EUSIPCO 2013 1569744203

ROBUST MIXTURE POPULATION MONTE CARLO SCHEME WITH ADAPTATION OF THE NUMBER OF COMPONENTS

Eugenia Koblents, Joaquín Míguez

Universidad Carlos III de Madrid, Department of Signal Theory and Communications
Avda. de la Universidad 30, 28911 Leganés, Spain, E-mail: ekoblents,jmiguez@tsc.uc3m.es

ABSTRACT

We address the Monte Carlo approximation of probability distributions in high-dimensional spaces. In particular, we investigate the population Monte Carlo (PMC) scheme, which is based on an iterative importance sampling approach, and its extension the mixture-PMC method (MPMC), which models the importance functions as mixtures of kernels. We propose an extension of the MPMC method which incorporates adaptation of the number of mixture components, and applies a nonlinear transformation to the importance weights in order to smooth their variations and avoid degeneracy problems. We present numerical results that illustrate the performance improvement attained by the new method.

Index Terms— Importance sampling, population Monte Carlo, mixture-PMC

1. INTRODUCTION

Computational inference in high-dimensional spaces is a challenging problem. Various techniques based on the Monte Carlo methodology [1] have been successfully applied to a large variety of complex problems. In this work we investigate the population Monte Carlo (PMC) scheme [2]. PMC algorithms perform iterative importance sampling (IS) by adapting the proposal probability density functions (pdf's, also known as importance functions) according to the latest samples and importance weights (IW's) available, so that they "approach" the static target pdf over the iterations. PMC methods are sensitive to the selection of the proposal pdf and may often perform poorly even in low-dimensional problems.

A well known extension of PMC is the mixture-PMC (MPMC) algorithm presented in [3], which constructs the sequence of importance functions as mixtures of kernels. In [3], a set of rules are put forward for selecting the parameters of the importance function that minimize the Kullback-Leibler divergence (KLD) between the target and the proposal pdf at each iteration. The rules are developed for mixtures of Gaussian and Student's *t* distributions.

In [4], a simple modification of the PMC scheme was proposed that consists in updating the proposal density based on nonlinearly transformed IWs (TIWs). This single modification can increase the efficiency of the method drastically. In [5], a MPMC algorithm with TIW's has also been investigated, showing a considerable improvement in performance with respect to the original scheme.

In this paper we introduce an extension of the MPMC that incorporates both nonlinear transformations to the IWs, in order to mitigate weight degeneracy, and an adaptation step to dynamically select the number of mixture components in the proposal pdf's. This modification of the MPMC scheme provides valuable information about the number of components required to represent the target pdf and can also alleviate the computational demands of the algorithm (as it is simpler to draw samples from mixtures with less components). We have applied MPMC and the proposed extension to the approximation of the banana-like target density [6], which is relevant in inference problems of cosmological parameters. Recently, the MPMC has been widely applied to the problem of parameter estimation in cosmological applications [6] and is the base of the tool CosmoPMC [7].

The rest of the paper is organized as follows. In Section 2, we review IS and the PMC method. In Section 2.2 we describe the MPMC method and its Gaussian and Student's *t* extensions. In Section 3 we propose a MPMC algorithm with TIWs and adaptation of the number of mixture components. In Section 4 we present numerical results. Finally, in Section 5 we summarize and discuss the contributions of this paper.

2. BACKGROUND

Let $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_K]^\top$ be a vector of K real random variables with pdf $\pi(\boldsymbol{\theta})$, termed the *target* pdf. The Monte Carlo framework allows to approximate the probability measure $\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ and its moments, by means of empirical sums, i.e.,

$$E_\pi[f(\boldsymbol{\theta})] = \int f(\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \approx \frac{1}{M}\sum_{i=1}^{M} f(\boldsymbol{\theta}^{(i)}),$$

E. K. acknowledges the support of *Ministerio de Educación, Cultura y Deporte* of Spain (*Programa de Formación de Profesorado Universitario*, ref. AP2008-00469). This work has been partially supported by *Ministerio de Economía y Competitividad* of Spain (program Consolider-Ingenio 2010 CSD2008-00010 COMONSENS, project DEIPRO TEC2009-14504-C02-01 and project COMPREHENSION TEC2012-38883-C02-01).

where $f : \mathbb{R}^K \rightarrow \mathbb{R}$ is a real, integrable function of $\boldsymbol{\theta}$, $E_\pi[f(\boldsymbol{\theta})]$ denotes the expectation of f with respect to $\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ and $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^M$ is a random i.i.d. (independent and identically distributed) sample drawn from $\pi(\boldsymbol{\theta})$.

2.1. Importance sampling

In many practical cases it is not possible to sample from $\pi(\boldsymbol{\theta})$ directly. The importance sampling [1] approach consists in drawing the samples $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^M$ from a (simpler) proposal pdf, or importance function, $q(\boldsymbol{\theta})$, and then computing normalized IW's as $w^{(i)} \propto w^{(i)*} = \pi(\boldsymbol{\theta}^{(i)})/q(\boldsymbol{\theta}^{(i)})$ with $\sum_{i=1}^M w^{(i)} = 1$. The integral $E_\pi[f(\boldsymbol{\theta})]$ is then approximated by the weighted sum $E_\pi[f(\boldsymbol{\theta})] \approx \sum_{i=1}^M w^{(i)} f(\boldsymbol{\theta}^{(i)})$.

In order to ensure the asymptotic convergence of this approximation when M is large enough, it is sufficient to select $q(\boldsymbol{\theta})$ such that $q(\boldsymbol{\theta}) > 0$ whenever $\pi(\boldsymbol{\theta}) > 0$, and guarantee that $q(\boldsymbol{\theta})$ has heavier tails than $\pi(\boldsymbol{\theta})$ [1].

2.2. Mixture population Monte Carlo algorithm

The PMC method [2] is an iterative IS scheme that generates a sequence of proposal pdfs $q_\ell(\boldsymbol{\theta})$, $\ell = 1, \ldots, L$, such that every new proposal is "closer" to the target density $\pi(\boldsymbol{\theta})$ than the previous importance function, based on the set of samples and weights at the $(\ell-1)$-th iteration ($\ell \geq 2$).

The original PMC scheme in [2], however, does not provide a universal update scheme for the proposals $q_\ell(\boldsymbol{\theta})$, $\ell = 1, ..., L$. A powerful extension is the mixture-PMC algorithm [3], which constructs the sequence of proposal pdf's as mixtures of D kernels of the form

$$q_\ell(\boldsymbol{\theta}) = \sum_{d=1}^{D} \alpha_{\ell,d} q_{\ell,d}(\boldsymbol{\theta}; \boldsymbol{\beta}_{\ell,d}), \tag{1}$$

where the mixture weights $\alpha_{\ell,d}$ and the kernel parameters $\boldsymbol{\beta}_{\ell,d}$ of each component are adapted along the iterations in order to minimize the KLD between the target and the proposal pdf, i.e., $\mathcal{D}(\pi||q_\ell) = \int \log(\frac{\pi(\boldsymbol{\theta})}{q_\ell(\boldsymbol{\theta})})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$. In [3], updating rules for the parameters $\boldsymbol{\beta}_{\ell,d}$ are provided for the case of mixtures of Gaussian and Student's t distributions. The MPMC algorithm is outlined in Table 1. In [6] the authors suggest to discard those mixture components with a very low weight $\alpha_{\ell,d}$, to avoid numerical problems in the updating of the parameters.

2.2.1. Gaussian mixture importance functions

Assume that the proposal pdf $q_\ell(\boldsymbol{\theta})$ at iteration ℓ is a mixture of D, K-dimensional Gaussian kernels of the form

$$q_{\ell,d}(\boldsymbol{\theta}; \boldsymbol{\beta}_{\ell,d}) = \mathcal{N}_K(\boldsymbol{\theta}; \boldsymbol{\mu}_{\ell,d}, \boldsymbol{\Sigma}_{\ell,d}), \quad d = 1, \ldots, D,$$

where $\boldsymbol{\mu}_{\ell,d}$ and $\boldsymbol{\Sigma}_{\ell,d}$ are the mean vector and covariance matrix of each component, respectively. These parameters are updated for the next iteration $\ell + 1$ as [3]

$$\boldsymbol{\mu}_{\ell+1,d} = \frac{\sum_{i=1}^M w_\ell^{(i)} \rho_{\ell,d}^{(i)} \boldsymbol{\theta}_\ell^{(i)}}{\alpha_{\ell+1,d}} \quad \text{and}$$

$$\boldsymbol{\Sigma}_{\ell+1,d} = \frac{\sum_{i=1}^M w_\ell^{(i)} \rho_{\ell,d}^{(i)} (\boldsymbol{\theta}_\ell^{(i)} - \boldsymbol{\mu}_{\ell+1,d})(\boldsymbol{\theta}_\ell^{(i)} - \boldsymbol{\mu}_{\ell+1,d})^\top}{\alpha_{\ell+1,d}}.$$

Table 1. Mixture-PMC algorithm [3]

Iteration ($\ell = 1, \ldots, L$)**:**

1. Generate an i.i.d. sample $\{\boldsymbol{\theta}_\ell^{(i)}\}_{i=1}^M$ from the current mixture proposal $q_\ell(\boldsymbol{\theta})$ in Eq. (1).

2. For $i = 1, \ldots, M$, compute normalized IWs $w_\ell^{(i)} \propto \pi(\boldsymbol{\theta}_\ell^{(i)})/q_\ell(\boldsymbol{\theta}_\ell^{(i)})$ and mixture posterior probabilities

$$\rho_{\ell,d}^{(i)} = \frac{\alpha_{\ell,d} q_{\ell,d}(\boldsymbol{\theta}_\ell^{(i)}; \boldsymbol{\beta}_{\ell,d})}{\sum_{k=1}^D \alpha_{\ell,k} q_{\ell,k}(\boldsymbol{\theta}_\ell^{(i)}; \boldsymbol{\beta}_{\ell,k})}. \tag{2}$$

3. For $d = 1, \ldots, D$, update the weights and the parameters of each component as

$$\alpha_{\ell+1,d} = \sum_{i=1}^M w_\ell^{(i)} \rho_{\ell,d}^{(i)} \quad \text{and} \tag{3}$$

$$\boldsymbol{\beta}_{\ell+1,d} = \arg\max_{\boldsymbol{\beta}_{\ell,d}} \left[\sum_{i=1}^M w_\ell^{(i)} \rho_{\ell,d}^{(i)} \log q_{\ell,d}(\boldsymbol{\theta}_\ell^{(i)}; \boldsymbol{\beta}_{\ell,d}) \right].$$

2.2.2. Student's t mixture importance functions

The *t* mixture has been suggested for importance sampling, opposite to the Gaussian mixture, because its heavier tails may capture a wider range of non-Gaussian targets with a smaller number of components.

Thus, assume that the proposal pdf at iteration ℓ is a mixture of D, K-dimensional Student's *t* kernels (with a fixed number ν_d of degrees of freedom) denoted

$$q_{\ell,d}(\boldsymbol{\theta}; \boldsymbol{\beta}_{\ell,d}) = \mathcal{T}_K(\boldsymbol{\theta}; \boldsymbol{\mu}_{\ell,d}, \boldsymbol{\Sigma}_{\ell,d}, \nu_d), \quad d = 1, ..., D.$$

We compute $\gamma_{\ell,d}^{(i)} = \frac{\nu_d + K}{\nu_d + (\boldsymbol{\theta}_\ell^{(i)} - \boldsymbol{\mu}_{\ell,d})^\top \boldsymbol{\Sigma}_{\ell,d}^{-1} (\boldsymbol{\theta}_\ell^{(i)} - \boldsymbol{\mu}_{\ell,d})}$ and the mean and covariance parameters are updated as [3]

$$\boldsymbol{\mu}_{\ell+1,d} = \frac{\sum_{i=1}^M w_\ell^{(i)} \rho_{\ell,d}^{(i)} \gamma_{\ell,d}^{(i)} \boldsymbol{\theta}_\ell^{(i)}}{\sum_{i=1}^M w_\ell^{(i)} \rho_{\ell,d}^{(i)} \gamma_{\ell,d}^{(i)}} \quad \text{and}$$

$$\boldsymbol{\Sigma}_{\ell+1,d} = \frac{\sum_{i=1}^M w_\ell^{(i)} \rho_{\ell,d}^{(i)} \gamma_{\ell,d}^{(i)} (\boldsymbol{\theta}_\ell^{(i)} - \boldsymbol{\mu}_{\ell+1,d})(\boldsymbol{\theta}_\ell^{(i)} - \boldsymbol{\mu}_{\ell+1,d})^\top}{\alpha_{\ell+1,d}}.$$

3. ADAPTIVE NONLINEAR MIXTURE-PMC

In this section we introduce a modification of the MPMC method that incorporates an adaptation mechanism for the number of mixture components, and performs nonlinear transformations to the IW's in order to mitigate the weight degeneracy phenomenon. The new algorithm is termed adaptive nonlinear-MPMC (NMPMC).

3.1. Adaptation of the number of components

The MPMC algorithm assumes a fixed number of components D (which needs to be overestimated in general), hence the final outcome of the algorithm does not provide any information about the number of components required to adequately approximate a target pdf $\pi(\boldsymbol{\theta})$. In this paper we propose an extension of the MPMC which incorporates an update step of the number of components D, along the iterations. We consider an initial number of components D_1 and perform pruning and merging operations to the mixture components, reducing D_ℓ over the iterations $\ell = 2, ..., L$.

The pruning operation consists in removing the d-th mixture component when its associated weight falls below a prescribed threshold μ_{prn}, i.e., $\alpha_{\ell+1,d} < \mu_{prn}$, as suggested in [6]. The merging operation allows to fuse two similar mixture components $q_{\ell+1,i}$ and $q_{\ell+1,j}$ when the distance $\mathcal{D}_{i,j} = \mathcal{D}(q_{\ell+1,i}||q_{\ell+1,j}) + \mathcal{D}(q_{\ell+1,j}||q_{\ell+1,i})$ is less than a second threshold μ_{mrg}. The parameters of the resulting component are obtained as the average of the parameters of the original components. The KLD can be computed exactly in the case of Gaussian mixtures, and can be approximated by exact Monte Carlo sampling in the case of *t* mixtures. Up to one merging and any number of pruning operations are performed at each iteration of the algorithm. The thresholds μ_{prn} and μ_{mrg} are set a priori.

3.2. Nonlinear transformation of the IWs

IS and PMC methods usually suffer from degeneracy of the IWs, which present extreme variations leading to a very low number of "effective" samples (those with non-negligible IW's). In order to avoid this problem, we perform a nonlinear transformation of the IWs by "clipping" them (as described in detail in [4, 5]). Choosing an integer $M_T < M$, the unnormalized TIWs $\bar{w}_\ell^{(i)*}$ are computed as

$$\bar{w}_\ell^{(i)*} = \min(w_\ell^{(i)*}, \mathcal{T}_\ell^{M_T}), \quad i = 1, \ldots, M, \qquad (4)$$

where $w_\ell^{(i)*} = \pi(\boldsymbol{\theta}_\ell^{(i)})/q_\ell(\boldsymbol{\theta}_\ell^{(i)})$ is an unnormalized IW and the threshold value $\mathcal{T}_\ell^{M_T}$ corresponds to the M_T-th highest IW[1]. This transformation leads to flat TIW's in the region of interest of $\boldsymbol{\theta}$, and guarantees a baseline of M_T effective samples. The proposed method is outlined in Table 2.

[1]Let $i_1, ..., i_M$ be an ordering of the samples at the ℓ-th iteration such that $w_\ell^{(i_1)*} \geq w_\ell^{(i_2)*} \geq \cdots \geq w_\ell^{(i_M)*}$. Then, $\mathcal{T}_\ell^{M_T} = w_\ell^{(i_{M_T})*}$.

Table 2. Adaptive nonlinear MPMC algorithm

Iteration ($\ell = 1, \ldots, L$)**:**

1. Generate a sample $\{\boldsymbol{\theta}_\ell^{(i)}\}_{i=1}^M$ from the current mixture proposal $q_\ell(\boldsymbol{\theta})$ in Eq. (1) with $D = D_\ell$ components.
2. For $i = 1, \ldots, M$, compute unnormalized IWs $w_\ell^{(i)*} = \pi(\boldsymbol{\theta}_\ell^{(i)})/q_\ell(\boldsymbol{\theta}_\ell^{(i)})$ and mixture posterior probabilities $\rho_{\ell,d}^{(i)}$ as in Eq. (2), with $D = D_\ell$.
3. For $i = 1, \ldots, M$, compute unnormalized TIW's using Eq. (4) and normalize them as $\bar{w}_\ell^{(i)} = \bar{w}_\ell^{(i)*} / \sum_{j=1}^M \bar{w}_\ell^{(j)*}$.
4. Update the component weights $\alpha_{\ell+1,d}$, $d = 1, ..., D_\ell$, and parameters $\boldsymbol{\beta}_{\ell+1,d}$ of each component according to Eq. (3), but using TIW's $\bar{w}_\ell^{(i)}$ instead of IW's $w_\ell^{(i)}$.
5. Set $\tilde{D} = D_\ell$. Compute the distance $\mathcal{D}_{i,j}$ between each pair of mixture components $q_{\ell+1,i}$ and $q_{\ell+1,j}$, for $i, j = 1, \ldots, D_\ell$.

   If $\mathcal{D}_{i,j} < \mu_{mrg}$, merge components i and j. The overall weight es computed as $\alpha_{\ell+1,i} = \alpha_{\ell+1,i} + \alpha_{\ell+1,j}$ and the parameters as $\boldsymbol{\mu}_{\ell+1,i} = \boldsymbol{\mu}_{\ell+1,i}/2 + \boldsymbol{\mu}_{\ell+1,j}/2$ and $\boldsymbol{\Sigma}_{\ell+1,i} = \boldsymbol{\Sigma}_{\ell+1,i}/2 + \boldsymbol{\Sigma}_{\ell+1,j}/2$. Remove the j-th component setting $\alpha_{\ell+1,j} = 0$ and $\tilde{D} = \tilde{D} - 1$.
6. For $i = 1, \ldots, \tilde{D}$, if $\alpha_{\ell+1,i} < \mu_{prn}$, remove the i-th component setting $\alpha_{\ell+1,i} = 0$ and $\alpha_{\ell+1,j} = \alpha_{\ell+1,j} / \sum_{k=1}^{\tilde{D}} \alpha_{\ell+1,k}$, $j = 1, \ldots, \tilde{D}$.
7. Update $\tilde{D}$ according to the number of pruned components and set $D_{\ell+1} = \tilde{D}$.

4. COMPUTER SIMULATIONS

To illustrate the performance of the original MPMC and the new adaptive NMPMC, we apply both schemes to the approximation of a 10-dimensional target pdf $\pi(\boldsymbol{\theta})$, by means of a mixture of Gaussian and Student's *t* kernels.

4.1. Simulation setup

Following [6], we consider a target pdf $\pi(\boldsymbol{\theta})$ constructed from a Gaussian pdf $\pi(\boldsymbol{\theta}') = \mathcal{N}_{10}(\boldsymbol{\theta}'; \mathbf{0}, \boldsymbol{\Sigma})$ with covariance matrix $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, 1, \ldots, 1)$. The variable of interest $\boldsymbol{\theta}$ is constructed from the auxiliary variable $\boldsymbol{\theta}'$ by twisting the second coordinate according to $\theta_2 = \theta_2' - \beta(\theta_1'^2 - \sigma_1^2)$ and keeping the rest of the variables unchanged, i.e.,

$$\boldsymbol{\theta} = [\theta_1', \theta_2' - \beta(\theta_1'^2 - \sigma_1^2), \theta_3', \ldots, \theta_{10}']^\top.$$

We assume that the twist parameter is $\beta = 0.03$ and $\sigma_1^2 = 100$. This transformation results in a banana-shaped density

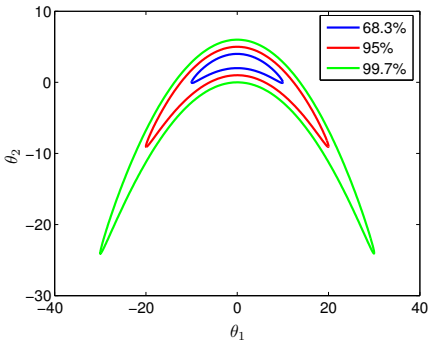

Fig. 1. Contour plot of the marginal target density $\pi(\theta_1, \theta_2)$.

in the first two dimensions, which is represented in Figure 1. This pdf is difficult to explore and provides a realistic scenario in many cosmological problems.

We have applied the MPMC and the adaptive NMPMC to the approximation of the described target pdf with importance functions built as Gaussian and *t* mixtures. In all the simulations, we consider an initial proposal pdf consisting of $D_1 = 10$ components $q_{1,d}$, $d = 1, ..., D_1$, with random mean vectors $\boldsymbol{\mu}_{1,d} \sim \mathcal{N}_{10}(\boldsymbol{\mu}_{1,d}; \mathbf{0}, \boldsymbol{\Sigma}_0/5)$, and a common covariance matrix $\boldsymbol{\Sigma}_{1,d} = \boldsymbol{\Sigma}_0$, where $\boldsymbol{\Sigma}_0 = \text{diag}(200, 50, 4, \ldots, 4)$. In the case of *t* mixtures, the number of degrees of freedom has been set to $\nu_d = 9$, for every d.

In each simulation run we have computed the normalized effective sample size (NESS) at all iterations as $M_\ell^e = [M \sum_{i=1}^M (w_\ell^{(i)})^2]^{-1}$ and $\bar{M}_\ell^e = [M \sum_{i=1}^M (\bar{w}_\ell^{(i)})^2]^{-1}$ for the MPMC and the NMPMC schemes, respectively. A NESS value close to 1 suggests a good agreement between the proposal and the target pdf. However, it is generally not enough to establish that the algorithm has converged to the true target, for example, when the target presents multiple modes.

As a measure of how well a set of samples $\{\boldsymbol{\theta}_\ell^{(i)}\}_{i=1}^M$ drawn from the mixture proposal pdf $q_\ell(\boldsymbol{\theta})$ represents the target density $\pi(\boldsymbol{\theta})$ we have computed the KLD between the corresponding Gaussian target pdf $\pi(\boldsymbol{\theta}')$ and the Gaussian approximation of the untwisted sample set $\{\boldsymbol{\theta}_\ell^{'(i)}\}_{i=1}^M$, obtained by the inverse transformation $\theta_{\ell,2}^{'(i)} = \theta_{\ell,2}^{(i)} + \beta[(\theta_{\ell,1}^{(i)})^2 - \sigma_1^2]$.

We have performed 10^4 independent simulation runs of each algorithm, both in the Gaussian and *t* cases. We have studied two settings with $L = 20$ iterations and a different number of samples per iteration, M.

4.2. Large sample size

The number of samples per iteration has been set to $M = 10^4$. The threshold parameter for the removal of a mixture component has been set to $\mu_{prn} = 0.002$. In the adaptive NMPMC scheme, the threshold parameter for the fusion of two components has been set to $\mu_{mrg} = 3$, and the clipping parameter to $M_T = 100$ samples.

In Figure 2 the evolution of the median KLD (*left*) and mean NESS (*center*) for the MPMC and the adaptive NMPMC algorithms are plotted, for the Gaussian and *t* cases. The median has been preferred to the mean because of its robustness against outliers. The adaptive NMPMC scheme obtains a lower KLD and a higher NESS, with both mixture families. The Gaussian mixture provides better results, in terms of KLD and NESS, than the *t* mixture for both algorithms. This occurs because samples drawn from the tails of the *t* components are not usually representative and obtain low IWs. The evolution of the number of components D_ℓ (Figure 2, *right*) is similar in all the schemes, converging in average to a value between 6 and 7.

In Table 3 statistics on the values of KLD, NESS and D_ℓ after the last iteration $\ell = L$ are displayed. It can be observed that the original MPMC schemes present an extremely high KLD variance and also a higher variance of NESS and D_ℓ than the proposed NMPMC techniques.

4.3. Reduced sample size

In this case, the number of samples per iteration has been set to $M = 2000$. The threshold parameters have been set to $\mu_{prn} = 0.01$ and $\mu_{mrg} = 2$, and $M_T = 100$ samples.

Figure 3 displays the results obtained in this setting. In this scenario, the MPMC performs poorly with both mixture families, obtaining an increasing KLD, a NESS close to 0 and a mean D_L close to 1. On the contrary, the proposed NMPMC performs similarly to the $M = 10^4$ case, with a slightly higher KLD due to the fact that with a lower number of samples, the tails of the target pdf are less accurately represented. The number of components D_ℓ still attains a similar final value and the NESS converges to a high value as well. In Table 4 statistics on the final values of the KLD, NESS and D_ℓ for the NMPMC are also displayed for comparison. It can be seen that the proposed scheme presents stable results even in this low data scenario.

5. CONCLUSIONS

In this paper we have addressed the problem of Monte Carlo approximation of target probability distributions in high dimensional spaces. We have studied the recently proposed mixture-PMC scheme and proposed an extension which is stable (in terms of the NESS) in high-dimensional setups and provides information about the number of components required to adequately represent the pdf of interest. We have compared the performance of the original and the proposed schemes in the cases of Gaussian and *t* mixtures, in two scenarios with a different number of samples (hence, with a different computational effort). We present numerical results that show that the proposed scheme clearly outperforms the original one. We also show that the Gaussian mixture should be preferred for this problem.

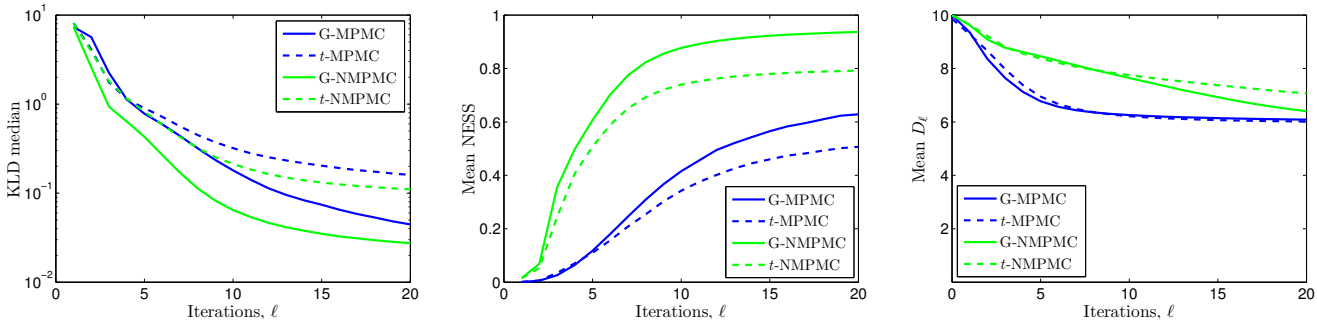

Fig. 2. Performance of MPMC vs NMPMC, with $M = 10^4$, Gaussian and t mixtures (labeled as G- and t-, respectively). Median KLD (*left*), mean NESS (*center*) and mean number of mixture components D_ℓ (*right*) along the iterations.

	Med KLD	Mean KLD	Std KLD	Mean NESS	Std NESS	Mean D_L	Std D_L
G-MPMC	0.0445	$1.63 \cdot 10^5$	$1.47 \cdot 10^7$	0.6286	0.2699	6.084	2.242
t-MPMC	0.1601	$1.06 \cdot 10^{26}$	$5.76 \cdot 10^{27}$	0.5069	0.2492	6.013	2.501
G-NMPMC	0.0275	0.0307	0.0139	0.9370	0.0176	6.401	1.167
t-NMPMC	0.1106	0.1147	0.0254	0.7923	0.0148	7.065	1.310

Table 3. Median, mean and standard deviation for KLD, NESS and D_ℓ, for MPMC and NMPMC with $M = 10^4$ and $\ell = L$.

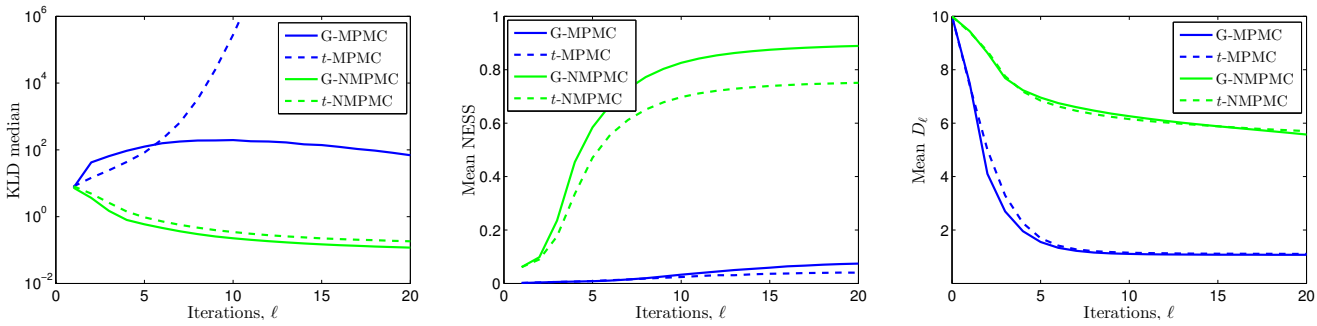

Fig. 3. MPMC vs NMPMC with $M = 2000$. Median KLD (*left*), mean NESS (*center*) and mean D_ℓ (*right*) along the iterations.

	Med KLD	Mean KLD	Std KLD	Mean NESS	Std NESS	Mean D_L	Std D_L
G-NMPMC	0.1182	0.1287	0.0514	0.8892	0.0153	5.573	1.276
t-NMPMC	0.1832	0.1949	0.0627	0.7512	0.0142	5.705	1.295

Table 4. Median, mean and standard deviation for KLD, NESS and D_ℓ, for NMPMC with $M = 2000$ and $\ell = L$.

6. REFERENCES

[1] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, 2004.

[2] O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert, "Population Monte Carlo," *Computational and Graphical Statistics*, vol. 13, no. 4, pp. 907–929, 2004.

[3] O. Cappé, R. Douc, A. Guillin, J. M. Marin, and C. P. Robert, "Adaptive importance sampling in general mixture classes," *Statistics and Computing*, vol. 18, no. 4, pp. 447–459, 2008.

[4] E. Koblents and J. Míguez, "A population Monte Carlo scheme with transformed weights and its application to stochastic kinetic models," *arXiv preprint arXiv:1208.5600*, 2012.

[5] E. Koblents and J. Míguez, "A population Monte Carlo scheme for computational inference in high dimensional spaces," ICASSP, 2013.

[6] D. Wraith et al, "Estimation of cosmological parameters using adaptive importance sampling," *Physical Review D*, vol. 80, no. 2, pp. 023507, 2009.

[7] M. Kilbinger et al, "CosmoPMC: Cosmology population Monte Carlo," *arXiv preprint arXiv:1101.0950v3*, 2012.