MATH97016/MATH97094 Computational Stochastic Processes

Urbain Vaes
u.vaes13@imperial.ac.uk
Imperial College London

Spring term 2019/2020

Welcome

- Lectures:
  - Mondays, 13:00–14:00, Huxley 145
  - Tuesdays, 12:00–13:00, Huxley 140
  - Fridays, 17:00–18:00, Huxley 213 (Clore)
- Office hours:
  - Tuesdays, 15:00–16:00, Huxley 737
  - or by appointment (send email)
- Course material will be made available on Blackboard.

The course

This is an introductory course on computational stochastic processes, aimed towards 4th year and MSc students in Applied Mathematics, Statistics and Theoretical Physics.

- Essential prerequisites:
  - Basic knowledge of probability: random variables, conditional expectation, law of large numbers, central limit theorem (CLT), etc.
  - Basic knowledge of stochastic processes in discrete and continuous time.
  - Familiarity with a (preferably free and open-source) programming language: Python, Julia, R, etc.

- Desirable prerequisites:
  - Ideally, you will have had a module on Stochastic Processes or similar.
  - Basic knowledge of stochastic differential equations (SDEs).
  - Basic knowledge of partial differential equations (PDEs).
  - Basic knowledge of numerical methods for ordinary differential equations (ODEs).

Lectures and Assesments

- Lectures:
  - Mostly on whiteboard / visualizer and sometimes slides.
  - Numerical examples in Python or Matlab.
  - I will hand out 2 or 3 problem sheets.
  - There will be a few problems classes: 1h every 2 weeks, approximately.

- Assessments:
  - Based on a coursework (25%) and a final exam (75%).
  - The coursework will be mostly computational and you will be free to use any "reasonable" language (Python, Julia, R, etc.).
  - The exam will deal mainly with theoretical aspects.

Syllabus I

1. Introduction
    1.1 Why study Computational Stochastic Processes?
    1.2 Motivation and Applications
2. Monte Carlo simulation
    2.1 Random number generation
    2.2 Monte Carlo simulation
    2.3 Variance reduction techniques
3. Simulation of continuous time Markov processes
    3.1 Brownian motion and related stochastic processes, and their simulation
    3.2 Gaussian stochastic processes
    3.3 Karhunen-Loeve expansion and simulation of Gaussian processes
4. Numerical solution of stochastic differential equations
    4.1 Stochastic integrals
    4.2 The Itô formula

Syllabus II

4.3 The Euler-Maruyama and Milstein schemes
4.4 Theoretical issues: convergence, consistency, stability

5. Markov Chain Monte Carlo (MCMC)
   5.1 Stationary processes, stationary distributions and ergodicity
   5.2 The Metropolis-Hastings and MALA algorithms
   5.3 Bias correction and variance reduction techniques

   Assessed coursework handed out

6. Inference for stochastic differential equations
   6.1 Inferring the diffusion coefficient
   6.2 The maximum likelihood estimator of drift coefficients in SDEs
   6.3 Nonparametric and Bayesian techniques

Bibliography

Books that cover (parts of) the contents of this course include the following:

- **pavliotis2011applied**
- **MR1214374**
- **MR3097957**
- **MR2410254**
- **MR2331321**

Why study Computational Stochastic Processes?

Introducing randomness into mathematical models or in computer algorithms is an extremely powerful and useful idea. It allows us to

- Model uncertainty in the parameters of various models in science, engineering and economics.
- Reflect structural uncertainty (ignorance about part of the model / underlying phenomena).
- Reduce complexity of existing models – many deterministic problems can be solved more efficiently using probabilistic techniques.
- Devise efficient algorithms, e.g. to escape local minima in global optimization.

Applications

Examples of applications that benefit from the addition of noise to their models include

- statistical physics
- cell biology (motion and growth of cells)
- epidemiology (spread of disease)
- medicine
- climate science (weather prediction)
- economy and finance

Stochastic Processes

We will focus on computational problems associated with stochastic processes, which describe dynamical systems whose evolution is of a probabilistic nature.

Definition (Stochastic process)

Let $T = \mathbb{N}$ or $T = [0, \infty)$ and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. A stochastic process is a function $X : T \times \Omega \to \mathbb{R}^d$ such that $X(t, \bullet)$ is a random variable for each $t \in T$.

- If $T = \mathbb{Z}$ or $T = \mathbb{N}$: discrete-time stochastic process.
- If $T = \mathbb{R}$ or $T = \mathbb{R}^+$: continuous-time stochastic process.

Markov processes

We will be mostly interested in Markov processes.

Definition (Markov chain)

A discrete time stochastic process is a Markov process or Markov chain if

$$\mathbb{P}[X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0] = \mathbb{P}[X_n = x_n | X_{n-1} = x_{n-1}].$$

- This concept can be generalised to continuous time processes naturally.
- The only continuous time Markov processes we will consider are diffusion processes, i.e. solutions of an SDE of the type

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t.$$

Motivating examples

1. Monte Carlo methods

Motivating example 1: Monte Carlo methods

When computing properties of statistical models, one frequently has to compute, over a high-dimensional state-space D, integrals of the form

$$I = \mathbb{E}_{X\sim\pi}[f(X)] := \int_D f(x)\,\pi(x)\,\mathrm{d}x.$$

Why do we need Monte Carlo methods?

Suppose that $D = [0,1]^d$ and that we want to employ a numerical quadrature:

1. Choose mesh of grid points within state space, with mesh-size h.
2. Evaluate $f(x_i)\,\pi(x_i)$ for every grid point x_i.
3. Use quadrature scheme to approximate integral.

Why do we need Monte Carlo method

With the standard quadrature-based approaches,

- the error is typically $O(h^k)$ for some $k \geq 2$ (e.g., $k = 2$ for the midpoint rule);
- the number of function evaluations scales as $M = O(h^{-d})$.

Therefore, the error scales as $O(M^{-k/d})$, i.e. the computational cost must grow exponentially as d increases in order to maintain the same error.

This is known as the curse of dimensionality.

Monte Carlo Methods

However, suppose we can generate i.i.d. samples $x_1, x_2, \ldots$ with distribution π. Then, using the Law of Large Numbers (LLN), we have

$$I_N := \frac{1}{N} \sum_{i=1}^{N} f(x_n) \xrightarrow{N \to \infty} \mathbb{E}[f(X)] \quad \text{almost surely.}$$

... so we can approximate I by I_N!

The rate of convergence is $O(N^{-1/2})$. This means that

$$\text{error} \sim \frac{1}{\sqrt{\text{computational cost}}}.$$

This is very slow, but does not suffer from the curse of dimensionality!

Monte Carlo Methods

However, suppose we can generate i.i.d. samples $x_1, x_2, \dots$ with distribution π. Then, using the Law of Large Numbers (LLN), we have

$$I_N := \frac{1}{N} \sum_{i=1}^{N} f(x_n) \xrightarrow{N \to \infty} \mathbb{E}[f(X)] \quad \text{almost surely.}$$

... so we can approximate I by I_N!

The rate of convergence is $O(N^{-1/2})$. This means that

$$\text{error} \sim \frac{1}{\sqrt{\text{computational cost}}}.$$

This is very slow, but does not suffer from the curse of dimensionality!

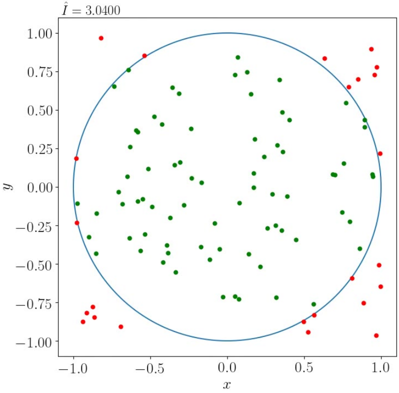
Some interesting questions

1. How can we generate i.i.d. samples from π? (Section 2)
2. How many samples do we need for a good approximation? (Section 2)
3. How do we measure performance of MC methods, and how to speed up convergence? (Section 2)
4. What to do if we cannot generate i.i.d. samples from π? (Section 5 – MCMC)

Monte Carlo simulation: an example

Suppose that we want to estimate

$$\pi = 4 \int_{-1}^{1} \int_{-1}^{1} I_{\{x^2+y^2 \leq 1\}} \underbrace{(1/4)}_{\pi(x,y)} \, \mathrm{d}x \, \mathrm{d}y.$$



Motivating examples

2. Statistical Physics

Motivation 2 – Statistical Physics

Consider a microscopic system of M particles.

- Position described by $\mathbf{q} = (q_1, \ldots, q_M)$.
- Interactions are characterised by a potential V.
- $\Rightarrow$ The evolution of the isolated system is governed by the Hamiltonian dynamics

$$\ddot{q}(t) = -\nabla V(q).$$

Example

A simple model for liquids is the case when

$$V(q) = \sum_{i,j=1}^{M} \Psi(|q_i - q_j|), \quad \text{where } \Psi(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right].$$

This is called the **Lennard–Jones potential**.

Statistical Physics: Langevin dynamics

The Langevin process is a model of a Hamiltonian system coupled to an infinite reservoir of energy via a thermostat.

The model arises through model reduction of a more complex molecular system (e.g. via Mori-Zwanzig formalism).

$$\ddot{q}_t = -\nabla V(q_t) - \gamma \dot{q}_t + \sqrt{2\gamma\beta^{-1}}\dot{W}_t$$

where

- W_t is a $3M-$dimensional Brownian motion.
- $\gamma > 0$ is the friction coefficient.
- β is the inverse temperature.

Statistical Physics

We can write it as a system of coupled first order equations:

$$\begin{cases} \mathrm{d}q_t = \boxed{p_t \, \mathrm{d}t}, \\ \mathrm{d}p_t = \boxed{-\boldsymbol{\nabla}_q V(q_t)\, \mathrm{d}t} \ \boxed{-\gamma p_t \, \mathrm{d}t + \sqrt{2\gamma\beta^{-1}}\, \mathrm{d}W_t}. \end{cases}$$

→ $\boxed{Hamiltonian}$ and $\boxed{fluctuation/dissipation}$ parts.

- in the $\gamma \to \infty$ limit, $q_\gamma(t) := q(\gamma t)$ converges to X_t, where

$$dX_t = -\nabla V(X_t)\ dt + \sqrt{2\beta^{-1}}\ dW_t$$

(overdamped Langevin equation).

- As $t \to \infty$, $\mathrm{Law}(X_t) \to \pi(dx)$ where

$$\pi(dx) = \frac{1}{Z} e^{-\beta V(x)}\ dx, \qquad Z = \int e^{-\beta V(x)}\ dx.$$

- Computing Z typically involves integrating a $10^3 - 10^6$ dimensional integral.

Langevin dynamics: simulation examples

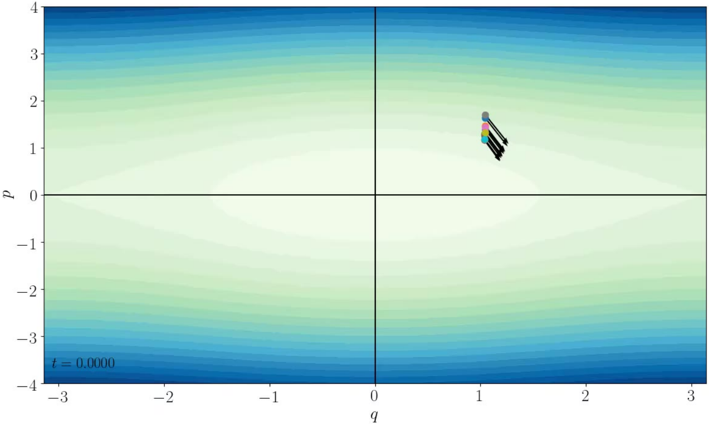


Figure: Langevin dynamics with friction $\gamma = 1$ (Inverse temperature $\beta = 1$)

Langevin dynamics: underdamped regime

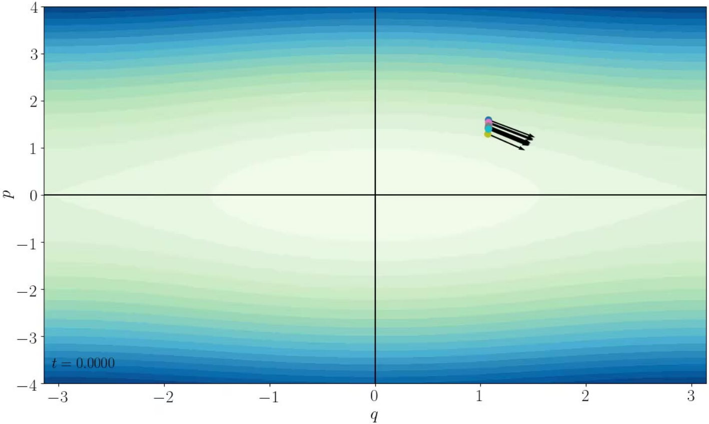


Figure: Langevin dynamics with friction $\gamma = .1$ (Inverse temperature $\beta = 1$)

Langevin dynamics: overdamped regime

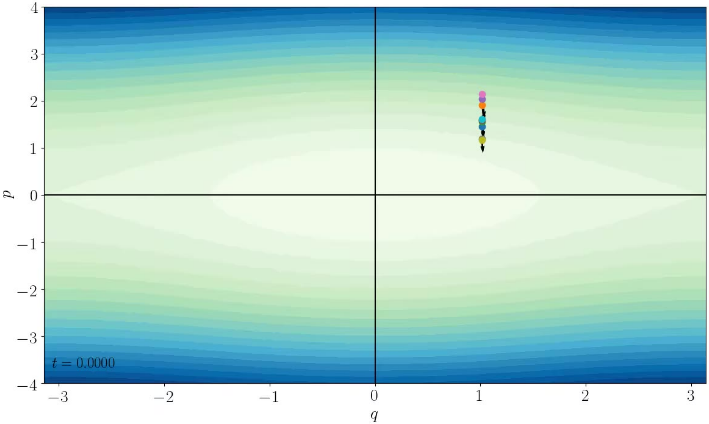


Figure: Langevin dynamics with friction $\gamma = 10$ (Inverse temperature $\beta = 1$)

Some interesting questions

In Section 4 we will answer the following questions:

- How do we simulate (q_t, p_t) and X_t? I.e., given a step size $\Delta t \ll 1$, can we derive numerical discretisations $X^{(n)}$ and $(q^{(n)}, p^{(n)})$ which are good approximations for $X_{n\Delta t}$ and $(q_{n\Delta t}, p_{n\Delta t})$?

- What conditions do we need to assume in order to ensure that the discretisation is stable?

- How can we generate samples from $\pi(\mathrm{d}x)$? Can we use $X^{(n)}$ to sample from π and/or compute expectations? If not, can we modify $X^{(n)}$ in order to do so?

Motivating examples

3. Inference

Motivation 3 – Inference

Suppose we have a statistical model for an experiment.

- The model depends on a parameter θ, which is unknown.
- By performing the experiment, we obtain observed data $\mathbf{y}$.

In Section 6 we will answer the question of how to identify the parameter θ that best explains $\mathbf{y}$.

Examples of applications include

- pharmacology (e.g., drug dosage)
- imaging
- ...

Inference – Maximum Likelihood Estimator

From the model, we can assume that, for parameters θ, the data $\mathbf{y}$ is distributed according to

$$\ell(\mathbf{y}|\theta).$$

This is called the likelihood function and is assumed to be completely known. It is often viewed as a function of the parameters θ given the data samples.

The goal is to find the value(s) of θ that is most compatible with the observed data $\mathbf{y}$. The maximum likelihood estimator (MLE) is the solution of

$$\hat{\theta} = \mathrm{argmax}_{\theta \in \Theta}\, \ell(\mathbf{y}|\theta).$$

Inference – Bayesian approach

Another approach is based on Bayes' rule:

- View θ, $\mathbf{y}$ as a coupled pair of random variables with density $\pi_0(\theta)\,\ell(\mathbf{y}|\theta)$.
  - $\pi_0(\theta)$ is the θ-marginal distribution.
  - $\ell(\mathbf{y}|\theta)$ is the conditional density of $\mathbf{y}$ given θ.
- Bayes' rule gives $\mathbb{P}(\theta|\mathbf{y}) = \frac{\ell(\mathbf{y}|\theta)\pi_0(\theta)}{\mathbb{P}(\mathbf{y})}$.
- $\pi_0(\theta)$ is the prior. It encodes prior beliefs about θ.
- This methodology gives a lot more information about the estimator.

Problem: Have to deal with the denominator

$$\mathbb{P}(\mathbf{y}) = \int_\Theta \ell(\mathbf{y}|\theta)\ \pi_0(\theta)\ d\theta.$$

Bayesian approach

How to deal with the denominator?

- Option 1: Clever choice of prior
- Option 2: Use the mode of the posterior distribution as best guess. This is known as Maximum a Posteriori estimator (MAP)
- Option 3: Use MCMC to generate approximate samples of posterior distribution.

Any questions?

Next week we will discuss Section 2:

- Random number generation
- Monte Carlo simulation
- Variance reduction techniques