An introduction to diffusion processes and Ito's stochastic calculus

Cédric Archambeau

University College, London
Centre for Computational Statistics and Machine Learning
c.archambeau@cs.ucl.ac.uk

Motivation

$$dx = \alpha(t, x)\ dt$$

$$dX_t = \alpha(t, X_t)\ dt + \beta(t, X_t)\ dW_t$$

stochastic term

- Tackle continuous time continuous state dynamical system
- Can we gain something?
- What is the impact of the stochastic terms?

Today's aim

- (Better) understand why there is need for stochastic calculus:

$$\int_0^t W_s \, dW_s = ?$$

- Understand the fundamental difference with non-stochastic calculus?

$$\int_0^t w(s) \, dw(s) = \frac{1}{2} w^2(t) \quad \text{if } w(0) = 0.$$

$$\int_0^t W_s \, dW_s = \frac{1}{2} W_t^2 - \frac{1}{2} t, \quad \text{w.p. } 1, \ (W_0 = 0, \text{w.p. } 1).$$

Outline

- Basic concepts:
  - Probability theory
  - Stochastic processes

- Diffusion Processes
  - Markov process
  - Kolmogorov forward and backward equations

- Ito calculus
  - Ito stochastic integral
  - Ito formula (stochastic chain rule)

Running example: the Wiener Process!

Basic concepts on probability theory

- A collection $\mathcal{A}$ of subsets of Ω is a σ-algebra if $\mathcal{A}$ contains Ω and $\mathcal{A}$ is closed under the set of operations of complementation and countable unions.

- The sequence $\{\mathcal{A}_t\,,\ \mathcal{A}_t \subseteq \mathcal{A}$ with $t \geq 0\}$ is an increasing family of σ- algebras of $\mathcal{A}$ if $\mathcal{A}_s$ is a subset of $\mathcal{A}_t$ for any $s \leq t$.

- A measure μ on the measurable space $(\Omega, \mathcal{A})$ is a nonnegative valued set function on $\mathcal{A}$ such that $\mu(\emptyset) = 0$ and which is additive under the countable union of disjoint sets.

- The probability measure P is a measure which is normalized with respect to the measure on the certain event $P(\Omega)$.

- Let $(\Omega, \mathcal{A}, P)$ be a probability space. A random variable is an $\mathcal{A}$-measurable function $X : \Omega \to \Re$, that is the pre-image $X^{-1}(B)$ of any Borel (or Lebesgue) subset B in the Borel σ-algebra $\mathcal{B}$ (or $\mathcal{L}$) is a subset of $\mathcal{A}$.

Basic concepts on stochastic processes

Let $(\Omega, \mathcal{A}, P)$ be a common probability space and T the time set. A stochastic process $X = \{X_t,\ t \in T\}$ is a function $X : T \times \Omega \rightarrow \Re$ such that

$X(t, .\ ) : \Omega \rightarrow \Re$ is a random variable for each $t \in T$,

$X(\ .\ , \omega) : T \rightarrow \Re$ is a sample path for each $\omega \in \Omega$.

- A Gaussian process is a stochastic process for which any joint distribution is Gaussian.

- A stochastic process is strictly stationary if it is invariant under time displacement and it is wide-sense stationary if there exist a constant μ and a function c such that

$$\mu_t = \mu, \quad \sigma_t^2 = c(0) \quad \text{and} \quad C_{s,t} = c(t-s),$$

  for all $s, t \in T$.

- A stochastic process is a martingale if

$$E\{X_t | \mathcal{A}_s\} = X_s, \quad \text{w.p. } 1,$$

  for any $0 \leq s \leq t$.

Example: the Wiener process

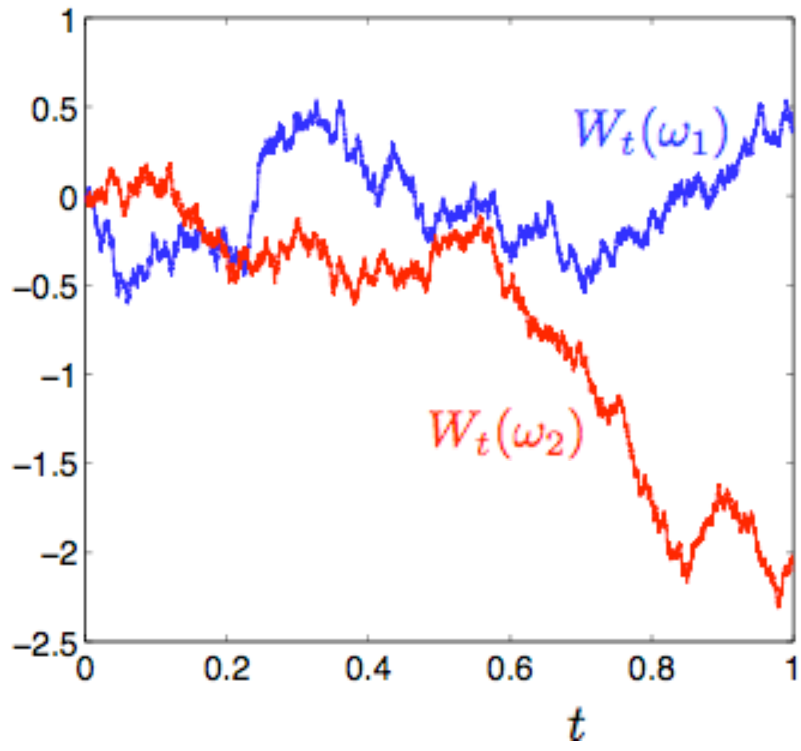
The standard Wiener process $W = \{W_t,\ t \geq 0\}$ is a continuous time continuous state stochastic process with independent Gaussian increments:

$$W_0 = 0 \quad \text{w.p. } 1,$$

$$E\{W_t\} = 0,$$

$$W_t - W_s \sim \mathcal{N}(0, t-s),$$

for all $0 \leq s \leq t$.



(Proposed by Wiener as mathematical description of Brownian motion.)

- The Wiener process is not wide-sense stationary since its (two-time) covariance is given by $C_{s,t} = \min\{s,t\}$:

For $s \leq t$, we have

$$\begin{aligned} C_{s,t} &= E\{(W_t - \mu_t)(W_s - \mu_s)\} \\ &= E\{W_t W_s\} \\ &= E\{(W_t - W_s + W_s)W_s\} \\ &= E\{W_t - W_s\}E\{W_s\} + E\{W_s^2\} \\ &= 0 \cdot 0 + s. \end{aligned}$$

- The Wiener process is a martingale:

$$\left.\begin{array}{l} E\{W_t - W_s | \mathcal{A}_s\} = 0 \\ E\{W_s | \mathcal{A}_s\} = W_s \end{array}\right\} \Rightarrow E\{W_t | \mathcal{A}_s\} = W_s, \quad \text{w.p. } 1.$$

Markov processes

The stochastic process $X = \{X_t\,,\ t \geq 0\}$ is a (continuous time continuous state) Markov process if it satisfies the Markov property:

$$P(X_t \in B | X_s = x) = P(X_t \in B | X_{r_1} = x_1, \ldots, X_{r_n} = x_n, X_s = x)$$

for all Borel subsets B of $\Re$ and time instants $0 \leq r_1 \leq \ldots \leq r_n \leq s \leq t$.

- The transition probability is a (probability) measure on the Borel σ-algebra $\mathcal{B}$ of the Borel subsets of $\Re$:
$$P(X_t \in B | X_s = x) = \int_B p(s, x; t, y)\, dy$$
- The Chapman-Kolmogorov equation follows from the Markov property:
$$p(s, x; t, y) = \int_{-\infty}^{\infty} p(s, x; \tau, z) p(\tau, z; t, y)\, dz \qquad \text{for } s \leq \tau \leq t.$$
- The Markov process X_t is homogeneous if all the transition densities depend only on the time difference.
- The Markov process X_t is ergodic if the time average on $[0, T]$ for $T \to \infty$ of any function $f(X_t)$ is equal to its space average with respect to (one of) its stationary probability densities.

Example: the Wiener process

- The standard Wiener process is a homogenous Markov process since its transition probability is given by

$$p(s, x; t, y) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left\{-\frac{(y-x)^2}{2(t-s)}\right\}.$$

- The transition density of the standard Wiener process satisfies the Chapman-Kolmogorov equation (convolution of two Gaussian densities).

Diffusion processes

The Markov process $X = \{X_t,\ t \geq 0\}$ is a diffusion process if the following limits exist:

$$\lim_{t\downarrow s} \frac{1}{t-s} \int_{|y-x|>\epsilon} p(s,x;t,y)\, dy = 0,$$

$$\lim_{t\downarrow s} \frac{1}{t-s} \int_{|y-x|<\epsilon} (y-x) p(s,x;t,y)\, dy = \alpha(s,x),$$

$$\lim_{t\downarrow s} \frac{1}{t-s} \int_{|y-x|<\epsilon} (y-x)^2 p(s,x;t,y)\, dy = \beta^2(s,x),$$

for all $\varepsilon > 0$, $s \geq 0$ and $x \in \Re$.

- Diffusion processes are *almost surely* continuous, but not necessarily differentiable.
- Parameter $\alpha(s,x)$is the drift at time s and position x.
- Parameter $\beta(s,x)$ is the diffusion coefficient at time s and position x.

Let $X = \{X_t, \ t \geq 0\}$ be a diffusion process.

The forward evolution of its transition density $p(s,x;t,y)$ is given by the Kolmogorov forward equation (or *Fokker-Planck equation*):

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial y}\{\alpha(t,y)p\} - \frac{1}{2}\frac{\partial^2}{\partial y^2}\{\beta^2(t,y)p\} = 0$$

for a fixed initial state (s,x).

The backward evolution is given by the Kolmogorov backward equation:

$$\frac{\partial p}{\partial s} + \alpha(s,x)\frac{\partial p}{\partial x} + \frac{1}{2}\beta^2(s,x)\frac{\partial^2 p}{\partial x^2} = 0$$

for a fixed final state (t,y).

Example: the Wiener process

- The standard Wiener process is a diffusion process with drift $\alpha(s,x) = 0$ and diffusion parameter $\beta(s,x) = 1$.

  For $W_s = x$ at a given time s, the transition density is given by $\mathcal{N}(y \,|\, x, t-s)$. Hence, we get

$$\alpha(x,s) = \lim_{t \downarrow s} \frac{E\{y-x|x\}}{t-s} = 0,$$
$$\beta^2(x,s) = \lim_{t \downarrow s} \frac{E\{(y-x)^2|x\}}{t-s} = \lim_{t \downarrow s} \frac{t-s}{t-s} = 1.$$

- Kolmogorov forward and backward equation for the standard Wiener process are given by

$$\frac{\partial p}{\partial t} - \frac{1}{2}\frac{\partial^2 p}{\partial y^2} = 0,$$
$$\frac{\partial p}{\partial s} + \frac{1}{2}\frac{\partial^2 p}{\partial x^2} = 0.$$

What makes the Wiener process so special?

- The sample paths of a Wiener process are *almost surely continuous* (see Kolmogorov criterion).

- However, they are almost surely nowhere differentiable.

  Consider the partition of a bounded time interval [*s*,*t*] into 2^n sub-intervals of length $(t\text{-}s)/2^n$. For each sample path $\omega \in \Omega$, it can be shown that

$$\lim_{n\to\infty} \sum_{k=0}^{2^n-1} \left(W_{\tau_{k+1}^{(n)}}(\omega) - W_{\tau_k^{(n)}}(\omega) \right)^2 = t - s, \quad \text{w.p. } 1.$$

  Hence, can write

$$t - s \leq \limsup_{n\to\infty} \max_k \left| W_{\tau_{k+1}^{(n)}}(\omega) - W_{\tau_k^{(n)}}(\omega) \right| \sum_{k=0}^{2^n-1} \left| W_{\tau_{k+1}^{(n)}}(\omega) - W_{\tau_k^{(n)}}(\omega) \right|.$$

  From the sample path continuity, we have

$$\max_k \left| W_{\tau_{k+1}^{(n)}}(\omega) - W_{\tau_k^{(n)}}(\omega) \right| \to 0, \quad \text{w.p. 1 when } n \to \infty,$$

  and thus

$$\sum_{k=0}^{2^n-1} \left| W_{\tau_{k+1}^{(n)}}(\omega) - W_{\tau_k^{(n)}}(\omega) \right| \to \infty, \quad \text{w.p. 1 as } n \to \infty.$$

  The sample paths do, almost surely, not have bounded variation on [*s*,*t*].

Why introducing stochastic calculus?

Consider the following stochastic differential:

$$dX_t = \alpha(t, X_t)\ dt + \beta(t, X_t)\xi_t\ dt$$

$$\xi_t \sim \mathcal{N}(0.1)$$

Or interpreted as an integral along a sample path:

$$X_t(\omega) = X_{t_0}(\omega) + \int_{t_0}^{t} \alpha(s, X_s(\omega))\ ds + \int_{t_0}^{t} \beta(s, X_s(\omega))\xi_s(\omega)\ ds$$

$$\approx dW_t$$

Problem: A Wiener process is almost surely nowhere differentiable!

$$\int_{t_0}^{t} \beta(s, X_s(\omega))dW_t(\omega) =?$$

Construction of the Ito integral

- Idea:

$$\int_{t_0}^{t} \beta \, dW_t(\omega) = \beta \{ W_t(\omega) - W_{t_0}(\omega) \}$$

- The integral of a random function *f* (mean square integrable) on the unit time interval is defined as

$$I[f](\omega) = \int_0^1 f(s, \omega) \, dW_s(\omega).$$

- Consider a partition of the unit time interval:

$$0 = t_1 \quad \dots \quad t_j \qquad\qquad t_{j+1} \quad \dots \quad t_n = 1$$

- Use properties of the standard Wiener process!

$$f(t,\omega) = f_j$$

1. The function f is a nonrandom step function:

$$I[f](\omega) = \sum_{j=1}^{n-1} f_j \{W_{t_{j+1}}(\omega) - W_{t_j}(\omega)\}, \quad \text{w.p. 1.}$$

- The mean of the stochastic integral:

$$E\{I[f]\} = 0$$

- The mean square fluctuation of the stochastic integral:

$$E\{I^2[f]\} = \sum_{j=1}^{n-1} f_j^2 (t_{j+1} - t_j)$$

$$f(t, \omega) = f_j(\omega)$$

2. *f* is a random step function:

$$I[f](\omega) = \sum_{j=1}^{n-1} f_j(\omega)\{W_{t_{j+1}}(\omega) - W_{t_j}(\omega)\}, \quad \text{w.p. 1.}$$

- The mean of the stochastic integral:

$$E\{I[f]\} = 0$$

- The mean square fluctuation of the stochastic integral:

$$E\{I^2[f]\} = \sum_{j=1}^{n-1} E\{f_j^2\}(t_{j+1} - t_j)$$

$$f(t,\omega)$$
$$\downarrow$$
$$f^{(n)}(t,\omega) = f(t_j^{(n)},\omega)$$

3. f is a general random function:

$$I[f](\omega) = \sum_{j=1}^{n-1} f(t_j^{(n)},\omega)\{W_{t_{j+1}}(\omega) - W_{t_j}(\omega)\}, \quad \text{w.p. 1.}$$

where $f^{(n)}$ is a sequence of random step functions converging to f.

- The mean of the stochastic integral:

$$E\{I[f]\} = 0$$

- The mean square fluctuation of the stochastic integral:

$$E\{I^2[f]\} = \sum_{j=1}^{n-1} E\{f^2(t_j^{(n)},\omega)\}(t_{j+1} - t_j)$$

Riemann sum!

Ito (stochastic) integral

$$I[f](\omega) = \underset{n\to\infty}{\text{m.s.}\lim} \sum_{j=1}^{n-1} f(t_j^{(n)}, \omega)\{W_{t_{j+1}}(\omega) - W_{t_j}(\omega)\}, \quad \text{w.p. } 1.$$

for a (mean square integrable) random function $f : T \times \Omega \to \Re$.

- The equality is interpreted in mean square sense!
- Unique solution for any sequence of random step functions converging to *f*.
- The time-dependent solution process is a martingale:

$$X_t(\omega) = \int_{t_0}^{t} f(s, \omega)\, dW_s(\omega)$$

- Linearity and additivity properties satisfied.
- Ito isometry:

$$E\{I^2[f]\} = \int_s^t E\{f^2(\tau, \cdot)\}\, d\tau$$

Ito formula (stochastic chain rule)

- Consider

$$Y_t = U(t, X_t) \qquad\qquad dX_t = f\, dW_t$$

- Taylor expansion:

$$\Delta Y_t = \underbrace{\left\{ \frac{\partial U}{\partial t} \Delta t + \frac{\partial U}{\partial x}\, \Delta x \right\}}_{\text{conventional calculus}} + \frac{1}{2} \left\{ \frac{\partial^2 U}{\partial t^2} \Delta t^2 + 2 \frac{\partial^2 U}{\partial t \partial x} \Delta t \Delta x + \frac{\partial^2 U}{\partial x^2} \underbrace{\Delta x^2}_{= \mathcal{O}(dt)} \right\} + \ldots .$$

- Ito formula:

$$Y_t - Y_s = \int_s^t \left\{ \frac{\partial U}{\partial t} + \frac{1}{2} f_u^2 \frac{\partial^2 U}{\partial x^2} \right\} du + \int_s^t \frac{\partial U}{\partial x}\, dX_u, \quad \text{w.p. } 1.$$

Concluding example:

- Consider:

$$U(x) = x^m \qquad\qquad X_t = W_t$$

- The Ito formula leads to

$$W_t^m - W_s^m = \int_s^t \frac{m(m-1)}{2} W_\tau^{(m-2)}\, d\tau + \int_s^t m W_\tau^{(m-1)}\, dW_\tau$$

- For $m = 2$:

$$W_t^2 = t + 2\int_0^t W_\tau\, dW_\tau \quad \Rightarrow \int_0^t W_\tau\, dW_\tau = \frac{1}{2} W_t^2 - \frac{1}{2} t$$

Stratonovich stochastic calculus

$$f(t, \omega)$$

$$\downarrow$$

$$f^{(n)}(t, \omega) = f(\tau_j^{(n)}, \omega)$$

- Consider different partition points:

$$\tau_j^{(n)} = (1 - \lambda)t_j^{(n)} + (1 - \lambda)t_{j+1}^{(n)}$$

- Mean square convergence with λ = 1/2:

$$\int_s^t f_t \circ dW_t$$

- No stochastic chain rule, but martingale property is lost.

Next reading groups…

- Stochastic Differential Equations!!!

- Who?
- When?
- Where?

- How?

References

- P. Kloeden and E. Platen: *Numerical Solutions to Stochastic Differential Equations*. Springer-Verlag, 1999 (3rd edition).

- B. Øksendael: *Stochastic Differential Equations*. Springer-Verlag, 2002 (6th edition).

- C.W. Gardiner: *Handbook of Stochastic Methods*. Springer-Verlag, 2004 (3rd edition).

- C. K. I. Williams, “A Tutorial Introduction to Stochastic Differential Equations: Continuous time Gaussian Markov Processes”, presented at *NIPS workshop on Dynamical Systems, Stochastic Processes and Bayesian Inference*, Dec. 2006.

- Lawrence E. Evans. An Introduction to Stochastic Differential Equations. Lecture notes (Department of Mathematics,UCBerkeley), available from http://math.berkeley.edu/evans/.