Mathematics and numerics for data assimilation and state estimation – Lecture 18

Mathematics for Uncertainty Quantification
RWTH Aachen University

Summer semester 2020

Overview

Summary lecture 17

- Introduced sequential importance sampling (SIS) and sequential importance resampling (SIR) particle filters, for dynamics generated by the classic kernel density.

- Proved convergence of the bootstrap particle filter.

- Plan for today: quick look on particle filtering with more general dynamics and introduction to stochastic processes.

Overview

Other sampling dynamics

In the SIS and SIR algorithms we have considered, given $\{(w_j^{(i)}, v_j^{(i)})\}$, the dynamics simulation for the next step reads

- "Simulate $\hat{v}_{j+1}^{(i)} = F(v_j^{(i)}, \xi_j^{(i)})$ with iid $\xi_j^{(i)}$"

  This could also have been written
- "Draw independent $\hat{v}_{j+1}^{(i)} \sim \pi_{V_{j+1}|V_j}(\cdot|v_j^{(i)})$ for $i = 1, \ldots, M$".

- For SIS, the particles $\hat{v}_j^{(i)}$ have precisely the same distribution as the true dynamics V_j for every $j \geq 0$, this ignore completely the information from observations and may lead to $n_{eff,j} \ll M$.

- To avoid degeneracy, one can sample from other "dynamics"/kernel density than $\pi_{V_{j+1}|V_j}(\cdot|v_j^{(i)})$ that takes $y_{1:j+1}$ into account.

- Generic notation for kernel density: $\rho(v_{j+1}|v_j, y_{1:j+1})$, it can for instance be

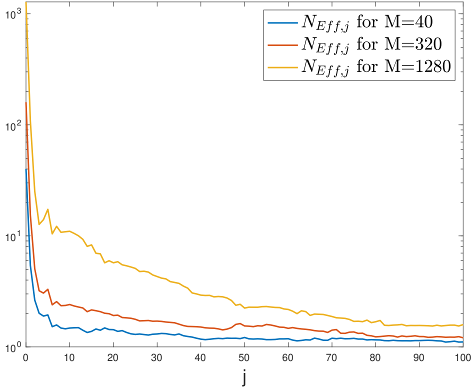
$$\rho(v_{j+1}|v_j, y_{1:j+1}) = \pi_{V_{j+1}|V_j,Y_{1:j+1}}(\cdot|v_j^{(i)}, y_{1:j+1})$$

Effective number of particles for SIS

...applied to linear-Gaussian problem

$$V_{j+1} = \begin{bmatrix} 1 & 0.1 \\ 0 & 1 \end{bmatrix} V_j + \xi_j, \qquad V_0 \sim N\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1/4 & 0 \\ 0 & 1/4 \end{bmatrix} \right)$$

where $\xi_j \overset{iid}{\sim} N(0, \Sigma)$ with $\Sigma = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.1 \end{bmatrix}$ (cf. Ubung 8.4).



Change of dynamics/kernel density

Recall that for the Bayes filter

$$\begin{aligned}\pi_{j+1}(v_{j+1}) &\propto \pi(y_{j+1}|v_{j+1})\pi(v_{j+1}|y_{1:j}) \\ &= \int_{\mathbb{R}^d} \underbrace{\pi(y_{j+1}|v_{j+1})}_{\text{“weight”}} \underbrace{\pi(v_{j+1}|v_j)}_{\text{“kernel density”}} \pi_j(v_j)dv_j,\end{aligned}$$

and for the particle filters this is approximated by

$$\pi_{j+1}^M = \mathscr{A}_{j+1}\mathcal{S}^M\mathscr{P}\pi_j^M = \sum_{i=1}^M w_{j+1}^{(i)}\delta_{\hat{v}_{j+1}^{(i)}}$$

with $\hat{v}_{j+1}^{(i)} \sim \int \pi_{V_{j+1}|V_j}(\cdot|v_j)\pi_j^M(v_j)dv_j$ and $w_{j+1}^{(i)} \propto \pi_{Y_{j+1}|V_{j+1}}(y_{j+1}|\hat{v}_{j+1}^{(i)})$

We replace the kernel density by $\rho(v_{j+1}|v_j, y_{1:j+1})$ as follows

$$\pi_{j+1}(v_{j+1}) \propto \int \pi(y_{j+1}|v_{j+1})\pi(v_{j+1}|v_j)\pi_j(v_j)dv_j$$

$$= \int \underbrace{\frac{\pi(y_{j+1}|v_{j+1})\pi(v_{j+1}|v_j)}{\rho(v_{j+1}|v_j, y_{1:j+1})}}_{\text{“weight”}} \underbrace{\rho(v_{j+1}|v_j, y_{1:j+1})}_{\text{“dynamics”}} \pi_j(v_j)dv_j$$

Constraint for the kernel density: Given $y_{1:j+1}$, it must hold for any $v_j, v_{j+1} \in \mathbb{R}^d$ such that

$$\pi(y_{j+1}|v_{j+1})\pi(v_{j+1}|v_j) > 0, \quad \text{also} \quad \rho(v_{j+1}|v_j, y_{1:j+1}) > 0.$$

Essential idea for the modified particle filter:

$$\pi_{j+1}^M = \sum_{i=1}^M w_{j+1}^{(i)} \delta_{\hat{v}_{j+1}^{(i)}}, \qquad \text{with } \hat{v}_{j+1}^{(i)} \sim \int \rho(\cdot|v_j, y_{1:j+1})\pi_j^M(v_j)dv_j$$

$$\text{and} \qquad w_{j+1}^{(i)} \propto \frac{\pi_{Y_{j+1}|V_{j+1}}(y_{j+1}|\hat{v}_{j+1}^{(i)})\pi_{V_{j+1}|V_j}(\hat{v}_{j+1}^{(i)}|v_j^{(i)})}{\rho(\hat{v}_{j+1}^{(i)}|v_j^{(i)}, y_{1:j+1})}$$

More general sequential importance resampling algorithm

- **Input:** Initial distribution π_0 (which we also write π_0^M), obs sequence $y_1, y_2, \ldots$, and M.
- **Particle generation:** For $j = 0, 1, \ldots,$,

  1. **Resampling** Draw $v_j^{(i)} \overset{iid}{\sim} \pi_j^M$ for $i = 1, \ldots, M$.
  2. Draw independent $\hat{v}_{j+1}^{(i)} \sim \rho(\cdot | v_j^{(i)}, y_{1:j+1})$ for $i = 1, \ldots, M$.
  3. Set

$$\bar{w}_{j+1}^{(i)} = \frac{\pi_{Y_{j+1}|V_{j+1}}(y_{j+1}|\hat{v}_{j+1}^{(i)})\pi_{V_{j+1}|V_j}(\hat{v}_{j+1}^{(i)}|v_j^{(i)})}{\rho(\hat{v}_{j+1}^{(i)}|v_j^{(i)}, y_{1:j+1})}$$

  4. and $w_{j+1}^{(i)} = \bar{w}_{j+1}^{(i)} / \sum_{k=1}^M \bar{w}_{j+1}^{(k)}$.
  5. Set $\pi_{j+1}^M = \sum_{i=1}^M w_{j+1}^{(i)} \delta_{\hat{v}_{j+1}^{(i)}}$.

- **Output:** π_j^M approximating the distribution of $V_j | Y_{1:j} = y_{1:j}$.

Modified Sequential importance sampling algorithm

- **Input:** Initial distribution π_0, obs sequence $y_1, y_2, \ldots$, and M.
- **Initialization:** Draw $\hat{v}_j^{(i)} \overset{iid}{\sim} \pi_0$ and set $w_0^{(i)} = 1/M$ for $i = 1, \ldots, M$. (Hat notation here is formally "wrong" but practical.)
- **Particle and weight dynamics:** For $j = 0, 1, \ldots,$,

1. Draw independent $\hat{v}_{j+1}^{(i)} \sim \rho(\cdot|\hat{v}_j^{(i)}, y_{1:j+1})$ for $i = 1, \ldots, M$.
2. Set

$$\bar{w}_{j+1}^{(i)} = \frac{\pi_{Y_{j+1}|V_{j+1}}(y_{j+1}|\hat{v}_{j+1}^{(i)})\pi_{V_{j+1}|V_j}(\hat{v}_{j+1}^{(i)}|\hat{v}_j^{(i)})}{\rho(\hat{v}_{j+1}^{(i)}|\hat{v}_j^{(i)}, y_{1:j+1})}$$

3. and $w_{j+1}^{(i)} = \bar{w}_{j+1}^{(i)} / \sum_{k=1}^M \bar{w}_{j+1}^{(k)}$.
4. Set $\pi_{j+1}^M = \sum_{i=1}^M w_{j+1}^{(i)} \delta_{\hat{v}_{j+1}^{(i)}}$.

- **Output:** π_j^M approximating the distribution of $V_j | Y_{1:j} = y_{1:j}$.

Sampling from a different kernel density

- Sampling from the kernel density

$$\pi_{V_{j+1}|V_j,Y_{j+1}}(\cdot|v_j, y_{j+1}) \tag{1}$$

in SIS gives you the so called optimal particle filter. Meaning

$$\mathrm{Var}^{\pi_{V_{j+1}|V_j,Y_{j+1}}(\cdot|\hat{v}_j^{(i)},y_{j+1})}[\bar{w}_{j+1}^{(i)}] = \inf_{\rho(\cdot|\hat{v}_j^{(i)},y_{1:j+1})} \mathrm{Var}^{\rho(\cdot|\hat{v}_j^{(i)},y_{1:j+1})}[\bar{w}_{j+1}^{(i)}]$$

- In other words, of all possible kernel densities $\rho(\cdot|v_j, y_{1:j+1})$, sampling from (1) leads to the minimum variance in $\bar{w}_j^{(i)}$.

- See [SST 12.3] for a setting where it actually is possible to sample from $\pi_{V_{j+1}|V_j,Y_{j+1}}(\cdot|v_j, y_{j+1})$.

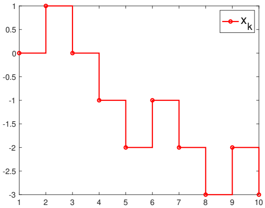
Overview

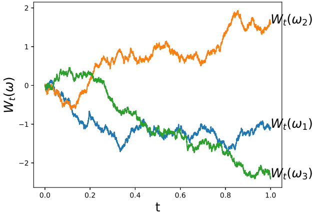
Stochastic processes

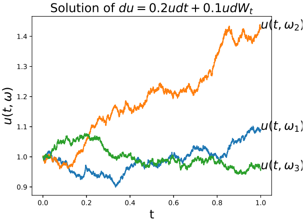
Definition 1

A stochastic process on $\mathbb{R}^d$ is family of rv $\{X_t\}_{t\in\mathbb{T}}$ all taking values in $\mathbb{R}^d$, all defined on $(\Omega, \mathcal{F}, \mathbb{P})$, for some parameter set $\mathbb{T}$, typically $\mathbb{T} = \mathbb{N}$, or $[0, T]$ or $[0, \infty)$.
For any fixed $t \in \mathbb{T}$, the mapping $X_t : \Omega \to \mathbb{R}^d$ is an rv.
For any fixed $\omega \in \Omega$, the mapping $X.(\omega)\mathbb{T} \to \mathbb{R}^d$ is a d-dimensional path.

Examples on $\mathbb{R}$: (Simple walk, Wiener process, Geometric Brownian motion)







Construction of a stochastic process

Defining probability measures on spaces of stochastic processes is subtle: Consider the fair coin-tossing process

$$X = (X_1, X_2, \dots) \in \{0, 1\}^{\mathbb{N}}$$

where $X_n(T) = 0$ and $X_n(H) = 1$.

- We assume $X_m \perp X_n$ when $m \neq n$.
- Implication: all events

  $$\{X = k\} = \{X_1 = k_1, X_2 = k_2, \dots\}$$

  for $k \in \{0, 1\}^{\mathbb{N}}$ are equally likely.
- Problem: there are infinitely many of equally likely events, so

  $$\mathbb{P}(X = k) = \mathbb{P}(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n, \dots) = \lim_{n\to\infty} \left(\frac{1}{2}\right)^n = 0,$$

  which means it is difficult to construct the probability measure bottom up using the probability of individual paths (the "atoms" of the probability space) to derive the probability of unions of paths.

Probability on $\mathcal{F}$ generated by cylinder sets

We define the probability space by

$$\Omega = \{0,1\}^{\mathbb{N}}$$

and for any **fintie subsequence** $\{i_k\}_{k=1}^{m} \subset \mathbb{N}$, the finite projection of a paths have positive measure:

$$\mathbb{P}(\{\omega_{i_k}\}_{k=1}^{m}) = \mathbb{P}(X_{i_1} = \omega_{i_1}, \ldots X_{i_m} = \omega_{i_m}) = 2^{-m}, \quad \omega_{i_k} \in \{0,1\}.$$

An idea is therefore to let $\mathcal{F}$ be the defined as the smallest σ-algebra containing all events of the form

$$\{X_{i_1} = \omega_{i_1}, X_{i_2} = \omega_{i_2}, \ldots, X_{i_m} = \omega_{i_m}\} \quad \textbf{cylinder sets} \tag{2}$$

for any $1 \leq i_1 < i_2 < \ldots < i_m$, $\omega_{i_k} \in \{0,1\}$ and $m \in \mathbb{N}$: $\mathcal{F}$ is called the product $\sigma - algebra$.

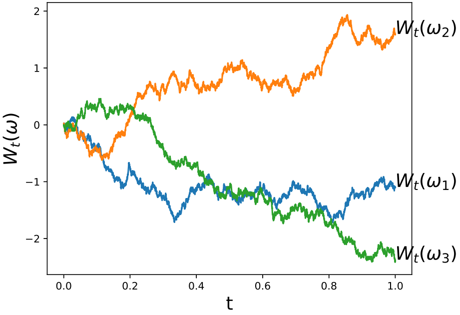
Question: We know value of $\mathbb{P}$ on every cylinder set, but is it possible to extend $\mathbb{P}$ so that we can apply it to any $C \in \mathcal{F}$?

Continuous state-space stochastic process - and measure

More generally, for $\{X_t\}_{t \in \mathbb{R}}$ on $(\mathbb{R}^d, \mathcal{B}^d)$ cylinder sets generating $\mathcal{F}$ can be defined by

$$\{X_{i_1} \in F_{i_1}, X_{i_2} \in F_{i_2}, \ldots, X_{i_m} \in F_{i_m}\}$$

for any $1 \leq i_1 < i_2 < \ldots < i_m$, $F_{i_k} \in \mathcal{B}^d$ and $m \in \mathbb{N}$.



$$C_1 = \{W_{0.5} \in (0,2), W_1 \in (1,2)\} \text{ and } C_2 = \{W_{0.1} \in (-1,1), W_1 < -2\}$$

we have $\omega_2 \in C_1$ and $\omega_1, \omega_3 \notin C_1$. And $\omega_3 \in C_2$ and $\omega_1, \omega_2 \notin C_2$.

Question: We know value of $\mathbb{P}$ on every cylinder set, but is it possible to extend $\mathbb{P}$ so that we can apply it to any $C \in \mathcal{F}$? Yes!:

Theorem 2 (Kolmogorov's extension theorem [ELV-E 5.2])

Let $\{\mu_{t_1,\ldots,t_m}\}$ be a family of finite-dimensional distributions satisfying for any $t_1,\ldots,t_m \in \mathbb{T}$, $F_1,\ldots,F_k \in \mathcal{B}^d$ and $m \in \mathbb{N}$ that

(i) *For any permutation σ of $\{1,2,\ldots,m\}$,*

$$\mu_{t_{\sigma(1)},t_{\sigma(2)},\ldots,t_{\sigma(m)}}(F_1 \times F_2 \times \ldots \times F_m) = \mu_{t_1,t_2,\ldots,t_m}(F_{\sigma(1)} \times F_{\sigma(2)} \times \ldots \times F_{\sigma(m)}).$$

(ii) *For any $k \in \mathbb{N}$,*

$$\mu_{t_1,\ldots,t_m}(F_1 \times \ldots \times F_m) = \mu_{t_1,\ldots,t_m,t_{m+1},\ldots,t_{m+k}}(F_1 \times \ldots \times F_m \times \mathbb{R}^d \times \ldots \times \mathbb{R}^d).$$

Then there exists a space $(\Omega, \mathcal{F}, \mathbb{P})$, with $\mathcal{F}$ being the product σ-algebra, and a proces $\{X_t\}_{t \in \mathbb{T}}$ s.t.

$$\mu_{t_1,t_2,\ldots,t_m}(F_1 \times F_2 \times \ldots \times F_m) = \mathbb{P}(X_{t_1} \in F_1, X_{t_2} \in F_2, \ldots, X_{t_m} \in F_m)$$

for any $t_1,\ldots,t_m \in \mathbb{T}$, $F_i \in \mathcal{B}^d$ and and $m \in \mathbb{N}$.

Filtrations

To simplify the presentation, assume $\mathbb{T} = [0, \infty)$, but the below easily extends to other $\mathbb{T}$-sets

Definition 3

Given $(\Omega, \mathcal{F}, \mathbb{P})$ a filtration is a non-increasing family of σ-algebras $\{\mathcal{F}\}_{t \geq 0}$ such that $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ for any $0 \leq s < t$.

A stochastic process $\{X_t\}_{t \geq 0}$, say on $(\mathbb{R}^d, \mathcal{B}^d)$, is called $\mathcal{F}_t$-adapted if for any $t \geq 0$

$$X_t^{-1}(B) \in \mathcal{F}_t \quad \forall B \in \mathcal{B}^d.$$

Given $\{X_t\}$ the filtration generated by the process

$$\mathcal{F}_t^X = \sigma(\{X_s\}_{s \leq t})$$

provides all the information of the path up to time t and is the smallest filtration on which X_t is adapted.

Example filtration

Consider again the fair coin-tossing process

$$X = (X_1, X_2, \dots) \in \{0,1\}^{\mathbb{N}}$$

with $\Omega = \{0,1\}^{\mathbb{N}}$, $\mathcal{F}$ generated by the cylinder sets (2) and $\mathbb{P}$ existing by Kolmogorov's extension thm.

We associate the filtration $\{\mathcal{F}_n^X\}_{n\in\mathbb{N}}$ with

$$\mathcal{F}_1^X = \mathcal{F}(X_1) = \mathcal{F}(\{0\},\{1\}) = \Big\{\emptyset, \Omega, \underbrace{\{0\}}_{\{X_1=0\}}, \underbrace{\{1\}}_{\{X_1=1\}}\Big\}$$

$$\begin{aligned}
\mathcal{F}_2^X &= \mathcal{F}(X_1, X_2) = \mathcal{F}(\{00\},\{01\},\{10\},\{11\}) \\
&= \Big\{\color{red}{\emptyset, \Omega, \{0\cdot\}, \{1\cdot\}}, \{\cdot 0\}, \{\cdot 1\}, \\
&\{00\},\{01\},\{10\},\{11\},\{0\cdot\}\cup\{10\},\{0\cdot\}\cup\{11\}, \\
&\{1\cdot\}\cup\{00\},\{1\cdot\}\cup\{01\},\{00\}\cup\{11\},\{01\}\cup\{10\}\Big\}
\end{aligned}$$

Note: $\{0\cdot\} := \{X_1 = 0, X_2 \in \{0,1\}\} = \{0\}$ etc. And no sets of the form $\{01\color{red}{0}\} = \{01\} \cap \color{red}{\{X_3 = 0\}}$ are contained in $\mathcal{F}_2$, as that is in the future.

Filtration for a continuous-time stochastic process

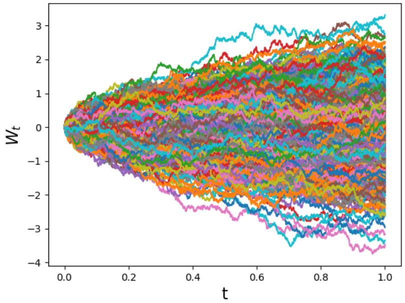
For a Wiener process $\{W_t\}_{t \in [0,T]}$ on $(\mathbb{R}, \mathcal{B})$,

$$\mathcal{F}_t^W = \sigma(\{W_s\}_{s \leq t})$$

where $\mathcal{F}_t^W$ is generated from all cylinder-sets

$$C = \{W_{t_1} \in F_{t_1}, W_{t_2} \in F_{t_2}, \ldots, W_{t_m} \in F_{t_m}\}$$

for any $0 \leq t_1 < t_2 < \ldots < t_m \leq t$, (this upper bound **is** the constraint on information), $F_{t_k} \in \mathcal{B}$ and $m \in \mathbb{N}$.
May be associated to a probability measure on the path-space $\Omega = C[0, T]$.



Overview

Markov processes

Definition 4 (Markov process)

Given $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\{\mathcal{F}_t\}_{t\geq 0}$, a stoch process X_t on $(\mathbb{R}^d, \mathcal{B}^d)$ is called a Markov process wrt $\mathcal{F}_t$ if

(i) X_t is $\mathcal{F}_t$-adapted

(ii) for any $t \geq s$ and $B \in \mathcal{B}^d$,

$$\underbrace{\mathbb{P}\left(X_t \in B | \mathcal{F}_s\right)}_{:=\mathbb{E}\left[\mathbb{1}_{X_t \in B} | \mathcal{F}_s\right]} = \underbrace{\mathbb{P}\left(X_t \in B | X_s\right)}_{:=\mathbb{E}\left[\mathbb{1}_{X_t \in B} | \sigma(X_s)\right]} \qquad \textbf{memorylessness}.$$

Connections to Markov chains:

$$\mathbb{P}\left(X_n \in B | X_{1:m} = x_{1:m}\right) = \mathbb{P}\left(X_n \in B | X_m = x_m\right)$$

Conditioning is either on (possibly more than the) full path history over $[0, s]$ or just on state at time s:

$$\mathcal{F}_s \supset \sigma(\{X_r\}_{r\leq s}) \supset \sigma(X_s).$$

The transition function

The transition function $p(B, t|x, s)$ of a Markov process X_t on $(\mathbb{R}^d, \mathcal{B}^d)$ is defined by

$$p(B, t|x, s) := \mathbb{P}(X_t \in B | X_s = x) \quad (:= \mathbb{P}(X_t \in B | X_s \in dx))$$

for $s \leq t$ and $B \in \mathcal{B}^d$.
The mapping $p : \mathcal{B}^d \times [0, \infty) | \mathbb{R}^d \times [0, \infty) \to [0, \infty)$ has the following properties:

(i) For any $t \geq s$ and $x \in \mathbb{R}^d$, $p(\cdot, t|x, s) : \mathcal{B}^d \to [0, 1]$ is a probability measure.

(ii) For any $t \geq s$ and $B \in \mathcal{B}^d$, $p(B, t|\cdot, s) : \mathbb{R}^d \to [0, 1]$ is a measureable function on $\mathbb{R}^d$.

(iii) p satisfies the "Chapman-Kolmogorov" analog:

$$p(B, t|x, s) = \int_{\mathbb{R}^d} p(B, t|y, u) p(dy, u|x, s), \quad s \leq u \leq t.$$

Transition kernel densities

We will restrict ourselves to stationary Markov processes:

Definition 5

$\{X_t\}_{t\geq 0}$ is stationary if for any $t_1, t_2, \ldots, t_m \geq 0$ and $m \in \mathbb{N}$ the joint distribution is translation invariant:

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_m}) \overset{D}{=} (X_{t_1+s}, X_{t_2+s}, \ldots, X_{t_m+s})$$

for any $s \in \mathbb{R}$ so that X_{t_k+s} are defined.

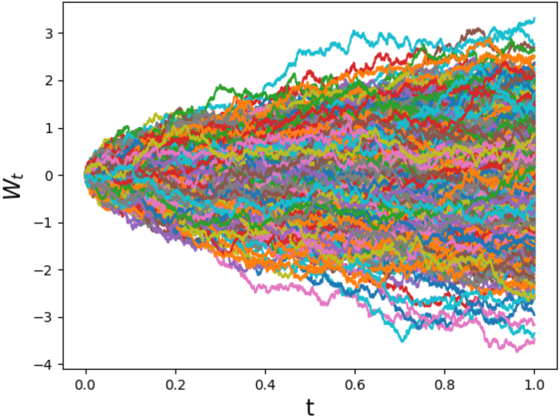
For stationary $\{X_t\}_{t\geq 0}$, the transition function then simplifies into the transition kernel:

$$p(t, x, B) = p(B, t+s|x, s) \qquad \text{for any } s, t \geq 0$$

and we refer to $p(t, x, y)$ as the kernel density if

$$p(t, x, B) = \int_B p(t, x, y) dy.$$

Example



Overview

Next steps

Our plan is to study filtering problems with stochastic differential equation (SDE) dynamics:

$$V_t = \Psi_t(V_0) = V_0 + \int_0^t b(V_s)ds + \int_0^t \sigma(V_s)dW_s$$

Here W_t is a Wiener process and V_t is a Markov process.
For this purpose, we need to describe:

- Wiener processes
- SDE (well-posedness, Itô integrals, etc)
- numerical methods for solving sde
- the probability density function of time-homogeneous SDE $p(t, y, x) = "\mathbb{P}(V_t = x | V_0 = y)"$.

Wiener processes

Definition 6

Wiener process $\{W_t\}_{t\geq 0}$ on $\mathbb{R}$ is a stochastic process described by

(i) $W_0 \overset{a.s.}{=} 0$, and or for any $0 \leq t_0 < t_1 < \ldots < t_m$, the increments $W_{t_0}, W_{t_1} - W_{t_0}, \ldots W_{t_m} - W_{t_{m-1}}$ are independent.

(ii) For any $s, t \geq 0$, $W_{t+s} - W_s \sim N(0, t)$,

(iii) With probability 1, the path $W_\cdot(\omega)$ is continuous.

Independent increments implies that $W_t - W_s \perp \{W_r\}_{r\leq s}$ and thus that W_t is Markovian wrt $\mathcal{F}_t^W$:

$$\mathbb{E}\left[\mathbb{1}_B(W_t)|\mathcal{F}_s^W\right] = \mathbb{E}\left[\mathbb{1}_B((W_t - W_s) + W_s)|\mathcal{F}_s^W\right]$$
$$= \mathbb{E}\left[\mathbb{1}_B(W_t)|\sigma(W_s)\right].$$

And by (ii), the kernel density of $W_{t+s} - W_s$ equals

$$p(t, x, y) = \frac{\exp\left(\frac{-(y-x)^2}{2t}\right)}{\sqrt{2\pi t}} \quad (\text{AKA } = \pi_{W_t|W_0}(y|x))$$

Construction of the Wiener process

By Kolmogorov's extension theorem: Using independent increments all finite-dimensional joint pdfs are computable:

$$
\begin{aligned}
&\pi_{W_{t_0} W_{t_1} \ldots W_{t_m}}(x_0, x_1, \ldots, x_m) \\
&= \pi_{W_{t_m}|W_{t_{m-1}}}(x_m|x_{m-1}) \pi_{W_{t_{m-1}}|W_{t_{m-2}}}(x_{m-1}|x_{m-2}) \ldots \pi_{W_{t_1}|W_{t_0}}(x_1|x_0) \pi_{W_{t_0}}(x_0) \\
&= \frac{\exp\left(-\sum_{k=1}^m \frac{(x_k - x_{k-1})^2}{2(t_k - t_{k-1})} - \frac{(x_0)^2}{2t_0} \right)}{\prod_{k=1}^m \sqrt{2\pi(t_k - t_{k-1})} \sqrt{2\pi t_0}}
\end{aligned}
$$

Consequently, we can compute all probabilities of the form

$$
\mu_{t_0, t_1, \ldots, t_m}(B_0 \times B_1 \times \ldots \times B_m) = \mathbb{P}(W_{t_0} \in B_0, W_{t_1} \in B_1, \ldots, W_{t_m} \in B_m)
$$

and the extension theorem ensures the existence of a prob space $(\Omega, \mathcal{F}, \mathbb{P})$ on which $\{W_t\}_{t \geq 0}$ is defined.

Gaussian processes

- Any stoch process $\{X\}_{t\geq 0}$ for which every finite-dimensional joint distribution $(X_{t_0}, \ldots, X_{t_m})$ is multivariate Gaussian, is called a Gaussian process.

- The Wiener process is a Gaussian process.

- For fixed y, the kernel density solves the Heat equation

$$\frac{\partial}{\partial t} p_t(\cdot, \cdot, y) = \frac{1}{2} p_{xx}(\cdot, \cdot, y) \quad (t, x) \in [0, \infty) \times \mathbb{R},$$

with initial condition $p(0, x, y) = \delta_y(x)$.

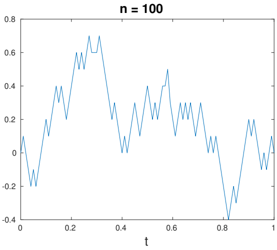
Second construction: limit of simple random walks

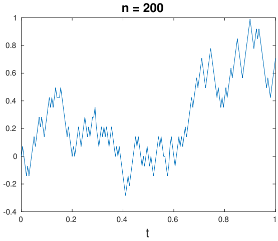
Theorem 7 (Random walk case of Donsker's theorem)

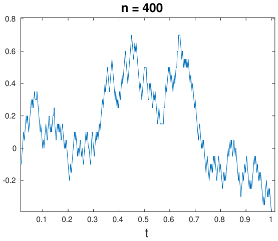
Let $\{X_n\}$ be a simple symmetric RW on $\mathbb{Z}$ with $X_0 = 0$ and consider

$$W^{(n)}(t) := \frac{X_{\lfloor nt \rfloor}}{\sqrt{n}} \quad t \in [0,1],$$

where $\lfloor x \rfloor := \max\{k \in \mathbb{Z} \mid k \leq x\}$. Then $\{W^{(n)}(t)\}_{t\in[0,1]}$ converges in distribution to a standard Brownian motion $\{W(t)\}_{t\in[0,1]}$.







Simulation of Wiener processes:

Suppose we want to simulate W_t **exactly** on a uniform mesh $t_k = k\Delta t$ covering $[0, T]$.

From property (ii) we know that

$$W_{t_{k+1}} = W_{t_k} + \underbrace{W_{t_{k+1}} - W_{t_k}}_{\sim N(0,\Delta t)}$$

Hence, one may simulate iteratively,

$$W_{t_{k+1}} = W_{t_k} + \sqrt{\Delta t}\xi_k$$

where ξ_k are iid standard normals.

We may approximate the full path e.g. by linear interpolation

$$W_s = LinInterp(s; \{(t_k, W_{t_k})\}), \qquad s \in [0, T].$$

or one may interpolate exactly by Brownian bridge refinement.

Regularity properties of the Wiener processes:

Over a compact interval $[0, T]$:

- α-Hölder continuity (almost one-half time differentiable):

$$\sup_{s,t\in[0,T]} \frac{|W(t) - W(s)|}{|t-s|^\alpha} \overset{a.s.}{<} \infty \iff 0 \le \alpha < 1/2$$

- Unbounded variation: Let $\Delta = \{t_k\}$ denote a mesh of $[0, T]$. Then

$$\sup_{\Delta} \sum_k |W_{t_{k+1}} - W_{t_k}| \overset{a.s.}{=} \infty$$

(Motivation for result: assume $t_k = k\Delta t$. Then

$$\sum_{k=0}^{T/\Delta t - 1} |W_{t_{k+1}} - W_{t_k}| = \sum_k \sqrt{\Delta t}|\xi_k|\text{''} \le \text{''}\, C\frac{T}{\Delta t}\sqrt{\Delta t} \xrightarrow{\Delta t \downarrow 0} \infty\ .)$$

Overview

Construction of stochastic integrals

Seeking to make sense of the SDE

$$V_t = V_0 + \int_0^t b(V_s)ds + \int_0^t \sigma(V_s)dW_s$$

we need to define the stochastic integral.
Riemann-Stieltjes approach: let $|\Delta|$ denote the largest timestep in a mesh over $[0, t]$ and

$$\int_0^t \sigma(V_s)dW_s = \lim_{|\Delta|\to 0} \sum_k \sigma(V_{t_k^*})(W_{t_{k+1}} - W_{t_k})$$

for some $t_k^* \in [t_k, t_{k+1}]$.

Problem: these integrals are well-defined provided $\sigma(V_t)$ is continuous (which is reasonable to assume) and W_t has bounded total variation – which almost surely is not the case for the Wiener process.

Implication: different choices of t_k^* may lead to different integral values (both pathwise and in expectation).

Example

Consider the integral $\int_0^t W_s dW_s$, and three different choices for integration point:

$$t_k^* = \begin{cases} \text{left: } t_k \quad \text{giving} \quad I^L = \sum_k W_{t_k}(W_{t_{k+1}} - W_{t_k}) \\ \text{right: } t_{k+1} \quad \text{giving} \quad I^R = \sum_k W_{t_{k+1}}(W_{t_{k+1}} - W_{t_k}) \\ \text{middle: } t_{k+1/2} \quad \text{giving} \quad I^M = \sum_k W_{t_{k+1/2}}(W_{t_{k+1}} - W_{t_k}) \end{cases}$$

And

$$\mathbb{E}\left[I^L \right] = \sum_k \mathbb{E}\left[W_{t_k}(W_{t_{k+1}} - W_{t_k}) \right] \overset{W_{t_k} \perp (W_{t_{k+1}} - W_{t_k})}{=} \sum_k \mathbb{E}\left[W_{t_k} \right] \mathbb{E}\left[W_{t_{k+1}} - W_{t_k} \right] = 0,$$

while

$$\begin{aligned} \mathbb{E}\left[I^R \right] &= \sum_k \mathbb{E}\left[W_{t_{k+1}}(W_{t_{k+1}} - W_{t_k}) \right] \\ &= \sum_k \mathbb{E}\left[\Big((W_{t_{k+1}} - W_{t_k}) + W_{t_k} \Big)(W_{t_{k+1}} - W_{t_k}) \right] \\ &= \sum_k \mathbb{E}\left[(W_{t_{k+1}} - W_{t_k})^2 \right] + I^L = \sum_k (t_{k+1} - t_k) = t \end{aligned}$$

and $\quad \mathbb{E}\left[I^M \right] = t/2.$

Itô integral

The Itô integral is defined by

$$\int_0^t \sigma(V_s)dW_s = \lim_{|\Delta|\to 0} \sum_k \sigma(V_{t_k})(W_{t_{k+1}} - W_{t_k})$$

i.e., the resolution limit integral using left integration points $t_k^* = t_k$.

More on the construction and properties of the Itô integral in the next lecture.

Summary and next lecture

- Degeneracy is an important issue for particle filters, particularly for high-dimensional problems. It is an ongoing research topic to understand this phenomenon and develop more robust particle filters, through e.g., adaptive resampling and alternative sampling dynamics.
- Have defined stochastic processes on filtered probability spaces $(\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\geq 0}, \mathbb{P})$, Markov processes, and particularly Wiener processes.
- Next time: Itô integrals, and theory and numerical integration of Itô SDE.