Probability

Note: Probability Theory

2024

Lecturer:

Typed by: Zhuohua Shen

Contents

References: MAT3280, STAT5005 and *Probability: Theory and Examples*, 4th edition, by Richard Durrett, published by Cambridge University Press.

1 Preliminary

1.1 Riemann–Stieltjes integrals

<u>**Goal**</u>: Consider a closed interval $[a, b], a < b$. Let $\alpha(x)$ be a non-decreasing function on $[a, b]$. $f \in \text{bdd[a, b]}$. Want to define RS integral $\int_a^b f(x)\mathrm{d}\alpha(x)$.

Let $P : a = x_0 < x_1 < \cdots < x_n = b$ be a partition of $[a, b]$. Let $M_k(m_k) := \sup(\inf)\{f(x) : x_{k-1} \le x \le x_k\}$, t_k is chosen arbitrarily in $[x_{k-1}, x_k]$, $k = 1, 2, \dots, n$, we define

- upper sum $U(P,f,\alpha) := \sum_{k=1}^{n} M_k \cdot (\alpha(x_k) - \alpha(\alpha_{k-1}))$,
- lower sum $L(P,f,\alpha) := \sum_{k=1}^{n} m_k \cdot (\alpha(x_k) - \alpha(\alpha_{k-1}))$,
- Riemann–Stieltjes sum $S(P,f,\alpha) := \sum_{k=1}^{n} f(t_k) \cdot (\alpha(x_k) - \alpha(\alpha_{k-1}))$.

Sandwiching inequalities: $L(P,f,\alpha) \leq S(P,f,\alpha) \leq U(P,f,\alpha)$, $\forall P$.

Definition 1.1 (Riemann–Stieltjes integrable)**.** If $\lim_{\|P\|\to\infty} L(P,f,\alpha) = U(P,f,\alpha)$, then we say f is Riemann–Stieltjes (RS) integrable on $[a,b]$ w.r.t. $\alpha(x)$, $f \in \mathscr{R}(\alpha)$. We denote the common limit as

$$\int_a^b f\mathrm{d}\alpha \quad \text{or} \quad \int_a^b f(x)\mathrm{d}(\alpha(x)).$$

An improper RS integral is defined by the double limit (provided exists and finite, i.e., converges)

$$\int_{-\infty}^{\infty} f\mathrm{d}\alpha := \lim_{(a,b)\to(-\infty,\infty)} \int_a^b f\mathrm{d}\alpha.$$

Note that if $\beta(x) = \alpha(x) + C$, then $\int_a^b f d\alpha = \int_a^b f d\beta$. RS integral can be defined more generally when $\alpha(x)$ is BV. RS integrals have the following properties:

1. (**linearity**) $f, g \in \mathscr{R}(\alpha)$, $\forall c, d \in \mathbb{R}$, $cf + dg \in \mathscr{R}(\alpha)$, $\int_a^b (cf+dg) d\alpha = c\int_a^b f d\alpha + d\int_a^b g d\alpha$;
2. (**monotonicity**) $f, g \in \mathscr{R}(\alpha)$, if $f \leq g, \forall x \in [a,b]$, then $\int_a^b f d\alpha \leq \int_a^b g d\alpha$;
3. (**additivity of integration**) Let $a < c < b$, if $f \in \mathscr{R}(\alpha)$ on $[a,c]$ and on $[c,b]$, then $f \in \mathscr{R}(\alpha)$ on $[a,b]$, and $\int_a^b f d\alpha = \int_a^c f d\alpha + \int_c^b f d\alpha$.
4. (**linearity of $\boldsymbol{\alpha}$**) Suppose $\alpha_1(x)$ and $\alpha_2(x)$ are two non-decreasing functions, $f \in \mathscr{R}(\alpha_1), f \in \mathscr{R}(\alpha_2)$,, then $f \in \mathscr{R}(\alpha_1 + \alpha_2)$, and $\int_a^b f d(\alpha_1 + \alpha_2) = \int_a^b f d\alpha_1 + \int_a^b f d\alpha_2$.

Theorem 1.2 (Finite discontinuities)**.** $f \in \mathrm{bdd}[a,b]$, and there are finitely many points $c_1, c_2, \dots, c_n \in [a,b]$ such that f is discontinuous at each c_i. If α is continuous on $c_i, \forall i$, then $f \in \mathscr{R}(\alpha)$.

Theorem 1.3 (Continuous functions)**.** Suppose $f \in \mathrm{C}[a,b]$ and let $\alpha \in \mathrm{C}^1[a,b]$ (or $\alpha'(x)$ Riemann integrable on $[a,b]$) be nondecreasing. If $f \in \mathscr{R}(\alpha)$, then the function $f\alpha'$ is Riemann integrable on $[a,b]$ and we have

$$\int_a^b f\mathrm{d}\alpha = \int_a^b f(\alpha)\alpha'(x)\mathrm{d}x.$$

There are two techniques to compute RS integral:

1. (integration by parts) $\int_a^b f\mathrm{d}\alpha = f\alpha\big|_a^b - \int_a^b \alpha\mathrm{d}f$;
2. (change of variables) if g is strictly increasing on $[c,d]$, $f \in \mathscr{R}(\alpha)$ on $[g(c), g(d)]$, let $h = f \circ g$, $\beta = \alpha \circ g$, then $h \in \mathscr{R}(\beta)$ on $[c,d]$, with $\int_{g(c)}^{g(d)} f\mathrm{d}\alpha = \int_c^d h\mathrm{d}\beta$.

2 Measure Theory

2.1 Probability spaces

Definition 2.1 (Concepts)**.**

- Sample space Ω: a set.
- Field/algebra $\mathscr{F}'$: collection of subsets of Ω satisfying: (i)$\Omega \in \mathscr{F}'$, (ii) $A \in \mathscr{F}$ implies $A^c \in \mathscr{F}'$, (iii) $A, B \in \mathscr{F}'$ implies $A \cup B \in \mathscr{F}'$.
- σ-field/algebra $\mathscr{F}$: collection of subsets of Ω satisfying: (i) $\Omega \in \mathscr{F}$, (ii) $A \in \mathscr{F}$ implies $A^c \in \mathscr{F}$, (iii) $A_i \in \mathscr{F}$ implies $\cup_{i=1}^{\infty} A_i \in \mathscr{F}$.
- $(\Omega, \mathscr{F})$ is a measurable space.
  - A set in $\mathscr{F}$ is called $\mathscr{F}$-measurable, or simply measurable.
  - If $\mathscr{G}$ is a subcollection of $\mathscr{F}$, we say that $\mathscr{G}$ is a sub-σ-field of $\mathscr{F}$ if $\mathscr{G}$ is also a σ-field.
- Measure $\mu : \mathscr{F} \to \mathbb{R}$, if (i) $\mu(A) \geq 0$, (ii) $\mu(\emptyset) = 0$, (iii) $A_i \in \mathscr{F}$ disjoint, then $\mu(\biguplus_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i)$.
  - A measure μ is said to be finite if $\mu(\Omega)$ is finite.
  - If $\mu(\Omega) = 1$, then μ is a probability measure.

Let μ be a measure on $(\Omega, \mathscr{F})$, it has the following properties:

(a) (**monotonicity**) if $A \subset B$, then $\mu(A) \leq \mu(B)$;
(b) (**addition law**) $\forall A.B \in \mathscr{F}$, $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$;
(c) (**sub-additivity**) $\mu(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$;

(d) (**lower semi-continuity/continuity from below**) if $A_n \uparrow A = \cup_{n=1}^{\infty} A_n$, then $\mu(A_n) \uparrow \mu(A)$;

(e) (**upper semi-continuity/continuity from above**) if $A_n \downarrow A = \cap_{n=1}^{\infty} A_n$ and $\mu(A_1) < \infty$, then $\mu(A_n) \downarrow \mu(A)$.

Definition 2.2 (liminf and limsup)**.** Let $E_1, E_2, \dots$ be an arbitrary sequence of subsets in a set Ω. The limit inferior and limit superior of $(E_i)_{i\geq 1}$ are defined respectively as

$$\liminf_i E_i := \bigcup_{j=1}^{\infty} \bigcap_{k\geq j} E_k = \{E_i\ i.o.\}, \qquad \limsup_i E_i := \bigcap_{j=1}^{\infty} \bigcup_{k\geq j} E_k = \{E_i\ e.v.\}$$

We have $\liminf_i E_i \subseteq \limsup_i E_i$. If equality holds, we say the limit of $(E_i)_{i\geq 1}$ exists, and is defined as $\liminf_i E_i$ or $\limsup_i E_i$.

Definition 2.3 (generated σ-field)**.** Let $\mathscr{A}$ be a collection of subsets of Ω, let $\sigma(\mathscr{A})$ be the σ-field generated by $\mathscr{A}$, defined as

$$\sigma(\mathscr{A}) = \bigcap_{\mathscr{A}\subseteq\mathscr{F},\mathscr{F}\,\sigma-\text{field}} \mathscr{F}.$$

It is well-defined because if $\{\mathscr{F}_i : i \in I\}$ are all σ-fields, I can be uncountable, then $\cap_{i\in I}\mathscr{F}_i$ is also a σ-field. The $\sigma(\mathscr{A})$ is the smallest σ-field containing $\mathscr{A}$. If $\Omega = \mathbb{R}$, $\mathscr{A} = \{(a, b) : -\infty < a \leq b < \infty\}$, then $\sigma(\mathscr{A})$ is called the Borel field/algebra, written as $\mathscr{B}(\mathbb{R})$. Likewise, define $\mathscr{B}(\mathbb{R}^d)$ the σ-algebra generated by the open balls in $\mathbb{R}^d$. A set in $\mathscr{B}(\mathbb{R})$ is called a Borel set.

- $\mathscr{B}(\mathbb{R}) = \sigma(\{(a, b) : a < b\})$, which can be $\{[a, b]\}$, $\{(a, b]\}, \dots$
- $\mathscr{B}(\mathbb{R}^d)$ can also be generated by open d-dimensional open boxes in the form $(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d)$.

2.1.1 Measure extension theorem

Definition 2.4 (Concepts)**.** A pre-measure $\mu_0 : \mathscr{F}_0 \to [0, \infty]$ defined on a field $\mathscr{F}_0$ is a set function satisfying (i) $\mu_0(\emptyset) = 0$ and (ii) if $A_i \in \mathscr{F}_0$ mutually disjoint sets in Ω and $\biguplus_i A_i \in \mathscr{F}_0$, then $\mu_0(\biguplus_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu_0(A_i)$.
Let $\mathscr{F}$ be a σ-field containing $\mathscr{F}_0$. We say that $\mu : \mathscr{F} \to [0, \infty]$ is an extension of μ_0 if μ is a measure and satisfies $\mu(E) = \mu_0(E)$, $\forall E \in \mathscr{F}_0$.
μ_0 is said to be finite if $\mu_0(\Omega) < \infty$. It is said to be σ-finite if $\exists$ at most countably many sets $\Omega_i \in \mathscr{F}_0$ such that $\Omega = \cup_{i=1}^{\infty}\Omega_i$ and $\mu_0(\Omega_i) < \infty$ for all i.

We note that a probability measure is automatically σ-finite. WLOG, we may assume Ω_i's form a partition of Ω. If Ω_i's are not mutually disjoint, set $\Omega_i = \Omega_i \setminus (\cup_{j=1}^{i-1}\Omega_j)$.

Theorem 2.5 (Measure extension theorem/Hahn-Kolmogorov-Carathéodory)**.** Let $\mathscr{F}_0$ be a field on Ω and μ_0 be a premeasure on $\mathscr{F}_0$. There is a measure μ defined on $\sigma(\mathscr{F}_0)$ that extends to μ_0. Moreover, if μ_0 is σ-finite, then the extension is unique.

- (Counting measure) $\Omega = \mathbb{N}^+$, $\mathscr{F}_0$ is the collection of all finite and co-finite sets. $\mu_0(E) = \infty$ if E is infinite and $|E|$ if E is finite. $E = \{2, 4, 6, 8, \dots\} \notin \mathscr{F}_0$, $\mu_0(E)$ is undefined. But it is σ-finite by taking $\Omega_i = \{i\}$. μ_0 can be extended to the power set $2^{\mathbb{N}}$, the counting measure on $\mathbb{N}^+$.
- If $\Omega = \mathbb{R}$, $\mathscr{F}_0 = \{\text{finite union of } (a, b], a \in \mathbb{R} \cup \{-\infty\}, b \in \mathbb{R} \cup \{\infty\}\}$, then $\mathscr{F}_0$ is a field. Define μ_0 on $\mathscr{F}_0$ by $\mu_0(\biguplus_{i=1}^{n}(a_i, b_i]) = \sum_{i=1}^{n}(b_i - a_i)$ the total length of the set, if all a_i and $b_i < \infty$. Otherwise, the length is ∞. It is σ-finite by taking $\Omega_i = (-i, i]$. μ_0 can be extended to $\mathscr{B}(\mathbb{R})$, and the extension is unique, called the Borel measure on $\mathbb{R}$, denoted by λ.

2.1.2 Lebesgue–Stieltjes measures

Stieltjes measure function

Definition 2.6 (distribution function)**.** Let $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathbb{P})$ be a probability space. The distribution function induced by P is defined as

$$F(x) := \mathbb{P}((-\infty, x]), \quad x \in \mathbb{R}.$$

df vs cdf: cdf is for rv, while df is for a probability measure.

Theorem 2.7 (Properties of df)**.** The df $F(x)$ of a probability measure satisfies the following properties:

1. $F(x)$ is non-decreasing;
2. $F(x)$ is right-continuous;
3. $\lim_{x\to\infty} F(x) = 1$;
4. $\lim_{x\to-\infty} F(x) = 0$.

Definition 2.8 (Stieltjes measure function)**.** A function $F : \mathbb{R} \to \mathbb{R}$ is called a Stieltjes measure function if

1. F is non-decreasing;
2. F is right-continuous.

Lebesgue-Stieltjes measures Given a Stieltjes measure function, we can apply the measure extension theorem 2.5 to construct a measure on $\mathscr{B}(\mathbb{R})$. Intuitively, the Lebesgue–Stieltjes measure μ assigns a length to each Borel set on the real line, where the length of $(a, b]$ is given by $F(b) - F(a)$.

Theorem 2.9. Let F be a Stieltjes measure function defined on $\mathbb{R}$. Then $\exists$ a unique measure μ defined on $\mathscr{B}(\mathbb{R})$ such that $\mu((a, b]) = F(b) - F(a)$. The measure μ is called the Lebesgue-Stieltjes (LS) measure induced by F.

For example,

- (Lebesgue measure) take $F(x) = x$, then the LS measure on $\mathscr{B}(\mathbb{R})$ is Lebesgue measure, denoted by λ, $\lambda([a, b]) = b - a$. It is translation-invariant;
- (Uniform distribution) $F(x) = x\mathbf{1}_{\{0\le x\le 1\}} + \mathbf{1}_{\{1<x\}}$, then the LS measure μ is the uniform distribution on $[0, 1]$, $\mu((a, b]) = b - a$ if $0 < a < b < 1$;
- (Continuous distribution) Let $f(x) \ge 0$ be a pdf, i.e., Riemann-integrable on R, and $\int_{\mathbb{R}} f(x)\mathrm{d}x = 1$. Then $F(x) := \int_{-\infty}^{x} f(t)\mathrm{d}t$ is a Stieltjes measure function. Denote the LS measure by P, $P((a, b)) = \int_a^b f(t)\mathrm{d}t$, $a < b$;
- (Dirac measure, point mass) $F(x) = \mathbf{1}_{\{x\ge x_0\}}$. $P(A) = \mathbf{1}_{\{x_0\in A\}}$;
- (Discrete distribution) $i \in \mathbb{N}^+$, $p_i \ge 0$, $\sum_{i=1}^{\infty} p_i = 1$. Define $F(x) = \sum_{i=1}^{\infty} p_i\mathbf{1}_{\{x\ge x_i\}}$. Then LS measure $P(\{x\}) = \sum_{i=1}^{\infty} p_i\mathbf{1}_{\{x=x_i\}}$.

2.1.3 Null sets and complete measures

Consider a probability space $(\Omega, \mathscr{F}, \mu)$. If $\mu(E) = 0$, we hope $\forall E' \subseteq E$, $\mu(E') = 0$. But E' may not be measurable.

Definition 2.10. Given a measure space $(\Omega, \mathscr{F}, \mu)$, we define a set $N \subseteq \Omega$ as a null set if $\exists E \in \mathscr{F}$ s.t. $N \subseteq E$ and $\mu(E) = 0$. The measure space $(\Omega, \mathscr{F}, \mu)$ is said to be complete if all null sets are indeed $\mathscr{F}$-measurable.

Theorem 2.11 (Completion)**.** *Let $(\Omega, \mathscr{F}, \mu)$ be a measure space and $\mathscr{N}$ be the set of all null sets. We can define a new collection of sets*

$$\mathscr{F}' := \{A \cup N : A \in \mathscr{F},\ N \in \mathscr{N}\}.$$

We can extend the measure μ to a measure on $\mathscr{F}'$, denoted by μ', by

$$\mu'(A \cup N) := \mu(A),$$

and the extended measure space $(\Omega, \mathscr{F}', \mu')$ is complete.

By enlarging the σ-field in this way, we may assume that the measure is complete without loss of generality.

2.1.4 π-λ theorem and uniqueness of measure extension

We will use the term set system to refer to a collection of subsets in Ω.

Definition 2.12. We define a π-system in Ω as a set system $\mathscr{P}$ that satisfies

$$A, B \in \mathscr{P} \Rightarrow A \cap B \in \mathscr{P}.$$

A λ-system/Dynkin system in Ω is a set system $\mathscr{L}$ that satisfies:

(i) $\Omega \in \mathscr{L}$;
(ii) $A \in \mathscr{L} \Rightarrow \Omega \setminus A \in \mathscr{L}$;
(iii) If $A_i \in \mathscr{L}$, $A_i \uparrow A$, then $A \in \mathscr{L}$.

- (iii) is equivalent to: If $A_i \in \mathscr{L}$ mutually disjoint, then $\cup_i A_i \in \mathscr{L}$.
- If $\mathscr{A}$ is both π and λ-system, then $\mathscr{A}$ is a σ-field.

Theorem 2.13 (Dynkin's π-λ theorem)**.** If $\mathscr{P}$ is a π-system and $\mathscr{L}$ is a λ-system, and $\mathscr{P} \subset \mathscr{L}$, then $\sigma(\mathscr{P}) \subset \mathscr{L}$.

We can use the π-λ theorem to prove the uniqueness of measure extension.

Theorem 2.14 (Uniqueness of measure extension)**.** Suppose $\mathscr{F}_0$ is a field on a sample space Ω, and μ_1 and μ_2 are two measures on $\sigma(\mathscr{F}_0)$ such that $\mu_1(A) = \mu_2(A)$ for all $A \in \mathscr{F}_0$. Furthermore, suppose $\exists$ a sequence of disjoint sets $\Omega_i \in \mathscr{F}_0$, for $i = 1, 2, \ldots$, such that $\biguplus_i \Omega_i = \Omega$ and $\mu_1(\Omega_i) = \mu_2(\Omega_i) < \infty$ for all i. Then $\mu_1(B) = \mu_2(B)$ for all $B \in \sigma(\mathscr{F}_0)$.

As an application of the uniqueness result, we prove that a probability measure on $\mathbb{R}$ is uniquely determined by its Stieltjes measure function.

Theorem 2.15. Let P be a probability measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, and let $F(x)$ be its induced distribution function. If $\exists$ another probability Q on $\mathscr{B}(\mathbb{R})$ such that

$$P((-\infty, x]) = Q((-\infty, x]), \quad \forall x \in \mathbb{R},$$

then $P(B) = Q(B)$, $\forall B \in \mathscr{B}(\mathbb{R})$.

Multivariate version: On $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$, P is uniquely determined by multivariate distribution function $F(\boldsymbol{x})$.

2.2 Measurable functions

Given a function $f : \Omega \to \Omega'$, the inverse image of a set $A \subseteq \Omega'$ via the function f is defined as

$$f^{-1}(A) := \{x \in \Omega : f(x) \in A\}.$$

- $f^{-1}(A^c) = (f^{-1}(A))^c$.
- If $(A_i)_{i\in I}$ is any collection of sets in Ω', then $f^{-1}(\cap_{i\in I} A_i) = \bigcap_{i\in I} f^{-1}(A_i)$ and $f^{-1}(\cup_{i\in I} A_i) = \bigcup_{i\in I} f^{-1}(A_i)$.
- If $h = g \circ f$, then $h^{-1}(B) = f^{-1}(g^{-1}(B))$ for any subset B of the codomain of g.

Definition 2.16 (Measurable functions). Let $(\Omega, \mathscr{F})$ and $(\Omega', \mathscr{G})$ be two measurable spaces. A function $f : \Omega \to \Omega'$ is called $(\mathscr{F}, \mathscr{G})$-measurable if $\forall B \in \mathscr{G}$, $f^{-1}(B) \in \mathscr{F}$. When $\mathscr{G}$ is understood from the context, we say that f is $\mathscr{F}$-measurable. If both $\mathscr{F}$ and $\mathscr{G}$ are understood, we say that f is measurable.
When $(\Omega, \mathscr{F}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$, we will refer to an $\mathscr{F}$-measurable function as a Borel measurable function.
When $(\Omega', \mathscr{G}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$, a $(\mathscr{F}, \mathscr{B}(\mathbb{R}))$-measurable function $X : \Omega \to \mathbb{R}$ is called a (real-valued) measurable function. If $(\Omega, \mathscr{F}, \mu)$ is a probability space, we refer to X as a random variable. Furthermore, if $(\Omega', \mathscr{G}) = (\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$, then $\boldsymbol{X}$ is called a random vector.

Examples:
- If $\mathscr{F} = 2^\Omega$ or $\mathscr{G} = \{\emptyset, \Omega'\}$, then $\forall f : \Omega \to \Omega'$ is $(\mathscr{F}, \mathscr{G})$-measurable.
- If $\mathscr{F} = \{\emptyset, \Omega\}$, and $\mathscr{G}$ is the σ-field in which all singletons are $\mathscr{G}$-measurable, then an $(\mathscr{F}, \mathscr{G})$-measurable function is a constant function.
- Let $A \subseteq \Omega$. The indicator function $\mathbf{1}_A(\omega)$ is $(\mathscr{F}, \mathscr{B}(\mathbb{R}))$-measurable iff A is measurable.
  - Let $(\Omega, \mathscr{F}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$, since $\mathbb{Q}$ and Cantor set $C \in \mathscr{B}(\mathbb{R})$, $\mathbf{1}_\mathbb{Q}$ and $\mathbf{1}_C$ is Borel measurable. But Vitali set $V \notin \mathscr{B}(\mathbb{R})$, so $\mathbf{1}_V$ is not Borel measurable.

σ-field generated by f: $\sigma(f) := \{f^{-1}(B) : B \in \mathscr{G}\}$ is the smallest σ-field in Ω to make f measurable.
- If $f(\omega) = a \in \Omega'$, $\forall \omega \in \Omega$, then $\sigma(f) = \{\Omega, \emptyset\}$.
- If $f(\omega) = a\mathbf{1}_A + b\mathbf{1}_{A^c}$, $a \neq b$, then $\sigma(f) = \{\Omega, \emptyset, A, A^c\} = \sigma(A)$.

Induced σ-field: $\{B \subset \Omega' : f^{-1}(B) \in \mathscr{F}\}$ is the largest σ-field in Ω' to make f measurable.

Theorem 2.17 (Composition of measurable functions). Suppose $f : (\Omega, \mathscr{F}) \to (\Omega', \mathscr{G})$ is $(\mathscr{F}, \mathscr{G})$-measurable, and $g : (\Omega', \mathscr{G}) \to (\Omega'', \mathscr{H})$ is $(\mathscr{G}, \mathscr{H})$-measurable. Then the composed function $h = g \circ f$ is $(\mathscr{F}, \mathscr{H})$-measurable.

Theorem 2.18 (Check measurability). Let $f : (\Omega, \mathscr{F}) \to (\Omega', \mathscr{G})$ be a function, $\sigma(\mathscr{C}) = \mathscr{G}$. Then f is $(\mathscr{F}, \mathscr{G})$-measurable iff $f^{-1}(A) \in \mathscr{F}$, $\forall A \in \mathscr{C}$.

Theorem 2.19 (Measurability on $\mathbb{R}^m$). A continuous function $f : \mathbb{R}^m \to \mathbb{R}^n$ is $(\mathscr{B}(\mathbb{R}^m), \mathscr{B}(\mathbb{R}^n))$-measurable. In particular, a continuous real-valued function $f : \mathbb{R}^m \to \mathbb{R}$ is measurable.

Theorem 2.20 (Random vector). $\boldsymbol{X} = (X_1, \ldots, X_d)^T$ is a random vector iff X_i is a rv, $\forall i = 1, \ldots, d$.

$\sigma(X_1, \ldots, X_d)$: the smallest σ-field to make any X_i measurable, $\forall i = 1, \ldots, d$, which is $\sigma(\sigma(X_1), \ldots, \sigma(X_d))$; it is also the smallest σ-field to make $\boldsymbol{X} = (X_1, \ldots, X_d)^T$ measurable.
$\sigma(X_1, X_2, \ldots)$= $\sigma(\sigma(X_1), \sigma(X_2), \ldots)$ or $\sigma(\boldsymbol{X} : \Omega \to \mathbb{R}^\infty) = \sigma((-\infty, x_1] \times \cdots (-\infty, x_d],\ \boldsymbol{x} \in \mathbb{R}^d,\ d = 1, 2, \ldots,)$, that is, any finite collection of X_i is measurable.

2.2.1 Operations of measurable functions

If $X_1, \ldots, X_n$ are rvs, $f : (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n)) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is measurable, then $f(X_1, \ldots, X_n)$ is a rv by composition.

Theorem 2.21. If $f, g : \Omega \to \mathbb{R}$ are $\mathscr{F}$-measurable, then $f + g$, $f - g$, $c \cdot f$, and f/g are measurable, where c is a constant, and in f/g, we assume that $g(\omega) \neq 0$, $\forall \omega$.

Define the extended real line $\mathbb{R}^* = [-\infty, \infty]$, $\mathscr{B}(\mathbb{R}^*) = \sigma\{A, A \cup \{-\infty\}, A \cup \{\infty\}, A \cup \{\infty, -\infty\} : A \in \mathscr{B}(\mathbb{R})\} = \sigma((-\infty, x] : x \in \mathbb{R} \cup \{\infty\})$. By convention, $0 \cdot \infty = 0$ and $0 \cdot (-\infty) = 0$. If $f : \Omega \to \mathbb{R}^*$ is measurable, then it is called generalized random variable.

Theorem 2.22. Suppose $f_i : \Omega \to \mathbb{R}^*$ is measurable for $i \in \mathbb{N}^+$, then the functions

$$\inf_i f_i, \ \sup_i f_i, \ \liminf_i f_i, \ \limsup_i f_i, \ \lim_i f_i \text{ (if exists)}$$

are measurable, i.e., generalized rvs.

Let $\Omega_0 = \{\omega \in \Omega : \lim_n f_n(\omega) \text{ exists and finite}\} = \{\omega \in \Omega : \limsup_n f_n(\omega) - \liminf_n f_n(\omega) = 0\} \in \mathscr{F}$. If $P(\Omega_0) = 1$, then we say $\{f_n\}_{n=1}^{\infty}$ converges almost surely.

2.2.2 Random variables and distributions

Let $X : (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ be a rv. Then its induced measure on $\mathbb{R}$ is called the probability distribution of X, i.e., $\mathbb{P}(X \in B) := \mathbb{P}(X^{-1}(B))$, $\forall B \in \mathscr{B}(\mathbb{R})$. The distribution function (df) of X is defined to be $F_X(x) = \mathbb{P}(X \le x)$, $\forall x \in \mathbb{R}$. The properties in Theorem 2.7 hold.

- (Fact) If X has cdf F, F is cts, then $Y = F(X) \sim \text{Unif}(0, 1)$.

Theorem 2.23. Let $\Omega = (0, 1)$, $\mathscr{F} = \mathscr{B}((0, 1))$, and $\mathbb{P}$ is the Lebesgue measure on $(0, 1)$. Let F be an arbitrary distribution function. Define $X(\omega) = F^{-1}(\omega)$, $\omega \in (0, 1)$, where

$$F^{-1}(\omega) := \inf\{y \in \mathbb{R} : F(y) \ge \omega\} (= \sup\{y \in \mathbb{R} : F(y) < \omega\}).$$

Then X is regarded as a rv in Ω having df F.

It is useful to define several rvs on the same probability space, called coupling.

2.3 Statistical independence

2.4 Expectation

Lemma 2.24. *Let $X \ge 0$, $p > 0$, we have $\mathbb{E}X^p = \int_0^\infty p x^{p-1} \text{Unif}(X > x) \mathrm{d}x$.*

3 Law of Large Numbers

3.1 Almost Surely Convergence

This lemma gives an equivalent relation between expectation and sum of tail probability.

Lemma 3.1. *Let X_i iid and $\varepsilon > 0$, then $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > n\varepsilon) \le \varepsilon^{-1} \mathbb{E}|X_i| \le \sum_{n=0}^{\infty} \mathbb{P}(|X_n| > n\varepsilon)$.*

4 Central Limit Theorem

5 Random Walks

Random walk (RW): Let $\boldsymbol{X}_i$ be iid rvs in $\mathbb{R}^d$. Let $\boldsymbol{S}_n = \sum_{i=1}^n \boldsymbol{X}_i$. Then $\{\boldsymbol{S}_n : n \ge 1\}$ is called a RW. Take $\boldsymbol{S}_0 = \boldsymbol{0}$.
Simple random walk (SRW): If $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$, then $\{S_n\}$ is called a SRW in $\mathbb{R}^1$. If $\mathbb{P}(\boldsymbol{X}_i = (1, 1)) = \mathbb{P}(\boldsymbol{X}_i = (1, -1)) = \mathbb{P}(\boldsymbol{X}_i = (-1, 1)) = \mathbb{P}(\boldsymbol{X}_i = (-1, -1)) = 1/4$, then called a SRW in $\mathbb{R}^2$.

5.1 Stopping Times (A.1.1)

Long-term behavior of RW

Permutable (or exchangeable): An event that does not change under finite permutation of $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots\}$.

- All events in the tail σ-field $\mathcal{T}$ are permutable.
- $\{\omega : \boldsymbol{S}_n(\omega) \in B \text{ i.o.}\}$ is permutable but not tail event.
- $\{\omega : \limsup_{n\to\infty} \boldsymbol{S}_n(\omega)/c_n \ge 1\}$.

Theorem 5.1 (Hewitt-Savage 0-1 law)**.** *If $\boldsymbol{X}_i$ iid and event A is permutable, then $\mathbb{P}(A) = 0$ or 1.*

Theorem 5.2 (Long-term behavior of RW)**.** *For a RW in $\mathbb{R}$, one of the following has probability 1:*

(i) $S_n = 0$ *for all* n ;
(ii) $S_n \to \infty$ *as* $n \to \infty$;
(iii) $S_n \to -\infty$ *as* $n \to \infty$;

(iv) $-\infty = \liminf_n S_n < \limsup_n S_n = \infty$.

For two levels $a < b$, find the probability that RW reaches b before a

Filtration: Let $\boldsymbol{X}_i$ be a sequence of rvs, $\{\mathcal{F}_n := \sigma(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)\}_{n=1}^{\infty}$ as an increasing sequence of σ-fields, is called a filtration. We usually take $\mathcal{F}_0 = \{\phi, \Omega\}$.

Stopping time/optional random variable/optimal time/Markov time: $\tau \in \mathbb{N}^+ \cup \{\infty\}$ is a stopping time w.r.t. $\{\mathcal{F}_n\}$ if $\{\tau = n\} \in \mathcal{F}_n$, $\forall n \in \mathbb{N}^+$. (Equivalent def: $\{\tau \le n\} \in \mathcal{F}_n$ or $\{\tau \ge n+1\} \in \mathcal{F}_n$ for $n \in \mathbb{N}^+$)

- Constant $\tau = n$ is a stopping time.
- If τ_1, τ_2 are stopping time, then $\tau_1 \wedge \tau_2$, $\tau_1 \vee \tau_2$, $\tau_1 + \tau_2$ are stopping times.
- Hitting time of A: let A measurable, then $\tau = \inf\{n \ge 1 : \boldsymbol{S}_n \in A\}$ is a stopping time.
- σ-field $\mathcal{F}_N$=the information known at time N. Def: $\mathcal{F}_N$ is the collection of sets A that have $A \cup \{N = n\} \in \mathcal{F}_n$, $\forall n < \infty$. Example: $\{N \le n\} \in \mathcal{F}_N$, i.e., N is $\mathcal{F}_N$-measurable.

Theorem 5.3 (Wald's equation)**.** *Let X_i iid and τ be a stopping time.*

1. *(Wald's first equation) If $\mathbb{E}|X_1| < \infty$ and $\mathbb{E}\tau < \infty$, then $\mathbb{E}S_\tau = \mathbb{E}X_1\mathbb{E}\tau$.*
2. *(Wald's second equation) If $\mathbb{E}X_1 = 0$, $\mathbb{E}X_1^2 = \sigma^2 < \infty$, $\mathbb{E}\tau < \infty$, then $\mathbb{E}S_\tau^2 = \sigma^2\mathbb{E}\tau$.*

Example 5.4 (Results for 1-d SRW)**.** *For 1-d SRW, let $a, b \in \mathbb{Z}$, $a < 0 < b$. Let $N = \inf\{n : S_n \notin (a, b)\} = \inf\{n : S_n = a \text{ or } b\}$. Then*

1. *$\mathbb{E}N < \infty$,*
2. *$S_N = a$ or b,*
3. *$\mathbb{P}(S_N = a) = b/(b-a)$, $\mathbb{P}(S_N = b) = -a/(b-a)$,*
4. *$\mathbb{E}N = \mathbb{E}S_N^2 = (-a)b$.*

5.2 Recurrence vs. Transience (A.1.2)

When RW return to 0? We consider SRW on $\mathbb{R}^d$ and define its first, second, ..., nth returning time to the origin to be

$$\tau_1 = \inf\{m \ge 1 : \boldsymbol{S}_m = \boldsymbol{0}\},$$
$$\tau_n = \inf\{m > \tau_{n-1} : \boldsymbol{S}_m = \boldsymbol{0}\}.$$

Theorem 5.5. *For any RW, the following are equivalent:*

(i) $\mathbb{P}(\tau_1 < \infty) = 1$

(ii) $\mathbb{P}(\tau_n < \infty) = 1, \forall n = 1, 2, 3, \ldots$

(iii) $\mathbb{P}(\boldsymbol{S}_m = \boldsymbol{0} \text{ i.o.}) = 1$

(iv) $\sum_{m=1}^{\infty} \mathbb{P}(\boldsymbol{S}_m = \boldsymbol{0}) = \infty$.

- $\mathbb{P}(\tau_n < \infty) = (\mathbb{P}(\tau_1 < \infty))^n$.

Recurrent: If $\mathbb{P}(\tau_1 < \infty) = 1$, then the RW is called recurrent.

Transient: If $\mathbb{P}(\tau_1 < \infty) < 1$, then the RW is called Transient.

Theorem 5.6 (Recurrence of SRW)**.** *SRW is recurrent in $d \le 2$ and transient in $d \ge 3$.*

- We define the first time a random walk starting from $\boldsymbol{a}$ reaches $\boldsymbol{b}$: $\tau_{\boldsymbol{a}\to\boldsymbol{b}} := \inf\{m \ge 1 : \boldsymbol{a} + \boldsymbol{S}_m = \boldsymbol{b}\}$. It can be proved that $\mathbb{P}(\tau_1 < \infty) = 1$ iff $\mathbb{P}(\tau_{\boldsymbol{a}\to\boldsymbol{b}} < \infty) = 1$, $\forall \boldsymbol{a}, \boldsymbol{b}$.

5.3 Reflection Principle and Arcsine Distribution (A.1.3)

What is the distribution of the time spent above 0? We consider the SRW, $d = 1$, and think of the sequence $S_1, \ldots, S_n$ as being represented by a polygonal line with segments $(k-1, S_{k-1}) \to (k, S_k)$.

Theorem 5.7 (Reflection Principle)**.**

- *(Reflection principle for numbers) If $x, y > 0$, then the number of paths from $(0, x)$ to (n, y) that are 0 at some time is equal to the number of paths from $(0, -x)$ to (n, y).*
- *(Reflection principle for SRW) Let X_i be SRW with $d = 1$. Then $\forall b \in \mathbb{N}^+$,*

$$\mathbb{P}(\max_{1\le k\le n} S_k \ge b) = 2\mathbb{P}(S_n > b) + \mathbb{P}(S_n = b).$$

Theorem 5.8 (Hit 0 time)**.** $\mathbb{P}(S_1 > 0, \ldots, S_{2n} > 0) = \frac{1}{2}\mathbb{P}(S_{2n} = 0)$, *and* $\mathbb{P}(S_1 \ne 0, \ldots, S_{2n} \ne 0) = \mathbb{P}(S_{2n} = 0)$.

Arcsine distribution: a continuous distribution with density $\frac{1}{\pi\sqrt{x(1-x)}}$, $x \in (0,1)$. Define

$$L_{2n} := \sup\{m \le 2n : S_m = 0\}, \qquad \text{(last time at 0)}$$
$$F_n := \inf\{0 \le m \le n : S_m = \max_{0\le k\le n} S_k\}, \qquad \text{(first time at maximum)}$$

$\pi_{2n} :=$ number of $k : 1 \le k \le 2n$ such that the line $(k-1, S_{k-1}) \to (k, S_k)$ is above the x-axis.

Theorem 5.9 (Arcsine law). $\frac{L_{2n}}{2n}, \frac{F_n}{n}, \frac{\pi_{2n}}{2n}$ *all converge in distribution to the arcsine distribution.*

6 Martingales

6.1 Conditional expectation A.2.1

Definition 6.1 (Conditional expectation). X is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}|X| < \infty$. $\mathcal{A}$ is a σ-field and $\mathcal{A} \subset \mathcal{F}$. We define the conditional expectation of X given $\mathcal{A}$, $\mathbb{E}(X|\mathcal{A})$, to be any random variable Y satisfying

(i) Y is $\mathcal{A}$-measurable, and

(ii) $\forall A \in \mathcal{A}$, $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Y\mathbf{1}_A]$.

- For rvs X and Y, $\mathbb{E}(X|Y) := \mathbb{E}(X|\sigma(Y))$.
- For set A, $\mathbb{P}(A|\mathcal{A}) := \mathbb{E}(\mathbf{1}_A|\mathcal{A})$.
- For set A, B, $\mathbb{P}(A|B) := \mathbb{P}(A \cap B)/\mathbb{P}(B)$ given $\mathbb{P}(B) > 0$.

- If Y satisfies (i) and (ii), then $\mathbb{E}|Y| \le \mathbb{E}|X| < \infty$.
- $\mathbb{E}Y = \mathbb{E}X$, i.e., $\mathbb{E}[\mathbb{E}(X|\mathcal{A})] = \mathbb{E}X$.
- **Uniqueness**: If Y' also satisfies (i) and (ii), then $Y = Y'$ a.s. Any such Y is said to be a version of $\mathbb{E}(X|\mathcal{A})$.
- **Existence**: By Radon-Nikodym theorem.

Example 6.2.

- If X is $\mathcal{A}$-measurable, then $\mathbb{E}(X|\mathcal{A}) = X$. So a constant $c = \mathbb{E}(c|\mathcal{A})$. (Know X, the "best guess" is X)
- If $\sigma(X) \perp\!\!\!\perp \mathcal{A}$, then $\mathbb{E}(X|\mathcal{A}) = \mathbb{E}X$. (Don't know anything about X, the best guess is its mean)
- Suppose $\Omega_1, \Omega_2, \ldots$ is a finite or infinite partition of Ω into disjoint sets with positive probability, and let $\mathcal{A} = \sigma(\Omega_1, \ldots)$, then $\mathbb{E}(X|\mathcal{A}) = \mathbb{E}(X\mathbf{1}_{\Omega_i})/\mathbb{P}(\Omega_i)$ on Ω_i.
  - Let $\mathcal{A} = \{\emptyset, \Omega\}$, then $\mathbb{E}(X|\mathcal{A}) = \mathbb{E}X$.
- (Bayes's formula) Let $G \in \mathcal{G}$, then $\mathbb{P}(G|A) = \int_G \mathbb{P}(A|\mathcal{G})\mathrm{d}\mathbb{P} \Big/ \int_\Omega \mathbb{P}(A|\mathcal{G})\mathrm{d}\mathbb{P}$. When $\mathcal{G}$ is a σ-field generated by a partition, this reduces to the usual Bayes' formula $\mathbb{P}(G_i|A) = \mathbb{P}(A|G_i)P(G_i) \Big/ \sum_j \mathbb{P}(A|G_j)\mathbb{P}(G_j)$.

Theorem 6.3 (Properties). If $\mathbb{E}|X|, \mathbb{E}|X_n|, \mathbb{E}|Y| < \infty$, then

(a) (linearity) $\mathbb{E}(aX + Y|\mathcal{A}) = a\mathbb{E}(X|\mathcal{A}) + \mathbb{E}(Y|\mathcal{A})$.

(b) (monotonicity) If $X \le Y$, then $\mathbb{E}(X|\mathcal{A}) \le \mathbb{E}(Y|\mathcal{A})$.

(c) (MCT) If $X_n \ge 0$, $X_n \uparrow X$, and $\mathbb{E}X < \infty$, then $E(X_n|\mathcal{A}) \uparrow E(X|\mathcal{A})$.

(d) (Fatou) If $X_n \ge 0$, $\mathbb{E}X_n < \infty$, and $\mathbb{E}[\liminf_n X_n] < \infty$, then $\liminf_n \mathbb{E}(X_n|\mathcal{A}) \le \mathbb{E}[\liminf_n X_n|\mathcal{A}]$.

(e) (DCT) If $X_n \to X$ a.s., $|X_n| \le Y$, $\mathbb{E}|Y| < \infty$, then $\mathbb{E}(X_n|\mathcal{A}) \to \mathbb{E}(X|\mathcal{A})$.

(f) If X is $\mathcal{A}$-measurable, $\mathbb{E}|XY| < \infty$, then $\mathbb{E}(XY|\mathcal{A}) = X\mathbb{E}(Y|\mathcal{A})$.

(g) (Tower property) If $\mathcal{A}_1 \subset \mathcal{A}_2$, then $\mathbb{E}[\mathbb{E}(X|\mathcal{A}_1)|\mathcal{A}_2] = \mathbb{E}(X|\mathcal{A}_1)$, and $\mathbb{E}[\mathbb{E}(X|\mathcal{A}_2)|\mathcal{A}_1] = \mathbb{E}(X|\mathcal{A}_1)$.

Theorem 6.4 (Inequalities). If $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$, then

(i) (Jensen's inequality) If φ is convex, $\mathbb{E}|\varphi(X)| < \infty$, then $\varphi(\mathbb{E}[X|\mathcal{A}]) \le \mathbb{E}[\varphi(X)|\mathcal{A}]$.

(ii) (Markov's inequality) If $X \ge 0$, $a > 0$, then $\mathbb{P}(X \ge a|\mathcal{A}) \le a^{-1}\mathbb{E}(X|\mathcal{A})$.

(iii) (Chebyshev's inequality) If $a > 0$, then $\mathbb{P}(|X| \ge a|\mathcal{A}) \le a^{-2}\mathbb{E}(X^2|\mathcal{A})$.

(iv) (Hölder's Inequality) If $p \ge 1$, $p^{-1} + q^{-1} = 1$, and $\mathbb{E}|X|^p$, $\mathbb{E}|Y|^p < \infty$, then
$$|\mathbb{E}(XY|\mathcal{A})| \le \{\mathbb{E}(|X|^p\,|\mathcal{A})\}^{1/p}\{\mathbb{E}(|Y|^q\,|\mathcal{A})\}^{1/q}.$$

(v) (Minkowski inequality) If $p \ge 1$, $\mathbb{E}|X|^p$, $\mathbb{E}|Y|^p < \infty$, then
$$\{\mathbb{E}(|X+Y|^p\,|\mathcal{A})\}^{1/p} \le \{\mathbb{E}(|X|^p\,|\mathcal{A})\}^{1/p} + \{\mathbb{E}(|Y|^p\,|\mathcal{A})\}^{1/p}.$$

(vi) (Triangular inequality) If $\mathbb{E}X^2 < \infty$. Then for any $\mathcal{A}$-measurable Y with $\mathbb{E}Y^2 < \infty$, we have
$$\|X - \mathbb{E}(X|\mathcal{A})\|^2 \le \|X - Y\|^2.$$

By (vi), $\mathbb{E}(X|\mathcal{A})$ is the projection of X onto $\mathcal{L}^2(\mathcal{A})$, that is, $\mathbb{E}(X|\mathcal{A}) = \arg\min_Y \|X - Y\|^2$, for $\mathcal{A}$-measurable Y.

6.2 Martingales

Definition 6.5. Let $\{\mathcal{F}_n\}$ be a filtration, an increasing sequence of σ-fields. A sequence $\{S_n\}$ is said to be adapted to $\{\mathcal{F}_n\}$ if S_n is $\mathcal{F}_n$-measurable. $\{S_n\}$ is called a martingale w.r.t. $\{\mathcal{F}_n\}$ if
(i) $\mathbb{E}|S_n| < \infty$.
(ii) $\{S_n\}$ is adapted to $\{\mathcal{F}_n\}$.
(iii) $\mathbb{E}(S_n|\mathcal{F}_{n-1}) = S_{n-1}$.
If in (iii), $\mathbb{E}(S_n|\mathcal{F}_{n-1}) \le S_{n-1}$ (or $\ge$), then $\{S_n\}$ is said to be a supermartingale (or submartingale).

Simple facts:
- If $\{S_n\}$ is a martingale, then $\mathbb{E}S_1 = \cdots = \mathbb{E}S_n = \cdots$, and $\mathbb{E}|S_1| \le \mathbb{E}|S_2| \le \cdots$.
- If $\{S_n\}$ is a supermartingale, then $\mathbb{E}S_1 \ge \mathbb{E}S_2 \ge \cdots$; if submartingale, $\mathbb{E}S_1 \le \mathbb{E}S_2 \le \cdots$.
- If $\{S_n\}$ is a supermartingale, then $\{-S_n\}$ is a submartingale, and vice versa.
- Let $n > m$, if $\{S_n\}$ is a
  - martingale $\Longrightarrow \mathbb{E}(S_n|\mathcal{F}_m) = S_m$
  - supermartingale $\Longrightarrow \mathbb{E}(S_n|\mathcal{F}_m) \le S_m$
  - submartingale $\Longrightarrow \mathbb{E}(S_n|\mathcal{F}_m) \ge S_m$

Theorem 6.6 (Martingale transforms).
(1) If $\{S_n\}$ is a martingale and φ is a convex (concave) function such that $\mathbb{E}|\varphi(S_n)| < \infty$, then $\varphi(S_n)$ is a submartingale (supermartingale).
(2) If S_n is a submartingale and φ is an increasing convex (concave) function such that $\mathbb{E}|\varphi(S_n)| < \infty$, then $\varphi(S_n)$ is a submartingale (supermartingale).

- From (1), if $p \ge 1$ and $\mathbb{E}|S_n|^p < \infty$, then $|S_n|^p$ is a submartingale.
- From (2), if S_n is a submartingale, then $(X_n - a)^+$ is a submartingale; if X_n is a supermartingale, then $X_n \wedge a$ is a supermartingale.

6.3 Martingale convergence

Predictable sequence: H_n, $n \ge 2$, which is $\mathcal{F}_{n-1}$-measurable. $(H \cdot S)_n = \sum_{m=1}^n H_m(S_m - S_{m-1})$.
- If $\{S_n\}$ is a supermartingale (submartingale), H_n $(n \ge 2)$ is predictable, $H_n \ge 0$, and H_n is bounded. Then $(H \cdot S)_n$ is a supermartingale (submartingale). For martingale, it is true without assuming $H_n \ge 0$.
- Let N be a stopping time, and $H_n = \mathbf{1}_{\{n \le N\}}$, then $S_{n\wedge N} = (H \cdot S)_n + S_0$ is a supermartingale/submartingale/-martingale as S_n is.

Let $a < b$, $N_0 = 1$, for $k \ge 1$, let $N_{2k-1} = \inf\{m > N_{2k-2} : S_m \le a\}$ and $N_{2k} = \inf\{m > N_{2k-1} : S_m \ge b\}$, all are stopping times. Let $H_m = \mathbf{1}\{N_{2k-1} < m < N_{2k} \text{ for some } k\}$ be the indicator of climbing. Let $U_n = \sup\{k : N_{2k} \le n\}$ be the number of upcrossings by time n. From the picture we have
(1) H_m is predictable;
(2) $(b-a)U_n \le \sum_{m=2}^n H_m(S_m - S_{m-1}) \Rightarrow (b-a)\mathbb{E}U_n \le \mathbb{E}(S_n - S_1)$.

Theorem 6.7 (Upcrossing inequality)**.** $\{S_n\}$ is a submartingale. $a < b$ are two constants. For U_n defined above,
$$\mathbb{E}U_n \le \frac{1}{b-a}\left[\mathbb{E}(S_n - a)^+ - \mathbb{E}(S_1 - a)^+\right]$$

Theorem 6.8 (Martingale convergence theorem). Suppose S_n is a submartingale and $\liminf \mathbb{E}S_n^+ < \infty$ (or $\sup \mathbb{E}S_n^+ < \infty$), then $S_n \to S$ a.s. with $\mathbb{E}|S| < \infty$.

Theorem 6.9. Suppose $\{S_n\}$ is a supermartingale. If $S_n \ge 0$, then $S_n \to S$ a.s. and $\mathbb{E}S \le \mathbb{E}S_1 < \infty$.

Corollary 6.10 (WLLN for martingales)**.** Suppose X_i are identically distributed and $\mathbb{E}|X_1| < \infty$. Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Let $S_1 = X_1$, and $S_n = S_{n-1} + X_n - \mathbb{E}(X_n|\mathcal{F}_{n-1})$, $n \ge 2$. Then $\{S_n\}$ is a martingale and $S_n/n \xrightarrow{\text{P}} 0$.

Example 6.11 (Branching processes)**.** Let $\xi_i^n \in \mathbb{N}$, $i, n \ge 1$ be iid (n: time, i: the ith parent). Define Z_n by $Z_0 = 1$, and
$$Z_{n+1} = \begin{cases} \xi_1^{n+1} + \cdots + \xi_{Z_n}^{n+1} & \text{if } Z_n > 0 \\ 0 & \text{if } Z_n = 0. \end{cases}$$

Z_n is called a Galton-Watson process. Let $\mathcal{F}_n = \sigma(\xi_i^m : i \geq 1, 1 \leq m \leq n)$ and $\mu = \mathbb{E}\xi_i^m \in (0, \infty)$, then $\{Z_n/\mu^n\}$ is a martingale w.r.t. $\mathcal{F}_n$.

- If $\mu < 1$, then $Z_n = 0$ for all n sufficiently large, so $Z_n/\mu^n \to 0$.
- If $\mu = 1$ and $\mathbb{P}(\xi_i^m = 1) < 1$, then $Z_n = 0$ for all n sufficiently large.

6.4 Doob's inequality; $\mathcal{L}^p$ convergence; CLT

Theorem 6.12 (Doob's inequality). $\{S_n\}$ is a submartingale w.r.t. $\{\mathcal{F}_n\}$. Then $\forall x > 0$,

$$\mathbb{P}\left(\max_{1\leq k\leq n} S_k \geq x\right) \leq \frac{1}{x}\mathbb{E}\left[S_n \mathbf{1}\{\max_{1\leq k\leq n} S_k \geq x\}\right] \leq \frac{\mathbb{E}S_n^+}{x}.$$

7 Techniques

7.1 Convergence

7.1.1 Convergence of random series

Let X_i be a sequence of rvs. $S_n = \sum_{i=1}^n X_i$. By 0-1 law, $\mathbb{P}(\lim_n S_n \text{ exists}) = 0$ or 1.

To show the convergence of random series:

1.

To show the divergence of random series:

1. If SLLN holds, $S_n/n \to \mu$ a.s., if $\mu > 0$, then $S_n \to \infty$ a.s.
2. Use S_n', $S_n' = S_n$ e.v. a.s., that is, $\mathbb{P}(S_n \neq S_n' \ i.o.) = 0$. And show $S_n' \to \infty$ a.s.

Next, we consider $S_n/f(n)$.

A Proofs

A.1 Proofs - 5

A.1.1 Proofs - 5.1

Proof of Theorem 5.2. By the 0-1 law 5.1, $\{\limsup_n S_n \geq c\}$ has probability 0 or 1, meaning that $\limsup_n S_n = c \in [-\infty, \infty]$ w.p.1. Since $S_n \overset{\text{d}}{=} S_{n+1} - X_1$, we have $c = c - X_1$.

(i) If $c \in \mathbb{R}$, then $X_1 \equiv 0$ a.s., so $S_n = 0$ for all n a.s.

If $X_1 \neq 0$ a.s., then $c = -\infty$ or ∞,

(ii) If $c = \infty$, and $\liminf_n S_n = \infty$, then case (ii);

(iii) If $c = -\infty$, and $\liminf_n S_n = -\infty$, then case (iii);

(iv) If $c = \infty$, and $\liminf_n S_n = -\infty$, then case (iv).

□

Proof of Theorem 5.3. Prove 1: First suppose $X_i \geq 0$. We have

$$\mathbb{E}S_\tau = \mathbb{E}\sum_{i=1}^{\tau} X_i = \mathbb{E}\sum_{i=1}^{\infty} X_i \mathbf{1}_{\{\tau\geq i\}} = \sum_{i=1}^{\infty} \mathbb{E}X_i \mathbf{1}_{\{\tau\geq i\}} = \sum_{i=1}^{\infty} \mathbb{E}X_i \mathbb{E}\mathbf{1}_{\{\tau\geq i\}} = \mathbb{E}X_1 \mathbb{E}\tau,$$

where the 3rd equality uses Fubini by $X_i \geq 0$, and the 4th uses $\{\tau \geq i\} \in \mathcal{F}_{i-1}$. For general case, since $\sum_{i=1}^\infty \mathbb{E}|X_i| \mathbf{1}_{\{\tau\geq i\}} = \sum_{i=1}^\infty \mathbb{E}|X_i| \mathbb{E}\mathbf{1}_{\{\tau\geq i\}} < \infty$, we can still use the Fubini.

Prove 2: If $\tau < n$, then $\tau \wedge n = \tau \wedge (n-1)$, so $S_{\tau\wedge n}^2 = S_{\tau\wedge(n-1)}^2$; if $\tau \geq n$, we have $\tau \wedge n = n$ and $\tau \wedge (n-1) = n-1$, so $S_{\tau\wedge n}^2 = S_n^2 = (S_{n-1} + X_n)^2 = (S_{\tau\wedge(n-1)} + X_n)^2$. Hence write

$$S_{\tau\wedge n}^2 = S_{\tau\wedge(n-1)}^2 + (2X_n S_{n-1} + X_n^2)\mathbf{1}_{\{\tau\geq n\}}.$$

Note that all the following expectations exist,

$$\begin{aligned}
\mathbb{E}S_{\tau\wedge n}^2 &= \mathbb{E}S_{\tau\wedge(n-1)}^2 + \mathbb{E}\left(2X_n S_{n-1} 1_{\{\tau\geq n\}}\right) + \mathbb{E}[X_n^2 1_{\{\tau\geq n\}}] \\
&= \mathbb{E}S_{\tau\wedge(n-1)}^2 + \sigma^2 \mathbb{P}(\tau \geq n) && \text{(stopping time, independence, and } \mathbb{E}X_i = 0) \\
&= \dots && \text{(reduce to } n-2, n-3, \dots) \\
&= \sigma^2 \sum_{i=1}^n \mathbb{P}(\tau \geq i).
\end{aligned}$$

In the 1st line, the expectation $\mathbb{E}X_n S_{n-1}$ exists since both rvs are in $\mathcal{L}^2$. By the last line, $\|S_{\tau\wedge n} - S_{\tau\wedge m}\|^2 = \sigma^2 \sum_{i=m+1}^n \mathbb{P}(\tau \geq i) \to 0$ as $n, m \to \infty$, $\{S_{\tau\wedge n}\}_n$ is a Cauchy sequence in $\mathcal{L}^2$, so letting $n \to \infty$ gives the result. □

Proof of Example 5.4. 1. For any positive integer k, by dividing the interval $(0, k(b-a))$ into k subintervals of equal length and considering an extreme case behavior (keep going upwards) of the random walk within each subinterval, we obtain

$$\begin{aligned}\mathbb{E}N = \sum_{i=0}^{\infty}\mathbb{P}(N > i) &\le (b-a)\sum_{k=0}^{\infty}\mathbb{P}(N > k(b-a)) \\ &\le (b-a)\sum_{k=0}^{\infty}\mathbb{P}((X_{(j-1)(b-a)+1},\ldots,X_{j(b-a)}) \neq (1,\ldots,1), j=1,\ldots,k) \\ &\le (b-a)\sum_{k=0}^{\infty}\left(1-\frac{1}{2^{b-a}}\right)^k < \infty.\end{aligned}$$

2. It is obvious.
3. By Wald's first equation 5.3, $0 = \mathbb{E}S_N = a\mathbb{P}(S_N = a) + b\mathbb{P}(S_N = b)$, we also have $1 = \mathbb{P}(S_N = a) + \mathbb{P}(S_N = b)$, so solve for the result.
4. By Wald's second equation 5.3 and $\sigma = 1$, we have $\mathbb{E}N = \mathbb{E}S_N^2 = a^2\mathbb{P}(S_N = a) + b^2\mathbb{P}(S_N = b)$, and use 3. □

A.1.2 Proofs - 5.2

Proof of Theorem 5.5. We have

$$\begin{aligned}\mathbb{P}(\tau_2 < \infty) &= \mathbb{P}(\tau_1 < \infty, \tau_2 - \tau_1 < \infty) \\ &= \sum_{m,n=1}^{\infty}\mathbb{P}(\tau_1 = m, \tau_2 - \tau_1 = n) \\ &= \sum_{m,n=1}^{\infty}\mathbb{P}(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_m = \mathbf{0}, \boldsymbol{X}_1 + \cdots + \boldsymbol{X}_u \neq \mathbf{0}, \forall 1 \le u < m; \\ &\qquad \boldsymbol{X}_{m+1} + \cdots + \boldsymbol{X}_{m+n} = \mathbf{0}, \boldsymbol{X}_{m+1} + \cdots + \boldsymbol{X}_{m+v} \neq \mathbf{0}, \forall 1 \le v < n) \\ &= \sum_{m,n=1}^{\infty}\mathbb{P}(\tau_1 = m)\mathbb{P}(\tau_1 = n) \qquad \text{(iid)} \\ &= (\mathbb{P}(\tau_1 < \infty))^2.\end{aligned}$$

Similarly, we can prove $\mathbb{P}(\tau_n < \infty) = (\mathbb{P}(\tau_1 < \infty))^n$. So (i) and (ii) are equivalent. They are equivalent to (iii) by examining their meanings. Finally,

$$\begin{aligned}\sum_{m=0}^{\infty}\mathbb{P}(\boldsymbol{S}_m = \mathbf{0}) = \sum_{m=0}^{\infty}\mathbb{E}\mathbf{1}_{\{\boldsymbol{S}_m=\mathbf{0}\}} = \mathbb{E}\sum_{m=0}^{\infty}\mathbf{1}_{\{\boldsymbol{S}_m=\mathbf{0}\}} &= \mathbb{E}\sum_{n=0}^{\infty}1_{\{\tau_n<\infty\}} \\ = \sum_{n=0}^{\infty}\mathbb{P}(\tau_n < \infty) = \sum_{n=0}^{\infty}\mathbb{P}(\tau_1 < \infty)^n &= \frac{1}{1-\mathbb{P}(\tau_1 < \infty)}.\end{aligned}$$

So (i) and (iv) are equivalent. □

Proof of Theorem 5.6. In $d = 1$, use (iv) in Theorem 5.5 to show.

$$\begin{aligned}\sum_{m=1}^{\infty}P(S_m = 0) &= \sum_{n=1}^{\infty}P(S_{2n} = 0) && \text{(can only return to 0 at even steps)} \\ &= \sum_{n=1}^{\infty}\binom{2n}{n}(\frac{1}{2})^{2n} && \text{(combinatorics)} \\ &\sim \sum_{n=1}^{\infty}\frac{\sqrt{2\pi 2n}(\frac{2n}{e})^{2n}}{(\sqrt{2\pi n}(\frac{n}{e})^n)^2}\frac{1}{2^{2n}} && \text{(Stirling's formula)} \\ &= \sum_{n=1}^{\infty}\frac{1}{\sqrt{\pi n}} = \infty.\end{aligned}$$

In $d = 2$, note that in order for $S_{2n} = 0$, we must for some $0 \le m \le n$ have m up steps, m down steps, $n-m$ to the left, and $n-m$ to the right, so

$$\begin{aligned}\mathbb{P}(\boldsymbol{S}_{2n} = \mathbf{0}) &= \frac{1}{4^{2n}}\sum_{m=0}^{n}\binom{2n}{m}\binom{2n-m}{m}\binom{2n-2m}{n-m} = \frac{1}{4^{2n}}\sum_{m=0}^{n}\frac{(2n)!}{m!m!(n-m)!(n-m)!} \\ &= \frac{1}{4^{2n}}\binom{2n}{n}\sum_{m=0}^{n}\binom{n}{m}\binom{n}{n-m} = 4^{-2n}\binom{2n}{n}^2 \asymp n^{-1}.\end{aligned}$$

by Stirling's formula. So its sum is ∞, still recurrent.

For $d = 3$, more complicated combinatorics give $\mathbb{P}(\boldsymbol{S}_{2n} = 0) \asymp \frac{1}{n^{3/2}}$, summing up to a finite number; hence transient. In even higher dimensions, the probabilities become even smaller; hence all transient. $\square$

A.1.3 Proofs - 5.3

Proof of Theorem 5.7. To show the first result, suppose $(0, s_0), (1, s_1), \ldots, (n, s_n)$ is a path from $(0, x)$ to $(0, y)$. Let $K = \inf\{k : s_k = 0\}$. Let $s'_k = -s_k$ for $k \leq K$ and $s'_k = s_k$ for $K \leq k \leq n$. Then (k, s'_k), $0 \leq k \leq n$, is a path from $(0, -x)$ to (n, y). Conversely, given a path from $(0, -x)$ to (n, y), we can also construct a reflected path. We have a one-toone correspondence between the two classes of paths, so their numbers must be equal.

Then, we have

$$\begin{aligned}
\mathbb{P}(\max_{1\leq k\leq n} S_k \geq b) &= P(\max_{1\leq k\leq n} S_k \geq b, S_n > b) + P(\max_{1\leq k\leq n} S_k \geq b, S_n < b) + P(\max_{1\leq k\leq n} S_k \geq b, S_n = b) \\
&= \mathbb{P}(S_n > b) + \mathbb{P}(\max_{1\leq k\leq n} S_k \geq b, S_n > b) + \mathbb{P}(S_n = b) \\
&= 2\mathbb{P}(S_n > b) + \mathbb{P}(S_n = b),
\end{aligned}$$

which prove the second result. $\square$

Proof of Theorem 5.8. To count the number of paths from $(0, 0)$ to (n, x), denote $a, b \in \mathbb{N}$ be the number of positive steps and b negative steps, respectively. $n = a + b$, and $x = a - b$, where $x \in [-n, n]$, and $n - x$ is even. The number of paths from $(0, 0)$ to (n, x) is $N_{n,x} = \binom{n}{a}$.

Since

$$\mathbb{P}(S_1 > 0, \ldots, S_{2n} > 0) = \sum_{r=1}^{\infty} \mathbb{P}(S_1 > 0, \ldots, S_{2n-1} > 0, S_{2n} = 2r).$$

Now we count the number of paths of $(1, 1) \to (2n, 2r)$, that are never 0. Since the total number of paths of $(1, 1) \to (2n, 2r)$ is $N_{2n-1,2r-1}$, the number of these paths touching 0 is the number of paths of $(1, -1) \to (2n, 2r)$, i.e., $N_{2n-1,2r+1}$, by reflection principle, we have the number of paths of $(1, 1) \to (2n, 2r)$ never touching 0 is $N_{2n-1,2r-1} - N_{2n-1,2r+1}$. Hence

$$\sum_{r=1}^{\infty} \mathbb{P}(S_1 > 0, \ldots, S_{2n-1} > 0, S_{2n} = 2r) = \sum_{r=1}^{\infty} \frac{1}{2} \frac{1}{2^{2n-1}} (N_{2n-1,2r-1} - N_{2n-1,2r+1}) = \frac{1}{2^{2n}} N_{2n-1,1},$$

where the $1/2$ in the 2nd term guarantees $S_1 > 0$. Since $\mathbb{P}(S_{2n} = 0) = \frac{1}{2^{2n}} N_{2n-1,-1} + \frac{1}{2^{2n}} N_{2n-1,1} = 2 \cdot \frac{1}{2^{2n}} N_{2n-1,1}$,

$$\mathbb{P}(S_1 > 0, \ldots, S_{2n} > 0) = \frac{1}{2^{2n}} N_{2n-1,1} = \frac{1}{2} \mathbb{P}(S_{2n} = 0).$$

Symmetry implies $\mathbb{P}(S_1 < 0, \ldots, S_{2n} < 0) = (1/2)\mathbb{P}(S_{2n} = 0)$. Then the proof is completed. $\square$

A.2 Proofs - 6

A.2.1 Proofs - 6.1

References