Topics in probability theory: Independence

Zhenyao Sun

Beijing Institute of Technology

September 11, 2024

Borel-Cantelli Lemma

Theorem 4.8

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. For any $A_1, A_2, \cdots \in \mathcal{A}$,

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \implies \mathbb{P}(A_n \text{ i.o.}) = 0,$$

and equivalence holds when $(A_n)_{n=1}^{\infty}$ are independent.

- Probability space?
- i.o.?
- $(A_n)_{n=1}^{\infty}$ independent?

Probability space

- A measure μ on a measurable space $(S, \mathcal{S})$ is called a probability measure if $\mu(S) = 1$.
- Probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is a measure space with $\mathbb{P}(\Omega) = 1$.

i.o.?

- i.o. means infinitely often.
- More precisely,

$$\{A_n \ i.o.\} := \bigcap_n \bigcup_{k \geq n} A_k.$$

Independence

- For a family of events $(A_t)_{t\in T}$ in $\mathcal{A}$, we say they are independent if for any distinct $t_1, \ldots, t_n \in T$,

$$\mathbb{P}(\bigcap_{k=1}^{n} A_{t_k}) = \prod_{k=1}^{n} \mathbb{P}(A_{t_k}).$$

Borel-Cantelli Lemma

Theorem 4.8

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. For any $A_1, A_2, \dots \in \mathcal{A}$,

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \implies \mathbb{P}(A_n \text{ i.o.}) = 0,$$

and equivalence holds when $(A_n)_{n=1}^{\infty}$ are independent.

- Borel-Cantelli Lemma is a basic tool in analyzing limiting theory in Probability.
- It is used in the proof of law of large numbers.

Sequence of independent random elements

Theorem 4.19

For any probability measures $\mu_1, \mu_2, \ldots$ on the Polish spaces $S_1, S_2, \ldots$, there exist some independent random elements $\xi_1, \xi_2, \ldots$ on a common probability space with distributions $\mu_1, \mu_2, \ldots$

- Random elements?
- Independent random elements?
- Distribution of a given random element?

Random elements

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
- Let $(S, \mathcal{S})$ be a measurable space. If S is a topological space, then we require $\mathcal{S} = \mathcal{B}_S$.
- If ξ is a measurable map from $(\Omega, \mathcal{F})$ to $(S, \mathcal{S})$, then we say ξ is an S-valued random element.
- In particular, if $S = \mathbb{R}$, then ξ is called a random variable.

Distribution of a random element

- Let ξ be a random element from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to a measurable space $(S, \mathcal{S})$.
- Recall the definition of the measurable map: for every $A \in \mathcal{S}$,

$$\xi^{-1}(A) := \{\omega \in \Omega : \xi(\omega) \in A\} \in \mathcal{F}.$$

- For every $A \in \mathcal{S}$, define $\mu(A) = \mathbb{P}(\xi^{-1}(A))$.
- It can be verified that μ is a probability measure on $(S, \mathcal{S})$.
- We say μ is the distribution of ξ.

σ-field generated by a random element

- Let ξ be a random element from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to a measurable space $(S, \mathcal{S})$.
- Define $\sigma(\xi) := \{\xi^{-1}(A) : A \in \mathcal{S}\}$.
- It can be verified that $\sigma(\xi)$ is a σ-field of the space Ω.
- We say $\sigma(\xi)$ is the σ-field generated by ξ.
- If $(\xi_\lambda)_{\lambda\in\Lambda}$ is a family of random elements. Then we say

$$\sigma(\xi_\lambda : \lambda \in \Lambda) = \sigma\left(\cup_{\lambda\in\Lambda}\sigma(\xi_\lambda)\right)$$

  is the σ-field generated by $(\xi_\lambda)_{\lambda\in\Lambda}$.

Independent random elements

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
- Let T be an arbitrary set (for example $T = \{1, 2, \dots, \}$), and for every $t \in T$, let X_t be a random element taking values in a measurable space $(S_t, \mathcal{S}_t)$.
- For every $t \in T$, let $\sigma(\xi_t)$ be the σ-field generated by the random element ξ_t.
- We say $(\xi_t)_{t \in T}$ are independent, if for every family of events $(B_t)_{t \in T}$ in $\mathcal{F}$:

$$B_t \in \sigma(X_t), \forall t \in T \implies (B_t)_{t \in T} \text{ are independent.}$$

Sequence of independent random elements

Theorem 4.19

For any probability measures $\mu_1, \mu_2, \ldots$ on the Polish spaces $S_1, S_2, \ldots$, there exist some independent random elements $\xi_1, \xi_2, \ldots$ on a common probability space with distributions $\mu_1, \mu_2, \ldots$

- This theorem allows us to always construct one more independent random element.
- This allows us to study the infinite sequence of random elements.
- The Polish space condition is important.

Strong Law of Large numbers

Theorem 5.23

Let $\xi, \xi_1, \xi_2, \ldots$ be i.i.d. random variables, put $S_n = \sum_{k=1}^{n} \xi_k$, and fix a $p \in (0, 2)$. Then $n^{-1/p} S_n$ converges a.s. iff these conditions hold, depending on the value of p:

- for $p \in (0, 1]$: $\xi \in L^p$,
- for $p \in (1, 2)$: $\xi \in L^p$ and $\mathbb{E}[\xi] = 0$.

In the case that $n^{-1/p} S_n$ converges, the limit equals to $\mathbb{E}[\xi]$ when $p = 1$ and is otherwise equals to 0.

- i.i.d.?
- Converges a.s.?
- The expecation $\mathbb{E}[\xi]$?
- L^p space?

i.i.d.

- i.i.d. means independent and identically distributed.
- Let $(\xi_i)_{i=1}^{\infty}$ be a sequence of random variable defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
- We say $(\xi_i)_{i=1}^{\infty}$ are i.i.d. random variable:
    - if $(\xi_i)_{i=1}^{\infty}$ are independent,
    - $\mu_{\xi_i}(A) = \mathbb{P}(\xi_i \in A) = \mathbb{P}(\xi_j \in A) = \mu_{\xi_j}(A)$ for every $i, j \in \mathbb{N}$ and $A \in \mathcal{B}_{\mathbb{R}}$.

Converges a.s.

- Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variable defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
- For any random variable Z on Ω, it can be verified that the following event is measurable:

$$\{\omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = Z(\omega)\}.$$

- We say $(X_n)_{n=1}^{\infty}$ convergence almost surely (a.s.) if there exists a random variable Z on Ω, such that

$$\mathbb{P}(\{\omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = Z(\omega)\}) = 1.$$

Expectation

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
- Let $\xi : \Omega \to \mathbb{R}$ be a random variable.
- Define $\mathbb{E}[\xi] = \int \xi(\omega)\mathbb{P}(\mathrm{d}\omega)$ if ξ is integrable w.r.t. the measure $\mathbb{P}$.
- If ξ is not integrable, then there are three possibilities:
  - $\mathbb{E}[\xi^+] < \infty$ but $\mathbb{E}[\xi^-] = \infty$. In this case, define $\mathbb{E}[\xi] = -\infty$.
  - $\mathbb{E}[\xi^+] = \infty$ but $\mathbb{E}[\xi^-] < \infty$. In this case, define $\mathbb{E}[\xi] = \infty$.
  - $\mathbb{E}[\xi^+] = \infty$ and $\mathbb{E}[\xi^-] = \infty$. In this case, we say the expectation of ξ does not exist.
- Here, $\xi^+ := \max\{\xi, 0\}$ and $\xi^- := (-\xi)^+$.
- (Question: Why not define the expectation of the symmetric Cauchy distribution as 0?)

L^p space

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
- Let $\xi : \Omega \to \mathbb{R}$ be a random variable.
- Let $p \in (0, \infty)$. Define

$$\|\xi\|_p := (\mathbb{E}[|\xi|^p])^{1/p}.$$

- Define $L^p := \{\text{random variable } \xi : \|\xi\|_p < \infty\}$.
- (Warning: when $p < 1$, L^p is not a Banach space.)

Strong Law of Large numbers

Theorem 5.23

Let $\xi, \xi_1, \xi_2, \dots$ be i.i.d. random variables, put $S_n = \sum_{k=1}^{n} \xi_k$, and fix a $p \in (0, 2)$. Then $n^{-1/p} S_n$ converges a.s. iff these conditions hold, depending on the value of p:

- for $p \in (0, 1]$: $\xi \in L^p$,
- for $p \in (1, 2)$: $\xi \in L^p$ and $\mathbb{E}[\xi] = 0$.

In the case that $n^{-1/p} S_n$ converges, the limit equals to $\mathbb{E}[\xi]$ when $p = 1$ and is otherwise equals to 0.

- Law of large number is the foundation of Statistics.
- When $p = 2$, the result does not hold.

Central limit theorem

Theorem 6.18

Let $\xi, \xi_1, \xi_2, \ldots$ be i.i.d., non-degenerate random variables, and let ζ be a standard normal distribution N(0,1). Then,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \xi_k \xrightarrow{d} \zeta \quad \Longleftrightarrow \quad \mathbb{E}[\xi] = 0, \mathbb{E}[\xi^2] = 1.$$

- non-degenerate?
- standard normal distribution?
- $\xrightarrow{d}$, convergence in distribution?

Non-degenerate random variable

- We say a random variable ξ is non-degenerate, if it is not a constant.
- In other word, $\mathbb{P}(\xi = c) < 1$ for any $c \in \mathbb{R}$.

Standard normal distribution

- We say a probability measure μ on $\mathbb{R}$ is a standard Gaussian distribution, if

$$\mu(\mathrm{d}x) = ce^{-\frac{1}{2}x^2}\mathrm{d}x$$

  where c is a constant such that μ is a probtbiliay measure.
- Actually, $c = \frac{1}{\sqrt{2\pi}}$.
- We say a random variable ζ is standard normal N(0,1) if its distribution is given by the standard Gaussian distribution.

Convergence in distribution

- Suppose that S is Polish space.
- Let μ and $\mu_1, \mu_2 \ldots$ be probability measures on S.
- We say μ_n converges weakly to μ if for every continuous bounded function f, $\mu_n(f)$ converges to $\mu(f)$.
- Let $\xi, \xi_1, \xi_2 \ldots$ be random elements taking values in S.
- We say ξ_n converges in distribution to ξ, if the distributions of ξ_n converges weakly to the distribution of ξ.
- If μ is the distribution of ξ, then it is known that $\mu(f) = \mathbb{E}[f(\xi)]$ for any bounded measurable f.
- Therefore, equivalently speaking, we say ξ_n converges in distribution to ξ, if for every continuous bounded function f on S, $\mathbb{E}[f(\xi_n)]$ converges to $\mathbb{E}[f(\xi)]$.

Central limit theorem

Theorem 6.18

Let $\xi, \xi_1, \xi_2, \ldots$ be i.i.d., non-degenerate random variables, and let ζ be a standard normal distribution N(0,1). Then,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \xi_k \xrightarrow{d} \zeta \quad \Longleftrightarrow \quad \mathbb{E}[\xi] = 0, \mathbb{E}[\xi^2] = 1.$$

- CLT is one of the cornerstone results in probability theory and statistics.
- It emphasis the importance of Gaussian/Normal distributions.

Domain of attraction of Gaussian

Theorem 6.18

Let $\xi, \xi_1, \xi_2, \ldots$ be i.i.d., non-degenerate random variables, and let ζ be a standard normal distribution N(0,1). Then these conditions are equivalent:

1. there exist some constants a_n and m_n, such that

$$a_n \sum_{k=1}^{n} (\xi_k - m_n) \xrightarrow{d} \zeta,$$

2. the function $L(x) = \mathbb{E}[\xi^2 \mathbf{1}_{|\xi| \leq x}]$ varies slowly at ∞.

- Slowly varying function?

Slowly varying

- Let L be a non-decreasing and non-negative function on $[0, \infty)$.
- We say L is vary slowly at ∞, if $\sup_x L(x) > 0$ and for each $c > 0$,

$$\lim_{x \to \infty} \frac{L(cx)}{L(x)} = 1.$$

- In particular, if L is bounded then it is slowly varing.
- $L(x) = \log(x + 1)$ is also slowly varing.

Domain of attraction of Gaussian

Theorem 6.18

Let $\xi, \xi_1, \xi_2, \dots$ be i.i.d., non-degenerate random variables, and let ζ be a standard normal distribution N(0,1). Then these conditions are equivalent:

1. there exist some constants a_n and m_n, such that
$$a_n \sum_{k=1}^{n} (\xi_k - m_n) \xrightarrow{d} \zeta,$$
2. the function $L(x) = \mathbb{E}[\xi^2 \mathbf{1}_{|\xi| \le x}]$ varies slowly at ∞.

- Given (2), (1) holds with
$$m_n := \mathbb{E}[\xi]; \quad a_n := \frac{1}{1 \vee \sup\{x > 0 : nL(x) \ge x^2\}}.$$

Levy's continuity theorem

Theorem 6.23

Let $\mu_1, \mu_2, \ldots$ be probability measures on $\mathbb{R}^d$ with characteristics functions $\hat{\mu}_n$. Then these conditions are equivalent:

1. $\hat{\mu}_n(t) \to \varphi(t), t \in \mathbb{R}^d$ for a function φ which is continuous at 0,
2. $\mu_n \to \mu$ weakly for a probability measure μ on $\mathbb{R}^d$.

In that case, μ has characteristic function φ.

- Characteristic functions?

Characteristic function

- Let μ be a probability measure on $\mathbb{R}^d$.
- Define a function $\hat{\mu}$ on $\mathbb{R}^d$ such that

$$\hat{\mu}(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} \mu(\mathrm{d}x)$$

  where

$$\langle t, x\rangle := \sum_{k=1}^{d} t_i x_i, \quad t, x \in \mathbb{R}^d.$$

- We call $\hat{\mu}$ the characteristic function of μ.
- As the name suggested, $\hat{\mu} = \hat{\nu}$ implies $\mu = \nu$.

Levy's continuity theorem

Theorem 6.23

Let $\mu_1, \mu_2, \dots$ be probability measures on $\mathbb{R}^d$ with characteristics functions $\hat{\mu}_n$. Then these conditions are equivalent:

1. $\hat{\mu}_n(t) \to \varphi(t), t \in \mathbb{R}^d$ for a function φ which is continuous at 0,
2. $\mu_n \to \mu$ weakly for a probability measure μ on $\mathbb{R}^d$.

In that case, μ has characteristic function φ.

- Levy's continuity theorem is a powerful tool while establishing CLT.
- The characteristic function of standard Gaussian μ is $\hat{\mu}(t) = e^{-\frac{1}{2}t^2}$.

Thanks!