Class contents:

1. Probability background

2. Wiener process

3. Ito integral

Probability Background

A probability space is a triple $(\Omega, \mathcal{F}, P)$, where Ω is the set of outcomes, $\mathcal{F}$ is the set of events and $P : \mathcal{F} \to [0, 1]$ is a function that assigns probabilities to events satisfying certain rules.

Definition 1 (Measurable Space) *If Ω is a given non empty set, then a σ-*algebra *$\mathcal{F}$ on Ω is a collection $\mathcal{F}$ of subsets of Ω that satisfy:*

(1) $\Omega \in \mathcal{F}$;

(2) $F \in \mathcal{F} \Rightarrow F^c \in \mathcal{F}$, *where $F^c = \Omega - F$ is the complement set of F in Ω; and*

(3) $F_1, F_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{+\infty} F_i \in \mathcal{F}$.

Definition 2 (Probability Measure) *A probability measure on* $(\Omega, \mathcal{F})$ *is a set function* $P : \mathcal{F} \to [0, 1]$ *such that:*

(1) $P(\emptyset) = 0$, $P(\Omega) = 1$; *and*

(2) *If* $A_1, A_2, \ldots \in \mathcal{F}$ *are mutually disjoint sets then*

$$P\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} P(A_i).$$

Question 1: Give an example of a probability space and distinguish clearly the events $F \in \mathcal{F}$ from the outcomes $\omega \in \Omega$.

Question 2: Give an example of two different σ-algebras, $\mathcal{G} \subset \mathcal{F}$ for the same set of outcomes Ω. Can you give an intuitive interpretation of the relation $\mathcal{G} \subset \mathcal{F}$?

Question 3: Is the intersection of σ-algebras still a σ-algebra?

Question 4: What about the union of σ-algebras?

Definition 3 (generated σ-algebra) *Given a family of sets, $\{A_n\}$, there exists a unique σ-algebra, $\sigma(\{A_n\})$, s.t.*

1. $\{A_n\} \subset \sigma(\{A_n\})$,

2. *if $\mathcal{F}$ is a $\sigma - algebra$,*

$$\{A_n\} \subset \mathcal{F} \Rightarrow \sigma(\{A_n\}) \subset \mathcal{F}$$

Definition 4 *A random variable X, in the probability space $(\Omega, \mathcal{F}, P)$, is a function*

$$X : \Omega \to \mathbb{R}^d,$$

such that the inverse image

$$X^{-1}(A) \equiv \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F},$$

for all open subsets A of $\mathbb{R}^d$.

Equivalently, we may say that X is an $\mathcal{F}$-measurable function and write $X \in \mathcal{F}$.

Example: Consider a finite family of disjoint sets,

$$\{A_n\}_{n=1}^{N}$$

and let $\Omega \equiv \cup_{1 \leq n \leq N} A_n$, $\mathcal{F} \equiv \sigma(\{A_n\})$. What condition has to satisfy

$$X : \Omega \to \mathbb{R}$$

in order to be a random variable in $(\Omega, \mathcal{F})$?

Definition 5 (Independence) *Two sets $A, B \in \mathcal{F}$ are said to be independent if*

$$P(A \cap B) = P(A)P(B).$$

Two independent random variables X, Y in $\mathbb{R}^d$ are independent if for all open sets $A, B \subseteq \mathbb{R}^d$ we have that the events

$$X^{-1}(A) \text{ and } Y^{-1}(B) \text{ are independent }. \quad (9)$$

Definition 6 (Expected value) *Let* $X : \Omega \to \mathbb{R}$ *be a random variable and suppose that the density function*

$$p'(x) = \frac{P(X \in dx)}{dx}$$

is integrable. The expected value of X *is then defined by the integral*

$$E[X] = \int_{-\infty}^{\infty} x p'(x) dx, \qquad (10)$$

which also can be written

$$E[X] = \int_{-\infty}^{\infty} x dp(x). \qquad (11)$$

The last integral makes sense also in general when the density function is a measure, e.g. by successive approximation with random variables possessing integrable densities. A point mass, i.e. a Dirac delta measure, is an example of a measure.

Definition 7 (Stochastic Process) *A stochastic process* $X : [0,T] \times \Omega \rightarrow \mathbb{R}^d$ *in the probability space* $(\Omega, \mathcal{F}, P)$ *is a function such that* $X(t,\cdot)$ *is a random variable in* $(\Omega, \mathcal{F}, P)$ *for all* $t \in (0,T)$. *We will often write* $X(t) \equiv X(t,\cdot)$.

The t variable will usually be associated with the notion of time.

Definition 8 (Wiener process) *The one dimensional* Wiener process $W : [0, \infty) \times \Omega \to \mathbb{R}$, *also known as the Brownian motion, has the following properties:*

(1) *with probability 1, the mapping* $t \mapsto W(t)$ *is continuous and* $W(0) = 0$*;*

(2) *if* $0 = t_0 < t_1 < \ldots < t_N = T$, *then the Wiener increments*

$$W(t_N) - W(t_{N-1}), \ldots, W(t_1) - W(t_0)$$

are independent*; and*

(3) *for all* $t > s$ *the increment* $W(t) - W(s)$ *has the* normal *distribution, with* $E[W(t) - W(s)] = 0$ *and* $E[(W(t) - W(s))^2] = t - s$, *i.e. for real intervals* Γ *we have*

$$P(W(t) - W(s) \in \Gamma) = \frac{1}{\sqrt{2\pi(t-s)}} \int_\Gamma e^{\frac{-y^2}{2(t-s)}} dy.$$

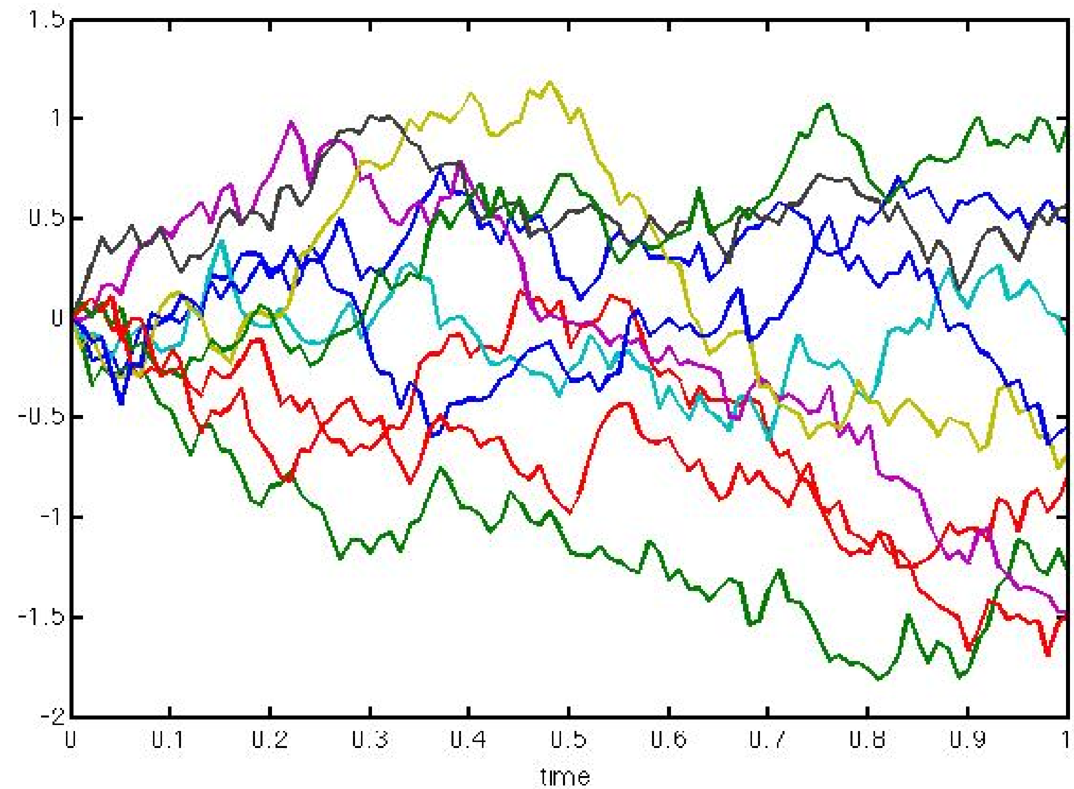
Problem: Given the times points

$$0 = t_0 < t_1 < \ldots < t_N = T.$$

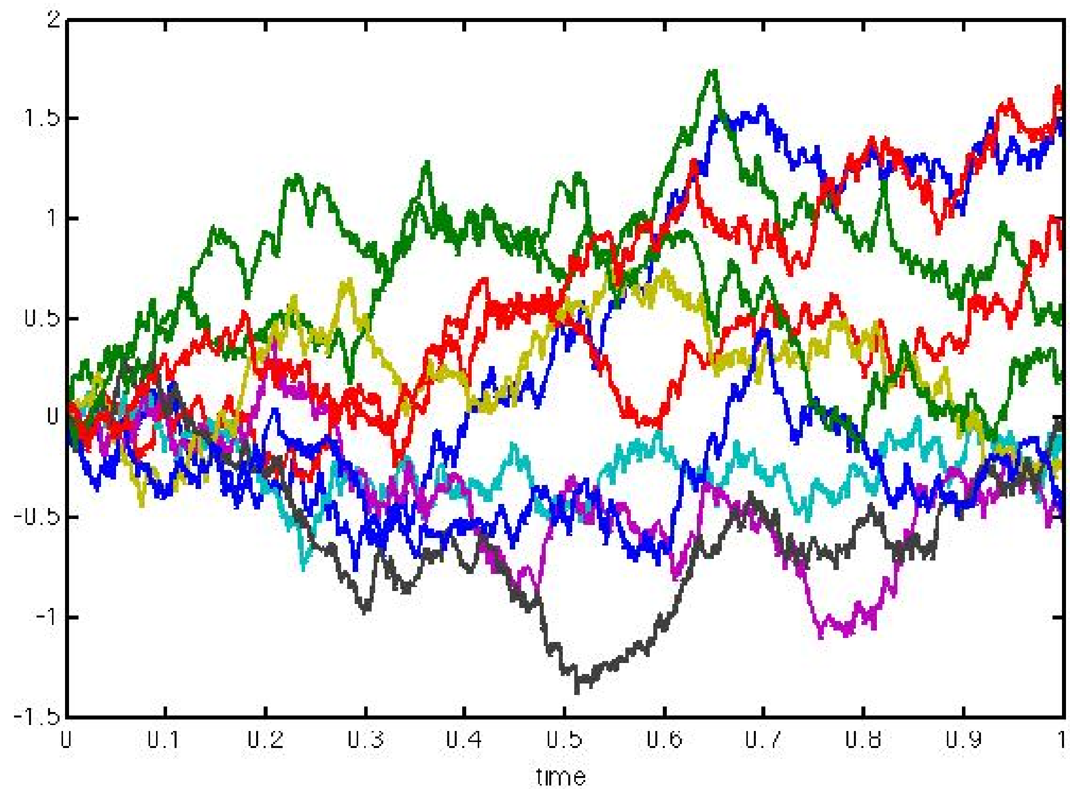
How do we sample realizations of $W(t_n)$, for $n = 0, \ldots, N$?

Sampling W at discrete times

```
>> N = 100;
>> dt = 1/N;
>> M = 10;
>> dW = sqrt(dt)*randn(N,M);
>> W = cumsum(dW);
>> W = [zeros(1,M);W];
>> t = linspace(0,1,N+1);
>> plot(t,W)
```



Sampling W on $[0,1]$, $N = 10^2$ uniform time steps, $M = 10$ realizations



Sampling W on $[0,1]$, $N = 10^3$ uniform time steps, $M = 10$ realizations

Question: Which changes are needed to extend the previous code to variable step size, i.e. to sample W at given times, not necessarily evenly distributed,

$$0 = t_0 < \ldots < t_N?$$

Observe that the previous code is vectorized (no for loops) so it runs faster in MATLAB. What about its memory use?

Approximation and Definition of Stochastic Integrals

Questions on the definition of a stochastic integral

Remark 1 Problem: *How to define the stochastic integral* $\int_0^T W(t)dW(t)$, *where* $W(t)$ *is the Wiener process.*

Can we use the same approach as with Riemann integrals, taking sums

$$\sum_{n=0}^{N-1} W(\xi_n)(W(t_{n+1}) - W(t_n))$$

with $\xi_n \in [t_n, t_{n+1}]$*?*

As a first step, use the forward Euler *discretization*

$$\sum_{n=0}^{N-1} W(t_n)\underbrace{(W(t_{n+1}) - W(t_n))}_{=\Delta W_n}.$$

Taking expected values we obtain (why?)

$$\begin{aligned} E[\sum_{n=0}^{N-1} W(t_n)\Delta W_n] &= \sum_{n=0}^{N-1} E[W(t_n)\Delta W_n] \\ &= \sum_{n=0}^{N-1} E[W(t_n)]\underbrace{E[\Delta W_n]}_{=0} \\ &= 0. \end{aligned}$$

Now let us use instead the backward Euler *discretization*

$$\sum_{n=0}^{N-1} W(t_{n+1})\Delta W_n.$$

Taking expected values yields a different re-

sult:

$$
\begin{aligned}
&\sum_{n=0}^{N-1} E[W(t_{n+1})\Delta W_n] \\
&= \sum_{n=0}^{N-1} E[W(t_n)\Delta W_n] + E[(\Delta W_n)^2] \\
&= \sum_{n=0}^{N-1} \Delta t \\
&= T \neq 0.
\end{aligned}
$$

Moreover, if we use the trapezoidal *method the result is*

$$\sum_{n=0}^{N-1} E\left[\frac{W(t_{n+1}) + W(t_n)}{2}\Delta W_n\right]$$
$$= \sum_{n=0}^{N-1} E[W(t_n)\Delta W_n] + E[(\Delta W_n)^2/2]$$
$$= \sum_{n=0}^{N-1} \frac{\Delta t}{2} = T/2 \neq 0.$$

□

Conclusion: we need more information to define $\int_0^T W(s)dW(s)$ than to define a deterministic integral!

In fact, limits of the forward Euler define the so called *Itô integral*, while the trapezoidal method yields the so called *Stratonovich integral*.

Strong and weak convergence

Depending on the application, we focus either on

- *strong convergence*, where approximation of the outcomes of $X(T)$ is relevant,

- or *weak convergence*, where only the distribution (law) of $X(T)$ needs to be approximated.

Definition. The sequence of random variables $\{Y_n\}_{n\in\mathbb{N}}$ converges *strongly* to the random variable Y if

$$\|Y - Y_n\|_{L^2_P(\Omega)} \equiv \sqrt{E[(Y - Y_n)^2]} \to 0$$

Obs: By Chebychev we have

$$P(|Y - Y_n| \geq \epsilon) \leq \frac{E[(Y - Y_n)^2]}{\epsilon^2} \to 0$$

for ay fixed $\epsilon > 0$.

Definition. The sequence of random variables $\{Y_n\}_{n\in\mathbb{N}}$ converges *weakly* to the random variable Y if $E[g(Y)] - E[g(Y_n)] \to 0$, for all bounded continuous functions g.

Observe: strong convergence $\Rightarrow$ weak convergence, but the converse is in general not true.

Strong and weak convergence

Counterexample. Let random variables $\{Y_n\}_{n\in\mathbb{N}}$ be *iid* in $(\Omega, \mathcal{F}, P)$, and $Y_n \sim N(0,1), n = 1, \ldots$.
Verify that Y_n converges weakly but not strongly!

Proof of ($\Rightarrow$) for Lipschitz functions:

$$\begin{aligned} |E[g(X) - g(Y_n)]| &\le E[|g(X) - g(Y_n)|] \\ &\le C_g E[|X - Y_n|] \\ &\le C_g \underbrace{\sqrt{E[|X - Y_n|^2]}}_{=\|X-Y_n\|_{L^2_P(\Omega)}} \to 0. \end{aligned}$$

Obs: The previous estimate may not be optimal. There are cases where the weak error goes to zero much faster than the strong one.

Ito Integrals

Theorem 1 *Suppose that there exists* $C > 0$ *s.t.* $f : [0,T] \times \mathbb{R} \to \mathbb{R}$ *satisfies*

$$|f(t+\Delta t, W+\Delta W) - f(t,W)| \leq C(\Delta t + |\Delta W|)$$

then the forward Euler (left point quadrature) approximations

$$I_{\Delta t} = \sum_{n=0}^{N-1} f(t_n, W(t_n))\Delta W_n,$$

with $0 = t_0 < t_1 < \ldots < t_N = T$, *satisfy*

$$\|I_{\Delta t} - I_{\Delta t'}\|_{L^2_P(\Omega)} = E[I_{\Delta t} - I_{\Delta t'}]^{1/2} \leq \mathcal{O}\left(\sqrt{\Delta t_{\text{max}}}\right) \quad (12)$$

Ito integrals

Remark 2 *The previous theorem implies that $I_{\Delta t}$ is Cauchy in $L^2_P(\Omega)$ and its limit defines the Ito integral*

$$\sum_{n=0}^{N-1} f(t_n, W(t_n))\Delta W_n = \int_0^T f(s, W(s))dW(s) + \mathcal{O}\left(\sqrt{\Delta t_{\max}}\right)$$

The previous estimate should be understood *as*

$$E\Big[\Big(\sum_{n=0}^{N-1} f(t_n, W(t_n))\Delta W_n - \int_0^T f(s, W(s))dW(s)\Big)^2\Big]$$
$$= \mathcal{O}\left(\Delta t_{\text{max}}\right)$$

Question: What is the computational work to reach an accuracy ϵ in $L^2_P(\Omega)$ sense using uniform time steps?

Information generated by a process.

Definition 9 *The symbol* $\mathcal{F}_t^W$ *denotes* the information generated by W on the interval $[0, t]$. *If, based on the observation of the trajectory* $\{W(s), 0 \leq s \leq t\}$ *it is possible to decide if an event* $A \in \mathcal{F}$ *has occurred or not, then we write* $A \in \mathcal{F}_t^W$*.*

If the value of a random variable Z can be completely determined by the observations $\{W(s), 0 \leq s \leq t\}$ then we write $Z \in \mathcal{F}_t^W$.

A stochastic process g is called **adapted to the filtration** $\{\mathcal{F}_t^W\}_{t\geq 0}$ if $g(t) \in \mathcal{F}_t^W$ for all $t \geq 0$.

Obs Math Grads: The filtration $\{\mathcal{F}_t^W\}_{t\geq 0}$ is actually an increasing family of σ-algebras. See Øksendal's book, Chapter 3, for precise definition.

Examples

1. $A = \{W(10) < 5\}$

2. $Z = \int_0^1 W(s)ds$

3. $f(t) = \sup_{s \le t} W(s)$

4. $g(t) = \sup_{s \le t+1} W(s)$