Lecture 12. Stochastic Process and Brownian Motion

Tiejun Li[1,2]

[1]School of Mathematical Sciences (SMS),
&
[2]Center for Machine Learning Research (CMLR),
Peking University,
Beijing 100871,
P.R. China
tieli@pku.edu.cn

Office: No. 1 Science Building, Room 1376E

Table of Contents

Fair coin tossing process

Example (Coin tossing game)

Consider the independent fair coin tossing process described by the sequence

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

where $X_n = 0$ *or* 1 *if the* n*th output is 'Tail' (T) or 'Head' (H), respectively. Different trials are assumed to be independent and* $\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 1) = 1/2$.

Fair coin tossing process

Example (Coin tossing game)

Consider the independent fair coin tossing process described by the sequence

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

where $X_n = 0$ *or* 1 *if the* n*th output is 'Tail' (T) or 'Head' (H), respectively. Different trials are assumed to be independent and* $\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 1) = 1/2$*.*

- The number of all possible outputs is uncountable. One can not define the probability of an event through summation of the probability of each atom as the case of discrete random variables.

Fair coin tossing process

Example (Coin tossing game)

Consider the independent fair coin tossing process described by the sequence

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

where $X_n = 0$ *or* 1 *if the* n*th output is 'Tail' (T) or 'Head' (H), respectively. Different trials are assumed to be independent and* $\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 1) = 1/2$*.*

- The number of all possible outputs is uncountable. One can not define the probability of an event through summation of the probability of each atom as the case of discrete random variables.
- In fact, if we define $\Omega = \{H, T\}^{\mathbb{N}}$, the probability of an *atom* $\omega = (\omega_1, \omega_2, \ldots, \omega_n, \ldots) \in \{H, T\}^{\mathbb{N}}$ is 0, i.e.

$$\mathbb{P}(X_1(\omega) = k_1, \ldots, X_n(\omega) = k_n, \ldots) = \lim_{n\to\infty} \left(\frac{1}{2}\right)^n = 0.$$

Events like $\{X_n(\omega) = 1\}$ involve uncountably many atoms.

Fair coin tossing process: Construction

Now we set up a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for this process.

- It is natural to take $\Omega = \{H, T\}^{\mathbb{N}}$.

Fair coin tossing process: Construction

Now we set up a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for this process.

- It is natural to take $\Omega = \{H, T\}^{\mathbb{N}}$.
- The σ-algebra $\mathcal{F}$ as the smallest σ-algebra containing all events of the form:

$$C = \Big\{\omega | \omega \in \Omega, (\omega_j)_{j=1:m} \in C_m\Big\}, C_m \subset \{H, T\}^m, m \in \mathbb{N},$$

  i.e. the sets whose finite time projections are specified. These sets are called *cylinder sets*.

Fair coin tossing process: Construction

Now we set up a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for this process.

- It is natural to take $\Omega = \{H, T\}^{\mathbb{N}}$.
- The σ-algebra $\mathcal{F}$ as the smallest σ-algebra containing all events of the form:

  $$C = \Big\{\omega | \omega \in \Omega, (\omega_j)_{j=1:m} \in C_m\Big\}, C_m \subset \{H, T\}^m, m \in \mathbb{N},$$

  i.e. the sets whose finite time projections are specified. These sets are called *cylinder sets*.
- The probability measure $\mathbb{P}$ of a cylinder set is defined to be

  $$\mathbb{P}(C) = \frac{1}{2^m}|C|.$$

Fair coin tossing process: Construction

Now we set up a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for this process.

- It is natural to take $\Omega = \{H, T\}^{\mathbb{N}}$.
- The σ-algebra $\mathcal{F}$ as the smallest σ-algebra containing all events of the form:

$$C = \Big\{\omega | \omega \in \Omega, (\omega_j)_{j=1:m} \in C_m\Big\}, C_m \subset \{H, T\}^m, m \in \mathbb{N},$$

  i.e. the sets whose finite time projections are specified. These sets are called *cylinder sets*.
- The probability measure $\mathbb{P}$ of a cylinder set is defined to be

$$\mathbb{P}(C) = \frac{1}{2^m}|C|.$$

- Denote $\mathcal{C}$ the set of cylinder sets. $\mathcal{C}$ is an algebra which is only closed under finite union/intersection operation. To extend the probability measure $\mathbb{P}$ from $\mathcal{C}$ to $\mathcal{F}$, we need to verify that $\mathbb{P}$ is countably additive on $\mathcal{C}$.

An Important Lemma

Lemma

If $A_n \downarrow A$ *and* $A_n \in \mathcal{C}$ *is non-empty, then* A *is non-empty.*

An Important Lemma

Lemma

If $A_n \downarrow A$ and $A_n \in \mathcal{C}$ is non-empty, then A is non-empty.

Proof. Denote

$$A_n = \{\omega | (\omega_1, \omega_2, \ldots, \omega_{m_n}) \in C_n\}$$

where $\omega_k \in \{H, T\}$. From the non-empty condition of A_n, there exists $\omega^n \in A_n$. Consider

$$\begin{array}{cccc} \omega_1^1 & \omega_1^2 & \omega_1^3 & \cdots \\ \omega_2^1 & \omega_2^2 & \omega_2^3 & \cdots \\ \omega_3^1 & \omega_3^2 & \omega_3^3 & \cdots \\ \vdots & \vdots & \vdots & \ddots, \end{array}$$

then there exists an infinite subsequence $\{n_k^1\}_{k \in \mathbb{N}}$ such that $\omega_1^{n_k^1} = H$ or T always in the first row.

An Important Lemma

Pick up the sub-columns according to $\{n_k^1\}_{k\in\mathbb{N}}$. Then the similar argument can be applied to the continued rows by an subsequence trick. Take the diagonal indices and define $n_k := n_k^k$ and $u_k := \omega_k^{n_k}$ for $k = 1, 2, \ldots$. Denote $u = (u_1, u_2, \ldots)$.

An Important Lemma

Pick up the sub-columns according to $\{n_k^1\}_{k\in\mathbb{N}}$. Then the similar argument can be applied to the continued rows by an subsequence trick. Take the diagonal indices and define $n_k := n_k^k$ and $u_k := \omega_k^{n_k}$ for $k = 1, 2, \ldots$. Denote $u = (u_1, u_2, \ldots)$.

For any r, if $k \geq r$, one has $\omega_j^{n_k} = u_j$ for $1 \leq j \leq r$.
For any n, if $k \geq n$, then $n_k \geq n$, and $\omega^{n_k} \in A_{n_k} \subset A_n$. So $(\omega_1^{n_k}, \omega_2^{n_k}, \ldots, \omega_{m_n}^{n_k}) \in C_n$.
Take $k \geq m_n$. We get $\omega_j^{n_k} = u_j$ for $1 \leq j \leq m_n$, i.e. $u \in A_n$ for any n.

An Important Lemma

Pick up the sub-columns according to $\{n_k^1\}_{k\in\mathbb{N}}$. Then the similar argument can be applied to the continued rows by an subsequence trick. Take the diagonal indices and define $n_k := n_k^k$ and $u_k := \omega_k^{n_k}$ for $k = 1, 2, \ldots$. Denote $u = (u_1, u_2, \ldots)$.

For any r, if $k \geq r$, one has $\omega_j^{n_k} = u_j$ for $1 \leq j \leq r$.
For any n, if $k \geq n$, then $n_k \geq n$, and $\omega^{n_k} \in A_{n_k} \subset A_n$. So $(\omega_1^{n_k}, \omega_2^{n_k}, \ldots, \omega_{m_n}^{n_k}) \in C_n$.
Take $k \geq m_n$. We get $\omega_j^{n_k} = u_j$ for $1 \leq j \leq m_n$, i.e. $u \in A_n$ for any n.

In summary, $u \in A$ and we are done. □

Fair coin tossing process: Construction

Theorem (Measure extension)

A finite measure μ, *i.e.,* $\mu(\Omega) < \infty$, *on an algebra* $F_0 \subset F$ *can be uniquely extended to a measure on* $\sigma(F_0)$.

Fair coin tossing process: Construction

Theorem (Measure extension)

A finite measure μ, i.e., $\mu(\Omega) < \infty$, on an algebra $F_0 \subset F$ can be uniquely extended to a measure on $\sigma(F_0)$.

- With the previous lemma, we obtain if $A_n \downarrow \emptyset$, then $\mathbb{P}(A_n) \downarrow 0$, which is equivalent to the countable additivity. This shows $\mathbb{P}$ is a measure on the cylinder sets $\mathcal{C}$.

Fair coin tossing process: Construction

Theorem (Measure extension)

A finite measure μ*, i.e.,* $\mu(\Omega) < \infty$*, on an algebra* $F_0 \subset F$ *can be uniquely extended to a measure on* $\sigma(F_0)$*.*

- With the previous lemma, we obtain if $A_n \downarrow \emptyset$, then $\mathbb{P}(A_n) \downarrow 0$, which is equivalent to the countable additivity. This shows $\mathbb{P}$ is a measure on the cylinder sets $\mathcal{C}$.
- From the extension theorem of measures, the probability measure $\mathbb{P}$ is well-defined on $\mathcal{F} = \sigma(\mathcal{C})$, i.e., the σ-algebra generated by $\mathcal{C}$.

General stochastic process: State Space

- A stochastic process is a parameterized random variables $\{X_t\}_{t\in\mathbf{T}}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in $\mathbb{R}$. $\mathbf{T}$ can be $\mathbb{N}$, $[0, +\infty)$ or some finite interval.

General stochastic process: State Space

- A stochastic process is a parameterized random variables $\{X_t\}_{t\in\mathbf{T}}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in $\mathbb{R}$. $\mathbf{T}$ can be $\mathbb{N}$, $[0, +\infty)$ or some finite interval.
- For any fixed $t \in \mathbf{T}$, we have a random variable

$$X_t : \Omega \to \mathbb{R} \qquad \omega \rightarrowtail X_t(\omega).$$

For any fixed $\omega \in \Omega$, we have a real-valued measurable function on $\mathbf{T}$

$$X.(\omega) : \mathbf{T} \to \mathbb{R} \qquad t \rightarrowtail X_t(\omega),$$

which is called a trajectory or sample path of X.

General stochastic process: State Space

- A stochastic process is a parameterized random variables $\{X_t\}_{t\in\mathbf{T}}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in $\mathbb{R}$. $\mathbf{T}$ can be $\mathbb{N}$, $[0, +\infty)$ or some finite interval.
- For any fixed $t \in \mathbf{T}$, we have a random variable

$$X_t : \Omega \to \mathbb{R} \qquad \omega \rightarrowtail X_t(\omega).$$

  For any fixed $\omega \in \Omega$, we have a real-valued measurable function on $\mathbf{T}$

$$X.(\omega) : \mathbf{T} \to \mathbb{R} \qquad t \rightarrowtail X_t(\omega),$$

  which is called a trajectory or sample path of X.
- As a bi-variate function, a stochastic process can also be viewed as a measurable function from $\Omega \times \mathbf{T}$ to $\mathbb{R}$

$$(\omega, t) \rightarrowtail X(\omega, t) := X_t(\omega),$$

  with the σ-algebra in $\Omega \times \mathbf{T}$ been chosen as $\mathcal{F} \times \mathcal{T}$, and $\mathcal{T}$ is the Borel σ-algebra on $\mathbf{T}$.

Cylinder Sets

- The largest probability space that one can take is the infinite product space $\Omega = \mathbb{R}^{\mathbf{T}}$, i.e. Ω is the space of all real-valued functions on $\mathbf{T}$. $\mathcal{F}$ can be taken as the infinite product σ-algebra $\mathcal{B}^{\mathbf{T}}$, which is the smallest σ-algebra containing all cylinder sets

$$C = \{\omega \in \mathbb{R}^{\mathbf{T}} | (\omega(t_1), \omega(t_2), \ldots, \omega(t_k)) \in A, A \in \mathcal{B}^k, t_i \in \mathbf{T}\},$$

where $\mathcal{B}, \mathcal{B}^k$ is the Borel σ-algebra on $\mathbb{R}$ and $\mathbb{R}^k$, respectively.

Cylinder Sets

- The largest probability space that one can take is the infinite product space $\Omega = \mathbb{R}^{\mathbf{T}}$, i.e. Ω is the space of all real-valued functions on $\mathbf{T}$. $\mathcal{F}$ can be taken as the infinite product σ-algebra $\mathcal{B}^{\mathbf{T}}$, which is the smallest σ-algebra containing all cylinder sets

$$C = \{\omega \in \mathbb{R}^{\mathbf{T}} | (\omega(t_1), \omega(t_2), \ldots, \omega(t_k)) \in A, A \in \mathcal{B}^k, t_i \in \mathbf{T}\},$$

  where $\mathcal{B}, \mathcal{B}^k$ is the Borel σ-algebra on $\mathbb{R}$ and $\mathbb{R}^k$, respectively.
- When $\mathbf{T} = \mathbb{N}$ and X_t only takes values in $\{0, 1\}$, we are back to the setting of the Fair coin tossing example.

Finite Dimensional Distribution

- Finite dimensional distributions are particularly interesting for a stochastic process, since they are the ones we can really observe.

Finite Dimensional Distribution

- Finite dimensional distributions are particularly interesting for a stochastic process, since they are the ones we can really observe.
- Let

$$\mu_{t_1,\ldots,t_k}(F_1 \times F_2 \times \cdots \times F_k) = \mathbb{P}[X_{t_1} \in F_1, \ldots, X_{t_k} \in F_k]$$

for all $F_1, F_2, \ldots, F_k \in \mathcal{B}$. $\mu_{t_1,\ldots,t_k}$ is called the finite dimensional distributions of $\{X_t\}_{t \in \mathbf{T}}$ at the time slice $(t_1, \ldots, t_k)$, where $t_i \in \mathbf{T}$ for $i = 1, 2, \ldots, k$.

Finite Dimensional Distribution

- Finite dimensional distributions are particularly interesting for a stochastic process, since they are the ones we can really observe.
- Let

$$\mu_{t_1,\ldots,t_k}(F_1 \times F_2 \times \cdots \times F_k) = \mathbb{P}[X_{t_1} \in F_1, \ldots, X_{t_k} \in F_k]$$

  for all $F_1, F_2, \ldots, F_k \in \mathcal{B}$. $\mu_{t_1,\ldots,t_k}$ is called the finite dimensional distributions of $\{X_t\}_{t\in\mathbf{T}}$ at the time slice $(t_1, \ldots, t_k)$, where $t_i \in \mathbf{T}$ for $i = 1, 2, \ldots, k$.
- Kolmogorov's extension theorem states that an abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$ can be established for a stochastic process X by knowing its all finite dimensional distributions with suitable consistency conditions.

Kolmogorov's extension theorem

Theorem (Kolmogorov's extension theorem)

Assume that a family of finite dimensional distributions $\{\mu_{t_1,\ldots,t_k}\}$ *satisfy the following two consistency conditions for arbitrary sets of* $t_1, t_2, \ldots, t_k \in T, k \in \mathbb{N}$*:*
(i) For any permutation σ *of* $\{1, 2, \ldots, k\}$*,*

$$\mu_{t_{\sigma(1)},\ldots,t_{\sigma(k)}}(F_1 \times \cdots \times F_k) = \mu_{t_1,\ldots,t_k}(F_{\sigma^{-1}(1)} \times \cdots \times F_{\sigma^{-1}(k)}).$$

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

$$\begin{aligned}&\mu_{t_1,\ldots,t_k}(F_1 \times F_2 \times \cdots \times F_k)\\ =&\mu_{t_1,\ldots,t_k,t_{k+1},\ldots,t_{k+m}}(F_1 \times \cdots \times F_k \times \mathbb{R} \times \cdots \times \mathbb{R}).\end{aligned}$$

Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *and a stochastic process* $\{X_t\}_{t\in\mathbf{T}}$ *such that*

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

for any $t_1, t_2, \ldots, t_m \in \mathbf{T}$*,* $m \in \mathbb{N}$*.*

Table of Contents

Filtration

The more we observe about a stochastic process, the more information we have at our disposal. This gives rise to a family of increasingly larger σ-algebras, which we call the filtration associated with the stochastic process.

Filtration

The more we observe about a stochastic process, the more information we have at our disposal. This gives rise to a family of increasingly larger σ-algebras, which we call the filtration associated with the stochastic process.

Definition (Filtration)

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

Filtration

The more we observe about a stochastic process, the more information we have at our disposal. This gives rise to a family of increasingly larger σ-algebras, which we call the filtration associated with the stochastic process.

Definition (Filtration)

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

- The filtration is the main conceptual difference between the random variables and and stochastic processes.

Filtration

The more we observe about a stochastic process, the more information we have at our disposal. This gives rise to a family of increasingly larger σ-algebras, which we call the filtration associated with the stochastic process.

Definition (Filtration)

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

- The filtration is the main conceptual difference between the random variables and and stochastic processes.
- A stochastic process $\{X_t\}$ is called *$\mathcal{F}_t$-adapted* if X_t is $\mathcal{F}_t$-measurable, i.e. $X_t^{-1}(B) \in \mathcal{F}_t$, for any $t \geq 0$ and $B \in \mathcal{B}$.

Filtration: Intuition

- Given a stochastic process $\{X_t\}$, one can define the filtration generated by this process by: $\mathcal{F}_t^X = \sigma(X_s, s \leq t)$, which is the smallest σ-algebra such that the $\{X_s\}_{s \leq t}$ are measurable.

Filtration: Intuition

- Given a stochastic process $\{X_t\}$, one can define the filtration generated by this process by: $\mathcal{F}_t^X = \sigma(X_s, s \leq t)$, which is the smallest σ-algebra such that the $\{X_s\}_{s \leq t}$ are measurable.
- $\mathcal{F}_t^X$ is the smallest filtration such that the process $\{X_t\}$ is adapted.

Filtration: Intuition

- Given a stochastic process $\{X_t\}$, one can define the filtration generated by this process by: $\mathcal{F}_t^X = \sigma(X_s, s \leq t)$, which is the smallest σ-algebra such that the $\{X_s\}_{s \leq t}$ are measurable.
- $\mathcal{F}_t^X$ is the smallest filtration such that the process $\{X_t\}$ is adapted.
- The filtration $\mathcal{F}_t^X$ can be thought of as the information supplied by the process up to time t.

Filtration: Example

Taking the independent coin tossing as an example:

- $\Omega = \{H, T\}^{\mathbb{N}}$. $\mathbf{T} = \mathbb{N}$ and the filtration is $\{\mathcal{F}_n^X\}_{n \geq 0}$.

Filtration: Example

Taking the independent coin tossing as an example:

- $\Omega = \{H, T\}^{\mathbb{N}}$. $\mathbf{T} = \mathbb{N}$ and the filtration is $\{\mathcal{F}_n^X\}_{n \geq 0}$.
- When $n = 0$, the σ-algebra is trivial

$$\mathcal{F}_0^X = \{\emptyset, \Omega\},$$

which means that we do not know any information about the output of the coin tossing.

Filtration: Example

Taking the independent coin tossing as an example:

- $\Omega = \{H, T\}^{\mathbb{N}}$. $\mathbf{T} = \mathbb{N}$ and the filtration is $\{\mathcal{F}_n^X\}_{n \geq 0}$.
- When $n = 0$, the σ-algebra is trivial

  $$\mathcal{F}_0^X = \{\emptyset, \Omega\},$$

  which means that we do not know any information about the output of the coin tossing.
- When $n = 1$, the σ-algebra is

  $$\mathcal{F}_1^X = \{\emptyset, \Omega, \{H\}, \{T\}\}$$

  since the first output gives either Head or Tail and we only know this information about the first output.

Filtration: Example

- When $n = 2$, we have

$$\mathcal{F}_2^X = \{\emptyset, \Omega, \{H\cdot\}, \{T\cdot\}, \{\cdot H\}, \{\cdot T\}, \{HH\}, \{HT\}, \{TH\}, \{TT\}, \ldots\},$$

  which contains all possible combinations of the outputs for the first two rounds of experiments.

Filtration: Example

- When $n = 2$, we have

$$\mathcal{F}_2^X = \{\emptyset, \Omega, \{H\cdot\}, \{T\cdot\}, \{\cdot H\}, \{\cdot T\}, \{HH\}, \{HT\}, \{TH\}, \{TT\}, \ldots\},$$

  which contains all possible combinations of the outputs for the first two rounds of experiments.

- Sets like

$$\{HH\cdots T\} \quad \text{or} \quad \{HH\cdots H\}$$

  are not contained in $\mathcal{F}_0^X$, $\mathcal{F}_1^X$ or $\mathcal{F}_2^X$ since the first two outputs can not tell such information.

Filtration: Example

- When $n = 2$, we have

$$\mathcal{F}_2^X = \{\emptyset, \Omega, \{H\cdot\}, \{T\cdot\}, \{\cdot H\}, \{\cdot T\}, \{HH\}, \{HT\}, \{TH\}, \{TT\}, \dots\},$$

  which contains all possible combinations of the outputs for the first two rounds of experiments.
- Sets like

$$\{HH \cdots T\} \quad \text{or} \quad \{HH \cdots H\}$$

  are not contained in $\mathcal{F}_0^X$, $\mathcal{F}_1^X$ or $\mathcal{F}_2^X$ since the first two outputs can not tell such information.
- It is obvious that $\mathcal{F}_n^X$ becomes finer and finer as n increases.

Stopping Time: Discrete Case

Definition (Stopping time: Discrete case)

A random variable T taking values in $\{1, 2, \ldots\} \cup \{\infty\}$ is said to be a stopping time if for any $n < \infty$

$$\{T \leq n\} \in \mathcal{F}_n.$$

Stopping Time: Discrete Case

Definition (Stopping time: Discrete case)

A random variable T taking values in $\{1, 2, \ldots\} \cup \{\infty\}$ is said to be a stopping time if for any $n < \infty$

$$\{T \leq n\} \in \mathcal{F}_n.$$

- One simple example of stopping time for the coin tossing process is

$$T = \inf \left\{ n : \text{ there exists three consecutive } 0 \text{ in } \{X_k\}_{k \leq n} \right\}.$$

Stopping Time: Discrete Case

Definition (Stopping time: Discrete case)

A random variable T taking values in $\{1, 2, \ldots\} \cup \{\infty\}$ is said to be a stopping time if for any $n < \infty$

$$\{T \leq n\} \in \mathcal{F}_n.$$

- One simple example of stopping time for the coin tossing process is

$$T = \inf \left\{ n : \text{ there exists three consecutive } 0 \text{ in } \{X_k\}_{k \leq n} \right\}.$$

- It is easy to show that the condition $\{T \leq n\} \in \mathcal{F}_n$ is equivalent to $\{T = n\} \in \mathcal{F}_n$ for discrete time processes.

Stopping Time: Simple Properties

Proposition (Properties of stopping times)

For the Markov process $\{X_n\}_{n\in\mathbb{N}}$*, we have*

1. *If* T_1, T_2 *are stopping times, then* $T_1 \wedge T_2$*,* $T_1 \vee T_2$ *and* $T_1 + T_2$ *are also stopping times.*
2. *If* $\{T_k\}_{k\geq 1}$ *are stopping times, then*

$$\sup_k T_k, \quad \inf_k T_k, \quad \limsup_k T_k, \quad \liminf_k T_k$$

are stopping times.

Stopping Time: Continuous Case

Definition (Stopping time: Continuous case)

A random variable T taking values in $\bar{\mathbb{R}}^+$ is said to be a stopping time if for any $t \in \mathbb{R}^+$

$$\{T \leq t\} \in \mathcal{F}_t.$$

Stopping Time: Continuous Case

Definition (Stopping time: Continuous case)

A random variable T taking values in $\bar{\mathbb{R}}^+$ is said to be a stopping time if for any $t \in \mathbb{R}^+$

$$\{T \leq t\} \in \mathcal{F}_t.$$

- In this case we no longer have the equivalence between $\{T \leq t\} \in \mathcal{F}_t$ and $\{T = t\} \in \mathcal{F}_t$. Previous proposition also holds for the continuous time case if the filtration is right continuous, i.e. $\mathcal{F}_t = \mathcal{F}_{t+} := \cap_{s>t} \mathcal{F}_s$.

Table of Contents

Gaussian Distribution

- Any Gaussian vector $X = (X_1, X_2, \ldots, X_n)^T$ is completely determined by its first moment $m = \mathbb{E}X$ and second moment $K = \mathbb{E}(X - m)(X - m)^T$, where $m_i = \mathbb{E}X_i$ and $K_{ij} = \mathbb{E}(X_i - m_i)(X_j - m_j)$.

Gaussian Distribution

- Any Gaussian vector $X = (X_1, X_2, \ldots, X_n)^T$ is completely determined by its first moment $m = \mathbb{E}X$ and second moment $K = \mathbb{E}(X - m)(X - m)^T$, where $m_i = \mathbb{E}X_i$ and $K_{ij} = \mathbb{E}(X_i - m_i)(X_j - m_j)$.
- If K is invertible, the corresponding pdf is

$$p(\boldsymbol{x}) = \frac{1}{Z} e^{-\frac{1}{2}(\boldsymbol{x}-m)^T K^{-1}(\boldsymbol{x}-m)},$$

where Z is a normalization constant.

Gaussian Distribution

- Any Gaussian vector $X = (X_1, X_2, \ldots, X_n)^T$ is completely determined by its first moment $m = \mathbb{E}X$ and second moment $K = \mathbb{E}(X - m)(X - m)^T$, where $m_i = \mathbb{E}X_i$ and $K_{ij} = \mathbb{E}(X_i - m_i)(X_j - m_j)$.
- If K is invertible, the corresponding pdf is

$$p(\boldsymbol{x}) = \frac{1}{Z} e^{-\frac{1}{2}(\boldsymbol{x}-m)^T K^{-1}(\boldsymbol{x}-m)},$$

  where Z is a normalization constant.
- For the general case, we can represent X via the characteristic function

$$\mathbb{E}e^{i\boldsymbol{\xi}\cdot X} = e^{i\boldsymbol{\xi}\cdot m - \frac{1}{2}\boldsymbol{\xi}^T K \boldsymbol{\xi}}.$$

Gaussian Process

Definition

A Gaussian process *means that all of the finite dimensional distributions* $\mu_{t_1,\ldots,t_k}$ *are Gaussian for any* $t_1, t_2, \ldots, t_k \in T$.

Gaussian Process

Definition

A Gaussian process means that all of the finite dimensional distributions $\mu_{t_1,\ldots,t_k}$ are Gaussian for any $t_1, t_2, \ldots, t_k \in T$.

- From the properties of Gaussian vectors, a Gaussian process is uniquely determined by the mean function $m(t) = \mathbb{E}X_t$ and the covariance function $K(s,t) = \mathbb{E}(X_s - m(s))(X_t - m(t))$.

Gaussian Process

Definition

A Gaussian process means that all of the finite dimensional distributions $\mu_{t_1,\ldots,t_k}$ are Gaussian for any $t_1, t_2, \ldots, t_k \in T$.

- From the properties of Gaussian vectors, a Gaussian process is uniquely determined by the mean function $m(t) = \mathbb{E}X_t$ and the covariance function $K(s,t) = \mathbb{E}(X_s - m(s))(X_t - m(t))$.
- If we consider the finite dimensional distribution at the time slice $(t_1, t_2, \ldots, t_n)$, then $m(t)$ and $K(s,t)$ give the first moment

  $$M = \big(m(t_1), m(t_2), \ldots, m(t_n)\big)$$

  and second moment

  $$K = \begin{bmatrix} K(t_1,t_1) & K(t_1,t_2) & \cdots & K(t_1,t_n) \\ K(t_2,t_1) & K(t_2,t_2) & \cdots & K(t_2,t_n) \\ \vdots & \vdots & \ddots & \vdots \\ K(t_n,t_1) & K(t_n,t_2) & \cdots & K(t_n,t_n) \end{bmatrix}.$$

Gaussian Process: Characteristic Functional

- For any $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, we have

$$\sum_{i,j} K(t_i, t_j) x_i x_j = \mathbb{E}\Big(\sum_i (X_{t_i} - m(t_i)) x_i \Big)^2 \geq 0.$$

Thus we may view $m(t)$ as an infinite dimensional vector, and $K(s, t)$ as an infinite dimensional positive semi-definite matrix.

Gaussian Process: Characteristic Functional

- For any $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, we have

$$\sum_{i,j} K(t_i, t_j) x_i x_j = \mathbb{E}\Big(\sum_i (X_{t_i} - m(t_i)) x_i\Big)^2 \geq 0.$$

  Thus we may view $m(t)$ as an infinite dimensional vector, and $K(s,t)$ as an infinite dimensional positive semi-definite matrix.

- The Gaussian process X can be explained as a Gaussian random element in an infinite dimensional space $L^2(T)$ since

$$\mathbb{E}e^{i(\xi,X)} = e^{i(\xi,m) - \frac{1}{2}(\xi,K\xi)},$$

  where $(\xi, m) = \int_a^b \xi(t) m(t) dt$, and $(K\xi)(t) = \int_a^b K(t,s)\xi(s) ds$ is the action of the kernel function K on the function ξ.

Gaussian Process: Characteristic Functional

- For any $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, we have

$$\sum_{i,j} K(t_i, t_j) x_i x_j = \mathbb{E}\Big(\sum_i (X_{t_i} - m(t_i)) x_i\Big)^2 \geq 0.$$

  Thus we may view $m(t)$ as an infinite dimensional vector, and $K(s,t)$ as an infinite dimensional positive semi-definite matrix.

- The Gaussian process X can be explained as a Gaussian random element in an infinite dimensional space $L^2(T)$ since

$$\mathbb{E}e^{i(\xi,X)} = e^{i(\xi,m) - \frac{1}{2}(\xi,K\xi)},$$

  where $(\xi, m) = \int_a^b \xi(t) m(t) dt$, and $(K\xi)(t) = \int_a^b K(t,s)\xi(s) ds$ is the action of the kernel function K on the function ξ.

- Based on the Kolmogorov's extension theorem, we can construct a Gaussian process X from a given mean function $m(t)$ and covariance function $K(s,t)$.

Covariance Kernel

The covariance function K is obviously symmetric, i.e. $K(t,s) = K(s,t)$, by definition. In addition, we have the semi-positivity of K in the following sense.

Theorem

Assume the Gaussian process $(X_t)_{t\in[0,T]}$ possesses the regularity $X \in L^2_\omega L^2_t$ in the sense that $X \in L^2(\Omega; L^2[0,T])$, i.e.

$$\mathbb{E}\int_0^T X_t^2 dt < \infty.$$

We have $m \in L^2_t$ and the operator

$$\mathcal{K}f(s) := \int_0^T K(s,t)f(t)dt, \quad s \in [0,T]$$

is a positive, compact operator on L^2_t.

Covariance Kernel

Proof. The mean function $m \in L_t^2$ is obvious since

$$\int_0^T m^2(t)dt = \int_0^T (\mathbb{E}X_t)^2 dt \le \int_0^T \mathbb{E}X_t^2 dt < \infty.$$

In addition, we have

$$\int_0^T \int_0^T K^2(s,t)dsdt = \int_0^T \int_0^T \Big(\mathbb{E}(X_t - m(t))(X_s - m(s))\Big)^2 dsdt$$
$$\le \int_0^T \int_0^T \mathbb{E}(X_t - m(t))^2 \mathbb{E}(X_s - m(s))^2 dsdt \le \Big(\int_0^T \mathbb{E}X_t^2 dt\Big)^2,$$

which means $K \in L^2([0,T] \times [0,T])$. Thus $\mathcal{K}$ is a compact operator on L_t^2.

It is easy to find that the adjoint operator of $\mathcal{K}$ is

$$\mathcal{K}^* f(s) := \int_0^T K(t,s) f(t) dt, \quad s \in [0,T].$$

Covariance Kernel

From the symmetry of $K(s,t)$, we know that $\mathcal{K}$ is self-adjoint.
To show the positivity of $\mathcal{K}$, we have

$$
\begin{aligned}
(\mathcal{K}f, f) &= \int_0^T \int_0^T \mathbb{E}(X_t - m(t))(X_s - m(s)) f(t) f(s) ds dt \\
&= \mathbb{E}\Big(\int_0^T (X_t - m(t)) f(t) dt \Big)^2 \geq 0.
\end{aligned}
$$

□

Closure Property

Theorem (Closure property for Gaussian random variables)

Suppose $X_1, X_2, \ldots$ are a sequence of Gaussian random variables and X_n converges to X in probability. Then X is also Gaussian.

Closure Property

Theorem (Closure property for Gaussian random variables)

Suppose $X_1, X_2, \ldots$ are a sequence of Gaussian random variables and X_n converges to X in probability. Then X is also Gaussian.

Proof.

Let us denote

$$m_k = \mathbb{E}X_k, \quad \sigma_k^2 = \mathrm{var}X_k.$$

Then by dominated convergence theorem we have

$$e^{i\xi m_k - \frac{1}{2}\sigma_k^2\xi^2} = \mathbb{E}e^{i\xi X_k} \to \mathbb{E}e^{i\xi X} \quad \text{for any } \xi \in \boldsymbol{R}.$$

From the existence of the limit of the above equation, there are numbers m and σ^2 such that

$$m = \lim m_k, \quad \sigma^2 = \lim \sigma_k^2$$

and $\mathbb{E}e^{i\xi X} = e^{i\xi m - \frac{1}{2}\sigma^2\xi^2}$. □

Table of Contents

Wiener Process: Definition

Definition (Brownian motion)

The one dimensional Brownian motion (or Wiener process) W_t *is defined as*

Wiener Process: Definition

Definition (Brownian motion)

The one dimensional Brownian motion (or Wiener process) W_t is defined as

1. *It is a Gaussian process.*

Wiener Process: Definition

Definition (Brownian motion)

The one dimensional Brownian motion (or Wiener process) W_t is defined as

1. *It is a Gaussian process.*
2. *It has mean function $m(t) = 0$, and covariance function $K(s,t) = s \wedge t = \min(s,t)$.*

Wiener Process: Definition

Definition (Brownian motion)

The one dimensional Brownian motion (or Wiener process) W_t is defined as

1. *It is a Gaussian process.*
2. *It has mean function $m(t) = 0$, and covariance function $K(s,t) = s \wedge t = \min(s,t)$.*
3. *With probability one, $t \mapsto W_t$ is continuous.*

Wiener Process: Definition

Definition (Brownian motion)

The one dimensional Brownian motion (or Wiener process) W_t is defined as

1. *It is a Gaussian process.*
2. *It has mean function $m(t) = 0$, and covariance function $K(s,t) = s \wedge t = \min(s,t)$.*
3. *With probability one, $t \mapsto W_t$ is continuous.*

- The m-dimensional Brownian motion $\boldsymbol{W}_t$ has the form $\boldsymbol{W}_t = (W_t^1, W_t^2, \ldots, W_t^m)$, where each component W_t^j is a Brownian motion and they are independent each other.

Wiener Process: Definition

Definition (Brownian motion)

The one dimensional Brownian motion (or Wiener process) W_t is defined as

1. *It is a Gaussian process.*
2. *It has mean function $m(t) = 0$, and covariance function $K(s,t) = s \wedge t = \min(s,t)$.*
3. *With probability one, $t \mapsto W_t$ is continuous.*

- The m-dimensional Brownian motion $\boldsymbol{W}_t$ has the form $\boldsymbol{W}_t = (W_t^1, W_t^2, \dots, W_t^m)$, where each component W_t^j is a Brownian motion and they are independent each other.
- The Brownian motion (m-dimensional Brownian motion) is usually denoted as W_t or B_t ($\boldsymbol{W}_t$ or $\boldsymbol{B}_t$).

Wiener Process: Equivalent Definition

It is not difficult to prove that the three conditions are equivalent to the following definition.

$1'$. For any $t_0 < t_1 < \cdots < t_n$, the random variables $W_{t_0}, W_{t_1} - W_{t_0}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent.

Wiener Process: Equivalent Definition

It is not difficult to prove that the three conditions are equivalent to the following definition.

$1'$. For any $t_0 < t_1 < \cdots < t_n$, the random variables $W_{t_0}, W_{t_1} - W_{t_0}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent.

$2'$. For any $s, t \geq 0$, $W_{s+t} - W_s \sim N(0, t)$.

Wiener Process: Equivalent Definition

It is not difficult to prove that the three conditions are equivalent to the following definition.

1'. For any $t_0 < t_1 < \cdots < t_n$, the random variables $W_{t_0}, W_{t_1} - W_{t_0}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent.

2'. For any $s, t \geq 0$, $W_{s+t} - W_s \sim N(0, t)$.

3. With probability one, $t \mapsto W_t$ is continuous.

Wiener Process: Equivalent Definition

It is not difficult to prove that the three conditions are equivalent to the following definition.

1′. For any $t_0 < t_1 < \cdots < t_n$, the random variables $W_{t_0}, W_{t_1} - W_{t_0}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent.
2′. For any $s, t \geq 0$, $W_{s+t} - W_s \sim N(0, t)$.
3. With probability one, $t \mapsto W_t$ is continuous.

- Then we obtain the joint probability distribution density for $(W_{t_1}, W_{t_2}, \ldots, W_{t_n})$ $(t_1 < t_2 < \cdots < t_n)$ as

$$p_n(w_1, w_2, \ldots, w_n)$$

$$= \frac{1}{\sqrt{2\pi t_1}} e^{-\frac{w_1^2}{2t_1}} \frac{1}{\sqrt{2\pi(t_2 - t_1)}} e^{-\frac{(w_2 - w_1)^2}{2(t_2 - t_1)}} \cdots \frac{1}{\sqrt{2\pi(t_n - t_{n-1})}} e^{-\frac{(w_n - w_{n-1})^2}{2(t_n - t_{n-1})}}.$$

Wiener Process: Equivalent Definition

It is not difficult to prove that the three conditions are equivalent to the following definition.

1′. For any $t_0 < t_1 < \cdots < t_n$, the random variables $W_{t_0}, W_{t_1} - W_{t_0}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent.
2′. For any $s, t \geq 0$, $W_{s+t} - W_s \sim N(0, t)$.
3. With probability one, $t \mapsto W_t$ is continuous.

- Then we obtain the joint probability distribution density for $(W_{t_1}, W_{t_2}, \ldots, W_{t_n})$ $(t_1 < t_2 < \cdots < t_n)$ as

$$p_n(w_1, w_2, \ldots, w_n)$$
$$= \frac{1}{\sqrt{2\pi t_1}} e^{-\frac{w_1^2}{2t_1}} \frac{1}{\sqrt{2\pi(t_2 - t_1)}} e^{-\frac{(w_2-w_1)^2}{2(t_2-t_1)}} \cdots \frac{1}{\sqrt{2\pi(t_n - t_{n-1})}} e^{-\frac{(w_n-w_{n-1})^2}{2(t_n-t_{n-1})}}.$$

- More compactly,

$$p_n(w_1, w_2, \ldots, w_n) = \frac{1}{Z_n} \exp(-I_n(w)).$$

Wiener Process: Basic Properties

- It's easy to show the stationarity and Markovianity of the Brownian motion with transition kernel function $p(x, t|y, s)$

$$\mathbb{P}(W_t \in B|W_s = y) = \int_B \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{(x-y)^2}{2(t-s)}} dx$$
$$= \int_B p(x, t|y, s) dx$$

  where $s < t$ and B is a Borel set on $\mathbf{R}$.

Wiener Process: Basic Properties

- It's easy to show the stationarity and Markovianity of the Brownian motion with transition kernel function $p(x, t|y, s)$

$$\mathbb{P}(W_t \in B|W_s = y) = \int_B \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{(x-y)^2}{2(t-s)}} dx$$
$$= \int_B p(x, t|y, s) dx$$

  where $s < t$ and B is a Borel set on $\mathbf{R}$.

- The transition probability density $p(x, t|y, s)$ satisfies the stationarity $p(x, t|y, s) = p(x - y, t - s|0, 0)$ and $p(x, t|0, 0)$ satisfies the PDE

$$\frac{\partial p}{\partial t} = \frac{1}{2} \frac{\partial^2 p}{\partial x^2}, \quad p(x, 0|0, 0) = \delta(x).$$

Wiener Process: Existence

Mathematically the first question is "*Is there a process with these properties?*"

- From Kolmogorov's extension theorem we can construct a probability space on $(\boldsymbol{R}^{[0,\infty)}, \mathcal{R}^{[0,\infty)})$ by the consistency of the finite dimensional distributions,

Wiener Process: Existence

Mathematically the first question is "*Is there a process with these properties?*"

- From Kolmogorov's extension theorem we can construct a probability space on $(\boldsymbol{R}^{[0,\infty)}, \mathcal{R}^{[0,\infty)})$ by the consistency of the finite dimensional distributions,
- But it is not straightforward that the condition 3 in the Definition must be satisfied automatically.

Wiener Process: Existence

Mathematically the first question is "*Is there a process with these properties?*"

- From Kolmogorov's extension theorem we can construct a probability space on $(\boldsymbol{R}^{[0,\infty)}, \mathcal{R}^{[0,\infty)})$ by the consistency of the finite dimensional distributions,
- But it is not straightforward that the condition 3 in the Definition must be satisfied automatically.
- In fact, define the set

$$C = \{\omega | \omega \in \boldsymbol{R}^T, \omega \text{ is continuous on } T\}.$$

we will show that C is not a measurable set in $\mathcal{R}^T$!

Wiener Process: Existence

Theorem

For any family of real functions $X_t : \Omega \to \boldsymbol{R}$, $t \in T$.

(i) *If* $A \in \sigma\{X_t, t \in T\}$ *and* $\omega \in A$, *and if* $X_t(\omega') = X_t(\omega)$ *for all* $t \in T$, *then we have* $\omega' \in A$.

(ii) *If* $A \in \sigma\{X_t, t \in T\}$, *then* $A \in \sigma\{X_t, t \in S\}$ *for some countable subset* $S \subset T$.

Wiener Process: Existence

Theorem

For any family of real functions $X_t : \Omega \to \boldsymbol{R}$, $t \in T$.

(i) *If $A \in \sigma\{X_t, t \in T\}$ and $\omega \in A$, and if $X_t(\omega') = X_t(\omega)$ for all $t \in T$, then we have $\omega' \in A$.*

(ii) *If $A \in \sigma\{X_t, t \in T\}$, then $A \in \sigma\{X_t, t \in S\}$ for some countable subset $S \subset T$.*

- To apply the above theorem, we take $T = [0, \infty)$ and S a countable dense subset of T. We will have $C \in \mathcal{R}^S$ if $C \in \mathcal{R}^T$ by the second statement.

Wiener Process: Existence

Theorem

For any family of real functions $X_t : \Omega \to \boldsymbol{R}$, $t \in T$.

(i) *If $A \in \sigma\{X_t, t \in T\}$ and $\omega \in A$, and if $X_t(\omega') = X_t(\omega)$ for all $t \in T$, then we have $\omega' \in A$.*

(ii) *If $A \in \sigma\{X_t, t \in T\}$, then $A \in \sigma\{X_t, t \in S\}$ for some countable subset $S \subset T$.*

- To apply the above theorem, we take $T = [0, \infty)$ and S a countable dense subset of T. We will have $C \in \mathcal{R}^S$ if $C \in \mathcal{R}^T$ by the second statement.
- From the first statement, C should contain all functions which have the same value with some $f \in C$ on S. This should contain lots of discontinuous functions. This contradicts with that C is the set of continuous functions.

Wiener Process: Existence

Theorem

For any family of real functions $X_t : \Omega \to \boldsymbol{R}$, $t \in T$.

(i) *If $A \in \sigma\{X_t, t \in T\}$ and $\omega \in A$, and if $X_t(\omega') = X_t(\omega)$ for all $t \in T$, then we have $\omega' \in A$.*

(ii) *If $A \in \sigma\{X_t, t \in T\}$, then $A \in \sigma\{X_t, t \in S\}$ for some countable subset $S \subset T$.*

- To apply the above theorem, we take $T = [0, \infty)$ and S a countable dense subset of T. We will have $C \in \mathcal{R}^S$ if $C \in \mathcal{R}^T$ by the second statement.
- From the first statement, C should contain all functions which have the same value with some $f \in C$ on S. This should contain lots of discontinuous functions. This contradicts with that C is the set of continuous functions.
- We need the concept "modification" of a process.

Modification

Definition (Modification)

Two processes X and X' defined on the same probability space are said to be modifications of each other if for each t,

$$X_t = X'_t \quad \textit{a.s.}$$

They are called indistinguishable if for almost all ω

$$X_t(\omega) = X'_t(\omega) \quad \textit{for every } t.$$

Modification

Definition (Modification)

Two processes X and X' defined on the same probability space are said to be modifications of each other if for each t,

$$X_t = X'_t \quad \text{a.s.}$$

They are called indistinguishable if for almost all ω

$$X_t(\omega) = X'_t(\omega) \quad \text{for every } t.$$

- It is clear that if X and X' are modifications of each other, they have the same finite dimensional distribution.

Modification

Definition (Modification)

Two processes X and X' defined on the same probability space are said to be modifications of each other if for each t,

$$X_t = X'_t \quad \text{a.s.}$$

They are called indistinguishable if for almost all ω

$$X_t(\omega) = X'_t(\omega) \quad \text{for every } t.$$

- It is clear that if X and X' are modifications of each other, they have the same finite dimensional distribution.
- If X and X' are modifications of each other and are almost surely continuous, they are indistinguishable.

Kolmogorov's continuity theorem: Wiener Path Continuity

Theorem (Kolmogorov's continuity theorem)

A real-valued process X for which there exist three strictly positive constants α, β, C such that

$$\mathbb{E}(|X_t - X_s|^\alpha) \leq C|t-s|^{1+\beta}$$

for any $s, t \geq 0$, then there is a modification $\tilde{X}$ of X which is almost-surely continuous.

For Brownian motion, the condition of the above theorem is satisfied with $\alpha = 4, \beta = 1$ and thus the continuity of Brownian motion can be ensured in the sense of modifications.