36-771 Fall 2018

Martingales 1 : Concentration inequalities

Filtrations, stopping times, conditional expectations

Lecturer : Aaditya Ramdas

1 Random walks

Let $X_1, X_2, \dots$ be i.i.d. taking values in $\mathbb{R}$. Then $S_n = X_1 + \dots + X_n$ is called a random walk. When $P(X_i = 1) = P(X_i = -1) = 1/2$, it's called a simple random walk.

Theorem 1 (Durrett, Thm 4.1.2) *For a random walk (S_n) on $\mathbb{R}$, there are only four possibilities, one of which has probability one.*

1. $S_n = 0$ *for all* n.
2. $S_n \to \infty$.
3. $S_n \to -\infty$.
4. $-\infty = \liminf S_n < \limsup S_n = +\infty$.

Any nondegenerate symmetric random walk (meaning that $P(X_i = 0) < 1$), such as the simple random walk, will satisfy case (iv).

2 Filtrations and stopping times

The sigma-field $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ is the information known at time n, and the sequence of sigma-fields $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \dots$ forms a "filtration" $(\mathcal{F}_n)$. A random variable τ taking values in $\{1, 2, \dots\} \cup \{\infty\}$ is called a stopping time if for all $n \in \mathcal{N}$, we have $\{\tau = n\} \in \mathcal{F}_n$, meaning that we can decide whether to stop the process at time n based only on the information known at time n. All constant times are stopping times, and if S, T are stopping times, then $S \vee T$ and $S \wedge T$ are also stopping times. $\mathcal{F}_\tau = \sigma(X_1, \dots, X_\tau)$ is the amount of information known at the stopping time τ. A simple example is

$$\tau = \inf\{k : S_k \geq x\} \text{ for some fixed } x.$$

If M, N are stopping times with $M \leq N$, then $\mathcal{F}_M \subseteq \mathcal{F}_N$, and if $Y_n \in \mathcal{F}_n$, then $Y_N \in \mathcal{F}_N$.

Theorem 2 (Durrett, Thm 4.1.3) *Let $X_1, X_2, \ldots$ be iid, and $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Let N be a stopping time with $P(N < \infty) > 0$. Conditional on $\{N < \infty\}$, the random variables $\{X_{N+n}, n \geq 1\}$ is independent of $\mathcal{F}_N$ and has the same distribution as the original sequence.*

Note that for any fixed n, a random walk with integrable X_i would satisfy $\mathbb{E}S_n = n\mathbb{E}X_1$, and a random walk with zero-mean and square-integrable increments would satisfy $\mathbb{E}S_n = n\mathbb{E}X_1^2$. Wald's identities extend these properties to stopping times with finite expectation.

Theorem 3 (Wald's identities, Durrett Thm 4.1.5 and 4.1.6) *Let $X_1, X_2, \ldots$ be iid and N be a stopping time with $\mathbb{E}N < \infty$. If $\mathbb{E}|X_i| < \infty$, then $\mathbb{E}S_N = \mathbb{E}N\mathbb{E}X_1$. Further, if $\mathbb{E}X_i = 0$ and $\mathbb{E}X_i^2 < \infty$, then $\mathbb{E}S_N^2 = \mathbb{E}N\mathbb{E}X_1^2$.*

> Exercise 4.1.12. Let $X_1, X_2, \ldots$ be i.i.d. uniform on $(0,1)$, let $S_n = X_1 + \cdots + X_n$ and let $T = \inf\{n : S_n > 1\}$. Show that $P(T > n) = 1/n!$, so $ET = e$ and $ES_T = e/2$.

3 Conditional Expectations

We start with a probability space $(\Omega, \mathcal{F}_0, P)$, a sigma-field $\mathcal{F} \subset \mathcal{F}_0$, and a random variable X that is measurable with respect to the sigma-field $\mathcal{F}_0$, denoted $X \in \mathcal{F}_0$. If X is integrable, meaning that $\mathbb{E}|X| = \int |X| dP < \infty$, recall that the conditional expectation of X given $\mathcal{F}$, denoted $\mathbb{E}(X|\mathcal{F})$, is any $\mathcal{F}$-measurable random variable Y such that for all $A \in \mathcal{F}$, we have $\int_A X dP = \int_A Y dP$. The conditional expectation is unique, in the sense that all "versions" that satisfy the above definition are equal almost surely. Quoting from Durrett, section 5.1:

> Intuitively, we think of $\mathcal{F}$ as describing the information we have at our disposal - for each event $A \in \mathcal{F}$, we know whether or not A has occurred. $\mathbb{E}(X|\mathcal{F})$ is then our "best guess" of the value of X given the information we have.

If $X \in \mathcal{F}$ (perfect information), then $\mathbb{E}(X|\mathcal{F}) = X$, meaning that if X is contained in the available information $\mathcal{F}$, then our best guess of X is X itself. If $\mathcal{F} = \varnothing$ (no information), then $\mathbb{E}(X|\mathcal{F}) = \mathbb{E}X$, and if X is independent of $\mathcal{F}$ (useless information), then $\mathbb{E}(X|\mathcal{F}) = \mathbb{E}X$.

Conditional expectations are monotone and linear, meaning that if $X \leq Y$, then $\mathbb{E}(X|\mathcal{F}) \leq \mathbb{E}(Y|\mathcal{F})$, and also that $\mathbb{E}(aX + bY|\mathcal{F}) = a\mathbb{E}(X|\mathcal{F}) + b\mathbb{E}(Y|\mathcal{F})$. Further, if $\mathcal{F}_1 \subset \mathcal{F}_2$, then $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_1)|\mathcal{F}_2) = \mathbb{E}(X|\mathcal{F}_1)$, and also $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1) = \mathbb{E}(X|\mathcal{F}_1)$.

If $X \in \mathcal{F}$ and Y, XY are integrable, then $\mathbb{E}(XY|\mathcal{F}) = X\mathbb{E}(Y|\mathcal{F})$. Of course, a special case of this is that $\mathbb{E}(cX) = c\mathbb{E}X$ for any constant c. Define $\text{Var}(X|\mathcal{F}) = \mathbb{E}(X^2|\mathcal{F}) - \mathbb{E}(X|\mathcal{F})^2$, so that $\text{Var}(X) = \text{Var}(\mathbb{E}(X|\mathcal{F})) + \mathbb{E}(\text{Var}(X|F))$.

Geometric interpretation: Let $(\Omega, \mathcal{F}_0, P)$ be a probability space with $X \in \mathcal{F}_0$. If $\mathcal{F} \subset \mathcal{F}_0$ and $\mathbb{E}X^2 < \infty$, then $\mathbb{E}(X|\mathcal{F}) = \arg\min_{Y \in \mathcal{F}} \mathbb{E}(X - Y)^2$. In other words, $\mathbb{E}(X|\mathcal{F})$ is the projection of X onto the closed subspace $\mathcal{L}_2(\mathcal{F}) = \{Y \in \mathcal{F} : \mathbb{E}Y^2 < \infty\}$ of the Hilbert space $L_2(\mathcal{F}_0)$.

4 Standard inequalities

Chebyshev's inequality:

$$P(|X| \geq a|\mathcal{F}) \leq \frac{\mathbb{E}(X^2|\mathcal{F})}{a^2}.$$

Jensen's inequality: if ϕ is convex, and $\mathbb{E}|X| < \infty, \mathbb{E}|\phi(X)| < \infty$, then

$$\phi(\mathbb{E}(X|\mathcal{F})) \leq \mathbb{E}(\phi(X)|\mathcal{F})$$

Cauchy-Schwarz inequality:

$$E(XY|\mathcal{G})^2 \leq E\left(X^2|\mathcal{G}\right) E\left(Y^2|\mathcal{G}\right)$$