x}. \end{aligned}$$

- Since $\widehat{Z} \to Z$, for $M \to \infty$, then $\widehat{I} \to \widetilde{I}$, is consistent.
- There are several possible combinations of sampling ($\mathbf{x}$) and weighting ($\mathbf{w}$) strategies (this is only the classical approach).

Proposal densities - performance

- The performance depends strictly on the choice of $q(\mathbf{x})$ (in any MC method).
- If we consider a specific function f, in IS:
  - Optimal choice $q(\mathbf{x}) \propto |f(\mathbf{x})|\bar{\pi}(\mathbf{x})$.
- If we consider a generic function f:
  - Optimal choice $q(\mathbf{x}) = \bar{\pi}(\mathbf{x})$.

Proposal densities - performance

- The performance depends strictly on the choice of $q(\mathbf{x})$ (in any MC method).
- If we consider a specific function f, in IS:
  - Optimal choice $q(\mathbf{x}) \propto |f(\mathbf{x})|\bar{\pi}(\mathbf{x})$.
- If we consider a generic function f:
  - Optimal choice $q(\mathbf{x}) = \bar{\pi}(\mathbf{x})$.

- Hence, we need:
  - $q(\mathbf{x})$ as closer as possible to $\bar{\pi}(\mathbf{x})$.
  - proper tuning of the parameters;
  - adaptive methods.
- Another strategy for increasing the robustness:
  - Combined use of several proposal pdfs $q_1, \ldots, q_N$.

- LAYERED ADAPTIVE IMPORTANCE SAMPLING (LAIS)

In this work: brief sketch - contributions

- We design a class Adaptive Importance Sampling schemes using a population of different proposals $q_1, \ldots, q_N$.
- We focus on the adaptation of the means (location parameters) $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_N$ of the proposals $q_1, \ldots, q_N$.

In this work: brief sketch - contributions

- We mix the benefits of IS and MCMC methods:
  - with MCMC $\rightarrow$ good explorative behavior.
  - with IS $\rightarrow$ easy to estimate Z.

In this work: brief sketch - contributions

- We mix the benefits of IS and MCMC methods:
  - with MCMC $\rightarrow$ good explorative behavior.
  - with IS $\rightarrow$ easy to estimate Z.
- Two layers of Monte Carlo:
  1. Upper level - MCMC adaptation: The location parameters of the proposal pdfs are updated via MCMC transitions.
  2. Lower level - IS estimation: Different weighting strategies yielding consistent IS estimators.

General LAIS algorithm

Choose $\{q_{n,0}\}_{n=1}^{N}$, $\{\boldsymbol{\mu}_{n,0}\}_{n=1}^{N}$, and the covariance matrices $\{\mathbf{C}_n\}_{n=1}^{N}$.

1. For $t = 1, \ldots, T$:
   1.1 **Adaptation:** Given $\{\boldsymbol{\mu}_{n,t-1}\}_{n=1}^{N}$ apply MCMC transitions (with invariant pdf $\bar{\pi}$), obtaining $\{\boldsymbol{\mu}_{n,t}\}_{n=1}^{N}$.
   1.2 **Generation:** Draw M samples from each proposal,

$$\mathbf{x}_{n,t}^{(m)} \sim q_{n,t}(\mathbf{x}|\boldsymbol{\mu}_{n,t}, \mathbf{C}_n),$$

   with $m = 1, \ldots, M$ and $n = 1, \ldots, N$.
   1.3 **Weighting:** Assign to each sample the weight

$$w_{n,t}^{(m)} = \frac{\pi(\mathbf{x}_{n,t}^{(m)})}{\boldsymbol{\Phi}_{n,t}(\mathbf{x}_{n,t}^{(m)})}$$

2. **Output:** Return all the pairs $\{\mathbf{x}_{n,t}^{(m)}, w_{n,t}^{(m)}\}$, for all m, n and t.

LAIS algorithms: observations

1. A specific LAIS scheme is determined by the specific choices of the MCMC strategies for adapting $\boldsymbol{\mu}_{n,t}$, and the function $\boldsymbol{\Phi}_{n,t}$.

LAIS algorithms: observations

1. A specific LAIS scheme is determined by the specific choices of the MCMC strategies for adapting $\boldsymbol{\mu}_{n,t}$, and the function $\boldsymbol{\Phi}_{n,t}$.
2. The outputs of the MCMC steps ($\boldsymbol{\mu}_{n,t}$'s) are not included in the estimators. They are only used for adapting the location parameters of the proposal pdfs.

LAIS algorithms: observations

1. A specific LAIS scheme is determined by the specific choices of the MCMC strategies for adapting $\boldsymbol{\mu}_{n,t}$, and the function $\boldsymbol{\Phi}_{n,t}$.
2. The outputs of the MCMC steps ($\boldsymbol{\mu}_{n,t}$'s) are not included in the estimators. They are only used for adapting the location parameters of the proposal pdfs.
3. Important feature: the MCMC adaptation (upper layer) is independent from the estimation part (lower layer).

LAIS algorithms: observations

1. A specific LAIS scheme is determined by the specific choices of the MCMC strategies for adapting $\boldsymbol{\mu}_{n,t}$, and the function $\boldsymbol{\Phi}_{n,t}$.
2. The outputs of the MCMC steps ($\boldsymbol{\mu}_{n,t}$'s) are not included in the estimators. They are only used for adapting the location parameters of the proposal pdfs.
3. Important feature: the MCMC adaptation (upper layer) is independent from the estimation part (lower layer).
4. Important consideration: the function $\boldsymbol{\Phi}_{n,t}$ must produce consistent IS estimators, (at least) in a static non-adaptive scenario.

Examples of adaptive strategy

- Use N parallel Metropolis-Hastings methods:

$$\{\boldsymbol{\mu}_{n,t-1}\}_{n=1}^{N} \rightarrow \{\boldsymbol{\mu}_{n,t}\}_{n=1}^{N}.$$

- For $n = 1, \ldots, N$:
  1. Draw $\boldsymbol{\mu}' \sim \varphi_n(\boldsymbol{\mu}|\boldsymbol{\mu}_{n,t-1})$,
  2. Set $\boldsymbol{\mu}_{n,t} = \boldsymbol{\mu}'$ with probability

$$\alpha = \min\left[1, \frac{\pi(\boldsymbol{\mu}')\varphi_n(\boldsymbol{\mu}_{n,t-1}|\boldsymbol{\mu}')}{\pi(\boldsymbol{\mu}_{n,t-1})\varphi_n(\boldsymbol{\mu}'|\boldsymbol{\mu}_{n,t-1})}\right]$$

  otherwise set $\boldsymbol{\mu}_{n,t} = \boldsymbol{\mu}_{n,t-1}$ (with prob. $1-\alpha$).

Examples of proper weighting strategies

- Proposal pdfs spread in time-space.

1. $\boldsymbol{\Phi}_{n,t}(\mathbf{x}) = \psi(\mathbf{x}) = \frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T} q_{n,t}(\mathbf{x})$ (full deterministic mixture),
2. $\boldsymbol{\Phi}_{n,t}(\mathbf{x}) = \xi_n(\mathbf{x}) = \frac{1}{T}\sum_{t=1}^{T} q_{n,t}(\mathbf{x})$ (partial deterministic mixture (1)) ,
3. $\boldsymbol{\Phi}_{n,t}(\mathbf{x}) = \phi_t(\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N} q_{n,t}(\mathbf{x})$ (partial deterministic mixture (2)).
4. $\boldsymbol{\Phi}_{n,t}(\mathbf{x}) = q_{n,t}(\mathbf{x})$ (standard IS).

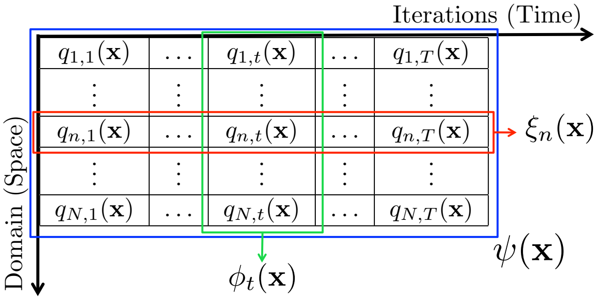


Figure: NT proposal pdfs, spread through the state space $\mathcal{X}$ $(n = 1, \ldots, N)$ and adapted over time $(t = 1, \ldots, T)$.

Choice of the weighting strategies

- All of them provide consistent estimators (in a static scenario).
- Full DM: best performance - highest computational cost.
- Partial DM (1): computational cost depending on T.
- Partial DM (2): fixed computational cost, depending on N.
- Standard IS: worst performance - lowest computational cost.

- CONSISTENCY OF THE ESTIMATORS (IN LAIS)

Consistency

- The weights are proper (if $q_{n,t}$'s have heavier tails than $\bar{\pi}$), providing consistent estimators in a static non-adaptive scenario.

Consistency

- The weights are proper (if $q_{n,t}$'s have heavier tails than $\bar{\pi}$), providing consistent estimators in a static non-adaptive scenario.
- However, the adaptation could jeopardize the consistency.

CONSISTENCY

- LAIS can always converted into a "static" IS algorithm.

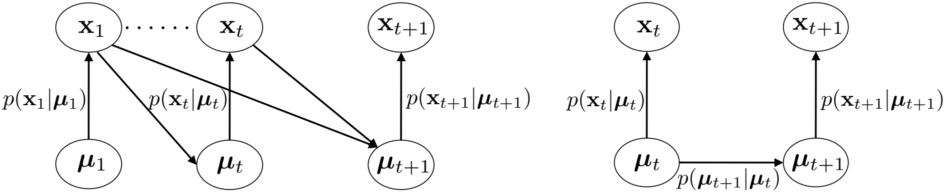
Consistency

- LAIS can always converted into a "static" IS algorithm.
- Indeed, the MCMC adaptation (upper layer) is independent from the estimation part (lower layer).
- We have described LAIS as an iterative IS method, repeating adaptation and estimation steps but:

Consistency

- LAIS can always converted into a "static" IS algorithm.
- Indeed, the MCMC adaptation (upper layer) is independent from the estimation part (lower layer).
- We have described LAIS as an iterative IS method, repeating adaptation and estimation steps but:
- We can first generate all $\{\boldsymbol{\mu}_{n,t}\}_{n=1}^{N}$ for all $t = 1, \ldots T$, and then perform the IS estimation (drawing and weighting all the $\mathbf{x}$'s).
- First all the adaptation part, then all the estimation part.

Consistency



(a) Other methods. (b) LAIS.

Figure: Graphical models: adaptation schemes.

- THEORETICAL MOTIVATION OF PROPOSED MARKOV ADAPTATION

AIS driven by MCMC, why?

- We control directly the (stationary) distribution of $\{\boldsymbol{\mu}_{n,t}\}_{n=1}^N$.
- $\{\boldsymbol{\mu}_{n,t}\}_{n=1}^N \Longrightarrow$ distributed around the modes of $\bar{\pi}$.
- We take advantage of the explorative behavior of the MCMC methods.

Prior for the location parameters

- Consider the following hierarchical procedure:
  For $n = 1, \ldots, N$:
  1. Draw $\boldsymbol{\mu}_n \sim h(\boldsymbol{\mu})$,
  2. Draw $\mathbf{x}_n \sim q(\mathbf{x}|\boldsymbol{\mu}_n, \mathbf{C})$.
- The equivalent proposal pdf is

$$\widetilde{q}(\mathbf{x}|\mathbf{C}) = \int_{\mathcal{X}} q(\mathbf{x} - \boldsymbol{\mu}|\mathbf{C}) h(\boldsymbol{\mu}) d\boldsymbol{\mu}, \quad (3)$$

i.e., $\mathbf{x}_n \sim \widetilde{q}(\mathbf{x}|\mathbf{C})$.

Hierarchical procedure in LAIS

- Prior $h(\boldsymbol{\mu}) = \bar{\pi}(\boldsymbol{\mu})$.
- MCMC kernels $K(\boldsymbol{\mu}_{n,t}|\boldsymbol{\mu}_{n,t-1})$ yielding chains which converge to $\bar{\pi}(\boldsymbol{\mu})$.
- The mixtures $\boldsymbol{\Phi}_{n,t}$ are Monte Carlo approximations of $\widetilde{q}(\mathbf{x}|\mathbf{C})$.

Hierarchical procedure in LAIS

- Prior $h(\boldsymbol{\mu}) = \bar{\pi}(\boldsymbol{\mu})$.
- MCMC kernels $K(\boldsymbol{\mu}_{n,t}|\boldsymbol{\mu}_{n,t-1})$ yielding chains which converge to $\bar{\pi}(\boldsymbol{\mu})$.
- The mixtures $\boldsymbol{\Phi}_{n,t}$ are Monte Carlo approximations of $\widetilde{q}(\mathbf{x}|\mathbf{C})$.

- The prior $h(\boldsymbol{\mu}) = \bar{\pi}(\boldsymbol{\mu})$ is not *optimal*.

Hierarchical procedure in LAIS

- Prior $h(\boldsymbol{\mu}) = \bar{\pi}(\boldsymbol{\mu})$.
- MCMC kernels $K(\boldsymbol{\mu}_{n,t}|\boldsymbol{\mu}_{n,t-1})$ yielding chains which converge to $\bar{\pi}(\boldsymbol{\mu})$.
- The mixtures $\boldsymbol{\Phi}_{n,t}$ are Monte Carlo approximations of $\widetilde{q}(\mathbf{x}|\mathbf{C})$.

- The prior $h(\boldsymbol{\mu}) = \bar{\pi}(\boldsymbol{\mu})$ is not *optimal*.
- But it can justify using a kernel density estimation (KDE) argument:

Hierarchical procedure in LAIS

- Prior $h(\boldsymbol{\mu}) = \bar{\pi}(\boldsymbol{\mu})$.
- MCMC kernels $K(\boldsymbol{\mu}_{n,t}|\boldsymbol{\mu}_{n,t-1})$ yielding chains which converge to $\bar{\pi}(\boldsymbol{\mu})$.
- The mixtures $\boldsymbol{\Phi}_{n,t}$ are Monte Carlo approximations of $\widetilde{q}(\mathbf{x}|\mathbf{C})$.

- The prior $h(\boldsymbol{\mu}) = \bar{\pi}(\boldsymbol{\mu})$ is not *optimal*.
- But it can justify using a kernel density estimation (KDE) argument:
  - when $h(\boldsymbol{\mu}) = \bar{\pi}(\boldsymbol{\mu})$, $\widetilde{q}$ is a KDE of $\bar{\pi}$.
  - there exists an optimal scale parameter $\mathbf{C}^*$ such that $\widetilde{q}(\mathbf{x}|\mathbf{C}^*)$ is unbiased estimator of $\bar{\pi}(\mathbf{x})$.

- NUMERICAL SIMULATIONS

Multimodal target distribution

- Consider the target pdf

$$\bar{\pi}(\mathbf{x}) = \frac{1}{5}\sum_{i=1}^{5} \mathcal{N}(\mathbf{x}; \nu_i, \mathbf{\Sigma}_i), \quad \mathbf{x} \in \mathbb{R}^2, \tag{4}$$

with means $\nu_1 = [-10, -10]^\top$, $\nu_2 = [0, 16]^\top$, $\nu_3 = [13, 8]^\top$, $\nu_4 = [-9, 7]^\top$, $\nu_5 = [14, -14]^\top$, and covariance matrices $\mathbf{\Sigma}_1 = [2,\ 0.6; 0.6,\ 1]$, $\mathbf{\Sigma}_2 = [2,\ -0.4; -0.4,\ 2]$, $\mathbf{\Sigma}_3 = [2,\ 0.8; 0.8,\ 2]$, $\mathbf{\Sigma}_4 = [3,\ 0; 0,\ 0.5]$ and $\mathbf{\Sigma}_5 = [2,\ -0.1; -0.1,\ 2]$.

- The main challenge is the ability in discovering the 5 different modes of $\bar{\pi}(\mathbf{x}) \propto \pi(\mathbf{x})$.
- Since we know the moments of $\bar{\pi}(\mathbf{x})$ (in this toy example), we can easily compare the performance of the different techniques.
- We consider the problem of approximating via Monte Carlo the expected value $E[\mathbf{X}] = [1.6, 1.4]^\top$ and the normalizing constant $Z = 1$.

Proposal densities

- We compare LAIS with different alternative methods (using the same number of target evaluations).
- We use Gaussian proposal densities for all the techniques: for the IS estimation (lower layer of LAIS), we have

$$q_{n,t}(\mathbf{x}|\boldsymbol{\mu}_{n,t}, \mathbf{C}_n) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_{n,t}, \mathbf{C}_n),$$

  with covariance matrices $\mathbf{C}_n = \sigma^2 \mathbf{I}_2$ and $\sigma \in \{0.5, 1, 2, 5, 10, 20, 70\}$.
- For the upper layer of LAIS (adaptation), we consider

$$\varphi_n(\mathbf{x}|\boldsymbol{\mu}_{n,t}, \boldsymbol{\Lambda}_n) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_{n,t}, \boldsymbol{\Lambda}_n),$$

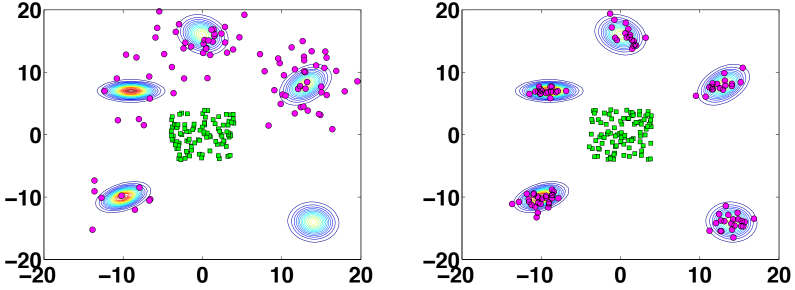
  with $\boldsymbol{\Lambda}_n = \lambda^2 \mathbf{I}_2$ and $\lambda \in \{5, 10, 70\}$.

Multimodal target distribution

Algorithm			$\sigma = 0.5$	$\sigma = 1$	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$	$\sigma = 70$
LAIS ($N = 100$)	$\lambda = 5$	$M = 99,\ T = 20$	1.2760	0.5219	0.5930	0.0214	0.0139	0.1815
		$M = 19,\ T = 100$	0.2361	0.1205	0.0422	0.0087	0.0140	0.1868
		$M = 1,\ T = 1000$	0.1719	0.0019	0.0155	0.0103	0.0273	0.3737
	$\lambda = 10$	$M = 99,\ T = 20$	1.0195	0.1546	0.2876	0.0178	0.0133	0.1789
		$M = 19,\ T = 100$	0.1750	0.0120	0.0528	0.0086	0.0136	0.1856
		$M = 1,\ T = 1000$	0.1550	**0.0021**	**0.0020**	0.0095	0.0252	0.3648
	$\lambda = 70$	$M = 99,\ T = 20$	16.9913	5.5790	1.4925	0.0382	0.0128	0.1834
		$M = 19,\ T = 100$	2.6693	0.9182	0.1312	0.0147	0.0143	0.1844
		$M = 1,\ T = 1000$	0.3014	0.1042	0.0136	0.0115	0.0267	0.3697
	$\lambda_{n,j} \sim \mathcal{U}([1, 10])$	$M = 99,\ T = 20$	1.0707	0.5364	0.3523	0.0199	**0.0121**	0.1919
		$M = 19,\ T = 100$	0.2481	0.0595	0.1376	**0.0075**	0.0144	0.1899
		$M = 1,\ T = 1000$	**0.1046**	0.0037	0.0045	0.0099	0.0274	0.3563
AMIS	(best results)		124.22	121.21	100.23	0.8640	**0.0121**	**0.0136**
	(worst results)		125.43	123.38	114.82	16.92	0.0128	18.66
PMC	$N = 100,\ T = 2000$		112.99	114.11	47.97	2.34	0.0559	2.41
variant-PMC			111.92	107.58	26.86	0.6731	0.0744	2.42
Mixture PMC			110.17	113.11	50.23	2.75	0.0521	2.57

Table: MSE obtained by different methods with the same number of evaluations of the target pdf.

LAIS ADAPTATION VERSUS PMC ADAPTATION



(a) PMC ($N = 100$, $\sigma = 5$)

(b) LAIS ($N = 100$, $\lambda = 5$)

FIGURE: Initial (squares) and final (circles) configurations of the location parameters of the proposal densities for the standard PMC and the PI-MAIS methods, in a specific run.

Conclusions

- LAIS is a very efficient algorithm which combines the benefits of MCMC and IS.

Conclusions

- LAIS is a very efficient algorithm which combines the benefits of MCMC and IS.
- Tested in different scenarios/applications (with dimension until 80); LAIS outperforms state-of-the-art methods.

Conclusions

- LAIS is a very efficient algorithm which combines the benefits of MCMC and IS.
- Tested in different scenarios/applications (with dimension until 80); LAIS outperforms state-of-the-art methods.
- LAIS works particularly well addressing multimodal posterior distributions.

Conclusions

- LAIS is a very efficient algorithm which combines the benefits of MCMC and IS.
- Tested in different scenarios/applications (with dimension until 80); LAIS outperforms state-of-the-art methods.
- LAIS works particularly well addressing multimodal posterior distributions.
- We obtain similar results only using additional information about π, like the gradient.

Conclusions

- LAIS is a very efficient algorithm which combines the benefits of MCMC and IS.
- Tested in different scenarios/applications (with dimension until 80); LAIS outperforms state-of-the-art methods.
- LAIS works particularly well addressing multimodal posterior distributions.
- We obtain similar results only using additional information about π, like the gradient.
- We are working in order to provide a "clean" and optimized free-code in Matlab and R.

- Thank you very much!
- Any questions?

Main references

[Owen00]: A.Owen, Y.Zhou. **Safe and effective importance sampling**. *Journal of the American Statistical Association*, 95 (449):135-143. 2000.

[Elvira15]: V. Elvira, L. Martino, D. Luengo, and M. Bugallo. **Efficient multiple importance sampling estimators**. *IEEE Signal Processing Letters*, 22 (10):1757-1761, 2015.

[Cornuet12]: J.M. Cornuet, J.M. Marin, A. Mira, C.P. Robert. **Adaptive multiple importance sampling**. *Scandinavian Journal of Statistics*, 39 (4):798-812, 2012.

[Martino15]: L. Martino, V. Elvira, D. Luengo, J. Corander. **An adaptive population importance sampler: Learning from the uncertainty**. *IEEE Transactions on Signal Processing* (In Press), 2015.

[Cappe04]: O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert. **Population Monte Carlo**. *Journal of Computational and Graphical Statistics*, 13 (4):907-929, 2004.