Math Tools

Note: Probability and measure

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 $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \dots\}$.

- 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, $\{\mathcal{F}_n := \sigma(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)\}_{n=1}^{\infty}$ 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

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

Theorem 4.5. *For any RW, the following are equivalent:*

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

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

(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 4.6 (Recurrence of SRW). *SRW is recurrent in $d \leq 2$ and transient in $d \geq 3$.*

- We define the first time a random walk starting from $\boldsymbol{a}$ reaches $\boldsymbol{b}$: $\tau_{\boldsymbol{a}\to\boldsymbol{b}} := \inf\{m \geq 1 : \boldsymbol{a} + \boldsymbol{S}_m = \boldsymbol{b}\}$. 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

$$L_{2n} := \sup\{m \le 2n : S_m = 0\}, \qquad \text{(last time at 0)}$$
$$F_n := \inf\{0 \le m \le n : S_m = \max_{0\le k\le n} S_k\}, \qquad \text{(first time at maximum)}$$
$$\pi_{2n} := \text{ number of } k : 1 \le k \le 2n \text{ such that the line } (k-1, S_{k-1}) \to (k, S_k) \text{ is above the } x\text{-axis.}$$

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 Techniques

5.1 Convergence

5.1.1 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.

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, $\{\limsup_n S_n \ge c\}$ has probability 0 or 1, meaning that $\limsup_n S_n = c \in [-\infty, \infty]$ w.p.1. Since $S_n \overset{\mathrm{d}}{=} S_{n+1} - X_1$, 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 $X_1 \ne 0$ 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 4.3. Prove 1: First suppose $X_i \ge 0$. We have

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

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

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

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

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, $\|S_{\tau\wedge n}-S_{\tau\wedge m}\|^2 = \sigma^2\sum_{i=m+1}^n \mathbb{P}(\tau\geq i)\to 0$ 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 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. □

A.1.2 Proofs - 4.2

Proof of Theorem 4.5. We have

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

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,

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

So (i) and (iv) are equivalent. □

Proof of Theorem 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

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

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 $\mathbb{P}(S_{2n} = 0) = \frac{1}{2^{2n}} N_{2n-1,-1} + \frac{1}{2^{2n}} N_{2n-1,1} = 2 \cdot \frac{1}{2^{2n}} N_{2n-1,1}$,

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

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

References