Perfect simulation using atomic regeneration with application to Sequential Monte Carlo

Anthony Lee, University of Warwick

Joint work with Arnaud Doucet (Oxford), Krzysztof Łatuszyński (Warwick)

February 11th, 2015
Gatsby Computational Neuroscience Unit, UCL

arXiv:1407.5770

Outline

Outline

Introduction

- Given a target probability measure (distribution) π on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$, we would like to obtain exact samples from π.
- Some methodology: inverse transform, special relationships, rejection sampling.
- In many practical situations, none of the above are suitable.
- Markov chain Monte Carlo: define a Markov kernel P with stationary distribution π
  - One can obtain samples whose asymptotic distribution is π.
- In the 90s the ability to sample exactly from the stationary distribution of a Markov chain was investigated.

Sampling from the stationary distribution

- Asmussen et al. (1992) investigated methods to sample from the stationary distribution of a Markov chain.
  - These methods were prohibitively expensive to implement on general-state spaces.
- Propp and Wilson (1996) proposed an alternative method: Coupling From The Past (CFTP).
  - This has had notable successes on discrete state spaces.
- Murdoch and Green (1998) proposed a CFTP method for general state spaces: the multigamma coupler.
  - One of the algorithms we propose is a multigamma coupler.

Structure of the talk

- We will overview, for a Markov kernel P with stationary distribution π,
  - regeneration and the split chain,
  - mixture representations of π,
  - perfect simulation.
- We will then discuss the special case where P is intractable but admits a singleton atom and defines a uniformly ergodic Markov chain.
  - Solutions to Bernoulli factory problems provide implementations of perfect simulation in this context.
- We then discuss how one can introduce an artificial singleton atom, as in Brockwell and Kadane (2005).
- Finally, we use the methodology to sample from a Feynman–Kac path measure.

Motivation

- The primary methodology we propose is for uniformly ergodic Markov chains.
- However, the transition kernel P can be intractable in the sense that we cannot compute, e.g., $p(x, y)$ where

$$P(x, A) = \int_A p(x, y)\mathrm{d}y, \qquad A \in \mathcal{B}(\mathsf{X}), x \notin A.$$

- There is no barrier, e.g., to letting P be an iterate of another kernel.
- In our primary example the Markov kernel is intractable but "almost" a perfect sampler.
- Of course, there can be difficulties in applying the method, which we will discuss.

Primary example: Feynman–Kac path measures

- We consider a generic discrete-time Feynman–Kac model with time horizon n.
- Let $(\mathsf{Z}, \mathcal{B}(\mathsf{Z}))$ be a measurable space and define
  - a probability measure $\mu : \mathcal{B}(\mathsf{Z}) \to [0, 1]$,
  - some Markov kernels $M_p : \mathsf{Z} \times \mathcal{B}(\mathsf{Z}) \to [0, 1]$ for $p \in \{2, \ldots, n\}$ and
  - non-negative $\mathcal{B}(\mathsf{Z})$-measurable functions $G_p : \mathsf{Z} \to \mathbb{R}_+$ for $p \in \{1, \ldots, n\}$.
- We define for any $p \in \{1, \ldots, n\}$, the measure γ_p by

$$\gamma_p(A) := \int_A \left[\prod_{q=1}^{p} G_q(z_q) \right] \mu(\mathrm{d}z_1) \prod_{q=2}^{p} M_q(z_{q-1}, \mathrm{d}z_q), \quad A \in \mathcal{B}(\mathsf{Z}^p),$$

  and its associated probability measure $\pi_p := \gamma_p(\mathsf{Z}^p)^{-1} \gamma_p$.
- With $\mathsf{X} := \mathsf{Z}^n$ the Feynman–Kac path measure of interest is the probability measure $\pi := \pi_n$ on $\mathcal{B}(\mathsf{X})$.

Hidden Markov models

- One direct application is for inferring the distribution of the latent variables in a hidden Markov model.
- Here μ and $\mathcal{M} := (M_p)_{p\in\{2,\dots,n\}}$ determine the unconditional distribution of the latent variables.
- $\mathcal{G} := (G_p)_{p\in\{1,\dots,n\}}$ encode the probability densities of the observed data, i.e.,

$$G_p(x_p) = g(x_p, y_p),$$

where $(y_1, \dots, y_n)$ is the sequence of observed data at the times $1, \dots, n$.

Sequential Monte Carlo

- Expressing π using this Feynman–Kac formalism has immediate methodological consequences.
- One can approximate $\pi(f) := \int_{\mathsf{X}} f(x)\pi(\mathrm{d}x)$ using SMC or particle filtering methods.
- The following algorithm is a particle filter.

A particle filter

1. Simulate $\zeta_1^i \sim \mu$ for $i \in \{1, \ldots, N\}$.
2. For $p = 2, \ldots, n$:
   - For each $i \in \{1, \ldots, N\}$:
     - 2.1 Simulate $A_{p-1}^i \sim \mathcal{C}\left(G_{p-1}(\zeta_{p-1}^1), \ldots, G_{p-1}(\zeta_{p-1}^N)\right)$.
     - 2.2 Simulate $\zeta_p^i \sim M_p(\zeta_{p-1}^{A_{p-1}^i}, \cdot)$.
3. Set $V = (\zeta_1^1, \ldots, \zeta_n^N, A_1^1, \ldots, A_{n-1}^N)$.

- Here $\mathcal{C}$ denotes a categorical distribution, i.e.

$$\Pr(A_{p-1}^i = k) = \frac{G_{p-1}(\zeta_{p-1}^k)}{\sum_{j=1}^N G_{p-1}(\zeta_{p-1}^j)}.$$

Motivation for perfect simulation

- Following Andrieu et al. (2010), for any $k \in \{1, \ldots, N\}$, we define the ancestral lineage B^k to be the $\{1, \ldots, N\}^n$-valued random variable satisfying $B_n^k := k$ and $B_p^k := A_p^{B_{p+1}^k}$.
- The random variable

$$\zeta^k := (\zeta_1^{B_1^k}, \ldots, \zeta_n^{B_n^k})$$

  is then a path taking values in X.
- Let Q^N be the probability measure associated with the path ζ^K chosen by tracing an ancestral line after picking K with

$$\Pr(K = k) = \frac{G_n(\zeta_n^k)}{\sum_{k=1}^N G_n(\zeta_n^j)}.$$

- How close is Q^N to π?

Hope...

Proposition

Assume there exists $B < \infty$ such that for each $p \in \{1, \ldots, n\}$, $0 < G_p(z_p) < B$ for all $z_p \in Z$. Then there exists $F < \infty$ such that for any $N \geq 2$,

$$\sup_{x \in X} \frac{\pi(\mathrm{d}x)}{Q^N(\mathrm{d}x)} \leq \left(1 + \frac{F}{N}\right)^n.$$

- Great!
- Can we make these samples perfect somehow?

Iterated Conditional SMC

- This is a Markov kernel, defined by running a conditional SMC algorithm with a fixed path, followed by picking a new path.

$$P_N(x, A) := \int_{\mathsf{V}_N} \bar{Q}_x^N(\mathrm{d}v) Q_v^N(A), \quad x \in \mathsf{X}, A \in \mathcal{B}(\mathsf{X}).$$

- $\bar{Q}_x^N$ is the probability measure associated with the random variable V produced by conditional SMC with fixed path x.
- Q_v^N is the probability measure associated with the path ζ^K chosen by tracing an ancestral line after picking K with

$$\Pr(K = k) = \frac{G_n(\zeta_n^k)}{\sum_{k=1}^N G_n(\zeta_n^j)}.$$

- This is a reversible, π-invariant Markov kernel (Andrieu et al., 2010).

Uniform ergodicity of i-cSMC

- This Markov kernel has been studied in detail in Chopin and Singh (2013), Andrieu et al. (2013) and Lindsten et al. (2014).
- If assume π-essential boundedness of each G_p, then

$$P_N(x, \cdot) \geq \epsilon_N \pi(\cdot),$$

  where $\lim_{N \to \infty} \epsilon_N = 1$.
- Quantitative bounds provided in Andrieu et al. (2013) and Lindsten et al. (2014) can be used to bound ϵ_N under various assumptions.

Remarks

- In Andrieu et al. (2012), a perfect sampling method is proposed where the mechanism governing particle offspring is fundamentally changed from selection with a constant population size at each time to stochastic branching.
  - Computational guarantees are yet to be established.
- The only other perfect sampling method on a general state space is rejection in $\mathcal{O}(\exp(n))$ time.
- Some applications of our methodology are presented in the paper on arXiv, I will cover only the methodology here.

Outline

Notation

- Recall that π is a probability measure on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$.
- Let $\mathbf{X} := (X_n)_{n \geq 1}$ be a time-homogeneous, π-irreducible, Harris recurrent Markov chain with π-invariant transition kernel P, i.e.,

$$P(x, A) = \Pr(X_n \in A \mid X_{n-1} = x).$$

- We will use the notation, where $\mu : \mathcal{B}(\mathsf{X}) \to [0, 1]$,

$$\mu P(A) := \int_{\mathsf{X}} \mu(\mathrm{d}x) P(x, A), \qquad A \in \mathcal{B}(\mathsf{X}),$$

and for $n \in \mathbb{N}$,

$$P^n(x, A) := \int_{\mathsf{X}} P(x, \mathrm{d}y) P^{n-1}(y, A).$$

- We assume we can sample from $P(x, \cdot)$ for any $x \in \mathsf{X}$.

Atoms

- The set α is a proper atom for P is there exists a probability measure μ such that

$$P(x, A) = \mu(A), \qquad x \in \alpha, A \in \mathcal{B}(\mathsf{X}).$$

- A proper atom is accessible if $\pi(\alpha) > 0$ so that (ind. of X_1)

$$\Pr\left(\sum_{n \geq 1} \mathbb{I}(X_n \in \alpha) = \infty\right) = 1.$$

- Intuition: when a proper atom exists, the Markov chain occasionally visits α, at which point it regenerates.
- $\mathbf{X}$ can then be split into independent "tours".

The split chain

- On general state spaces, proper atoms are not guaranteed to exist.
- A key theoretical development was the split chain (Athreya and Ney, 1978; Nummelin, 1978).
- The key assumption is that P satisfies a minorization condition

$$P(x, \cdot) \geq s(x)\nu(\cdot),$$

for some function s with $\pi(s) = \int_{\mathsf{X}} s(x)\pi(\mathrm{d}x) > 0$ and a probability measure ν.

- This is a bivariate Markov chain $\tilde{\mathbf{X}}_{\nu,s} = (\tilde{X}_n^{(\nu,s)})_{n\geq 1}$ evolving on $\mathsf{X} \times \{0, 1\}$ whose first coordinate has identical law to $\mathbf{X}$.

The split chain

- When the minorization $P(x, \cdot) \geq s(x)\nu(\cdot)$ holds we can write

$$P(x, \cdot) = s(x)\nu(\cdot) + [1 - s(x)]R_{\nu,s}(x, \cdot)$$

where for $s(x) > 0$, the residual kernel is

$$R_{\nu,s}(x, \cdot) := \frac{P(x, \cdot) - s(x)\nu(\cdot)}{1 - s(x)}.$$

The split chain

- When the minorization $P(x, \cdot) \geq s(x)\nu(\cdot)$ holds we can write

$$P(x, \cdot) = s(x)\nu(\cdot) + [1 - s(x)]R_{\nu,s}(x, \cdot)$$

where for $s(x) > 0$, the residual kernel is

$$R_{\nu,s}(x, \cdot) := \frac{P(x, \cdot) - s(x)\nu(\cdot)}{1 - s(x)}.$$

- We then define $\tilde{P}$ as

$$\tilde{P}(x, \rho; \mathrm{d}y, \varrho) := \{\mathbb{I}(\rho = 1)\nu(\mathrm{d}y) + \mathbb{I}(\rho = 0)R_{\nu,s}(x, \mathrm{d}y)\}\, s(y)^{\varrho}[1-s(y)]^{1-\varrho},$$

and we can see that $\tilde{P}(x, 1; \cdot) = \tilde{\nu}_{\nu,s}(\cdot)$ where

$$\tilde{\nu}_{\nu,s}(\mathrm{d}y, \varrho) := \nu(\mathrm{d}y)s(y)^{\varrho}[1 - s(y)]^{1-\varrho}.$$

The split chain

- When the minorization $P(x, \cdot) \geq s(x)\nu(\cdot)$ holds we can write

$$P(x, \cdot) = s(x)\nu(\cdot) + [1 - s(x)]R_{\nu,s}(x, \cdot)$$

where for $s(x) > 0$, the residual kernel is

$$R_{\nu,s}(x, \cdot) := \frac{P(x, \cdot) - s(x)\nu(\cdot)}{1 - s(x)}.$$

- We then define $\tilde{P}$ as

$$\tilde{P}(x, \rho; \mathrm{d}y, \varrho) := \{\mathbb{I}(\rho = 1)\nu(\mathrm{d}y) + \mathbb{I}(\rho = 0)R_{\nu,s}(x, \mathrm{d}y)\}\, s(y)^{\varrho}[1-s(y)]^{1-\varrho},$$

and we can see that $\tilde{P}(x, 1; \cdot) = \tilde{\nu}_{\nu,s}(\cdot)$ where

$$\tilde{\nu}_{\nu,s}(\mathrm{d}y, \varrho) := \nu(\mathrm{d}y)s(y)^{\varrho}[1 - s(y)]^{1-\varrho}.$$

- That is, $\mathsf{X} \times \{1\}$ is a proper atom for $\tilde{P}$.

The split chain

- Since the law of the first coordinate of $\tilde{\mathbf{X}}_{\nu,s}$ is identical to $\mathbf{X}$ we will write $\tilde{\mathbf{X}}_{\nu,s} := (X_n, \rho_n^{(\nu,s)})_{n\geq 1}$.
- This emphasizes that it is the regeneration indicators $\rho_n^{(\nu,s)}$ that are affected by ν and s.
- We can call the times τ such that $\rho_\tau^{(\nu,s)} = 1$ the regeneration times.
- We can think of X_τ as being the sample just before regeneration, since $X_{\tau+1}|(\rho_\tau^{(\nu,s)} = 1) \sim \nu$.

Simulating the split chain

- There are two relatively simple ways.
- The first is the "direct approach" using $\tilde{P}$ when one can sample from ν and $R_{\nu,s}$ and flip an $s(x)$-coin.
- The second involves simulating $\mathbf{X}$ using P and then imputing the values of $(\rho_n^{(\nu,s)})_{n\geq 1}$, using

$$\Pr\left(\rho_{n-1}^{(\nu,s)} = 1 \mid X_{n-1} = x_{n-1}, X_n = x_n\right) = s(x_{n-1})\frac{\mathrm{d}\nu(\cdot)}{\mathrm{d}P(x_{n-1},\cdot)}(x_n),$$

  as observed in Mykland et al. (1995).
- In general, it may not be possible to sample from ν or access the Radon–Nikodym derivative above.
  - It may not be easy to detect regenerations!

General mixture representation

- Perfect simulation algorithms can be motivated using a mixture representation of π (Asmussen et al., 1992; Hobert and Robert, 2004; Hobert et al., 2006)

$$\pi(A) = \sum_{n \geq 1} \frac{\mathbb{P}_{\nu,s}(\tau_{\nu,s} \geq n)}{\mathbb{E}_{\nu,s}(\tau_{\nu,s})} \mathbb{P}_{\nu,s}(X_n \in A \mid \tau_{\nu,s} \geq n),$$

  where $A \in \mathcal{B}(\mathsf{X})$, $\tau_{\nu,s} := \inf\{n \geq 1 : \rho_n^{(\nu,s)} = 1\}$ is the first regeneration time and $\mathbb{P}_{\nu,s}$ and $\mathbb{E}_{\nu,s}$ are probabilities and expectations w.r.t. the law of $\tilde{\mathbf{X}}_{\nu,s}$ when $X_1 \sim \nu$.
- Asmussen et al. (1992) observed this, but the expected computational time of the algorithm is not finite.
- Implementation requires use of a Bernoulli factory, and inspired Keane and O'Brien (1994).
- See also Blanchet and Meng (2007) and Flegal and Herbei (2012).

A special case

- We consider now a special case where $s = \epsilon > 0$ is a constant function.
  - This implies that $\mathbf{X}$ is uniformly ergodic.
- The same general mixture representation then yields

$$\begin{aligned} \pi(A) &= \sum_{n \geq 1} \mathbb{P}_{\nu,\epsilon}(\tau_{\nu,\epsilon} = n) \nu R_{\nu,\epsilon}^{n-1}(A) \\ &= \sum_{n \geq 1} \epsilon(1-\epsilon)^{n-1} \nu R_{\nu,\epsilon}^{n-1}(A). \end{aligned}$$

- This mixture representation was highlighted in Hobert and Robert (2004).
- Key observation (in this special case): the sample just prior to regeneration, $X_{\tau_{\nu,\epsilon}}$, is an exact sample from π.

Two algorithms

- Algorithm 1: simulate the split chain $\tilde{\mathbf{X}}_{\nu,\epsilon}$, imputing the regeneration indicators using Mykland et al. (1995):

$$\Pr\left(\rho_{n-1}^{(\nu,\epsilon)} = 1 \mid X_{n-1} = x_{n-1}, X_n = x_n\right) = \epsilon \frac{\mathrm{d}\nu(\cdot)}{\mathrm{d}P(x_{n-1},\cdot)}(x_n).$$

  - We can stop as soon as $\rho_{n-1} = 1$ for some $n \geq 1$ and we then output X_{n-1}.
- Algorithm 2: simulate $N \sim \mathrm{Geometric}(\epsilon)$, and $Y \sim \nu R_{\nu,\epsilon}^{N-1}$.
  - This is the multigamma coupler of Murdoch and Green (1998), which can also be validated using a CFTP argument.

Two algorithms

- Algorithm 1: simulate the split chain $\tilde{\mathbf{X}}_{\nu,\epsilon}$, imputing the regeneration indicators using Mykland et al. (1995):

$$\Pr\left(\rho_{n-1}^{(\nu,\epsilon)} = 1 \mid X_{n-1} = x_{n-1}, X_n = x_n\right) = \epsilon \frac{\mathrm{d}\nu(\cdot)}{\mathrm{d}P(x_{n-1},\cdot)}(x_n).$$

  - We can stop as soon as $\rho_{n-1} = 1$ for some $n \geq 1$ and we then output X_{n-1}.

- Algorithm 2: simulate $N \sim \mathrm{Geometric}(\epsilon)$, and $Y \sim \nu R_{\nu,\epsilon}^{N-1}$.
  - This is the multigamma coupler of Murdoch and Green (1998), which can also be validated using a CFTP argument.

- Problem: how can we (in general)
  - calculate $\mathrm{d}\nu(\cdot)/\mathrm{d}P(x,\cdot)$, or
  - simulate from $R_{\nu,\epsilon}(x,\cdot)$?

Outline

Singleton atoms

- We will now assume that P admits a proper, accessible, singleton atom $\alpha = \{a\}$.
- We let $p(x) := P(x, \alpha)$, and assume that

$$\underline{p} := \inf_{x \in \mathsf{X}} p(x) \geq \beta > 0,$$

  for some known constant β.
- $\underline{p} \geq \beta > 0$ implies uniform ergodicity.
- Singleton atoms are rare.
- Later, we will introduce a generic modification of P to construct a Markov kernel $\check{P}$ for which this does occur.
- It is important that we can flip a $p(x)$-coin for any $x \in \mathsf{X}$.

Bernoulli factories

- Perfect simulation in this context hinges upon our ability to solve a Bernoulli factory.
  - Given the ability to flip a p-coin, can you flip an $f(p)$-coin?
- Existence of Bernoulli factories is shown in Keane and O'Brien (1994), but the proof is not constructive.
- Bernoulli factory algorithms have been provided in (Nacu and Peres, 2005; Łatuszyński et al., 2011; Thomas and Blanchet, 2011; Flegal and Herbei, 2012; Huber, 2014).
- Most attention is paid to the Bernoulli factory for f satisfying

$$f(p) = \begin{cases} cp & cp \leq \gamma, \\ ? & \text{otherwise.} \end{cases}$$

for a given $c > 0$ and $\gamma \in (0, 1)$.

Perfect simulation algorithms

- These will all involve choosing some $\beta \in (0, \underline{p}]$ and $\epsilon \in (0, \beta)$ and, simulation of the split chain $\tilde{\mathbf{X}}_{\nu,\epsilon}$, where

$$\nu = \delta_a.$$

- Note that this is a bit unnatural since a more obvious regeneration scheme would be to use $s(x) = P(x, \alpha)$.
  - This natural approach does not lead to efficient perfect simulation algorithms.
  - We will, however, have an interest in $R_{a,p}(x, \cdot)$ later.
- The only requirement will be that we know $\beta \leq \underline{p}$.
- We suggest to choose $\epsilon = \beta/2$ (with justification to follow).

Imputation-based algorithm

1. Set $X_1 = a$.
2. For $n = 2, 3, \ldots$:
   - 2.1 Simulate $X_n \sim P(X_{n-1}, \cdot)$.
   - 2.2 If $X_n = a$, sample $\rho_{n-1}^{(a,\epsilon)} \sim \text{Bernoulli}(\epsilon / p(X_{n-1}))$. Otherwise, set $\rho_{n-1}^{(a,\epsilon)} = 0$.
   - 2.3 If $\rho_n^{(a,\epsilon)} = 1$, stop and output X_{n-1}.

Imputation-based algorithm

1. Set $X_1 = a$.
2. For $n = 2, 3, \ldots$:
   - 2.1 Simulate $X_n \sim P(X_{n-1}, \cdot)$.
   - 2.2 If $X_n = a$, sample $\rho_{n-1}^{(a,\epsilon)} \sim \text{Bernoulli}(\epsilon / p(X_{n-1}))$. Otherwise, set $\rho_{n-1}^{(a,\epsilon)} = 0$.
   - 2.3 If $\rho_n^{(a,\epsilon)} = 1$, stop and output X_{n-1}.

- The important part is that

$$\Pr\left(\rho_{n-1}^{(a,\epsilon)} = 1 \mid X_{n-1} = x_{n-1}, X_n = x_n\right) = \epsilon \frac{\mathbb{I}(x_n = a)}{P(x_{n-1}, \alpha)}.$$

- We are all set if we can flip an (ϵ/p)-coin for arbitrary $p > \epsilon$.

Multigamma coupler I

1. Sample $N \sim \mathrm{Geometric}(\epsilon)$.
2. Set $X_1 = a$.
3. For $n = 2, 3, \ldots, N$:
   3.1 Sample $X_n \sim R_{a,\epsilon}(X_{n-1}, \cdot)$.
4. Output X_N.

Multigamma coupler I

1. Sample $N \sim \mathrm{Geometric}(\epsilon)$.
2. Set $X_1 = a$.
3. For $n = 2, 3, \ldots, N$:
   3.1 Sample $X_n \sim R_{a,\epsilon}(X_{n-1}, \cdot)$.
4. Output X_N.

- Great, but how can I sample from $R_{a,\epsilon}(X_{n-1}, \cdot)$?

Multigamma coupler II

- We need to sample from

$$\begin{aligned} R_{a,\epsilon}(x,\cdot) &= \frac{P(x,\cdot)-\epsilon\delta_a(\cdot)}{1-\epsilon} \\ &= \frac{1-p(x)}{1-\epsilon}R_{a,p}(x,\cdot)+\frac{p(x)-\epsilon}{1-\epsilon}\delta_a(\cdot). \end{aligned}$$

- So with probability $[1-p(x)]/[1-\epsilon]$ we simulate from $R_{a,p}(x,\cdot)$, otherwise output a.
- We can trivially sample from $R_{a,p}(x,\cdot)$ by rejection:

$$R_{a,p}(x,\mathrm{d}y) = \frac{P(x,\mathrm{d}y)\mathbb{I}(y\neq a)}{1-p(x)}.$$

Multigamma coupler III

1. Sample $N \sim \text{Geometric}(\epsilon)$.
2. Set $X_1 = a$.
3. For $n = 2, 3, \ldots, N$:
   - 3.1 Simulate $Y_n \sim \text{Bernoulli}([1 - p(X_{n-1})] \,/\, [1 - \epsilon])$.
   - 3.2 If $Y_n = 1$, sample $X_n \sim R_{a,p}(X_{n-1}, \cdot)$. Otherwise set $X_n = a$.
4. Output X_N.

Multigamma coupler III

1. Sample $N \sim \text{Geometric}(\epsilon)$.
2. Set $X_1 = a$.
3. For $n = 2, 3, \ldots, N$:
   3.1 Simulate $Y_n \sim \text{Bernoulli}([1 - p(X_{n-1})] \,/\, [1 - \epsilon])$.
   3.2 If $Y_n = 1$, sample $X_n \sim R_{a,p}(X_{n-1}, \cdot)$. Otherwise set $X_n = a$.
4. Output X_N.

- We are all set if we can flip a $([1 - p]/[1 - \epsilon])$-coin for arbitrary $p > \epsilon$.

Bernoulli factory algorithms

- The solution for $f(p) = [1-p]/[1-\epsilon]$ is solved by standard algorithms.
- For $f(p) = \epsilon/p$ we can flip an $f(p)$-coin by:
  - Simulating $K \sim \text{Geometric}(\epsilon)$.
  - Simulating a $[(1-p)/(1-\epsilon)]^{K-1}$-coin.
- From the Maclaurin series for $1/[1-(1-p)] = 1/p$:

$$\frac{\epsilon}{p} = \epsilon \sum_{k=1}^{\infty} (1-p)^{k-1} = \sum_{k=1}^{\infty} \epsilon (1-\epsilon)^{k-1} \left(\frac{1-p}{1-\epsilon} \right)^{k-1}.$$

- In practice, one can "stop early" if any of the $(1-p)/(1-\epsilon)$-coin flips are 0.

Cost of perfect simulation

Proposition

Assume $\beta \leq 0.5$ and $\epsilon = \beta/2$. Then the expected number of simulations from P order to obtain a perfect sample using either the imputation approach or the multigamma coupler is $12\epsilon^{-1}$.

- Expected number of $(1-p)/(1-\epsilon)$-coin flips required to simulate a single tour of the split chain $\tilde{\mathbf{X}}_{a,\epsilon}$ is $\epsilon^{-1} - 1$.
- Expected number of samples from P to additionally simulate the tour itself is ϵ^{-1}.
- With $\beta \leq 0.5$, $\epsilon = \beta/2$, and using the Bernoulli factory algorithm of Huber (2014), the expected number of p-coin flips to produce a $(1-p)/(1-\epsilon)$-coin flip is bounded above by 11.

Outline

Artificial singleton atoms

- We propose to use generic methodology along the lines of Brockwell and Kadane (2005).
- That is, we introduce a new transition kernel $\check{P}$ that evolves on $\check{\mathsf{X}} := \mathsf{X} \cup \alpha = \mathsf{X} \cup \{a\}$.
- We require that it is Harris recurrent and irreducible with unique invariant probability measure $\check{\pi}$ satisfying, for some $k \in (0, 1)$,

$$\check{\pi}(A) = k\pi(A \cap \mathsf{X}) + (1 - k)\mathbb{I}(a \in A), \quad A \in \mathcal{B}(\check{\mathsf{X}}).$$

- When this holds, it follows that $\check{\pi}(A) = k\pi(A)$ for any $A \in \mathcal{B}(\mathsf{X})$.
- We denote by $\check{\mathbf{X}} := (\check{X}_n)_{n \geq 1}$ the Markov chain with transition kernel $\check{P}$.

Definition of $\check{\pi}$

- In many applications π admits a density w.r.t. to a dominating measure λ on X and we can compute an unnormalized version $\gamma(x)$ of this density.
- In this case, we can choose a $b > 0$ and define an unnormalized version $\check{\gamma}(x)$ of the density of $\check{\pi}$ w.r.t. the dominating measure $\lambda + \delta_a$ on $\check{\mathsf{X}}$ through

$$\check{\gamma}(x) := \mathbb{I}(x \in \mathsf{X})\gamma(x) + \mathbb{I}(x = a)b.$$

- It follows that $\check{\pi}(\mathrm{d}x) = \check{\gamma}(x)\left\{\lambda\left(\mathrm{d}x\right) + \delta_a\left(\mathrm{d}x\right)\right\}/\check{\gamma}(\mathsf{X})$ satisfies our requirements with $k = \{1 + b/\gamma(\mathsf{X})\}^{-1}$.

Comments on the construction

- In practice, we would like $\check{\pi}(\{a\})$ to be not too close to either 0 or 1 so
  - $\check{P}(x, \{a\})$ can be fairly large, but
  - perfect samples from $\check{\pi}$ are often X-valued.
- An estimate of $\gamma(\mathrm{X})$ is necessary to be able to choose an appropriate value of b.

A simple example

- We define, for some $w \in (0, 1)$ and transition kernels Π_1 and Π_2,

$$\check{P}(x, \mathrm{d}y) := wP_1(x, \mathrm{d}y) + (1 - w)P_2(x, \mathrm{d}y),$$

  where $P_1(x, \mathrm{d}y) = \mathbb{I}(x \in \mathsf{X})P(x, \mathrm{d}y) + \mathbb{I}(x = a)\delta_a(\mathrm{d}y)$ and P_2 allows the chain to move between X and $\{a\}$.
- One choice of P_2, suggested by Brockwell and Kadane (2005), is a Metropolis–Hastings kernel with proposal

$$Q_x(\mathrm{d}y) = \mathbb{I}(x \in \mathsf{X})\delta_a(\mathrm{d}y) + \mathbb{I}(x = a)\mu(\mathrm{d}y),$$

  where μ is a "re-entry" distribution.

Useful results

Proposition

Assume that a generic Markov kernel $\check{P} : \check{\mathsf{X}} \times \mathcal{B}(\check{\mathsf{X}}) \to [0,1]$ *satisfies* $\check{P}(a, \mathsf{X}) > 0$ *and for some* $w > 0$,

$$\check{P}(x, A) \geq wP(x, A), \quad x \in \mathsf{X}, A \in \mathcal{B}(\mathsf{X}).$$

Then $\mathbf{X}$ *being uniformly ergodic implies that* $\check{\mathbf{X}}$ *is uniformly ergodic (although the converse does not hold).*

- Moreover, the existence of a $\beta > 0$ such that $\underline{p} \geq \beta$ is guaranteed in general for uniformly ergodic $\check{\mathbf{X}}$.
- This requires one to consider a m-step transition kernel since for some $m \in \mathbb{N}$ and $d > 0$,

$$\inf_{x \in \check{\mathsf{X}}} \check{P}^m(x, \{a\}) \geq d.$$

Outline

The problem & solution

- The goal is to use the iterated conditional SMC Markov kernel to sample from π, since

$$P_N(x, \cdot) \geq \epsilon_N \pi(\cdot).$$

- The problem(s):
  - We cannot evaluate $\epsilon_N \frac{\mathrm{d}\pi(\cdot)}{\mathrm{d}P_{\boldsymbol{N}}(x,\cdot)}$ pointwise.
  - We cannot sample according to ν (which is π in this case!) or the residual kernel $R_{\pi,\epsilon_{\boldsymbol{N}}}$.
  - There is no proper, accessible, singleton atom in general.
- The solution: we will introduce an artificial singleton atom.

Recap on discrete-time Feynman–Kac path measures

- We focus on a generic discrete-time Feynman–Kac model with time horizon n.
- Let $(\mathsf{Z}, \mathcal{B}(\mathsf{Z}))$ be a measurable space and consider
  - a probability measure $\mu : \mathcal{B}(\mathsf{Z}) \to [0, 1]$,
  - some Markov kernels $M_p : \mathsf{Z} \times \mathcal{B}(\mathsf{Z}) \to [0, 1]$ for $p \in \{2, \ldots, n\}$ and
  - non-negative $\mathcal{B}(\mathsf{Z})$-measurable functions $G_p : \mathsf{Z} \to \mathbb{R}_+$ for $p \in \{1, \ldots, n\}$.
- We define for any $p \in \{1, \ldots, n\}$, the measure γ_p by

$$\gamma_p(A) := \int_A \left[\prod_{q=1}^{p} G_q(z_q)\right] \mu(\mathrm{d}z_1) \prod_{q=2}^{p} M_q(z_{q-1}, \mathrm{d}z_q), \quad A \in \mathcal{B}(\mathsf{Z}^p),$$

  and its associated probability measure $\pi_p := \gamma_p(\mathsf{Z}^p)^{-1}\gamma_p$.
- With $\mathsf{X} := \mathsf{Z}^n$ the Feynman–Kac path measure of interest is the probability measure $\pi := \pi_n$ on $\mathcal{B}(\mathsf{X})$.

Atomic extension of a Feynman–Kac path measure

- Let $\check{\mathsf{Z}} := \mathsf{Z} \cup \alpha$, where $\alpha = \{a\}$ and a is a distinguished point.
- Let $\check{\mathsf{X}} := \check{\mathsf{Z}}^n$ and $a_n := (a, \ldots, a)$.
- We propose a generic way to define a new probability measure $\check{\pi}$ on $\check{\mathsf{X}}$ which satisfies for some $k \in (0, 1)$,

$$\check{\pi}(A) = k\pi(A \cap \mathsf{X}) + (1 - k)\mathbb{I}(a_n \in A), \quad A \in \mathcal{B}(\check{\mathsf{X}}).$$

Atomic extension of a Feynman–Kac path measure

- The extended Feynman–Kac model is defined by the initial distribution $\check{\mu}$, the Markov kernels $\check{\mathcal{M}} := (\check{M}_p)_{p\in\{2,\ldots,n\}}$ and potential functions $\check{\mathcal{G}} := (\check{G}_p)_{p\in\{1,\ldots,n\}}$ on $\check{\mathsf{Z}}$ which are given by

$$\begin{aligned}
\check{\mu}(A) &:= (1-b)\mu(A\cap \mathsf{Z}) + b\mathbb{I}\{a\in A\},\\
\check{M}_p(x,A) &:= M_p(x,A)\mathbb{I}\{x\in \mathsf{X}\} + \mathbb{I}\{x=a, A=\alpha\},\\
\check{G}_p(x) &:= G_p(x)\mathbb{I}\{x\in \mathsf{X}\} + \psi_p\mathbb{I}\{x=a\},
\end{aligned}$$

  for $A\in\mathcal{B}(\check{\mathsf{Z}})$ where $b\in(0,1)$ and $\psi_1,\ldots,\psi_n$ are user-defined positive constants.

Atomic extension of a Feynman–Kac path measure

- We define the measure $\check{\gamma}_n$ by

$$\check{\gamma}_n(A) := \int_A \left[\prod_{p=1}^{n} \check{G}_p(x_p)\right] \check{\mu}(\mathrm{d}x_1) \prod_{p=2}^{n} \check{M}_p(x_{p-1}, \mathrm{d}x_p), \quad A \in \mathcal{B}(\check{\mathsf{X}}),$$

  and its associated probability measure $\check{\pi} := \check{\gamma}_n(\check{\mathsf{X}})^{-1}\check{\gamma}_n$.

- It follows that

$$\check{\pi}(A) = k\pi(A \cap \mathsf{X}) + (1-k)\mathbb{I}(a_n \in A), \quad A \in \mathcal{B}(\check{\mathsf{X}})$$

  holds with

$$k = \frac{1-b}{1-b+b\gamma_n(\mathsf{X})^{-1}\prod_{p=1}^{n}\psi_p}.$$

Perfect simulation from a Feynman–Kac path measure

- If we assume π-essential boundedness of each G_p, then

$$\inf_{x \in \check{\mathsf{X}}} \check{P}_N(x, \{a_n\}) \geq \check{\epsilon}_N \check{\pi}(\{a_n\}),$$

  where $\lim_{N \to \infty} \check{\epsilon}_N = 1$.
- For our perfect simulation algorithms we need a lower bound on $\check{\epsilon}_N \times \check{\pi}(\{a_n\})$.
- Such bounds are typically difficult to obtain analytically, but it is straightforward to obtain conservative estimates.

Estimation ingredients

- The value of $\check{\pi}(\{a_n\}) = 1 - k$ depends only on b and $\prod_{p=1}^{n} \psi_p$:

$$k = \frac{1 - b}{1 - b + b\gamma_n(\mathrm{X})^{-1} \prod_{p=1}^{n} \psi_p}.$$

- We have $\lim_{N\to\infty} \check{\epsilon}_N = 1$, and performance is improved if

$$\psi_q \approx \frac{\gamma_q(\mathrm{X})}{\gamma_{q-1}(\mathrm{X})}.$$

- All of the ψ_q can be estimated using standard SMC.
- One can also use diagnostics that check probabilistically that $\inf_{x\in\check{\mathrm{X}}} \check{P}_N(x, \{a_n\}) \geq \beta$ for some $\beta > 0$.

Outline

Diagnostics and estimation of β

- Now we return to the general setting with $\mathbf{X} = (X_n)_{n \geq 1}$ the Markov chain with transition kernel P.
- Recall that all we require in general is knowing $\underline{p} = \inf_{x \in \mathsf{X}} P(x, \alpha) \geq \beta$ for some known $\beta > 0$.
- One approach could be to find $\underline{p}$ by some stochastic optimization procedure.
- Another approach is to simulate $\mathbf{X}$ for a long time and estimate $\underline{p}$ using the chain.
  - One could then impute regeneration indicators to obtain perfect samples.
- Yet another could be to monitor the validity of the assumption that a chosen β satisfies $\beta \leq \underline{p}$.

Monitoring diagnostic

- At each state x visited during the course of any of the algorithms, one can simply flip $p(x)$-coins until their average exceeds β.
  1. If $\pi - \operatorname{ess\,inf}_{x \in X} p(x) < \beta$, then the algorithm will not terminate with positive probability.
  2. If $\pi - \operatorname{ess\,inf}_{x \in X} p(x) = \beta$, then the algorithm has infinite expected running time.
  3. If $\pi - \operatorname{ess\,inf}_{x \in X} p(x) > \beta$, then the algorithm has finite expected running time.
- Quantitative results are also available for the expected running time.

Other remarks

- One can obtain quantitative bounds on the total variation distance between the target probability measure and the probability measure one samples from if $\beta < \underline{p}$.
- One can use a single perfect sample to obtain unbiased and consistent estimates using MCMC or SMC.
- Extensions to the general methodology for non-uniformly ergodic Markov chains are complicated but possible in principle.
- Theoretical work on establishing rigorous bounds on $\underline{p}$ is of practical interest.
- In an ideal case with a "forgetting" Feynman–Kac model, the overall perfect simulation procedure is (expected) $\mathcal{O}(n^2)$.
  - contrast with $\mathcal{O}(\exp(n))$ expected time for rejection sampling.

References I

Andrieu, C., N. Chopin, A. Doucet, and S. Rubenthaler (2012). Perfect simulation for the Feynman–Kac law on the path space. Preprint arXiv:1210.0376.

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B 72*(3), 269–342.

Andrieu, C., A. Lee, and M. Vihola (2013). Uniform ergodicity of the iterated conditional SMC and geometric ergodicity of particle Gibbs samplers. Preprint arXiv:1312.6432.

Asmussen, S., P. W. Glynn, and H. Thorisson (1992). Stationarity detection in the initial transient problem. *ACM Transactions on Modeling and Computer Simulation (TOMACS) 2*(2), 130–157.

Athreya, K. B. and P. Ney (1978). A new approach to the limit theory of recurrent Markov chains. *Transactions of the American Mathematical Society 245*, 493–501.

Blanchet, J. and X. Meng (2007). Exact sampling, regeneration and minorization conditions. Preprint, Dept. of Statistics, Harvard University.

Brockwell, A. E. and J. B. Kadane (2005). Identification of regeneration times in MCMC simulation, with application to adaptive schemes. *Journal of Computational and Graphical Statistics 14*(2).

Chopin, N. and S. S. Singh (2013). On the particle Gibbs sampler. Preprint arXiv:1304.1887v1.

Flegal, J. M. and R. Herbei (2012). Exact sampling for intractable probability distributions via a Bernoulli factory. *Electronic Journal of Statistics 6*, 10–37.

References II

Hobert, J. P., G. L. Jones, and C. P. Robert (2006). Using a Markov chain to construct a tractable approximation of an intractable probability distribution. *Scandinavian Journal of Statistics 33*(1), 37–51.

Hobert, J. P. and C. P. Robert (2004). A mixture representation of π with applications in Markov chain Monte Carlo and perfect sampling. *Annals of Applied Probability*, 1295–1305.

Huber, M. (2014). Nearly optimal Bernoulli factories for linear functions. *Combinatorics, Probability, and Computing*. To appear.

Keane, M. and G. L. O'Brien (1994). A Bernoulli factory. *ACM Transactions on Modeling and Computer Simulation (TOMACS) 4*(2), 213–219.

Łatuszyński, K., I. Kosmidis, O. Papaspiliopoulos, and G. O. Roberts (2011). Simulating events of unknown probabilities via reverse time martingales. *Random Structures & Algorithms 38*(4), 441–452.

Lindsten, F., R. Douc, and E. Moulines (2014). Uniform ergodicity of the particle Gibbs sampler. Preprint arXiv:1401.0683v1.

Murdoch, D. J. and P. J. Green (1998). Exact sampling from a continuous state space. *Scandinavian Journal of Statistics 25*(3), 483–502.

Mykland, P., L. Tierney, and B. Yu (1995). Regeneration in Markov chain samplers. *Journal of the American Statistical Association 90*(429), 233–241.

Nacu, Ş. and Y. Peres (2005). Fast simulation of new coins from old. *The Annals of Applied Probability 15*(1A), 93–115.

References III

Nummelin, E. (1978). A splitting technique for Harris recurrent Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 43*(4), 309–318.

Propp, J. G. and D. B. Wilson (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random structures and Algorithms 9*(1-2), 223–252.

Thomas, A. and J. H. Blanchet (2011). A practical implementation of the Bernoulli factory. Preprint arXiv:1106.2508.