Modern Discrete Probability

III - Stopping times and martingales

Review

Sébastien Roch
UW–Madison
Mathematics

January 2, 2015

1. **Conditioning**

2. Stopping times
   - Definitions and examples
   - Some useful results
   - Application: Hitting times and cover times

3. Martingales
   - Definitions and examples
   - Some useful results
   - Application: critical percolation on trees

Conditioning I

Theorem (Conditional expectation)

Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ *and* $\mathcal{G} \subseteq \mathcal{F}$ *a sub* σ*-field. Then there exists a (a.s.) unique* $Y \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ *(note the* $\mathcal{G}$*-measurability) s.t.*

$$\mathbb{E}[Y; G] = \mathbb{E}[X; G], \ \forall G \in \mathcal{G}.$$

Such a Y *is called a version of the* conditional expectation of X given $\mathcal{G}$ *and is denoted by* $\mathbb{E}[X \mid \mathcal{G}]$.

Theorem (Conditional expectation: L^2 case)

Let $\langle U, V \rangle = \mathbb{E}[UV]$*. Let* $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ *and* $\mathcal{G} \subseteq \mathcal{F}$ *a sub* σ*-field. Then there exists a (a.s.) unique* $Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ *s.t.*

$$\|X - Y\|_2 = \inf\{\|X - W\|_2 \,:\, W \in L^2(\Omega, \mathcal{G}, \mathbb{P})\},$$

and, moreover, $\langle Z, X - Y \rangle = 0, \ \forall Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$.

Conditioning II

In addition to linearity and the usual inequalities (e.g. Jensen's inequality, etc.) and convergence theorems (e.g. dominated convergence, etc.). We highlight the following three properties:

Lemma (Taking out what is known)

If $Z \in \mathcal{G}$ is bounded then $\mathbb{E}[ZX \,|\, \mathcal{G}] = Z\,\mathbb{E}[X \,|\, \mathcal{G}]$.

Lemma (Role of independence)

If $\mathcal{H}$ is independent of $\sigma(\sigma(X), \mathcal{G})$, then $\mathbb{E}[X \,|\, \sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X \,|\, \mathcal{G}]$.

Lemma (Tower property (or law of total probability))

We have $\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}]] = \mathbb{E}[X]$. *In fact, if $\mathcal{H} \subseteq \mathcal{G}$ is a σ-field*

$$\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}] \,|\, \mathcal{H}] = \mathbb{E}[X \,|\, \mathcal{H}].$$

Filtrations I

Definition

A *filtered space* is a tuple $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{Z}_+}, \mathbb{P})$ where:

- $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space
- $(\mathcal{F}_t)_{t\in\mathbb{Z}_+}$ is a *filtration*, i.e.,

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_\infty := \sigma(\cup\mathcal{F}_t) \subseteq \mathcal{F}.$$

  where each $\mathcal{F}_t$ is a σ-field.

Example

Let $X_0, X_1, \ldots$ be i.i.d. random variables. Then a filtration is given by

$$\mathcal{F}_t = \sigma(X_0, \ldots, X_t), \ \forall t \geq 0.$$

Filtrations II

Fix $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{Z}_+}, \mathbb{P})$.

Definition (Adapted process)

A process $(W_t)_t$ is *adapted* if $W_t \in \mathcal{F}_t$ for all t.

Example (Continued)

Let $(S_t)_t$ where $S_t = \sum_{i\leq t} X_i$ is adapted.

Stopping times I

Definition

A random variable $\tau : \Omega \to \overline{\mathbb{Z}}_+ := \{0, 1, \ldots, +\infty\}$ is called a *stopping time* if

$$\{\tau \leq t\} \in \mathcal{F}_t, \ \forall t \in \overline{\mathbb{Z}}_+,$$

or, equivalently, $\{\tau = t\} \in \mathcal{F}_t, \ \forall t \in \overline{\mathbb{Z}}_+$. (To see the equivalence, note $\{\tau = t\} = \{\tau \leq t\} \setminus \{\tau \leq t-1\}$, and $\{\tau \leq t\} = \cup_{i \leq t}\{\tau = i\}$.)

Example

Let $(A_t)_{t \in \mathbb{Z}_+}$, with values in $(E, \mathcal{E})$, be adapted and $B \in \mathcal{E}$. Then

$$\tau = \inf\{t \geq 0 \ : \ A_t \in B\},$$

is a stopping time.

Stopping times II

Definition (The σ-field $\mathcal{F}_\tau$)

Let τ be a stopping time. Denote by $\mathcal{F}_\tau$ the set of all events F such that $\forall t \in \overline{\mathbb{Z}}_+$ $F \cap \{\tau = t\} \in \mathcal{F}_t$.

Lemma

$\mathcal{F}_\tau = \mathcal{F}_t$ if $\tau \equiv t$, $\mathcal{F}_\tau = \mathcal{F}_\infty$ if $\tau \equiv \infty$ and $\mathcal{F}_\tau \subseteq \mathcal{F}_\infty$ for any τ.

Lemma

If (X_t) is adapted and τ is a stopping time then $X_\tau \in \mathcal{F}_\tau$.

Lemma

If σ, τ are stopping times then $\mathcal{F}_{\sigma\wedge\tau} \subseteq \mathcal{F}_\tau$.

Examples

Let (X_t) be a Markov chain on a countable space V.

Example (Hitting time)

The *first visit time* and *first return time* to $x \in V$ are

$$\tau_x := \inf\{t \geq 0 \,:\, X_t = x\} \quad \text{and} \quad \tau_x^+ := \inf\{t \geq 1 \,:\, X_t = x\}.$$

Similarly, τ_B and τ_B^+ are the first visit and first return to $B \subseteq V$.

Example (Cover time)

Assume V is finite. The *cover time* of (X_t) is the first time that all states have been visited, i.e.,

$$\tau_{\mathrm{cov}} := \inf\{t \geq 0 \,:\, \{X_0, \ldots, X_t\} = V\}.$$

Strong Markov property

Let (X_t) be a Markov chain and let $\mathcal{F}_t = \sigma(X_0, \ldots, X_t)$. The Markov property extends to stopping times. Let τ be a stopping time with $\mathbb{P}[\tau < +\infty] > 0$ and let $f_t : V^\infty \to \mathbb{R}$ be a sequence of measurable functions, uniformly bounded in t and let $F_t(x) := \mathbb{E}_x[f_t((X_t)_{t\geq 0})]$, then (see [D, Thm 6.3.4]):

Theorem (Strong Markov property)

$$\mathbb{E}[f_\tau((X_{\tau+t})_{t\geq 0}) \,|\, \mathcal{F}_\tau] = F_\tau(X_\tau) \qquad \text{on } \{\tau < +\infty\}$$

Proof: Let $A \in \mathcal{F}_\tau$. Summing over the value of τ and using Markov

$$\mathbb{E}[f_\tau((X_{\tau+t})_{t\geq 0}); A \cap \{\tau < +\infty\}] = \sum_{s\geq 0} \mathbb{E}[f_s((X_{s+t})_{t\geq 0}); A \cap \{\tau = s\}]$$

$$= \sum_{s\geq 0} \mathbb{E}[F_s(X_s); A \cap \{\tau = s\}] = \mathbb{E}[F_\tau(X_\tau); A \cap \{\tau < +\infty\}].$$

■

Reflection principle I

Theorem

Let $X_1, X_2, \ldots$ be i.i.d. with a distribution symmetric about 0 and let $S_t = \sum_{i \leq t} X_i$. Then, for $b > 0$,

$$\mathbb{P}\left[\sup_{i \leq t} S_i \geq b\right] \leq 2\,\mathbb{P}[S_t \geq b].$$

Proof: Let $\tau := \inf\{i \leq t \,:\, S_i \geq b\}$. By the strong Markov property, on $\{\tau < t\}$, $S_t - S_\tau$ is independent on $\mathcal{F}_\tau$ and is symmetric about 0. In particular, it has probability at least $1/2$ of being greater or equal to 0 (which implies that S_t is greater or equal to b). Hence

$$\mathbb{P}[S_t \geq b] \geq \mathbb{P}[\tau = t] + \frac{1}{2}\mathbb{P}[\tau < t] \geq \frac{1}{2}\mathbb{P}[\tau \leq t].$$

■

Reflection principle II

Theorem

Let (S_t) be simple random walk on $\mathbb{Z}$. Then, $\forall a, b, t > 0$,

$$\mathbb{P}_0[S_t = b + a] = \mathbb{P}_0\left[S_t = b - a,\ \sup_{i \le t} S_i \ge b\right].$$

Theorem (Ballot theorem)

In an election with n voters, candidate A gets α votes and candidate B gets $\beta < \alpha$ votes. The probability that A leads B throughout the counting is $\frac{\alpha-\beta}{n}$.

Recurrence I

Let (X_t) be a Markov chain on a countable state space V. The *time of k-th return to y* is (letting $\tau_y^0 := 0$)

$$\tau_y^k := \inf\{t > \tau_y^{k-1} \,:\, X_t = y\}.$$

In particular, $\tau_y^1 \equiv \tau_y^+$. Define $\rho_{xy} := \mathbb{P}_x[\tau_y^+ < +\infty]$. Then by the strong Markov property

$$\mathbb{P}_x[\tau_y^k < +\infty] = \rho_{xy}\rho_{yy}^{k-1}.$$

Letting $N_y := \sum_{t>0} \mathbb{1}_{\{X_t = y\}}$, by linearity $\mathbb{E}_x[N_y] = \frac{\rho_{xy}}{1-\rho_{yy}}$. So either $\rho_{yy} < 1$ and $\mathbb{E}_y[N_y] < +\infty$ or $\rho_{yy} = 1$ and $\tau_y^k < +\infty$ a.s. for all k.

Recurrence II

Definition (Recurrent state)

A state x is *recurrent* if $\rho_{xx} = 1$. Otherwise it is *transient*. A chain is recurrent or transient if all its states are. If x is recurrent and $\mathbb{E}_x[\tau_x^+] < +\infty$, we say that x is *positive recurrent*.

Lemma: If x is recurrent and $\rho_{xy} > 0$ then y is recurrent and $\rho_{yx} = \rho_{xy} = 1$.

A subset $C \subseteq V$ is *closed* if $x \in C$ and $\rho_{xy} > 0$ implies $y \in C$. A subset $D \subseteq V$ is *irreducible* if $x, y \in D$ implies $\rho_{xy} > 0$.

Theorem (Decomposition theorem)

Let $R := \{x \,:\, \rho_{xx} = 1\}$ be the recurrent states of the chain. Then R can be written as a disjoint union $\cup_j R_j$ where each R_j is closed and irreducible.

Recurrence III

Theorem

Let x be a recurrent state. Then the following defines a stationary measure

$$\mu_x(y) := \mathbb{E}_x\left[\sum_{0\le t<\tau_x^+} \mathbb{1}_{\{X_t=y\}}\right].$$

Theorem

If (X_t) is irreducible and recurrent, then the stationary measure is unique up to a constant multiple.

Theorem

If (X_t) is irreducible and has a stationary distribution π, then $\pi(x) = \frac{1}{\mathbb{E}_x\tau_x^+}$.

Recurrence IV

Example (Simple random walk on $\mathbb{Z}$)

Consider simple random walk on $\mathbb{Z}$. The chain is clearly irreducible so it suffices to check the recurrence type of 0. First note the periodicity. So we look at S_{2t}. Then by Stirling

$$\mathbb{P}_0[S_{2t} = 0] = \binom{2t}{t} 2^{-2t} \sim 2^{-2t} \frac{(2t)^{2t}}{(t^t)^2} \frac{\sqrt{2t}}{\sqrt{2\pi t}} \sim \frac{1}{\sqrt{\pi t}}.$$

So

$$\mathbb{E}_0[N_0] = \sum_{t>0} \mathbb{P}_0[S_t = 0] = +\infty,$$

and the chain is recurrent.

A useful identity I

Theorem (Occupation measure identity)

Consider an irreducible Markov chain $(X_t)_t$ with transition matrix P and stationary distribution π. Let x be a state and σ be a stopping time such that $\mathbb{E}_x[\sigma] < +\infty$ and $\mathbb{P}_x[X_\sigma = x] = 1$. Denote by $\mathscr{G}_\sigma(x, y)$ the expected number of visits to y before σ when started at x (the so-called Green function*). For any y,*

$$\mathscr{G}_\sigma(x, y) = \pi_y \, \mathbb{E}_x[\sigma].$$

A useful identity II

Proof: By the uniqueness of the stationary distribution, it suffices to show that $\sum_y \mathscr{G}_\sigma(x,y)P(y,z) = \mathscr{G}_\sigma(x,z), \forall z$, and use the fact that $\sum_y \mathscr{G}_\sigma(x,y) = \mathbb{E}_x[\sigma]$. To check this, because $X_\sigma = X_0$,

$$\mathscr{G}_\sigma(x,z) = \mathbb{E}_x\left[\sum_{0\le t<\sigma} \mathbb{1}_{X_t=z}\right] = \mathbb{E}_x\left[\sum_{0\le t<\sigma} \mathbb{1}_{X_{t+1}=z}\right] = \sum_{t\ge 0} \mathbb{P}_x[X_{t+1}=z, \sigma > t].$$

Since $\{\sigma > t\} \in \mathcal{F}_t$, applying the Markov property we get

$$\begin{aligned}
\mathscr{G}_\sigma(x,z) &= \sum_{t\ge 0}\sum_y \mathbb{P}_x[X_t = y, X_{t+1} = z, \sigma > t] \\
&= \sum_{t\ge 0}\sum_y \mathbb{P}_x[X_{t+1} = z \,|\, X_t = y, \sigma > t]\, \mathbb{P}_x[X_t = y, \sigma > t] \\
&= \sum_{t\ge 0}\sum_y P(y,z)\, \mathbb{P}_x[X_t = y, \sigma > t]
\end{aligned}$$

■

A useful identity III

Here is a typical application of this lemma.

Corollary

In the setting of the previous lemma, for all $x \neq y$,

$$\mathbb{P}_x[\tau_y < \tau_x^+] = \frac{1}{\pi_x(\mathbb{E}_x[\tau_y] + \mathbb{E}_y[\tau_x])}.$$

Proof: Let σ be the time of the first visit to x after the first visit to x. Then $\mathbb{E}_x[\sigma] = \mathbb{E}_x[\tau_y] + \mathbb{E}_y[\tau_x] < +\infty$, where we used that the network is finite and connected. The number of visits to x before the first visit to y is geometric with success probability $\mathbb{P}_x[\tau_y < \tau_x^+]$. Moreover the number of visits to x after the first visit to y but before σ is 0 by definition. Hence $\mathscr{G}_\sigma(x, y)$ is the mean of the geometric, namely $1/\mathbb{P}_x[\tau_y < \tau_x^+]$. Applying the occupation measure identity gives the result. ■

Exponential tail of hitting times I

Theorem

Let (X_t) be a finite, irreducible Markov chain with state space V and initial distribution μ. For $A \subseteq V$, there is $\beta_1 > 0$ and $0 < \beta_2 < 1$ depending on A such that

$$\mathbb{P}_\mu[\tau_A > t] \leq \beta_1 \beta_2^t.$$

In particular, $\mathbb{E}_\mu[\tau_A] < +\infty$ for any μ, A.

Proof: For any integer m, for some distribution θ,

$$\mathbb{P}_\mu[\tau_A > ms \,|\, \tau_A > (m-1)s] = \mathbb{P}_\theta[\tau_A > s] \leq \max_x \mathbb{P}_x[\tau_A > s] =: 1 - \alpha_s.$$

Choose s large enough that, from any x, there is a path to A of length at most s of positive probability. In particular $\alpha_s > 0$. By induction, $\mathbb{P}_\mu[\tau_A > ms] \leq (1-\alpha_s)^m$ or $\mathbb{P}_\mu[\tau_A > t] \leq (1-\alpha_s)^{\lfloor \frac{t}{s} \rfloor} \leq \beta_1 \beta_2^t$ for $\beta_1 > 0$ and $0 < \beta_2 < 1$ depending on α_s. ∎

Exponential tail of hitting times II

A more precise bound:

Theorem

Let (X_t) be a finite, irreducible Markov chain with state space V and initial distribution μ. For $A \subseteq V$, let $\bar{\mathrm{t}}_A := \max_x \mathbb{E}_x[\tau_A]$. Then

$$\mathbb{P}_\mu[\tau_A > t] \leq \exp\left(-\left\lfloor \frac{t}{\lceil e\,\bar{\mathrm{t}}_A \rceil} \right\rfloor\right).$$

Proof: For any integer m, for some distribution θ,

$$\mathbb{P}_\mu[\tau_A > ms \,|\, \tau_A > (m-1)s] = \mathbb{P}_\theta[\tau_A > s] \leq \max_x \mathbb{P}_x[\tau_A > s] \leq \frac{\bar{\mathrm{t}}_A}{s},$$

by the Markov property and Markov's inequality. By induction, $\mathbb{P}_\mu[\tau_A > ms] \leq \left(\frac{\bar{\mathrm{t}}_A}{s}\right)^m$ or $\mathbb{P}_\mu[\tau_A > t] \leq \left(\frac{\bar{\mathrm{t}}_A}{s}\right)^{\lfloor \frac{t}{s} \rfloor}$. By differentiating w.r.t. s, it can be checked that a good choice is $s = \lceil e\,\bar{\mathrm{t}}_A \rceil$. ∎

Application to cover times

Let (X_t) be a finite, irreducible Markov chain on V with $n := |V| > 1$. Recall that the cover time is $\tau_{\mathrm{cov}} := \max_y \tau_y$. We bound the mean cover time in terms of $\bar{t}_{\mathrm{hit}} := \max_{x,y} \mathbb{E}_x \tau_y$.

Theorem

$$\max_x \mathbb{E}_x \tau_{\mathrm{cov}} \leq (3 + \ln n) \lceil e \, \bar{t}_{\mathrm{hit}} \rceil$$

Proof: By a union bound over all states to be visited and our previous tail bound,

$$\max_x \mathbb{P}_x[\tau_{\mathrm{cov}} > t] \leq \min\left\{1, n \cdot \exp\left(-\left\lfloor \frac{t}{\lceil e \, \bar{t}_{\mathrm{hit}} \rceil} \right\rfloor\right)\right\}.$$

Summing over t and appealing to the sum of a geometric series,

$$\max_x \mathbb{E}_x \tau_{\mathrm{cov}} \leq (\ln(n) + 1) \lceil e \, \bar{t}_{\mathrm{hit}} \rceil + \frac{1}{1 - e^{-1}} \lceil e \, \bar{t}_{\mathrm{hit}} \rceil.$$

Matthews' cover time bounds

Let $\underline{t}_{\mathrm{hit}}^{A} := \min_{x,y\in A,\ x\neq y} \mathbb{E}_x\tau_y$ and $h_n := \sum_{m=1}^{n} \frac{1}{m}$.

Theorem

$$\max_x \mathbb{E}_x\tau_{\mathrm{cov}} \leq h_n\,\overline{t}_{\mathrm{hit}} \qquad \min_x \mathbb{E}_x\tau_{\mathrm{cov}} \geq \max_{A\subseteq V} h_{|A|-1}\,\underline{t}_{\mathrm{hit}}^{A}$$

Proof: We prove the lower bound for $A = V$. The other cases are similar. Let $(J_1,\ldots,J_n)$ be a uniform random ordering of V, let $C_m := \max_{i\leq J_m}\tau_i$, and let L_m be the last state visited among $J_1,\ldots,J_m$. Then

$$\mathbb{E}[C_m - C_{m-1}\,|\,J_1,\ldots,J_m,\{X_t, t\leq C_{m-1}\}] = \mathbb{E}_{L_{m-1}}[\tau_{J_m}]\,\mathbb{1}_{\{L_m=J_m\}} \geq \underline{t}_{\mathrm{hit}}^{V}\,\mathbb{1}_{\{L_m=J_m\}}.$$

By symmetry, $\mathbb{P}[L_m = J_m] = \frac{1}{m}$. Moreover $\mathbb{E}_x C_1 \geq (1-\frac{1}{n})\underline{t}_{\mathrm{hit}}^{V}$. Taking expectations above and summing over m gives the result. ■

Better lower bounds can be obtained by applying this technique to subsets of V.

Martingales I

Definition

An adapted process $\{M_t\}_{t\geq 0}$ with $\mathbb{E}|M_t| < +\infty$ for all t is a *martingale* if

$$\mathbb{E}[M_{t+1} \,|\, \mathcal{F}_t] = M_t, \qquad \forall t \geq 0$$

If the equality is replaced with $\leq$ or $\geq$, we get a supermartingale or a submartingale respectively. We say that a martingale in *bounded in L^p* if $\sup_n \mathbb{E}[|X_n|^p] < +\infty$.

Example (Sums of i.i.d. random variables with mean 0)

Let $X_0, X_1, \ldots$ be i.i.d. centered random variables, $\mathcal{F}_t = \sigma(X_0, \ldots, X_t)$ and $S_t = \sum_{i\leq t} X_i$. Note that $\mathbb{E}|S_t| < \infty$ by the triangle inequality and

$$\mathbb{E}[S_t \,|\, \mathcal{F}_{t-1}] = \mathbb{E}[S_{t-1} + X_t \,|\, \mathcal{F}_{t-1}] = S_{t-1} + \mathbb{E}[X_t] = S_{t-1}.$$

Martingales II

Example (Variance of a sum)

Same setup as previous example with $\sigma^2 := \mathrm{Var}[X_1] < \infty$. Define $M_t = S_t^2 - t\sigma^2$. Note that $\mathbb{E}|M_t| \leq 2t\sigma^2 < +\infty$ and

$$\begin{aligned}
\mathbb{E}[M_t \,|\, \mathcal{F}_{t-1}] &= \mathbb{E}[(X_t + S_{t-1})^2 - t\sigma^2 \,|\, \mathcal{F}_{t-1}] \\
&= \mathbb{E}[X_t^2 + 2X_t S_{t-1} + S_{t-1}^2 - t\sigma^2 \,|\, \mathcal{F}_{t-1}] \\
&= \sigma^2 + 0 + S_{t-1}^2 - t\sigma^2 = M_{t-1}.
\end{aligned}$$

Example (Accumulating data: Doob's martingale)

Let X with $\mathbb{E}|X| < +\infty$. Define $M_t = \mathbb{E}[X \,|\, \mathcal{F}_t]$. Note that $\mathbb{E}|M_t| \leq \mathbb{E}|X| < +\infty$, and $\mathbb{E}[M_t \,|\, \mathcal{F}_{t-1}] = \mathbb{E}[X \,|\, \mathcal{F}_{t-1}] = M_{t-1}$, by the tower property.

Convergence theorem I

Theorem (Martingale convergence theorem)

Let (X_t) be a supermartingale bounded in L^1. Then (X_t) converges a.s. to a finite limit X_∞. Moreover, $\mathbb{E}|X_\infty| < +\infty$.

Corollary

If (X_t) is a nonnegative martingale then X_t converges a.s.

Proof: (X_t) is bounded in L^1 since

$$\mathbb{E}|X_t| = \mathbb{E}[X_t] = \mathbb{E}[X_0], \ \forall t.$$

■

Convergence theorem II

Example (Polya's Urn)

An urn contains 1 red ball and 1 green ball. At each time, we pick one ball and put it back with an extra ball of the same color. Let R_t (resp. G_t) be the number of red balls (resp. green balls) after the tth draw. Let $\mathcal{F}_t = \sigma(R_0, G_0, R_1, G_1, \ldots, R_t, G_t)$. Define M_t to be the fraction of green balls. Then

$$\begin{aligned}
\mathbb{E}[M_t \,|\, \mathcal{F}_{t-1}] &= \frac{R_{t-1}}{G_{t-1} + R_{t-1}} \frac{G_{t-1}}{G_{t-1} + R_{t-1} + 1} \\
&\quad + \frac{G_{t-1}}{G_{t-1} + R_{t-1}} \frac{G_{t-1} + 1}{G_{t-1} + R_{t-1} + 1} \\
&= \frac{G_{t-1}}{G_{t-1} + R_{t-1}} = M_{t-1}.
\end{aligned}$$

Since $M_t \geq 0$ and is a martingale, we have $M_t \to M_\infty$ a.s.

Maximal inequality I

Theorem (Doob's submartingale inequality)

Let (M_t) be a nonnegative submartingale. Then for $b > 0$

$$\mathbb{P}\left[\sup_{1\le i\le t} M_t \ge b\right] \le \frac{\mathbb{E}[M_t]}{b}.$$

(Markov's inequality implies only $\sup_{1\le i\le t} \mathbb{P}[M_i \ge b] \le \frac{\mathbb{E}[M_t]}{b}$.)

Proof: Divide $F = \{\sup_{1\le i\le t} M_t \ge b\}$ according to the first time M_i crosses b: $F = F_0 \cup \cdots \cup F_t$, where

$$F_i = \{M_0 < b\} \cap \cdots \cap \{M_{i-1} < b\} \cap \{M_i \ge b\}.$$

Since $F_i \in \mathcal{F}_i$ and $\mathbb{E}[M_t \,|\, \mathcal{F}_i] \ge M_i$,

$$b\,\mathbb{P}[F_i] \le \mathbb{E}[M_i; F_i] \le \mathbb{E}[M_t; F_i].$$

Sum over i.

Maximal inequality II

A useful consequence:

Corollary (Kolmogorov's inequality)

Let $X_1, X_2, \ldots$ be independent random variables with $\mathbb{E}[X_i] = 0$ and $\mathrm{Var}[X_i] < +\infty$. Define $S_t = \sum_{i \le t} X_i$. Then for $\beta > 0$

$$\mathbb{P}\left[\max_{i \le t} |S_i| \ge \beta\right] \le \frac{\mathrm{Var}[S_t]}{\beta^2}.$$

Proof: (S_t) is a martingale. By Jensen's inequality, (S_t^2) is a submartingale. The result follows Doob's submartingale inequality. ■

Orthogonality of increments

Lemma (Orthogonality of increments)

Let (M_t) be a martingale with $M_t \in L^2$. Let $s \leq t \leq u \leq v$. Then,

$$\langle M_t - M_s, M_v - M_u \rangle = 0.$$

Proof: Use $M_u = \mathbb{E}[M_v \,|\, \mathcal{F}_u]$, $M_t - M_s \in \mathcal{F}_u$ and apply the L^2 characterization of conditional expectations. ■

Optional stopping theorem I

Definition

Let $\{M_t\}$ be an adapted process and σ be a stopping time. Then

$$M_t^\sigma(\omega) := M_{\sigma(\omega)\wedge t}(\omega),$$

is (M_t) *stopped at* σ.

Theorem

Let (M_t) be a supermartingale and σ be a stopping time. Then the stopped process (M_t^σ) is a supermartingale and in particular

$$\mathbb{E}[M_{\sigma\wedge t}] \leq \mathbb{E}[M_0].$$

The same result holds with equality if (M_t) is a martingale.

Optional stopping theorem II

Theorem

Let (M_t) be a supermartingale and σ be a stopping time. Then M_σ is integrable and

$$\mathbb{E}[M_\sigma] \leq \mathbb{E}[M_0].$$

if one of the following holds:

1. *σ is bounded*
2. *(M_t) is uniformly bounded and σ is a.s. finite*
3. *$\mathbb{E}[\sigma] < +\infty$ and (M_t) has bounded increments (i.e., there $c > 0$ such that $|M_t - M_{t-1}| \leq c$ a.s. for all t)*
4. *(M_t) is nonnegative and σ is a.s. finite.*

The first three imply equality above if (M_t) is a martingale.

Wald's identities

For $X_1, X_2, \ldots \in \mathbb{R}$, let $S_t = \sum_{i=1}^t X_i$.

Theorem (Wald's first identity)

Let $X_1, X_2, \ldots \in L^1$ be i.i.d. with $\mathbb{E}[X_1] = \mu$ and let $\tau \in L^1$ be a stopping time. Then

$$\mathbb{E}[S_\tau] = \mathbb{E}[X_1]\mathbb{E}[\tau].$$

Theorem (Wald's second identity)

Let $X_1, X_2, \ldots \in L^2$ be i.i.d. with $\mathbb{E}[X_1] = 0$ and $\mathrm{Var}[X_1] = \sigma^2$ and let $\tau \in L^1$ be a stopping time. Then

$$\mathbb{E}[S_\tau^2] = \sigma^2\mathbb{E}[\tau].$$

Gambler's ruin I

Example (Gambler's ruin: unbiased case)

Let (S_t) be simple random walk on $\mathbb{Z}$ started at 0 and let $\tau = \tau_a \wedge \tau_b$ where $a < 0 < b$. We claim that 1) $\tau < +\infty$ a.s., 2) $\mathbb{P}[\tau_a < \tau_b] = \frac{b}{b-a}$, 3) $\mathbb{E}[\tau] = -ab$, and 4) $\tau_a < +\infty$ a.s. but $\mathbb{E}[\tau_a] = +\infty$.

1) We first argue that $\mathbb{E}\tau < \infty$. Since $(b-a)$ steps to the right necessarily take us out of (a, b),

$$\mathbb{P}[\tau > t(b-a)] \leq (1 - 2^{-(b-a)})^t,$$

by independence of the $(b-a)$-long stretches, so that

$$\mathbb{E}[\tau] = \sum_{k \geq 0} \mathbb{P}[\tau > k] \leq \sum_t (b-a)(1 - 2^{-(b-a)})^t < +\infty,$$

by monotonicity. In particular $\tau < +\infty$ a.s.

Gambler's ruin II

2) By Wald's first identity, $\mathbb{E}[S_\tau] = 0$ or

$$a\,\mathbb{P}[S_\tau = a] + b\,\mathbb{P}[S_\tau = b] = 0,$$

that is (taking $b \to \infty$ in the second expression)

$$\mathbb{P}[\tau_a < \tau_b] = \frac{b}{b-a} \quad \text{and} \quad \mathbb{P}[\tau_a < \infty] \geq \mathbb{P}[\tau_a < \tau_b] \to 1.$$

3) Wald's second identity says that $\mathbb{E}[S_\tau^2] = \mathbb{E}[\tau]$ (by $\sigma^2 = 1$). Also

$$\mathbb{E}[S_\tau^2] = \frac{b}{b-a}a^2 + \frac{-a}{b-a}b^2 = -ab,$$

so that $\mathbb{E}\tau = -ab$.

4) Taking $b \to +\infty$ above shows that $\mathbb{E}[\tau_a] = +\infty$ by monotone convergence. (Note that this case shows that the L^1 condition on the stopping time is necessary in Wald's second identity.)

Gambler's ruin III

Example (Gambler's ruin: biased case)

The *biased simple random walk on* $\mathbb{Z}$ with parameter $1/2 < p < 1$ is the process $\{S_t\}_{t\geq 0}$ with $S_0 = 0$ and $S_t = \sum_{i\leq t} X_i$ where the X_is are i.i.d. in $\{-1,+1\}$ with $\mathbb{P}[X_1 = 1] = p$. Let $\tau = \tau_a \wedge \tau_b$ where $a < 0 < b$. Let $q := 1 - p$ and $\phi(x) := (q/p)^x$. We claim that 1) $\tau < +\infty$ a.s., 2) $\mathbb{P}[\tau_a < \tau_b] = \frac{\phi(b)-\phi(0)}{\phi(b)-\phi(a)}$, 3) $\mathbb{E}[\tau_b] = \frac{b}{2p-1}$, and 4) $\tau_a = +\infty$ with positive probability.

Let $\psi_t(x) := x - (p-q)t$. We use two martingales:

$$\mathbb{E}[\phi(S_t) \,|\, \mathcal{F}_{t-1}] = p(q/p)^{S_{t-1}+1} + q(q/p)^{S_{t-1}-1} = \phi(S_{t-1}),$$

and

$$\begin{aligned}\mathbb{E}[\psi_t(S_t) \,|\, \mathcal{F}_{t-1}] &= p[S_{t-1} + 1 - (p-q)t] + q[S_{t-1} - 1 - (p-q)t] \\ &= \psi_{t-1}(S_{t-1}).\end{aligned}$$

Claim 1) follows by the same argument as in the unbiased case.

Gambler's ruin IV

2) Now note that $(\phi(S_{\tau\wedge t}))$ is a bounded martingale and, therefore, by applying the martingale property at time t and taking limits as $t \to \infty$ (using dominated convergence) we get

$$\phi(0) = \mathbb{E}[\phi(S_\tau)] = \mathbb{P}[\tau_a < \tau_b]\phi(a) + \mathbb{P}[\tau_a > \tau_b]\phi(b),$$

or $\mathbb{P}[\tau_a < \tau_b] = \frac{\phi(b)-\phi(0)}{\phi(b)-\phi(a)}$. Taking $b \to +\infty$, by monotonicity $\mathbb{P}[\tau_a < +\infty] = \frac{1}{\phi(a)} < 1$ so $\tau_a = +\infty$ with positive probability.

3) By the martingale property

$$0 = \mathbb{E}[S_{\tau_b\wedge t} - (p-q)(\tau_b \wedge t)].$$

By monotone convergence, $\mathbb{E}[\tau_b \wedge t] \uparrow \mathbb{E}[\tau_b]$. Finally, $-\inf_t S_t \geq 0$ a.s. and for $x \geq 0$,

$$\mathbb{P}[-\inf_t S_t \geq x] = \mathbb{P}[\tau_{-x} < +\infty] = \left(\frac{q}{p}\right)^x,$$

so that $\mathbb{E}[-\inf_t S_t] = \sum_{x\geq 1} \mathbb{P}[-\inf_t S_t \geq x] < +\infty$. Hence, we can use dominated convergence with $|S_{\tau_b\wedge t}| \leq \max\{b, -\inf_t S_t\}$ to deduce that $\mathbb{E}[\tau_b] = \frac{\mathbb{E}[S_{\tau_b}]}{p-q} = \frac{b}{2p-1}$.

Critical percolation on $\mathbb{T}_d$

Consider bond percolation on $\mathbb{T}_d$ with density $p = \frac{1}{d-1}$. Let $X_n := |\partial_n \cap \mathcal{C}_0|$, where ∂_n are the n-th level vertices and $\mathcal{C}_0$ is the open cluster of the root. The first moment method does not work in this case because $\mathbb{E}X_n = d(d-1)^{n-1}p^n = \frac{d}{d-1} \nrightarrow 0$.

Theorem

$|\mathcal{C}_0| < +\infty$ *a.s.*

Proof: Let $b := d - 1$ be the branching ratio. Let Z_n be the number of vertices in the open cluster of the first child of the root n levels below it and let $\mathcal{F}_n = \sigma(Z_0, \ldots, Z_n)$. Then $Z_0 = 1$ and $\mathbb{E}[Z_n \,|\, \mathcal{F}_{n-1}] = bpZ_{n-1} = Z_{n-1}$. So (Z_n) is a nonnegative, integer-valued martingale and it converges to an a.s. finite limit. But, clearly, for any integer $k > 0$ and $N \geq 0$

$$\mathbb{P}[Z_n = k, \ \forall n \geq N] = 0,$$

so $Z_\infty \equiv 0$. ■

Critical percolation on $\mathbb{T}_d$: a tail estimate I

We give a more precise result that will be useful later. Consider the descendant subtree, T_1, of the first child, 1, of the root. Let $\widetilde{\mathcal{C}}_1$ be the open cluster of 1 in T_1. Assume $d \geq 3$.

Theorem

$\mathbb{P}\left[\left|\widetilde{\mathcal{C}}_1\right| > k\right] \leq \frac{4\sqrt{2}}{\sqrt{k}}$, *for k large enough*

Proof: Note first that $\mathbb{E}|\widetilde{\mathcal{C}}_1| = +\infty$ by summing over the levels. So we cannot use the first moment method directly to give a bound on the tail. Instead, we use Markov's inequality on a stopped process. We use an exploration process with 3 types of vertices:

- $\mathcal{A}_t$: *active* vertices
- $\mathcal{E}_t$: *explored* vertices
- $\mathcal{N}_t$: *neutral* vertices

We start with $\mathcal{A}_0 := \{1\}$, $\mathcal{E}_0 := \emptyset$, and $\mathcal{N}_0$ contains all other vertices in T_1.

Critical percolation on $\mathbb{T}_d$: a tail estimate II

Proof (continued): At time t, if $\mathcal{A}_{t-1} = \emptyset$ we let $(\mathcal{A}_t, \mathcal{E}_t, \mathcal{N}_t)$ be $(\mathcal{A}_{t-1}, \mathcal{E}_{t-1}, \mathcal{N}_{t-1})$. Otherwise, we pick a random element, a_t, from $\mathcal{A}_{t-1}$ and:

- $\mathcal{A}_t := \mathcal{A}_{t-1} \cup \{x \in \mathcal{N}_{t-1} \ : \ \{x, a_t\} \text{ is open}\} \backslash \{a_t\}$
- $\mathcal{E}_t := \mathcal{E}_{t-1} \cup \{a_t\}$
- $\mathcal{N}_t := \mathcal{N}_{t-1} \backslash \{x \in \mathcal{N}_{t-1} \ : \ \{x, a_t\} \text{ is open}\}$

Let $M_t := |\mathcal{A}_t|$. Revealing the edges as they are explored and letting $(\mathcal{F}_t)$ be the corresponding filtration, we have $\mathbb{E}[M_t \,|\, \mathcal{F}_{t-1}] = M_{t-1} + bp - 1 = M_{t-1}$ on $\{M_{t-1} > 0\}$ so (M_t) is a nonnegative martingale. Let $\sigma^2 := bp(1-p) \geq \frac{1}{2}$, $\tau := \inf\{t \geq 0 \ : \ M_t = 0\}$, and $Y_t := M_{t\wedge\tau}^2 - \sigma^2(t \wedge \tau)$. Then, on $\{M_{t-1} > 0\}$,

$$\begin{aligned}
\mathbb{E}[Y_t \,|\, \mathcal{F}_{t-1}] &= \mathbb{E}[(M_{t-1} + (M_t - M_{t-1}))^2 - \sigma^2 t \,|\, \mathcal{F}_{t-1}] \\
&= \mathbb{E}[M_{t-1}^2 + 2M_{t-1}(M_t - M_{t-1}) + (M_t - M_{t-1})^2 - \sigma^2 t \,|\, \mathcal{F}_{t-1}] \\
&= M_{t-1}^2 + 2M_{t-1} \cdot 0 + \sigma^2 - \sigma^2 t = Y_{t-1},
\end{aligned}$$

so (Y_t) is also a martingale. For $h > 0$, let

$$\tau_h' := \inf\{t \geq 0 \ : \ M_t = 0 \text{ or } M_t \geq h\}.$$

Critical percolation on $\mathbb{T}_d$: a tail estimate III

Proof (continued): Note that $\tau'_h \leq \tau = |\widetilde{\mathcal{C}}_1| < +\infty$ a.s. We use

$$\mathbb{P}[\tau > k] = \mathbb{P}[M_t > 0, \forall t \in [k]] \leq \mathbb{P}[\tau'_h > k] + \mathbb{P}[M_{\tau'_h} \geq h].$$

By Markov's inequality, $\mathbb{P}[M_{\tau'_h} \geq h] \leq \frac{\mathbb{E}[M_{\tau'_h}]}{h}$ and $\mathbb{P}[\tau'_h > k] \leq \frac{\mathbb{E}\tau'_h}{k}$. To compute $\mathbb{E}M_{\tau'_h}$, we use the optional stopping theorem

$$1 = \mathbb{E}[M_{\tau'_h \wedge s}] \to \mathbb{E}[M_{\tau'_h}],$$

as $s \to +\infty$ by bounded convergence ($|M_{\tau'_h \wedge s}| \leq h + b$). To compute $\mathbb{E}\tau'_h$, we use the optional stopping theorem again

$$1 = \mathbb{E}[M^2_{\tau'_h \wedge s} - \sigma^2(\tau'_h \wedge s)] = \mathbb{E}[M^2_{\tau'_h \wedge s}] - \sigma^2\mathbb{E}[\tau'_h \wedge s] \to \mathbb{E}[M^2_{\tau'_h}] - \sigma^2\mathbb{E}\tau'_h,$$

as $s \to +\infty$ by bounded convergence again and monotone convergence ($\tau'_h \wedge s \uparrow \tau'_h$) respectively.

Critical percolation on $\mathbb{T}_d$: a tail estimate IV

Proof (continued): Because

$$\mathbb{E}[M^2_{\tau'_h} \mid M_{\tau'_h} \geq h] \leq (h+b)^2,$$

we have

$$\mathbb{E}\tau'_h \leq \frac{1}{\sigma^2}\left\{\frac{1}{h}\mathbb{E}[M^2_{\tau'_h} \mid M_{\tau'_h} \geq h]\right\} \leq \frac{(h+b)^2}{\sigma^2 h} \leq \frac{2(h+b)^2}{h}.$$

Take $h := \sqrt{\frac{k}{8}}$. For k large enough, $h \geq b$ and

$$\mathbb{P}[\tau > k] \leq \mathbb{P}[\tau'_h > k] + \mathbb{P}[M_{\tau'_h} \geq h] \leq \frac{8h}{k} + \frac{1}{h} = 2\sqrt{\frac{8}{k}}.$$

■