Topics in probability theory: Markov Process

Zhenyao Sun

Beijing Institute of Technology

September 17, 2024

Disintegration

Theorem 8.5

Let ξ, η be random elements in S and T. Let T be a Polish space. Then $\mathcal{L}(\xi, \eta) = \mathcal{L}(\xi) \otimes \mu$ for a probability kernel μ from S to T. Furthermore,

$$\mathcal{L}(\eta|\xi) = \mu(\xi, \cdot), \quad a.s.$$

- Law of random elements, $\mathcal{L}(\xi)$ and $\mathcal{L}(\xi, \eta)$?
- Probability kernel μ?
- Outer product $\mathcal{L}(\xi) \otimes \mu$?
- Conditional distribution $\mathcal{L}(\eta|\xi)$?

Law of random elements

- Let ξ be a random element taking values in a measurable space $(S, \mathcal{S})$, defined on a probability space, say $(\Omega, \mathcal{A}, \mathbb{P})$.
- The law of ξ, denoted by $\mathcal{L}(\xi)$, is a probability measure on S s.t.

$$\mathcal{L}(\xi)(A) = \mathbb{P}(\xi \in A), \quad \forall A \in \mathcal{S}.$$

- Let η be another random element taking values in measure space $(T, \mathcal{T})$.
- Then (ξ, η) is a random element taking values in the product space $(S \times T, \mathcal{S} \otimes \mathcal{T})$.
- The law of (ξ, η) is denoted by $\mathcal{L}(\xi, \eta)$.

Probability kernels

- Let $(S, \mathcal{S})$ and $(X, \mathcal{X})$ be two measurable spaces.
- We say μ is a kernel from S to X, if
  - $\mu : S \times \mathcal{X} \to \mathbb{R}$;
  - for every $s \in S$, $\mu(s, \cdot) : \mathcal{X} \to \mathbb{R}$ is a measure of the measurable space $(X, \mathcal{X})$; and
  - for every $A \in \mathcal{X}$, $\mu(\cdot, A) : S \to \mathbb{R}$ is a measurable function on S.
- We say a kernel μ from S to X is a probability kernel, if for every $s \in S$, $\mu(s, \cdot) : \mathcal{X} \to \mathbb{R}$ is a probability measure.
- We say μ is a kernel on S, if it is a kernel from S to S.

Inner product of probability kernels

- Suppose that μ is a probability kernels from a measurable space $(S, \mathcal{S})$ to a measurable space $(X, \mathcal{X})$.
- Suppose that ν is a probability kernels from $(X, \mathcal{X})$ to a measurable space $(Y, \mathcal{Y})$.
- Define a map $\mu\nu : S \times \mathcal{Y} \to \mathbb{R}$ such that

$$(\mu\nu)(s, A) = \int_X \mu(s, \mathrm{d}x)\nu(x, A), \quad s \in S, A \in \mathcal{Y}.$$

- It can be verified that $\mu\nu$ is a probability kernel from $(S, \mathcal{S})$ to $(Y, \mathcal{Y})$.
- We say $\mu\nu$ is the inner product between μ and ν.

Outer product of probability kernels

- Suppose that μ is a probability kernels from a measurable space $(S, \mathcal{S})$ to a measurable space $(X, \mathcal{X})$.
- Suppose that ν is a probability kernels from $(X, \mathcal{X})$ to a measurable space $(Y, \mathcal{Y})$.
- Denote by $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$ the product measurable space of $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$.
- Define map $\mu \otimes \nu : S \times (\mathcal{X} \otimes \mathcal{Y}) \to \mathbb{R}$ such that for any $s \in S$, $A \in \mathcal{X}$ and $B \in \mathcal{Y}$,

$$(\mu \otimes \nu)(s, A \times B) = \int_A \mu(s, \mathrm{d}x)\nu(x, B).$$

- It can be verified that $\mu \otimes \nu$ is a probability kernel from $(S, \mathcal{S})$ to the product space $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$.
- We say $\mu \otimes \nu$ is the outer product between μ and ν.

Inner product and outer product of a probability measure and a probability kernel

- Suppose that μ is a probability measure of a measurable space $(X, \mathcal{X})$.
- Suppose that ν is a probability kernels from $(X, \mathcal{X})$ to a measurable space $(Y, \mathcal{Y})$.
- Define the inner product $\mu\nu$ as a probability measure on Y:

$$(\mu\nu)(A) = \int_X \mu(\mathrm{d}x)\nu(x, A), \quad A \in \mathcal{Y}.$$

- Define the outer product $\mu \otimes \nu$ as the unique probability measure on $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$ such that

$$(\mu \otimes \nu)(A, B) = \int_A \mu(\mathrm{d}x)\nu(x, B), \quad A \in \mathcal{X}, B \in \mathcal{Y}.$$

Conditional distribution

- Let η be a random element taking values in a measurable space $(T, \mathcal{T})$ defined on a probability space, say $(\Omega, \mathcal{A}, \mathbb{P})$.
- Let $\mathcal{F}$ be a sub σ-field.
- The conditional distribution $\mathcal{L}(\eta|\mathcal{F})$ is defined as a probability kernels from $(\Omega, \mathcal{A})$ to $(T, \mathcal{T})$ such that for every $A \in \mathcal{T}$,

$$\mathcal{L}(\eta|\mathcal{F})(\cdot, A) = \mathbb{P}(\eta \in A|\mathcal{F}) = \mathbb{E}[\mathbf{1}_{\{\eta \in A\}}|\mathcal{F}], \quad a.s.$$

- Warning: the existence of conditional distribution is not guaranteed!
- If $\mathcal{F}$ is generated by another random element ξ, then we define $\mathcal{L}(\eta|\xi) = \mathcal{L}(\eta|\mathcal{F})$.

Disintegration

Theorem 8.5

Let ξ, η be random elements in S and T. Let T be a Polish space. Then $\mathcal{L}(\xi, \eta) = \mathcal{L}(\xi) \otimes \mu$ for a probability kernel μ from S to T. Furthermore,

$$\mathcal{L}(\eta|\xi) = \mu(\xi, \cdot), \quad a.s.$$

- This gives the existence of conditional distribution.
- A key condition is that η takes values in a Polish space.

Existence of Markov processes

Theorem 11.4

Consider a time scale $T \subset \mathbb{R}$ starting at 0, a Polish space S, a probability measure ν on S, and a family of probability kernels $\{\mu_{s,t} | s \leq t, s, t \in T\}$ on S satisfying Chapman–Kolmogorov equations. Then there exists an S-valued Markov process $(X_t)_{t \in T}$ with intial distribution ν and transition kernels $\mu_{s,t}$.

- Chapman–Kolmogorov equations?
- Markov process?
- Initial distribution?
- Transition kernels?

Chapman–Kolmogorov equations

- Suppose that $T \subset \mathbb{R}$ and $0 \in T$.
- Let $\{\mu_{s,t} | s < t, s, t \in T\}$ be a family of probability kernels on a measurable space S.
- We say Chapman-Kolmogorov equation holds for $\{\mu_{s,t} | s < t, s, t \in T\}$, if for every $s < u < t$ in T,

$$\mu_{s,u}\mu_{u,t} = \mu_{s,t}.$$

Conditionally independent

- Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space.
- We say a family of sub σ-fields $\{\mathcal{F}_t | t \in T\}$ are independent given a sub σ-field $\mathcal{G}$, if for every distinct $t_1, t_2, \dots, t_n$ in T and $F_1 \in \mathcal{F}_{t_1}, F_2 \in \mathcal{F}_{t_2}, \dots, F_n \in \mathcal{F}_{t_n}$, it holds that almost surely

$$\mathbb{P}(F_1 \cap F_2 \cap \cdots \cap F_n | \mathcal{G}) = \mathbb{P}(F_1|\mathcal{G})\mathbb{P}(F_2|\mathcal{G}) \dots \mathbb{P}(F_n|\mathcal{G}).$$

- Notationally, if sub σ-fields $\mathcal{F}$ and $\mathcal{H}$ are conditionally independent given $\mathcal{G}$, then we write

$$\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{H}.$$

Markov process

- Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space.
- Let time scale $T \subset \mathbb{R}$ and $0 \in T$.
- Suppose that $(\mathcal{F}_t)_{t \in T}$ is a family of non-decreasing sub σ-fields, a.k.a. a filtration indexed by T.
- Let S be a Polish space.
- Let $(X_t)_{t \in T}$ be an S-valued stochastic process, a.k.a. a family of S-valued random elements.
- We say $(X_t)_{t \in T}$ is a Markov process w.r.t. filtration $(\mathcal{F}_t)_{t \in T}$ if for every $t \in T$, $\sigma(X_t) \in \mathcal{F}_t$ and

$$\mathcal{F}_t \perp\!\!\!\perp_{\sigma(X_t)} \sigma(X_u : u \geq t).$$

Initial distribution and the transition kernels

- Let $T \subset [0, \infty)$ and $0 \in T$.
- Let S be a Polish space.
- Let $(X_t)_{t \in T}$ be an S-valued Markov process w.r.t. a filtration $(\mathcal{F}_t)_{t \in T}$ of a probability sapce $(\Omega, \mathcal{A}, \mathbb{P})$.
- We say a probability measure ν on S is the initial distribution of $(X_t)_{t \in T}$, if it is the distribution of X_0.
- We say a family of probability kernels $\{\mu_{s,t} | s < t \text{ in } T\}$ on S are the transition kernels of $(X_t)_{t \in T}$, if for every $s < t$ in T and $B \in \mathcal{S}$, almost surely
$$\mathbb{P}(X_t \in B | \mathcal{F}_s) = \mu_{s,t}(X_s, B).$$
- Since S is Polish, transition kernels of a given Markov process always exists.

Existence of Markov processes

Theorem 11.4

Consider a time scale $T \subset \mathbb{R}$ starting at 0, a Polish space S, a probability measure ν on S, and a family of probability kernels $\{\mu_{s,t} | s \leq t, s, t \in T\}$ on S satisfying Chapman–Kolmogorov equations. Then there exists an S-valued Markov process $(X_t)_{t \in T}$ with intial distribution ν and transition kernels $\mu_{s,t}$.

- Starting point for further analysis of the properties of the Markov process.

Stationary and invariant

Lemma 11.11

Let $(X_t)_{t\geq 0}$ be a homogeneous Markov process in a Polish space S, with (homogeneous) transition kernels μ_t and initial distribution ν. Then

$$X \text{ is stationary} \iff \nu \text{ is invariant for } (\mu_t)_{t\geq 0}.$$

- Homogeneous Markov process and homogenous transition kernels?
- Stationary process?
- Invariant measure?

Homogeneous Markov process

- Let S be a Polish space.
- Suppose that $(X_t)_{t\geq 0}$ is an S-valued Markov process with transition kernels $\{\mu_{s,t}|s<t \text{ in } [0,\infty)\}$.
- Suppose that $\mu_{s,t}$ only depends on $t-s$.
- That is to say, there exists a family of probability kernels $\{\tilde{\mu}_t|t\geq 0\}$ of S, such that $\mu_{s,t}=\tilde{\mu}_{t-s}$ for every $0\leq s<t$.
- Then, we say $(X_t)_{t\geq 0}$ is a homogeneous Markov process with (homogeneous) transition kernels $(\tilde{\mu}_t)_{t\geq 0}$.
- It is known that $(\tilde{\mu}_t)_{t\geq 0}$ is a semigroup:

$$\tilde{\mu}_t\tilde{\mu}_s=\tilde{\mu}_{t+s}.$$

Stationary process

- We say a process $(X_t)_{t\geq 0}$ is stationary, if for every $0 \leq t_1 < t_2 < \cdots < t_n$ and $t \geq 0$, we have

$$\mathcal{L}(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) = \mathcal{L}(X_{t_1+t}, X_{t_2+t}, \ldots, X_{t_n+t}).$$

Invariant measure

- Let $(\mu_t)_{t\geq 0}$ is a family of probability kernels of a Polish space S.
- Suppose that $(\mu_t)_{t\geq 0}$ is a semigroup.
- We say a probability measure ν of S is invariant w.r.t. $(\mu_t)_{t\geq 0}$ if, for every $t \geq 0$, $\nu\mu_t = \nu$.

Stationary and invariant

Lemma 11.11

Let $(X_t)_{t\geq 0}$ be a homogeneous Markov process in a Polish space S, with (homogeneous) transition kernels μ_t and initial distribution ν. Then

$$X \text{ is stationary} \iff \nu \text{ is invariant for } (\mu_t)_{t\geq 0}.$$

- This result is foundational in the study of long-term behavior of Markov processes, particularly in:
  - Ergodic theory, where one studies conditions under which a process converges to a stationary distribution.
  - Statistical mechanics, where stationary distributions often describe equilibrium states

Strong Markov property

Theorem 11.14

Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration of a probability space. Let $(X_t)_{t\geq 0}$ be an $(\mathcal{F}_t)_{t\geq 0}$-adapted, càdlàg stochastic process taking values in a separable complete metric space (S, d_S). Let P be a probability kernel from S to $\mathbb{D}_S$. Then these conditions are equivalent:

1. X is strongly homogeneous at every bounded optional time τ w.r.t. kernel P.
2. X satisfies the strong Markov property at every optional time $\tau < \infty$ w.r.t. kernle P.

Similar result holds while replacing 'càdlàg' and '$\mathbb{D}_S$' by 'continuous' and 'C_S' respectively.

- Wiener space C_S? Skorokhod space $\mathbb{D}_S$?
- Strongly homogeneous?
- Strong Markov property?

Wiener space

- Let (S, d_S) be a seperable complete metric space.
- Denote by C_S the space of continuous paths in S indexed by $[0, \infty)$.
- In other word, $w \in C_S$ iff w is a continuous map from $[0, \infty)$ to S.
- For every x, y in C_S, define the pseudometrics

$$d_{C,n}(x, y) := \sup_{t \in [0,n]} d_S(x_t, y_t), \quad n \geq 0$$

  and metric

$$d_C(x, y) := \sum_{n=1}^{\infty} (d_{C,n}(x, y) \wedge 1)/2^n.$$

- C_S equipped with the topology generated by d_C is called the Wiener space.

Why do we need Wiener Space?

- Let $(X_t)_{t\geq 0}$ be a continuous process taking values in a separable complete metric space S, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.
- It can be verified that $X : \omega \mapsto (X_t(\omega))_{t\geq 0}$ is a measurable map from $(\Omega, \mathcal{A}, \mathbb{P})$ to C_S.
- In other word, X is a C_S-valued random element.
- With an abuse of notation, we do not distinguish the process $(X_t)_{t\geq 0}$ and X.
- Therefore, we can talk about the law of the process $(X_t)_{t\geq 0}$.
- Moreover, it is known that C_S is a Polish space.
- Therefore, we can talk about the conditional law of the process $(X_t)_{t\geq 0}$ given a sub σ-field $\mathcal{F}$.

Skorokhod space

- Let (S, d_S) be a seperable complete metric space.
- Denote by $\mathbb{D}_S$ the space of càdlàg paths in S indexed by $[0, \infty)$.
- In other word, $w \in \mathbb{D}_S$ iff $w : [0, \infty) \to S$,

$$w(t) = \lim_{r \downarrow t} w(r), \quad t \geq 0,$$

  and

$$\lim_{r \uparrow t} w(r) \text{ exist}, \quad t > 0.$$

- $\mathbb{D}_S$ equipped with the J_1-topology (defined in the next slide) is called the Skorokhod space.

The J_1-topology

- We say λ is a time change of $[0, n]$ if it is a strictly increasing continuous bijection from $[0, n]$ to itself.
- Let Λ_n be the collection of all time change of $[0, n]$.
- For any x and y in $\mathbb{D}_S$, define pseudometrics

$$d_{\mathbb{D},n}(x, y) = \inf_{\lambda \in \Lambda_n} \left(\sup_{t \in [0,n]} |\lambda(t) - t| + \sup_{t \in [0,n]} |x(\lambda(t)) - y(t)| \right)$$

  and metric

$$d_{\mathbb{D}}(x, y) = \sum_{n=1}^{\infty} (d_{\mathbb{D},n}(x, y) \wedge 1)/2^n.$$

- The topology generated by the metric $d_{\mathbb{D}}$ is called the J_1-topology of Skorokhod space.

Why do we need Skorokhod Space?

- Let $(X_t)_{t\geq 0}$ be a càdlàg process taking values in a separable complete metric space S, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.
- It can be verified that $X : \omega \mapsto (X_t(\omega))_{t\geq 0}$ is a measurable map from $(\Omega, \mathcal{A}, \mathbb{P})$ to $\mathbb{D}_S$.
- In other word, X is a $\mathbb{D}_S$-valued random element.
- With an abuse of notation, we do not distinguish the process $(X_t)_{t\geq 0}$ and X.
- Therefore, we can talk about the law of the process $(X_t)_{t\geq 0}$.
- Moreover, it is known that $\mathbb{D}_S$ is a Polish space.
- Therefore, we can talk about the conditional law of the process $(X_t)_{t\geq 0}$ given a sub σ-field $\mathcal{F}$.

Strong homogeneous Markov process

- Let S be a separable complete metric space.
- Let $\mathbb{D}_S$ be the Skorokhod space of S-valued càdlàg paths.
- For any $t \geq 0$ define the shift operator $\theta_t : \mathbb{D}_S \to \mathbb{D}_S$ such that

$$(\theta_t w)(s) = w(t+s), \quad w \in \mathbb{D}_S, s \geq 0.$$

- Suppose that $(X_t)_{t\geq 0}$ is an S-valued càdlàg Markov process.
- Suppose that there exists a probability kernels P from S to $\mathbb{D}_S$ such that, for every $t \geq 0$, almost surely

$$\mathcal{L}(\theta_t X | X_t) = P(X_t, \cdot).$$

- Then we say $(X_t)_{t\geq 0}$ is strong homogeneous.
- For càdlàg Markov processes, homogeneous and strong homogeneous are equivalent.

Strong homogeneous at an optional time

- Let $\mathbb{D}_S$ be the Skorokhod space w.r.t. a separable complete metric space S.
- Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ satisfying the usual hypothesis.
- Suppose that $(X_t)_{t\geq 0}$ is an S-valued adapted càdlàg process. (Not necessarily Markovian!)
- Let P be a probability kernel from S to $\mathbb{D}_S$.
- We say X is strong homogeneous at a optional time τ w.r.t. kernel P, if

$$\mathcal{L}(\theta_\tau X | X_\tau) = P(X_\tau, \cdot), \quad a.s.$$

Strongly Markov property at an optional time

- Let $\mathbb{D}_S$ be the Skorokhod space w.r.t. a separable complete metric space S.
- Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ satisfying the usual hypothesis.
- Suppose that $(X_t)_{t\geq 0}$ is an S-valued adapted càdlàg process. (Not necessarily Markovian!)
- Let P be a probability kernel from S to $\mathbb{D}_S$.
- We say X satisfies the strong Markov property at a optional time τ w.r.t. kernel P, if

$$\mathcal{F}_\tau \perp\!\!\!\perp_{\sigma(X_\tau)} \sigma(\theta_\tau X), \qquad \mathcal{L}(\theta_\tau X | X_\tau) = P_{X_\tau}, \quad a.s.$$

Strong Markov property

Theorem 11.14

Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration of a probability space. Let $(X_t)_{t\geq 0}$ be an $(\mathcal{F}_t)_{t\geq 0}$-adapted, càdlàg stochastic process taking values in a separable complete metric space (S, d_S). Let P be a probability kernel from S to $\mathbb{D}_S$. Then these conditions are equivalent:

1. X is strongly homogeneous at every bounded optional time τ w.r.t. kernel P.
2. X satisfies the strong Markov property at every optional time $\tau < \infty$ w.r.t. kernle P.

Similar result holds while replacing 'càdlàg' and '$\mathbb{D}_S$' by 'continuous' and 'C_S' respectively.

- This result is not obvious at all!

Thanks!