Piece-wise Deterministic Markov Processes (PDMP)

Pablo Ramsés Alonso Martín

July 2022

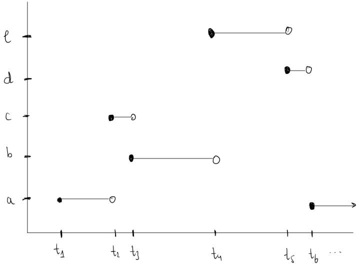
Index

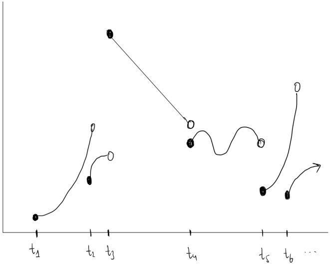
Some intuition...

- *"PDMP are continuous-time processes that evolve deterministically between a countable set of random event times".*

Some intuition...

- *"PDMP are continuous-time processes that evolve deterministically between a countable set of random event times"*.





Definition

Definition

A **Piecewise-Deterministic Markov Process** *is a continuous-time stochastic process whose dynamics involve random events with deterministic dynamics between events and random transition at events* $\{Z_t : t \geq 0\}$*. These dynamics are defined through the specification of three quantities:*

1. *The* **deterministic dynamics***:*

$$\frac{dZ_t^{(i)}}{dt} = \phi_i(Z_t)$$

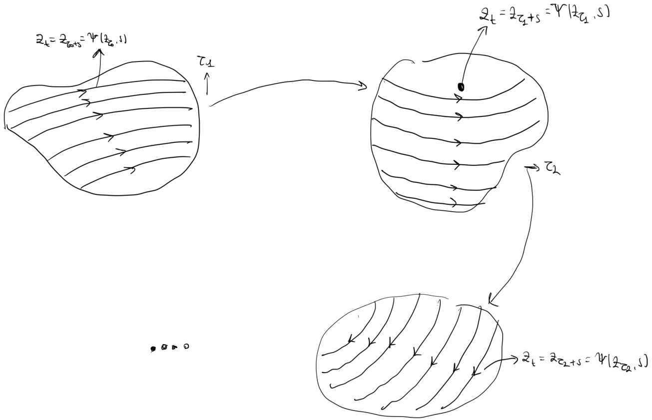
2. *The* **event rate***: events occur singularly at a rate* $\lambda(z_t)$ *that depends on the current state.*

3. **Transition kernel***: at any event time* τ*:*

$$z_\tau \sim q(\cdot|z_{\tau^-})$$

for some probability distribution.

Definition



Index

Why PDMPs

1. **Continuity**: well suited for Big Data, allows to target the posterior exactly even when subsampling.

Why PDMPs

1. **Continuity**: well suited for Big Data, allows to target the posterior exactly even when subsampling.
2. **Non-reversibility**: speeds up convergence to invariant distribution.

Why PDMPs

1. **Continuity**: well suited for Big Data, allows to target the posterior exactly even when subsampling.
2. **Non-reversibility**: speeds up convergence to invariant distribution.
3. **Designability**: generic schemes exist to fairly easily design desirable PDMPs.

Index

Generator

Definition

Generator *of a continuous-time stochastic process is an operator on functions with existing limit on the state-space:*

$$\mathcal{A}f(z) = \lim_{\delta \to 0} \frac{\mathbb{E}[f(Z_{t+\delta})|Z_t] - f(z)}{\delta}$$

Proposition

$$\frac{d\mathbb{E}(f(Z_t))}{dt} = \mathbb{E}(\mathcal{A}(f(Z_t))$$

Theorem (Davies 1984)

For a Piece-wise Deterministic Process:

$$\mathcal{A}f(x) = \phi(z) \cdot \nabla f(z) + \lambda(z) \cdot \int q(z^{'}|z) \cdot [f(z^{'}) - f(z)]dz^{'}$$

Adjoint

Definition

The **adjoint operator** *of the generator may be defined as the operator* $\mathcal{A}^*$ *such that*

$$\int g(z)\mathcal{A}f(z)dz = \int f(z)\mathcal{A}^*g(z)dz$$

Adjoint

Definition

The **adjoint operator** *of the generator may be defined as the operator $\mathcal{A}^*$ such that*

$$\int g(z)\mathcal{A}f(z)dz = \int f(z)\mathcal{A}^* g(z)dz$$

Proposition (Fokker-Plank Equation)

Let $p_t(z)$ the PDF of Z_t, then

$$\frac{\partial p_t(z)}{\partial t} = \mathcal{A}^* p_t(z)$$

Proposition

The adjoint operator of the generator of a PDMP can be written as:

$$\mathcal{A}^* g(z) = -\sum_{i=1}^{d} \frac{\partial(\phi_i(z) \cdot g(z))}{\partial z^i} + \int g(z^{'})\lambda(z^{'})q(z|z^{'})dz^{'} - g(z)\lambda(z)$$

Invariance

Using the Fokker-Plank equation, a probability distribution $\pi(z)$ is the invariant distribution of a PDMP if an only if

$$\mathcal{A}^* \pi(z) = 0$$

Putting all together:

Corollary

$\pi(z)$ is the invariant distribution of a PDMP if and only if:

$$-\sum_{i=1}^{d} \frac{\partial(\phi_i(z) \cdot \pi(z))}{\partial z^i} + \int \pi(z^{'})\lambda(z^{'})q(z|z^{'})dz^{'} - \pi(z)\lambda(z) = 0 \qquad (1)$$

Index

Data Augmentation

Most common approach is to consider $Z_t = (X_t, V_t)$ and choose the dynamics so that our distribution of interest $\pi(x)$ is the marginal distribution of X in the invariant distribution.

1. Choose some dynamics:

$$\frac{dx^i}{dt} = v_t^i \quad ; \quad \frac{dv_t^i}{dt} = 0 \tag{2}$$

2. Compute the rates and the kernel so that (1) is satisfied.

Choosing rates and kernel

- Under regular assumptions, (1) can be re-written as:

$$p(v) \cdot \lambda(x, v) - \int \lambda(x, v^{'}) \cdot q(v|x, v^{'}) \cdot p(v^{'}) dv^{'} = -p(v) \cdot v \cdot \nabla_x log(\pi(x)) \tag{3}$$

- Integrating both sides with respect to v yields:

$$\nabla_x log(\pi(x)) \cdot \mathbb{E}(V) = 0 \quad \forall x \Rightarrow \quad \mathbb{E}(V) = 0$$

- A **flip operator** F_x is therefore defined, satisfying $F_x(F_x(v)) = v$ and defining the transition kernel as a Dirac delta mass centred at $v^{'} = F_x(v)$

Choosing rates and kernel

- Including this latter condition, (3) becomes:

$$\lambda(x, v) - \lambda(x, v^{'}) = -v \cdot \nabla_x log(\pi(x)) \quad (4)$$

- The smallest rates compatible with (4) can be shown to be

$$\lambda(x, v) = max\{0, -v \cdot \nabla_x log(\pi(x))\}$$

and are known as **canonical rates**.

Example

The **Boomerang Sampler** *is defined by*

$$F_x(v) = v - 2 \cdot \frac{v \cdot \nabla_x log(\pi(x))}{\nabla_x log(\pi(x)) \cdot \nabla_x log(\pi(x))} \cdot \nabla_x log(\pi(x))$$

The **Zig-Zag sampler** *flips instead one component of the velocity at a time.*

Index

Simulating from a PDMP

Using the defining quantities:

1. Given Z_t, simulate the next event time τ

Simulating from a PDMP

Using the defining quantities:

1. Given Z_t, simulate the next event time τ
2. Calculate the state immediately before the event time $z_{\tau^-} = \psi(z_t, \tau - t)$

Simulating from a PDMP

Using the defining quantities:

1. Given Z_t, simulate the next event time τ
2. Calculate the state immediately before the event time $z_{\tau^-} = \psi(z_t, \tau - t)$
3. Draw the new value immediately after the event: $z_\tau \sim q(\cdot|z_{\tau^-})$

The non-homogeneous Poisson Process

Note that the rates:

$$\lambda(z_{t+s}) = \lambda(\psi(z_t, s)) = \tilde{\lambda}_{z_t}(s)$$

and thus can be analytically defined by a function of time starting at each event time. They change at each time t considered (which, recall it is considered over a continuous domain).

1. Event times can be simulated as arrival times of a Poisson Process with rates $\tilde{\lambda}_{z_t}(s)$.

The non-homogeneous Poisson Process

Note that the rates:

$$\lambda(z_{t+s}) = \lambda(\psi(z_t, s)) = \tilde{\lambda}_{z_t}(s)$$

and thus can be analytically defined by a function of time starting at each event time. They change at each time t considered (which, recall it is considered over a continuous domain).

1. Event times can be simulated as arrival times of a Poisson Process with rates $\tilde{\lambda}_{z_t}(s)$.
2. It is unclear how we can do that (and complicated) in general. Such Poisson Process is Non-Homogeneous and rates change continuously.

Index

Difficulties

1. *"Heisenberg Uncertainty Principle"* is not possible to simultaneously observe the current state and whether or not an event has occurred.

Difficulties

1. *"Heisenberg Uncertainty Principle"* is not possible to simultaneously observe the current state and whether or not an event has occurred.
2. Recall the canonical rates derived $\lambda(x, v) = max\{0, -v \cdot \nabla_x log(\pi(x))\}$. In the Bayesian Big Data setting, when using a subsample to estimate the gradient, λ becomes a random variable. We face here the challenge of simulating from a **Doubly-Stochastic** or **Cox** process.

Main methods to simulate event times

- If $\Lambda(t) = \int_0^t \lambda(u)du$ can be computed in a closed form, the following result can be used.

Theorem (Cinlar)

$T_1, ..., T_n$ are arrival times of a Poisson Process with intensity function $\lambda(t)$ if and only if $\Lambda(T_1), ..., \Lambda(T_n)$ are arrivals of a Poisson Process with rate 1.

For $n = 1, ...$

1. Compute $\Lambda(t) = \int_0^t \tilde{\lambda}_{z_{\tau_{n-1}}}(u)du$.
2. Simulate $T \sim Exp(1)$
3. Find τ_n such that $\Lambda(\tau_n) = T$

Then $\tau_1, ..., \tau_n$ are event times.

Main methods to simulate event times

- If $\tilde{\lambda}_{z_t}(s)$ cannot be integrated but instead it can be upper bounded along the domain: $\tilde{\lambda}_{z_t}(s) < \lambda^+$ then another result regarding the thinning property of Poisson Processes may be used:

Theorem (Lewis and Shedler 1979)

If t_0 is an arrival time of a Poisson Process with rate λ^+ then, it is also an arrival time of a coupled Poisson Process of rate $\tilde{\lambda}_{z_t}(s)$ with probability $\frac{\tilde{\lambda}_{z_t}(t_0)}{\lambda^+}$

Note that the tighter the bound the more efficient the sampling will be.

Main methods to simulate event times

Most common approach: combination of both.

1. Choose a simple function (commonly linear or piece-wisely linear) $\lambda^+(t)$ that upper bounds the rates.

Main methods to simulate event times

Most common approach: combination of both.

1. Choose a simple function (commonly linear or piece-wisely linear) $\lambda^+(t)$ that upper bounds the rates.
2. Use Cinlar's Theorem to simulate arrivals from the upper bound non-homogeneous process.

Main methods to simulate event times

Most common approach: combination of both.

1. Choose a simple function (commonly linear or piece-wisely linear) $\lambda^+(t)$ that upper bounds the rates.
2. Use Cinlar's Theorem to simulate arrivals from the upper bound non-homogeneous process.
3. Use the Thinning Theorem to simulate event times from out PDMP.

More shophisticated methods

1. For the **doubly-stochastic** process that arises in Bayesian inference for big data: use some available statistical model to estimate the rates. Note that in such cases the rates:

$$\lambda(z) = max0, -v \cdot \nabla_x \left[log(f(x)) + \sum_{i=1}^{N} log(p(y_i|x)) \right]$$

when using subsampling become a random quantity:

$$\hat{\lambda}(z) = max0, -v \cdot \nabla_x \left[log(f(x)) + \frac{N}{n} \sum_{i=1}^{n} log(p(y_{r_i}|x)) \right]$$

Example: Regression

Example (Pacman et al. 2014)

- *Model the rates using* **Linear Regression** *on previous steps:*

$$\hat{\lambda}_i = \beta_1 t_i + \beta_0 + \epsilon_{t_i}$$

where t_i *represent the previous observed event times.*

- *Then compute a confidence band* $[\tilde{\lambda}_L, \tilde{\lambda}_U]$ *for a given probability and use* $\tilde{\lambda}_U$ *as an upper bound to apply the combination of the first two methods.*
- *However this comes at a* **cost***: it is not an almost sure upper band and introduces bias (recall unbiasedness was one of the reasons underlying the whole construction).*

Index

Main Samplers

1. **Bouncy Particle Sampler**

Main Samplers

1. **Bouncy Particle Sampler**
2. **Zig-Zag Sampler**

Main Samplers

1. **Bouncy Particle Sampler**
2. **Zig-Zag Sampler**
3. **Boomerang Sampler**

Main Samplers

1. **Bouncy Particle Sampler**
2. **Zig-Zag Sampler**
3. **Boomerang Sampler**
4. Illustration of the samplers