Probability

Note: Probability Theory

Nov 2024

Lecturer:

Typed by: Zhuohua Shen

Contents

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

1 Measure Theory

1.1 Expectation

Lemma 1.1. *Let $X \geq 0$, $p > 0$, we have $\mathbb{E}X^p = \int_0^\infty px^{p-1}\mathbb{P}(X > x)\mathrm{d}x$.*

2 Law of Large Numbers

2.1 Almost Surely Convergence

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

Lemma 2.1. *Let X_i iid and $\varepsilon > 0$, then $\sum_{n=1}^\infty \mathbb{P}(|X_n| > n\varepsilon) \leq \varepsilon^{-1}\mathbb{E}\,|X_i| \leq \sum_{n=0}^\infty \mathbb{P}(|X_n| > n\varepsilon)$.*

3 Central Limit Theorem

4 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 $\{\boldsymbol{S}_n : n \geq 1\}$ is called a RW. Take $\boldsymbol{S}_0 = \mathbf{0}$.
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 $\mathbb{R}^2$.

4.1 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.
- $\{\omega : \limsup_{n\to\infty} \boldsymbol{S}_n(\omega)/c_n \geq 1\}$.

Theorem 4.1 (Hewitt-Savage 0-1 law). *If $\boldsymbol{X}_i$ iid and event A is permutable, then $\mathbb{P}(A) = 0$ or 1.*

Theorem 4.2 (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: $\{\tau \leq n\} \in \mathcal{F}_n$ or $\{\tau \geq n+1\} \in \mathcal{F}_n$ 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 $\tau = \inf\{n \geq 1 : \boldsymbol{S}_n \in A\}$ 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: $\{N \leq n\} \in \mathcal{F}_N$, i.e., N is $\mathcal{F}_N$-measurable.

Theorem 4.3 (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 4.4 (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. $\mathbb{E}N = \mathbb{E}S_N^2 = (-a)b$.

4.2 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

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

Theorem 4.5. *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) $\mathbb{P}(\boldsymbol{S}_m = \mathbf{0} \text{ i.o.}) = 1$
(iv) $\sum_{m=1}^\infty \mathbb{P}(\boldsymbol{S}_m = \mathbf{0}) = \infty$.

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

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 4.6 (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 $\mathbb{P}(\tau_{\boldsymbol{a}\to\boldsymbol{b}} < \infty) = 1$, $\forall \boldsymbol{a}, \boldsymbol{b}$.

4.3 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 $S_1, \dots, S_n$ as being represented by a polygonal line with segments $(k-1, S_{k-1}) \to (k, S_k)$.

Theorem 4.7 (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 4.8 (Hit 0 time). $\mathbb{P}(S_1 > 0, \dots, S_{2n} > 0) = \frac{1}{2}\mathbb{P}(S_{2n} = 0)$*, and* $\mathbb{P}(S_1 \ne 0, \dots, S_{2n} \ne 0) = \mathbb{P}(S_{2n} = 0)$.

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

5 Martingales

5.1 Conditional expectation A.2.1

Definition 5.1 (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 5.2.
- 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 $\Omega_1, \Omega_2, \dots$ is a finite or infinite partition of Ω into disjoint sets with positive probability, and let $\mathcal{A} = \sigma(\Omega_1, \dots)$, 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 $\mathbb{P}(G|A) = \int_G \mathbb{P}(A|\mathcal{G})d\mathbb{P} \Big/ \int_\Omega \mathbb{P}(A|\mathcal{G})d\mathbb{P}$. 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 5.3 (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 5.4 (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.

5.2 Martingales

Definition 5.5. 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) $\{\boldsymbol{S}_n\}$ 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 $\implies \mathbb{E}(S_n|\mathcal{F}_m) = S_m$
  - supermartingale $\implies \mathbb{E}(S_n|\mathcal{F}_m) \le S_m$
  - submartingale $\implies \mathbb{E}(S_n|\mathcal{F}_m) \ge S_m$

Theorem 5.6 (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.

5.3 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 $k \geq 1$, let $N_{2k-1} = \inf\{m > N_{2k-2} : S_m \leq a\}$ and $N_{2k} = \inf\{m > N_{2k-1} : S_m \geq b\}$, 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 $U_n = \sup\{k : N_{2k} \leq n\}$ be the number of upcrossings by time n. From the picture we have

(1) H_m is predictable;

(2) $(b-a)U_n \leq \sum_{m=2}^n H_m(S_m - S_{m-1}) \Rightarrow (b-a)\mathbb{E}U_n \leq \mathbb{E}(S_n - S_1)$.

Theorem 5.7 (Upcrossing inequality)**.** $\{S_n\}$ is a submartingale. $a < b$ are two constants. For U_n defined above,

$$\mathbb{E}U_n \leq \frac{1}{b-a}\left[\mathbb{E}(S_n - a)^+ - \mathbb{E}(S_1 - a)^+\right]$$

Theorem 5.8 (**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 5.9. Suppose $\{S_n\}$ is a supermartingale. If $S_n \geq 0$, then $S_n \to S$ a.s. and $\mathbb{E}S \leq \mathbb{E}S_1 < \infty$.

Corollary 5.10 (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 $S_n = S_{n-1} + X_n - \mathbb{E}(X_n|\mathcal{F}_{n-1})$, $n \geq 2$. Then $\{S_n\}$ is a martingale and $S_n/n \xrightarrow{\text{p}} 0$.

Example 5.11 (Branching processes)**.** Let $\xi_i^n \in \mathbb{N}$, $i, n \geq 1$ 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.

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

Theorem 5.12 (**Doob's inequality**)**.** $\{S_n\}$ is a submartingale w.r.t. $\{\mathcal{F}_n\}$. Then $\forall x > 0$,

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

6 Techniques

6.1 Convergence

6.1.1 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 - 4

A.1.1 Proofs - 4.1

Proof of Theorem 4.2. By the 0-1 law 4.1, $\{\limsup_n S_n \geq c\}$ has probability 0 or 1, meaning that $\limsup_n S_n = c \in [-\infty, \infty]$ w.p.1. Since $S_n \stackrel{\text{d}}{=} S_{n+1} - X_1$, 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).

$\square$

Proof of Theorem 4.3. 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 $\sum_{i=1}^{\infty}\mathbb{E}\left|X_i\right|\mathbf{1}_{\{\tau\geq i\}} = \sum_{i=1}^{\infty}\mathbb{E}\left|X_i\right|\mathbb{E}\mathbf{1}_{\{\tau\geq i\}} < \infty$, 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

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

Note that all the following expectations exist,

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

In the 1st line, the expectation $\mathbb{E}X_nS_{n-1}$ exists since both rvs are in $\mathcal{L}^2$. By the last line, $\left\|S_{\tau\wedge n} - S_{\tau\wedge m}\right\|^2 = \sigma^2\sum_{i=m+1}^{n}\mathbb{P}(\tau\geq i) \to 0$ 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. $\square$

Proof of Example 4.4. 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

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

2. It is obvious.
3. By Wald's first equation 4.3, $0 = \mathbb{E}S_N = a\mathbb{P}(S_N = a) + b\mathbb{P}(S_N = b)$, 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 4.3 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. $\square$

A.1.2 Proofs - 4.2

Proof of Theorem 4.5. We have

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

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,

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

So (i) and (iv) are equivalent. □

Proof of Theorem 4.6. In $d = 1$, use (iv) in Theorem 4.5 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

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

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. □

A.1.3 Proofs - 4.3

Proof of Theorem 4.7. To show the first result, suppose $(0, s_0), (1, s_1), \dots, (n, s_n)$ is a path from $(0, x)$ to $(0, y)$. Let $K = \inf\{k : s_k = 0\}$. Let $s'_k = -s_k$ for $k \le K$ and $s'_k = s_k$ for $K \le k \le n$. Then (k, s'_k), $0 \le k \le n$, 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

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

which prove the second result. □

Proof of Theorem 4.8. 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

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

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

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

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 - 5

A.2.1 Proofs - 5.1

References