An introduction to sequential Monte Carlo

Lecture 3: General SMC with Feynman-Kac models

Dr Joshua Bon

SMC Down Under 2023

Queensland University of Technology

Table of contents

Introduction

What's in a name?

Richard Feynman and Mark Kac

Motivation

- Describes many algorithms using interacting particle systems
- Pertains to both discrete, continuous, and mixed random variables
- Unified analysis with theoretical justification
- No “reinventing the wheel”
- Understand technical papers

Pierre Del Moral (2004). *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications.* New York: Springer-Verlag

Probability measures

Assumptions

- Measurable space $(X, \mathcal{X})$
- μ is a probability measure on $(X, \mathcal{X})$, $\mu(X) = 1$
- Consider sets $S \in \mathcal{X}$
- Integrable function $\varphi : X \to Z$

Probability measures

Probability measures assign a *unit* mass to sets in a coherent way

$$\mu(S) = \int_S \mu(\mathrm{d}x)$$

Probability measures

Probability measures assign a *unit* mass to sets in a coherent way

$$\mu(S) = \int_S \mu(\mathrm{d}x)$$

Or equivalently

$$\mu(S) = \int_X 1(x \in S)\mu(\mathrm{d}x)$$

Random variables

A "random variable" X has a probability law μ when

$$\mathbb{P}(X \in S) = \mu(S)$$

Random variables

A "random variable" X has a probability law μ when

$$\mathbb{P}(X \in S) = \mu(S)$$

We say X is distributed according to μ, or

$$X \sim \mu$$

Expectations

The expectation of $\varphi(X)$ when $X \sim \mu$ is

$$\mathbb{E}[\varphi(X)] = \mu(\varphi) = \int_{\mathsf{X}} \varphi(x)\mu(\mathrm{d}x)$$

Expectations

The expectation of $\varphi(X)$ when $X \sim \mu$ is

$$\mathbb{E}[\varphi(X)] = \mu(\varphi) = \int_{\mathsf{X}} \varphi(x)\mu(\mathrm{d}x)$$

Examples

$\varphi(x)$	$\mathbb{E}[\cdot]$	$\mu(\varphi)$
x	$\mathbb{E}[X]$	$\int_{\mathsf{X}} x\mu(\mathrm{d}x)$
x^2	$\mathbb{E}[X^2]$	$\int_{\mathsf{X}} x^2\mu(\mathrm{d}x)$
$(x-m)^2$	$\mathbb{E}[(X-m)^2]$	$\int_{\mathsf{X}}(x-m)^2\mu(\mathrm{d}x)$
$1(x \in S)$	$\mathbb{E}[1(x \in S)]$	$\int_{\mathsf{X}} 1(x \in S)\mu(\mathrm{d}x)$
$1(x \in S)$	$\mathbb{P}(x \in S)$	$\int_S \mu(\mathrm{d}x)$

Short hand notation

Writing $\mu(\mathrm{d}x)$ can be thought of as short hand for

$$\mu(\bullet) = \int_{\mathsf{X}} \bullet \, \mu(\mathrm{d}x)$$

Probability density functions

If you *squint* the right way, you can think of $\mu(\mathrm{d}x)$ as a pdf,

$$\mu(\mathrm{d}x) = p(x)\mathrm{d}x$$

where $\mathrm{d}x$ represents the base measure[1]

[1] Typically the Lebesgue measure

Probability density functions

For $(\mathsf{X}, \mathcal{X}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, a pdf exists and we have the Riemann integral interpretation

$$\mathbb{P}(X \in S) = \mu(S) = \int_S p(x)\mathrm{d}x$$

Probability density functions

For $(\mathsf{X}, \mathcal{X}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, a pdf exists and we have the Riemann integral interpretation

$$\mathbb{P}(X \in S) = \mu(S) = \int_{S} p(x) \mathrm{d}x$$

$$\mathbb{E}[\varphi(X)] = \mu(\varphi) = \int_{\mathsf{X}} \varphi(x) p(x) \mathrm{d}x$$

New measures from old measures

If we have integrable function $\psi : X \to [0, \infty)$ then

writing $\psi(x)\mu(\mathrm{d}x)$ is really defining the weighted measure ν

$$\nu(\varphi) = \int_X \varphi(x)\psi(x)\mu(\mathrm{d}x)$$

New measures from old measures

If we have integrable function $\psi : \mathsf{X} \to [0, \infty)$ then

writing $\psi(x)\mu(\mathrm{d}x)$ is really defining the weighted measure ν

$$\nu(\varphi) = \int_{\mathsf{X}} \varphi(x)\psi(x)\mu(\mathrm{d}x)$$

Or with short hand

$$\nu(\mathrm{d}x) = \psi(x)\mu(\mathrm{d}x)$$

Markov kernels

Markov kernels $M(\cdot,\cdot)$ are used as

- $M(v,\cdot)$ (conditional) probability measures for fixed v,
- $M(\cdot,\varphi)$ functions for fixed φ

Markov kernels

Markov kernels $M(\cdot, \cdot)$ are used as

- $M(v, \cdot)$ (conditional) probability measures for fixed v,
- $M(\cdot, \varphi)$ functions for fixed φ

We write

- $M(v, \mathrm{d}x) \Longleftrightarrow M(v, \bullet) = \int_{\mathsf{X}} \bullet \, M(v, \mathrm{d}x)$
- $M(\varphi)(\cdot) = M(\cdot, \varphi)$, i.e. $M(\varphi)(v)$ is a function of v

Feynman-Kac models

Inhomogeneous Markov chains

$$M_1(\mathrm{d}x_1)M_2(x_1, \mathrm{d}x_2)M_3(x_2, \mathrm{d}x_3) \cdots M_n(x_{n-1}, \mathrm{d}x_n)$$

for probability measure M_1 and Markov kernels $M_{2:n}$ with $x_t \in \mathsf{X}$

Inhomogeneous Markov chains

$$M_1(\mathrm{d}x_1)M_2(x_1, \mathrm{d}x_2)M_3(x_2, \mathrm{d}x_3) \cdots M_n(x_{n-1}, \mathrm{d}x_n)$$

for probability measure M_1 and Markov kernels $M_{2:n}$ with $x_t \in \mathsf{X}$

$$\mathbf{M}_n(\mathrm{d}x_{1:n}) = M_1(\mathrm{d}x_1) \prod_{t=2}^{n} M_t(x_{t-1}, \mathrm{d}x_t)$$

Inhomogeneous Markov chains

$$M_1(\mathrm{d}x_1)M_2(x_1, \mathrm{d}x_2)M_3(x_2, \mathrm{d}x_3)\cdots M_n(x_{n-1}, \mathrm{d}x_n)$$

for probability measure M_1 and Markov kernels $M_{2:n}$ with $x_t \in \mathsf{X}$

$$\mathbb{M}_n(\mathrm{d}x_{1:n}) = M_1(\mathrm{d}x_1)\prod_{t=2}^{n} M_t(x_{t-1}, \mathrm{d}x_t)$$

- State space: $x_t \in \mathsf{X}$
- Path space: $x_{1:n} \in \mathsf{X}^n$

Potential functions

$$G_t : \mathsf{X} \to [0, \infty)$$

Potential functions

$$G_t : \mathsf{X} \to [0, \infty)$$

$$\boldsymbol{G}_n(x_{1:n}) = \prod_{t=1}^{n} G_t(x_t)$$

Potential functions

$$G_t : \mathsf{X} \to [0, \infty)$$

$$\boldsymbol{G}_n(x_{1:n}) = \prod_{t=1}^{n} G_t(x_t)$$

Examples

Setting	Description	$G_t(x_t)$
General HMM	Likelihood function	$p(y_t \mid x_t)$
Normal HMM	Observation error	$\mathcal{N}(y_t; x_t, \Sigma)$
Static SMC	Incremental weighting function	$\frac{\pi_{t+1}(x_t)}{\pi_t(x_t)}$
ABC-SMC	Constraints	$1(x_t \in S_t)$

Path measures (predictive)

$$M_1(\mathrm{d}x_1)G_1(x_1)M_2(x_1, \mathrm{d}x_2)G_2(x_2) \cdots M_n(x_{n-1}, \mathrm{d}x_n)$$

Path measures (predictive)

$$M_1(\mathrm{d}x_1)G_1(x_1)M_2(x_1,\mathrm{d}x_2)G_2(x_2)\cdots M_n(x_{n-1},\mathrm{d}x_n)$$

$$\gamma_n(\mathrm{d}x_{1:n}) = \boldsymbol{M}_n(\mathrm{d}x_{1:n})\boldsymbol{G}_{n-1}(x_{1:n-1})$$

Path measures (predictive)

$$M_1(\mathrm{d}x_1)G_1(x_1)M_2(x_1, \mathrm{d}x_2)G_2(x_2)\cdots M_n(x_{n-1}, \mathrm{d}x_n)$$

$$\boldsymbol{\gamma}_n(\mathrm{d}x_{1:n}) = \boldsymbol{M}_n(\mathrm{d}x_{1:n})\boldsymbol{G}_{n-1}(x_{1:n-1})$$

$$\boldsymbol{\eta}_n(\mathrm{d}x_{1:n}) = \frac{\boldsymbol{\gamma}_n(\mathrm{d}x_{1:n})}{Z_n}$$

Path measures (predictive)

$$M_1(\mathrm{d}x_1)G_1(x_1)M_2(x_1, \mathrm{d}x_2)G_2(x_2)\cdots M_n(x_{n-1}, \mathrm{d}x_n)$$

$$\boldsymbol{\gamma}_n(\mathrm{d}x_{1:n}) = \boldsymbol{M}_n(\mathrm{d}x_{1:n})\boldsymbol{G}_{n-1}(x_{1:n-1})$$

$$\boldsymbol{\eta}_n(\mathrm{d}x_{1:n}) = \frac{\boldsymbol{\gamma}_n(\mathrm{d}x_{1:n})}{Z_n}$$

Example

- HMM joint predictive, $p(x_{1:n} \mid y_{1:n-1})$

Path measures (updated)

$$M_1(\mathrm{d}x_1)G_1(x_1)M_2(x_1, \mathrm{d}x_2)G_2(x_2) \cdots M_n(x_{n-1}, \mathrm{d}x_n)G_n(x_n)$$

$$\hat{\gamma}_n(\mathrm{d}x_{1:n}) = \boldsymbol{M}_n(\mathrm{d}x_{1:n})\boldsymbol{G}_n(x_{1:n})$$

Path measures (updated)

$$M_1(\mathrm{d}x_1)G_1(x_1)M_2(x_1, \mathrm{d}x_2)G_2(x_2) \cdots M_n(x_{n-1}, \mathrm{d}x_n)G_n(x_n)$$

$$\hat{\gamma}_n(\mathrm{d}x_{1:n}) = \boldsymbol{M}_n(\mathrm{d}x_{1:n})\boldsymbol{G}_n(x_{1:n})$$

$$\hat{\boldsymbol{\eta}}_n(\mathrm{d}x_{1:n}) = \frac{\hat{\gamma}_n(\mathrm{d}x_{1:n})}{Z_n}$$

Path measures (updated)

$$M_1(\mathrm{d}x_1)G_1(x_1)M_2(x_1, \mathrm{d}x_2)G_2(x_2)\cdots M_n(x_{n-1}, \mathrm{d}x_n)G_n(x_n)$$

$$\hat{\gamma}_n(\mathrm{d}x_{1:n}) = \boldsymbol{M}_n(\mathrm{d}x_{1:n})\boldsymbol{G}_n(x_{1:n})$$

$$\hat{\boldsymbol{\eta}}_n(\mathrm{d}x_{1:n}) = \frac{\hat{\gamma}_n(\mathrm{d}x_{1:n})}{Z_n}$$

Example

- HMM joint filtering, $p(x_{1:n} \mid y_{1:n})$

Normalised path measures

$$Z_n = \int \mathsf{G}_{n-1}(x_{1:n-1})\mathsf{M}_n(\mathrm{d}x_{1:n-1})$$
$$\hat{Z}_n = \int \mathsf{G}_n(x_{1:n})\mathsf{M}_n(\mathrm{d}x_{1:n})$$

Normalised path measures

$$
\begin{aligned}
Z_n &= \int \mathsf{G}_{n-1}(x_{1:n-1})\mathsf{M}_n(\mathrm{d}x_{1:n-1}) \\
\hat{Z}_n &= \int \mathsf{G}_n(x_{1:n})\mathsf{M}_n(\mathrm{d}x_{1:n})
\end{aligned}
$$

or alternatively

$$
\begin{aligned}
Z_n &= \mathbb{E}\left[\mathsf{G}_{n-1}(X_{1:n-1})\right], \\
\hat{Z}_n &= \mathbb{E}\left[\mathsf{G}_n(X_{1:n})\right], \qquad\qquad X_{1:n} \sim \mathsf{M}_n.
\end{aligned}
$$

Marginal measures from recursion

$$\gamma_1(\mathrm{d}x_1) = M_1(\mathrm{d}x_1)$$

$$\gamma_t(S) = \int_{\mathsf{X}} \gamma_{t-1}(\mathrm{d}x_{t-1}) G_{t-1}(x_{t-1}) M_t(x_{t-1}, S)$$

$$\hat{\gamma}_t(\mathrm{d}x_t) = \gamma_t(\mathrm{d}x_t) G_t(x_t)$$

Marginal probability measures

$$\eta_t(\mathrm{d}x_t) = Z_t^{-1}\gamma_t(\mathrm{d}x_t)$$

$$\hat{\eta}_t(\mathrm{d}x_t) = \hat{Z}_t^{-1}\hat{\gamma}_t(\mathrm{d}x_t)$$

The Bootstrap Particle Filter revisited

Bootstrap Particle Filter: Feynman-Kac formalism

Algorithm 1 The Bootstrap Particle Filter

1. Sample initial $\zeta_1^i \overset{\text{iid}}{\sim} M_1$ for $i \in \{1, \ldots, N\}$
2. For each time $t = 2, \ldots, n$
   (i) Sample ancestors $A_{t-1}^i \sim \text{Cat}\left(G_{t-1}(\zeta_{t-1}^1), \ldots, G_{t-1}(\zeta_{t-1}^N)\right)$ for $i \in \{1, \ldots, N\}$
   (ii) Sample prediction $\zeta_t^i \sim M_t(\zeta_{t-1}^{A_{t-1}^i}, \cdot)$ for $i \in \{1, \ldots, N\}$

Output: Particles $\zeta_t^{1:N}$ for $t = 1, \ldots, n$

Hidden Markov models

Markov process

Initial distribution	*Prior density*
$M_1(\mathrm{d}x_1)$	$p(x_1)$
Markov kernel	*Forward kernel density*
$M_t(x_{t-1}, \mathrm{d}x_t)$	$p(x_t \mid x_{t-1})$

Potential function

Potential function	*Likelihood function*
$G_t(x_t)$	$p(y_t \mid x_t)$

Marginal distributions

Predictive distribution	*Predictive density*
$\eta_t(\mathrm{d}x_t)$	$p(x_t \mid y_{1:t-1})$
Updated distribution	*Filtering density*
$\hat{\eta}_t(\mathrm{d}x_t)$	$p(x_t \mid y_{1:t})$

Static models (Resample-move SMC)

Markov process

Initial distribution	*Prior density*
$M_1(\mathrm{d}x_1)$	$\pi_1(x_1)$
Markov kernel	*MCMC kernel density*
$M_t(x_{t-1}, \mathrm{d}x_t)$	$q_t(x_t \mid x_{t-1})$

Potential function

Potential function	*Incremental weighting function*
$G_t(x_t)$	$\pi_{t+1}(x_t \mid y)/\pi_t(x_t \mid y)$

Static models (Resample-move SMC)

Marginal distributions

Predictive distribution	*Predictive density*
$\eta_t(\mathrm{d}x_t)$	$\pi_t(x_t)$
Updated distribution	*Filtering density*
$\hat{\eta}_t(\mathrm{d}x_t)$	$\pi_{t+1}(x_t)$

Feynman-Kac models in the wild

Waste-free SMC

Received: 12 November 2020 | Accepted: 13 September 2021

DOI: 10.1111/rssb.12475

ORIGINAL ARTICLE

Waste-free sequential Monte Carlo

Hai-Dang Dau | **Nicolas Chopin**

CREST-ENSAE and Institut Polytechnique de Paris, Palaiseau Cedex, France

Correspondence
Nicolas Chopin, CREST-ENSAE and Institut Polytechnique de Paris, 5 avenue Henry Le Chatelier, 91764 Palaiseau Cedex, France.
Email: nicolas.chopin@ensae.fr

Abstract

A standard way to move particles in a sequential Monte Carlo (SMC) sampler is to apply several steps of a Markov chain Monte Carlo (MCMC) kernel. Unfortunately, it is not clear how many steps need to be performed for optimal performance. In addition, the output of the intermediate steps are discarded and thus wasted somehow. We propose a new waste-free SMC algorithm

Iterated auxiliary particle filter

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2017, VOL. 112, NO. 520, 1636–1647, Theory and Methods
http://dx.doi.org/10.1080/01621459.2016.1222291

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS

The Iterated Auxiliary Particle Filter

Pieralberto Guarniero, Adam M. Johansen, and Anthony Lee

Department of Statistics, University of Warwick, Coventry, UK

ABSTRACT

We present an offline, iterated particle filter to facilitate statistical inference in general state space hidden Markov models. Given a model and a sequence of observations, the associated marginal likelihood L is central to likelihood-based inference for unknown statistical parameters. We define a class of "twisted" models: each member is specified by a sequence of positive functions $\boldsymbol{\psi}$ and has an associated $\boldsymbol{\psi}$-auxiliary particle filter that provides unbiased estimates of L. We identify a sequence $\boldsymbol{\psi}^*$ that is optimal in the sense that the $\boldsymbol{\psi}^*$-auxiliary particle filter's estimate of L has zero variance. In practical applications, $\boldsymbol{\psi}^*$ is unknown so the $\boldsymbol{\psi}^*$-auxiliary particle filter cannot straightforwardly be implemented. We use an iterative scheme to approximate $\boldsymbol{\psi}^*$ and demonstrate empirically that the resulting iterated auxiliary particle filter significantly outperforms the bootstrap particle filter in challenging settings. Applications include parameter estimation using a particle Markov chain Monte Carlo algorithm.

ARTICLE HISTORY
Received November 2015
Accepted July 2016

KEYWORDS
Hidden Markov models; Look-ahead methods; Particle Markov chain Monte Carlo; Sequential Monte Carlo; Smoothing; State-space models

SMC-ABC

Stat Comput (2012) 22:1009–1020
DOI 10.1007/s11222-011-9271-y

An adaptive sequential Monte Carlo method for approximate Bayesian computation

Pierre Del Moral · Arnaud Doucet · Ajay Jasra

Received: 30 May 2011 / Accepted: 7 July 2011 / Published online: 3 August 2011

Abstract Approximate Bayesian computation (ABC) is a popular approach to address inference problems where the likelihood function is intractable, or expensive to calculate. To improve over Markov chain Monte Carlo (MCMC) implementations of ABC, the use of sequential Monte Carlo (SMC) methods has recently been suggested. Most effective

1 Introduction

1.1 Background

Assume we are given a Bayesian model where $\pi(\theta)$ denotes the prior density of the parameter $\theta \in \Theta$ and $f(y|\theta)$

Group task

In groups of 2-4

1. Identify at least one Feynman-Kac model in each paper
   (a) What is the model state-space?
   (b) What are the potential functions?
   (c) What are the Markov kernels?
2. Choose one Feynman-Kac model to
   (a) Write an pseudo-code to sample from the mutation kernels, M_t
   (b) Describe the purpose of the potential functions, G_t
   (c) Simplify the chosen model to define a new (but equivalent) Feynman-Kac model

Waste-free SMC solution

Waste-free SMC

$$
\begin{aligned}
\mathsf{X}^{\mathrm{wf}} &= \mathcal{X}^P \\
G_t^{\mathrm{wf}}(x_t) &= \frac{1}{P}\sum_{i=1}^{P} G_t(x_{t,i}) \\
M_1^{\mathrm{wf}}(\mathrm{d}x_1) &= \prod_{i=1}^{P} \nu(\mathrm{d}x_{1,i}) \\
M_t^{\mathrm{wf}}(x_{t-1}, \mathrm{d}x_t) &= \left\{ \sum_{i=1}^{P} \frac{G_{t-1}(x_{t-1,i})}{\sum_{j=1}^{P} G_{t-1}(x_{t-1,j})} M_t(x_{t-1,i}, \mathrm{d}x_{t,1}) \right\} \prod_{i=2}^{P} M_t(x_{t,i-1}, \mathrm{d}x_{t,i})
\end{aligned}
$$

IAPF solution

IAPF model

$$
\begin{aligned}
\mathsf{X}^{\text{iapf}} &= \mathsf{X} \\
G_1^{\text{iapf}}(x_1) &= \frac{G_1(x_1)}{\psi_1(x_1)} M_2(\psi_2)(x_1) M_1(\psi_1) \\
G_t^{\text{iapf}}(x_t) &= \frac{G_t(x_t)}{\psi_t(x_t)} M_{t+1}(\psi_{t+1})(x_t) \\
M_1^{\text{iapf}}(\mathrm{d}x_1) &= \frac{M_1(\mathrm{d}x_1)\psi_1(x_1)}{M_1(\psi_1)} \\
M_t^{\text{iapf}}(x_{t-1}, \mathrm{d}x_t) &= \frac{M_t(x_{t-1}, \mathrm{d}x_t)\psi_t(x_t)}{M_t(\psi_t)(x_{t-1})}
\end{aligned}
$$

SMC-ABC solution

SMC-ABC model

$$
\begin{aligned}
\mathsf{X}^{\text{abc}} &= \mathsf{X} \\
G_1^{\text{abc}}(x_1) &= \frac{\pi_1(x_1)}{\eta_1(x_1)} \\
G_t^{\text{abc}}(x_{t-1}, x_t) &= \frac{\pi_t(x_t) L_{t-1}(x_t, x_{t-1})}{\pi_{t-1}(x_{t-1}) K_t(x_{t-1}, x_t)} \\
M_1^{\text{abc}}(\mathrm{d}x_1) &= \eta_1(x_1)\mathrm{d}x_1 \\
M_t^{\text{abc}}(x_{t-1}, \mathrm{d}x_t) &= K_t(x_{t-1}, x_t)\mathrm{d}x_t
\end{aligned}
$$

References i

Dau, Hai-Dang and Nicolas Chopin (2022). "Waste-free sequential Monte Carlo". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84.1, pp. 114–148.

Del Moral, Pierre (2004). *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. New York: Springer-Verlag.

Del Moral, Pierre, Arnaud Doucet, and Ajay Jasra (2012). "An adaptive sequential Monte Carlo method for approximate Bayesian computation". In: *Statistics and computing* 22, pp. 1009–1020.

Guarniero, Pieralberto, Adam M Johansen, and Anthony Lee (2017). "The iterated auxiliary particle filter". In: *Journal of the American Statistical Association* 112.520, pp. 1636–1647.