

1 Introduction

Monte Carlo methods have quickly become one of the most common methods of conducting statistical analysis. This was brought on partly due to the ease of access to powerful computing which previously made these methods unreachable from the layman. However, now that everyone is able to do basic MCMC, the problems researchers bring have become increasingly more difficult to analyze. In this review, we will be looking at the paper, *Particle Markov Chain Monte Carlo Methods*, by Andrieu et al.(2010)?? which provides a way to combat the complexity in incoming problems. This paper gives an introduction to Particle Markov Chain Monte Carlo (PMCMC) and gives some intuition behind the algorithm. For this review of the paper, I will focus on introducing the PMCMC; developing the algorithm and intuition behind PMCMC; and provide asymptotic results of the PMCMC algorithm.

The paper focuses on a new method called Particle Markov chain Monte Carlo or PMCMC for short. The basic idea behind PMCMC is to combine the best aspects of MCMC methods and Sequential Monte Carlo (SMC) methods. The goal of the algorithm is to use SMC to create proposal densities for the MCMC algorithms. The motivation for PMCMC comes from state space models (SSM), but the ideas and algorithm can be expanded to other high dimensional problems.

In general, the goal of a MCMC algorithm is to sample from the target distribution, say π . Usually, we cannot sample from π directly so we choose a proposal distribution which captures the essence of the target distribution and sample from the that. The hardest part of this is the balance between choosing a proposal distribution that adequately describes/captures the target distribution and choosing a proposal distribution that is easy to implement. PMCMC aims to provide a proposal distribution that balances these two features.

2 Background

2.1 State Space Models

Before we talk about the PMCMC algorithm, we first need to talk about SMC, but before we talk about SMC, we should introduce the motivating example of SMC which are state space models (SSMs).

Definition 2.1. State space models (also called hidden Markov models), $\{X_n; n \geq 1\}$, are be defined by:

- Initial density $X_1 \sim \mu_\theta(\cdot)$

- Transitional probability density

$$X_{n+1}|(X_n = x) \sim f_\theta(\cdot|x) \quad \exists \theta \in \Theta \quad (1)$$

Usually, we never see $\{X_n\}$, but we will “observe” them through $\{Y_n; n \geq 1\}$. These $\{Y_n\}$ are conditionally independent given $\{X_n\}$. We denote their marginal probability density function by

$$Y_n|(X_1, \dots, X_n = x, \dots, X_m) \sim g_\theta(\cdot|x) \text{ for } 1 \leq n \leq m \quad (2)$$

Notation note: For random variables, capitol letters will be used: X_n and lower case will be used for thier values. Lastly, we will denote a sequence, $\{z_n\}$, as follows $z_{i:j} = (z_i, z_{i+1}, \dots, z_j)$.

In Bayesian analysis, we want to conduct inference on two posteriors:

1. $p_\theta(x_{1:T}|y_{1:T}) \propto p_\theta(x_{1:T}, y_{1:T})$ assuming θ is known and where

$$p_\theta(x_{1:T}, y_{1:T}) = \mu_\theta(x_1) \prod_{i=2}^T f_\theta(x_i|x_{i-1}) \prod_{n=1}^T g_\theta(y_n|x_n) \quad (3)$$

2. When θ is unknown,

$$p(\theta, x_{1:T}|y_{1:T}) \propto p_\theta(x_{1:T}, y_{1:T})p(\theta) \quad (4)$$

Depending on the type of SSM, these posteriors may be difficult or failry simple to handle. In our case, we will consider the approximations that allow us to handle more complicated cases, and we will outline some background between SMC and MCMC for SSM that help in the development of the approximation.

2.2 Sequential Monte Carlo

The most common method of handling SSM in a bayesian context involves Sequential Monte Carlo methods (SMC). The big idea of SMC is it provides a method to sequentially approximate $p_\theta(x_{1:n}|y_{1:n})$ for $n \geq 1$ and $p_\theta(y_{1:n})$ using N weighted samples called *particles* i.e. approximate $p_\theta(x_1|y_1), p_\theta(y_1)$ first, then $p_\theta(x_{1:2}|y_{1:2}), p_\theta(y_{1:2})$ second, all the way to $p_\theta(x_{1:n}|y_{1:n}), p_\theta(y_{1:n})$.

The driving force behind SMC is to use importance sampling (IS) to approximate $p_\theta(x_1|y_1)$ by $q_\theta(x_1|y_1)$. For one iterations, we would start by generating N particles, $\{X_1^k\}$, from $q_\theta(x_1|y_1)$, next ascribe weights, $\{W_1^k\}$, to the particles which take into account the difference between the two densities. Since we are just interested in one iteration, we would sample N times from IS approx of $p_\theta(x_1|y_1)$,

$\hat{p}_\theta(dx_1|y_1)$ to get particles approx distributed from $p_\theta(x_1|y_1)$. Usually, we are interested in many iterations, so we would repeat the process as outlined below noting that

$$p_\theta(x_{1:2}|y_{1:2}) \propto p_\theta(x_1|y_1)f_\theta(x_2|x_1)g_\theta(y_2|x_2)$$

Here is the algrotihm:

Step 1: at time $n = 1$,

- (a) sample $X_1^k \sim q_\theta(\cdot|y_1)$
- (b) compute & normalize weights

$$\begin{aligned} w_1(X_1^k) &:= \frac{p_\theta(X_1^k, y_1)}{q_\theta(X_1^k|y_1)} = \frac{\mu_\theta(X_1^k)g_\theta(y_1|X_1^k)}{q_\theta(X_1^k|y_1)} \\ W_1^k &:= \frac{w_1(X_1^k)}{\sum_{m=1}^N w_1(X_1^m)} \end{aligned} \quad (5)$$

Step 2: at time $n = 2, \dots, T$,

- (a) sapmle $A_{n-1}^k \sim \mathcal{F}(\cdot|\mathbf{W}_{n-1})$
- (b) sample $X_n^k \sim q_\theta(\cdot|y_n, X_{n-1}^{A_{n-1}^k})$ and set $X_{1:n}^k := (X_{1:n-1}^{A_{n-1}^k}, X_n^k)$
- (c) compute and normalize the weights

$$\begin{aligned} w_1(X_{1:n}^k) &:= \frac{p_\theta(X_{1:n}^k, y_{1:n})}{p_\theta(X_{n-1}^{A_{n-1}^k}, y_{1:n-1})q_\theta(X_n^k|y_n, X_{n-1}^{A_{n-1}^k})} \\ &= \frac{f_\theta(X_n^k|X_{n-1}^{A_{n-1}^k})g_\theta(y_n|X_n^k)}{q_\theta(X_n^k|y_n, X_{n-1}^{A_{n-1}^k})} \\ W_n^k &:= \frac{w_n(X_{1:n}^k)}{\sum_{m=1}^N w_n(X_{1:n}^m)} \end{aligned} \quad (6)$$

There are a lot of different pieces in the algorithm; I'll limit my description to the big pieces. A_{n-1}^k is the 'parent' indicator at time $n - 1$ for particle $X_{1:n}^k$ for $2 \leq n \leq T$. $\mathbf{W}_n := (W_n^1, \dots, W_n^N)$ are the normalized important weights at time n , and $\mathcal{F}(\cdot|\mathbf{p})$ is a discrete probability distribution on $\{1, \dots, m\}$ typically a multinomial resampling procedure. The A_{n-1}^k variables are key because they allow us to keep track of 'genealogy' of the particles which will be needed for developing a Gibbs sampler for PMCMC.

The algorithm gives this approximation at time T :

$$\hat{p}_\theta(dx_{1:n}|y_{1:n}) := \sum_{k=1}^N W_n^K \delta_{X_{1:n}^k}(dx_{1:n}) \quad (7)$$

where δ is the Dirac measure.

The SMC also provides us with an approximation to the marginal density $p_\theta(y_{1:T})$:

$$\hat{p}_\theta(y_{1:T}) := \hat{p}_\theta(y_1) \prod_{n=2}^T \hat{p}_\theta(y_n|y_{1:n-1}) \quad (8)$$

where

$$\hat{p}_\theta(y_n|y_{1:n-1}) = \frac{1}{N} \sum_{k=1}^N w_n(X_{1:n}^k)$$

So through the use of SMC we are able to get a approx of the marginal and joint distributions, but there are some issues with SMC. The obvious question is what happens when T is very large? We can imagine when T is very large, the number of distinct x_n goes down. So basically in order to keep up with an increasing T we'd have to keep increasing k but that isn't practically feasible. This is becomes apparent when SMC is used to fit $p(\theta, x_{1:T}|y_{1:T})$ according to Andrieu et al.(1999), Fearnhead (2002) and Storvik (2002).

2.3 Markov chain Monte Carlo

MCMC methods for state space models focus on sampling from $p(\theta, x_{1:T}|y_{1:T})$ by switching in between sampling from $x_{1:T}|\theta$ and $\theta|x_{1:T}$. Typically, $\theta|x_{1:T}$ is easy to sample from but $\theta|x_{1:T}$ rarely is so there is a workaround that splits $x_{1:T}$ into K adjacent blocks and update block by block.

Ex Update $x_{n:n+K-1}$ like so:

$$p_\theta(x_{n:n+K-1}|y_{1:T}, x_{1:n-1}, x_{n+K:T}) \propto \prod_{k=n}^{n+K} f_\theta(x_k|x_{k-1}) \prod_{k=n}^{n+K-1} g_\theta(y_k|x_k) \quad (9)$$

This workaround is fine unless K becomes too large. If it becomes to large, then we will have trouble creating good approximations of Eq 9; so this will force us to keep K small and cause us to spend more computational time to explore the support of $x_{1:T}|\theta$. So the only way to move forward is to simulate from joint proposal, but this is less than ideal since as T increases, the proposal's performance doesn't get better since it doesn't use the information.

3 Particle Markov chain Monte Carlo

The Particle MCMC algorithm is the entire focus of this paper. This algorithm attempts to correct issues with both SMC and regular MCMC methods, by taking what these methods do best and combining them. Let's recall the ideal MCMC case; we have a target $p(\theta, x_{1:T}|y_{1:T})$ which we sample via $p_\theta(x_{1:T}|y_{1:T})$. Obviously this isn't feasible in practice, so what if we approximated this best case scenario by using the SMC approximation to $p_\theta(x_{1:T}|y_{1:T})$ as a proposal distribution in a Metropolis Hastings update? This sounds really good, but not really possible directly since the marginal density of a particle is not available and would be needed in the acceptance ratio. To tackle the issue, PMCMC includes the space of all the random variables generated by SMC in the target distributions considered. A nice feature of PMCMC is that it leaves the target density invariant; because of this it is considered an 'exact approximation.' The paper also presents three implementations of PMCMC which will follow.

3.0.1 Particle independent Metropolis-Hastings sampler

With Particle independent Metropolis-Hastings sampler the goal is still to $p_\theta(x_{1:T}|y_{1:T})$ except for now we are going to include SMC approximations in our proposal. Normally we choose a proposal that will leave $p_\theta(x_{1:T}|y_{1:T})$ invariant. In order for this to happen we need our proposal to suggest candidates $X_{1:T}^*$ that are accepted with probability

$$\min\left(1, \frac{p_\theta(X_{1:T}^*|y_{1:T}) q_\theta(X_{1:T}|y_{1:T})}{p_\theta(X_{1:T}|y_{1:T}) q_\theta(X_{1:T}^*|y_{1:T})}\right)$$

given current state $X_{1:T}$. (Also side note, the paper included some very unintuitive notation for the min operator.)

In the best case scenario, $q_\theta(X_{1:T}|y_{1:T}) = p_\theta(X_{1:T}|y_{1:T})$, but if that were the case then there would be no need for a sampler. The idea is to simply use the SMC approximation of $p_\theta(X_{1:T}|y_{1:T})$, but it's not possible to use it directly since the marginal distribution of $X_{1:T}^*$ is necessary to get the acceptance probability. To get over this bump, they suggest taking the expectation of the approximation over all the random variables given by SMC algorithm and consider this as the proposal i.e.

$$q_\theta(dx_{1:T}|y_{1:T}) = E(\hat{p}_\theta(dx_{1:T}|y_{1:T}))$$

This still doesn't give us a nice form to work with but this allows us to embed our sampling of $p_\theta(X_{1:T}|y_{1:T})$ into a sampling from another well-chosen distribution that covers fits our idea of having a space with all the random variables.

Step 1: $i=0$ & run SMC algorithm for $p_\theta(x_{1:T}|y_{1:T})$.

Sample $X_{1:T}(0) \sim \hat{p}_\theta(\cdot|y_{1:T})$ and $\hat{p}_\theta(y_{1:T})(0)$

Step 2: $i \geq 1$

(a) run SMC algorithm for $p_\theta(x_{1:T}|y_{1:T})$ & sample $X_{1:T}^* \sim \hat{p}_\theta(\cdot|y_{1:T})$ and $\hat{p}_\theta(y_{1:T})$

(b) With probability,

$$\min\left(1, \frac{\hat{p}_\theta(y_{1:T})^*}{\hat{p}_\theta(y_{1:T})(i-1)}\right)$$

set $X_{1:T}(i) = X_{1:T}^*$ and $\hat{p}_\theta(y_{1:T})(i) = \hat{p}_\theta(y_{1:T})^*$ otherwise stay the same.

These next few assumptions and theorems establish the invariance of $p_\theta(x_{1:T}|y_{1:T})$ after the PIMH update and ergodicity of the PIMH sampler.

Assumption 1. For $n = 1, \dots, T$, we have $\mathcal{S}_n \subseteq \mathcal{Q}_n$ where

$$\mathcal{S}_n = \{x_{1:T} \in (X)^n : p(x_{1:T}) > 0\} \quad n \geq 1$$

$$\mathcal{Q}_n = \{x_{1:T} \in (X)^n : p(x_{1:T-1})q(x_n|x_{1:T-1}) > 0\} \quad n \geq 1$$

Essentially, Assumption 1 is saying that we can use importance density $p(x_{1:T-1}) \times q(x_n|x_{1:T-1})$ to approximate $p(x_{1:T})$ and Assumption 2 basically has to do with resampling the $X_{1:T}$ according to the weights.

Assumption 2. For any $k = 1, \dots, N$ and $n = 1, \dots, T$, the resampling scheme satisfies

$$E(O_n^k | \mathbf{W}_n) = NW_n^K \quad (10)$$

and

$$r(A_n^K = m | \mathbf{W}_n) = W_n^m \quad (11)$$

Assumption 3. \exists a sequence of constants $\{C_n; n = 1, \dots, \bar{T}\}$ for some integer \bar{T} s.t. for any $x_{1:n} \in \mathcal{S}_n$

$$w_n(x_{1:n}) \leq C_n \quad (12)$$

Assumption 3 allows us to get a bound condition for the importance weights.

Theorem 1. Assume Assumption 2. Then for an $N \geq 1$, the PIMH update is a standard IMH update on the extended space \mathbf{X} with target density $\tilde{\pi}^N$ and proposal q^N .

Theorem 2. Assume Assumptions 1 and 2. Then

1. the PIMH sampler generates a sequence $\{X_{1:T}(i)\}$ whose marginal distributions $\{\mathcal{L}^N\{X_{1:T}(i) \in \cdot\}\}$ satisfy

$$\|\mathcal{L}^N\{X_{1:T}(i) \in \cdot\} - \pi(\cdot)\| \rightarrow 0 \quad \text{as } i \rightarrow \infty$$

2. if assumption 3 holds, then there exists $\rho_T \in [0, 1)$ s.t. $\forall i \geq 1$ and $x_{1:T} \in \mathcal{X}^T$

$$\|\mathcal{L}^N\{X_{1:T}(i) \in \cdot | X_{1:T}(0) = x_{1:T}\} - \pi(\cdot)\| \leq \rho_T^i$$

Also the authors claim that the acceptance probability converges to 1 as $N \rightarrow \infty$ since $\hat{p}_\theta(y_{1:T})^*$ and $\hat{p}_\theta(y_{1:T})(i-1)$ are consistent estimators of $p_\theta(y_{1:T})$

Note that if your only goal is to estimate $p_\theta(x_{1:T}|y_{1:T})$ then just using the SMC algorithm may be a better choice, but rather consider this an option as a step within a larger MCMC algorithm.

3.0.2 Particle marginal Metropolis-Hastings sampler

With Particle marginal Metropolis-Hastings, the goal is to sample from $p(\theta, x_{1:T}|y_{1:T})$. In general here we assume we can sample from $p_\theta(x_{1:T}|y_{1:T})$, and we can decompose $p(\theta, x_{1:T}|y_{1:T}) = p(\theta|y_{1:T})p_\theta(x_{1:T}|y_{1:T})$ so the focus is $p(\theta|y_{1:T})$. The intuitive proposal is

$$q\{(\theta^*, x_{1:T}^*) | (\theta, x_{1:T})\} = q(\theta^* | \theta) p_{\theta^*}(x_{1:T}^* | y_{1:T})$$

Then the acceptance ratio for the MH algorithm is

$$\frac{p(\theta^*, x_{1:T}^*) | (y_{1:T})}{p(\theta, x_{1:T}) | (y_{1:T})} \frac{q\{(\theta, x_{1:T}) | (\theta^*, x_{1:T}^*)\}}{q\{(\theta^*, x_{1:T}^*) | (\theta, x_{1:T})\}} = \frac{p_{\theta^*}(y_{1:T}) p(\theta^*)}{p_\theta(y_{1:T}) p(\theta)} \frac{q(\theta | \theta^*)}{q(\theta^* | \theta)} \quad (13)$$

The looking at the acceptance probability, it is natural to use SMC approximation for $p_\theta(x_{1:T}|y_{1:T})$ and $p_\theta(y_{1:T})$ in the PMMH update.

Step 1: $i = 0$

- (a) set $\theta(0)$ to some value
- (b) run SMC algorithm for $p_{\theta(0)}(x_{1:T}|y_{1:T})$, sample $X_{1:T}(0) \sim \hat{p}_{\theta(0)}(\cdot|y_{1:T})$ and $\hat{p}_\theta(y_{1:T})(0)$ is the marginal likelihood estimate

Step 2: $i \geq 1$

- (a) sample $\theta^* \sim q\{\cdot | \theta(i-1)\}$
- (b) run SMC algorithm for $p_{\theta^*}(x_{1:T}|y_{1:T})$
sample $X_{1:T}^* \sim \hat{p}_{\theta^*}(\cdot|y_{1:T})$ and $\hat{p}_{\theta^*}(y_{1:T})$ is the marginal likelihood estimate

(c) With probability,

$$\min \left(1, \frac{\hat{p}_{\theta^*}(y_{1:T})p(\theta^*)}{\hat{p}_{\theta(i-1)}(y_{1:T})p\{\theta(i-1)\}} \frac{q\{(\theta(i-1)|\theta^*)\}}{q\{(\theta^*|\theta(i-1))\}} \right)$$

set $\theta(i) = \theta^*$, $X_{1:T}(i) = X_{1:T}^*$ and $\hat{p}_{\theta(i)}(y_{1:T})(i) = \hat{p}_{\theta^*}(y_{1:T})$ otherwise stay the same.

The following theorem shows the invariance of $p_\theta(x_{1:T}|y_{1:T})$ after PMMH update and ergodicity of PMMH. Assumption 4 has the same idea as Assumption 1 in that it allows us to use an importance density in approximation. Assumption 5 is a key assumption that ensures we have a converging MH algorithm.

Assumption 4. For any $\theta \in \mathcal{S}$, we have $\mathcal{S}_n^\theta \subseteq \mathcal{Q}_n^\theta$ for $n = 1, \dots, T$.

Assumption 5. The MH sampler of target density $\pi(\theta)$ and proposal density $q(\theta^*|\theta)$ is irreducible and aperiodic (and hence converges for π almost all starting points).

Theorem 3. Assume assumption 2 holds. Then $\forall N \geq 1$

1. the PMMH update is an MH update defined on the extended space $\Theta \times X$ with a target density $\tilde{\pi}$
2. additionally, if assumptions 5 and 6 are met, the PMMH sampler generates a sequence $\{\theta(i), X_{1:T}(i)\}$ whose marginal distributions $\{\mathcal{L}^N\{\theta(i), X_{1:T}(i) \in \cdot\}\}$ satisfy

$$\|\mathcal{L}^N\{\theta(i), X_{1:T}(i) \in \cdot\} - \pi(\cdot)\| \rightarrow 0 \quad \text{as } i \rightarrow \infty$$

3.0.3 Particle Gibbs sampler

Lastly, a Particle Gibbs sampler is presented as an alternative to MMH for sampling $p(\theta, x_{1:T}|y_{1:T})$ by sampling from $p(\theta|x_{1:T}, y_{1:T})$ and $p_\theta(x_{1:T}|y_{1:T})$. Usually it is possible to sample from $p(\theta|x_{1:T}, y_{1:T})$ so it allows us to skip creating a proposal density to sample θ . The usual usage of the SMC approximation to $p_\theta(x_{1:T}|y_{1:T})$ is not possible here since it will not give us an invariant posterior. The authors present a fix to this problem using a *conditional* SMC update. The basic idea is similar except one $X_{1:T}$ with ancestry $B_{1:T}$ is ensured to be there through all the resampling steps. Here is the outline of the *conditional* SMC update:

Step 1 let $X_{1:T} = (X_1^{B_1}, X_1^{B_1}, \dots, X_T^{B_T})$ be a path with ancestry $B_{1:T}$

Step 2: $n = 1$

(a) for $k \neq B_1$, sample $X_1^k \sim q(\cdot|y_1)$

(b) compute weights $w_1(X_1^k)$ from equation (6) and normalize the weights creating W_1^k

Step 3: $2 \leq n \leq T$

- (a) for $k \neq B_n$, sample $A_{n-1}^k \sim \mathcal{F}(\cdot | \mathbf{W}_{n-1})$
- (b) for $k \neq B_n$, sample $X_n^k \sim q(\cdot | y_n, X_{n-1}^{A_{n-1}^k})$
- (c) compute weights $w_1(X_1^k)$ from equation (7) and normalize the weights creating W_1^k

Then the particle Gibbs sampler uses this like so:

Step 1 $i = 0$, set $\theta(0)$, $X_{1:T}(0)$, $B_{1:T}(0)$ to some values

Step 2: $i \geq 1$

- (a) sample $\theta(i) \sim p\{\cdot | y_{1:T}, X_{1:T}(i-1)\}$
- (b) run *conditional* SMC for $p_{\theta(i)}(x_{1:T} | y_{1:T})$ conditional on $X_{1:T}(i-1)$ and $B_{1:T}(i-1)$
- (c) sample $X_{1:T}(i) \sim \hat{p}_{\theta(i)}(\cdot | y_{1:T})$

This theorem shows invariance of $p_{\theta}(x_{1:T} | y_{1:T})$ holds after a Particle Gibbs update and is erogodic, and Assumption 6 is the Gibbs equivalent of Assumption 4.

Assumption 6. *The Gibbs sampler defined by conditionals $\pi(\theta | x_{1:T})$ and $p_{\theta}(x_{1:T})$ is irreducible and aperiodic (and hence converges for π almost all starting points).*

Theorem 4. *Assume Assumption 2. Then*

1. *the PG updates defines a transition kernel on the extended space $\Theta \times \mathbf{X}$ of invariant density $\tilde{\pi}^N \quad \forall N \geq 1$*
2. *if Assumption 4-6 hold, the PG sampler generates a sequence $\{\theta(i), X_{1:T}(i)\}$ whose marginal distributions $\{\mathcal{L}^N\{(\theta(i), X_{1:T}(i)) \in \cdot\}\}$ satisfy*

$$\|\mathcal{L}^N\{(\theta(i), X_{1:T}(i)) \in \cdot\} - \pi(\cdot)\| \rightarrow 0 \quad \text{as } i \rightarrow \infty \quad \exists \quad N \geq 2$$

Also for both PMMH and PG samplers, the author presents a thereom which ensures consistent estimators.

Theorem 5. *Assumer Assumptions 2 - 4 and let $f : \Theta \times \mathcal{X}^T \rightarrow \mathbb{R}$ be such that $E_{\pi}(|f|) < \infty$. Then when the PMMH or PG sampler is erogodic, the for any $N \geq 1$ or $N \geq 2$ respectively:*

$$\frac{1}{L} \sum_{i=1}^L \left[\sum_{k=1}^N W_T^k(i) f\{\theta(i), X_{1:T}^k(i)\} \right] \rightarrow E_{\pi}(f) \quad \text{a.s. as } L \rightarrow \infty$$

4 Conclusion

In this review, we have outlined and given some intuition behind PMCMC. We have limited the review to the simple case using SSMs as the motivation. PMCMC methods have been generalized to handle models much more complicated than SSMs. While these methods are better than using SMC or MCMC alone, there are some issues as well. At every iteration in the PMCMC, N particles need to be generated. This will lead to more computation and time in the algorithm. Overall, I feel like this is a minor concern given the benefits of the PMCMC. We also gave theorems which helped back the properties of PMCMC in its various settings.

References

- [1] C. ANDRIEU, A. DOUCET, AND R. HOLENSTEIN, *Particle markov chain monte carlo methods*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72 (2010), pp. 269–342.