REVIEW: MATH 547–STOCHASTIC PROCESSES

SHANSHAN DING

Spring 2010

Contents

1. Conditional Expectation

Definition 1.1. *Let $(\Omega, \mathcal{F})$ be a measurable space, then $\alpha : \mathcal{F} \to \mathbb{R} \cup \infty$ is called a* **signed measure** *if $\alpha(\emptyset) = 0$ and α is countably additive in the following sense: suppose G is the disjoint union of G_n,*

- *If $\alpha(G) < \infty$, then $\sum_n \alpha(G_n)$ converges absolutely and is equal to $\alpha(G)$, and*
- *If $\alpha(G) = \infty$, then $\sum_n \alpha(G_n)^- < \infty$ and $\sum_n \alpha(G_n)^+ = \infty$.*

Definition 1.2. *Let α be a signed measure on $(\Omega, \mathcal{F})$. A set A is* **positive** *(resp.* **negative***) if, for every measurable $B \subseteq A$, $\alpha(B) \geq 0$ (resp. $\alpha(B) \leq 0$). A is a* **null set** *if it is both positive and negative (note that $\alpha(A) = 0$ is not sufficient for A to be null).*

Theorem 1.3 (Hahn decomposition)**.** *Let α be a signed measure on $(\Omega, \mathcal{F})$. Then Ω can be partitioned into a disjoint union $A \cup B$, where A is a positive set and B is a negative set. Furthermore, if $\Omega = A' \cup B'$ is another such partition, then $A \Delta A'$ and $B \Delta B'$ are null sets.*

Definition 1.4. *Let μ, ν be measures on $(\Omega, \mathcal{F})$. If, for $A \in \mathcal{F}$, $\mu(A) = 0$ implies $\nu(A) = 0$, then we say that ν is* **absolutely continuous with respect to** *μ and write $\nu \ll \mu$. If there exists $A \in \mathcal{F}$ such that $\mu(A) = 0$ and $\nu(A^c) = 0$, then we say that μ and ν are* **(mutually) singular** *and write $\mu \perp \nu$.*

Theorem 1.5 (Lebesgue decomposition)**.** *Let μ and ν be σ-finite measures on $(\Omega, \mathcal{F})$ (i.e. Ω is a countable union of measurable sets of finite measures w.r.t. both μ and ν). Then there exists a (unique) decomposition $\nu = \nu_r + \nu_s$, where $\nu_s \perp \mu$ and $\nu_r(A) = \int_A g \, d\mu$ for all $A \in \mathcal{F}$ and some non-negative measurable function g.*

Theorem 1.6 (Radon-Nikodym)**.** *Let μ and ν be σ-finite measures on $(\Omega, \mathcal{F})$. If $\nu \ll \mu$, then there exists a non-negative measurable function g such that $\nu(A) = \int_A g \, d\mu$ for all $A \in \mathcal{F}$. We say that ν has* **Radon-Nikodym derivative** *g with respect to μ on $(\Omega, \mathcal{F})$ and write $d\nu/d\mu = g$.*

Remark. It can be shown that if h is another such function, then $g = h$ μ-a.e.

Proof. Let $\nu = \nu_r + \nu_s$ be the Lebesgue decomposition, and choose $A \in \mathcal{F}$ so that we have $\nu_s(A^c) = \mu(A) = 0$. Since $\nu \ll \mu$, $0 = \nu(A) \geq \nu_s(A)$, and hence $\nu_s \equiv 0$. □

The Radon-Nikodym derivative enables us to construct conditional expectation.

Definition 1.7. *Let $(\Omega, \mathcal{F}_0, \mathbb{P})$ be a probability space and let X be a random variable with $\mathbb{E}|X| < \infty$. Suppose $\mathcal{F}$ is a sub-σ-algebra of $\mathcal{F}_0$. We say that a random variable Y is a* **conditional expectation of** *X* **given** *$\mathcal{F}$, or a* **version of** *$\mathbb{E}(X|\mathcal{F})$, if*

(1) Y is measurable with respect to $\mathcal{F}$, and
(2) For all $A \in \mathcal{F}$, $\int_A Y d\mathbb{P} = \int_A X d\mathbb{P}$.

Right away, we should check that if Y is a version of $\mathbb{E}(X|\mathcal{F})$, then Y is integrable. Since $A = \{\omega : Y(\omega) > 0\} \in \mathcal{F}$ by (1), we get from (2) that

$$\int_A Y d\mathbb{P} = \int_A X d\mathbb{P} \leq \int_A |X| d\mathbb{P} \quad \text{and} \quad \int_{A^c} -Y d\mathbb{P} = \int_{A^c} -X d\mathbb{P} \leq \int_{A^c} |X| d\mathbb{P}, \tag{1.1}$$

which add up to give $\mathbb{E}|Y| \leq \mathbb{E}|X| < \infty$.

Theorem 1.8. *Conditional expectation exists. Furthermore, if Y, Y' are versions of $\mathbb{E}(X|\mathcal{F})$, then $Y = Y'$ a.s.*

Proof. Suppose $X \geq 0$. For $A \in \mathcal{F}$, define $\nu(A) = \int_A X d\mathbb{P}$. Clearly $\nu \ll \mathbb{P}$, so there exists Radon-Nikodym derivative $d\nu/d\mathbb{P}$ such that

$$\int_A X d\mathbb{P} = \nu(A) = \int_A \frac{d\nu}{d\mathbb{P}} d\mathbb{P} \tag{1.2}$$

for all $A \in \mathcal{F}$. As $\nu, \mathbb{P}$ are measures on $(\Omega, \mathcal{F})$, $d\nu/d\mathbb{P}$ is $\mathcal{F}$-measurable and therefore a version of $\mathbb{E}(X|\mathcal{F})$. For the general case, let $X = X^+ - X^-$, $Y_1 = \mathbb{E}(X^+|\mathcal{F})$, and $Y_2 = \mathbb{E}(X^-|\mathcal{F})$. Then $Y_1 - Y_2$ is $\mathcal{F}$-measurable and integrable, and

$$\int_A X d\mathbb{P} = \int_A X^+ d\mathbb{P} - \int_A X^- d\mathbb{P} = \int_A Y_1 d\mathbb{P} - \int_A Y_2 d\mathbb{P} = \int_A (Y_1 - Y_2) d\mathbb{P} \tag{1.3}$$

for all $A \in \mathcal{F}$. This shows that $Y_1 - Y_2$ is a version of $\mathbb{E}(X|\mathcal{F})$, which proves existence.

For uniqueness, let $A = \{Y - Y' \geq 0\}$, which is in $\mathcal{F}$ because Y, Y' are both $\mathcal{F}$-measurable by assumption. Then $\int_A Y d\mathbb{P} = \int_A X d\mathbb{P} = \int_A Y' d\mathbb{P}$, so $\int_A (Y - Y') d\mathbb{P} = 0$. Thus $Y - Y' = 0$ a.s. on A, which implies that $Y \leq Y'$ a.s. Similarly, integrating $Y - Y'$ on A^c shows that $Y' \leq Y$ a.s., and therefore $Y' = Y$ a.s. □

If $\mathcal{F} = \{\emptyset, \Omega\}$ is the trivial σ-field, then $\mathbb{E}(X|\mathcal{F}) = \mathbb{E}X$. At the other extreme, if X is $\mathcal{F}$-measurable, then $\mathbb{E}(X|\mathcal{F}) = X$. Moreover, since Ω is in every σ-field over Ω,

$$\mathbb{E}(\mathbb{E}(X|\mathcal{F})) = \int_\Omega \mathbb{E}(X|\mathcal{F}) d\mathbb{P} = \int_\Omega X d\mathbb{P} = \mathbb{E}X \tag{1.4}$$

for any $\mathcal{F}$ over Ω. Properties such as linearity, monotonicity, monotone convergence, and Jensen's inequality have unsurprising conditional analogues. Jensen's inequality also shows that conditional expectation is a contraction in L^p, $p \geq 1$:

$$\mathbb{E}(|\mathbb{E}(X|\mathcal{F})|^p) \leq \mathbb{E}(\mathbb{E}(|X|^p|\mathcal{F})) = \mathbb{E}|X|^p. \tag{1.5}$$

Theorem 1.9. *If X is $\mathcal{F}$-measurable and $\mathbb{E}|Y|$, $\mathbb{E}|XY| < \infty$, $\mathbb{E}(XY|\mathcal{F}) = X\mathbb{E}(Y|\mathcal{F})$.*

Theorem 1.10. *If $\mathcal{F}_1 \subseteq \mathcal{F}_2$, then $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_1)|\mathcal{F}_2) = \mathbb{E}(X|\mathcal{F}_1) = \mathbb{E}(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1)$. In words, the smaller σ-field always wins.*

Proof. The first equality follows once we note that $\mathbb{E}(X|\mathcal{F}_1)$ is $\mathcal{F}_2$-measurable. For the second equality, observe that $\mathbb{E}(X|\mathcal{F}_1)$ is $\mathcal{F}_1$-measurable, and that

$$\int_A \mathbb{E}(X|\mathcal{F}_1) d\mathbb{P} = \int_A X d\mathbb{P} = \int_A \mathbb{E}(X|\mathcal{F}_2) d\mathbb{P} \tag{1.6}$$

whenever $A \in \mathcal{F}_1 \subseteq \mathcal{F}_2$. □

On the other hand, if $\mathcal{F}_1$ and $\mathcal{F}_2$ are incompatible, then it is possible that

$$\mathbb{E}(\mathbb{E}(X|\mathcal{F}_1)|\mathcal{F}_2) \neq \mathbb{E}(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1). \tag{1.7}$$

For an example, consider $\Omega = \{a, b, c\}$, $\mathcal{F}_1 = \{\{a\}, \{b, c\}\}$, $\mathcal{F}_2 = \{\{a, b\}, \{c\}\}$, $X(b) = 1$, and $X(a) = X(c) = 0$. Recall that if $\Omega_1, \Omega_2, \ldots$ is a partion of Ω into disjoint sets each with positive probability and $\mathcal{F} = \sigma(\Omega_1, \Omega_2, \ldots)$, then

$$\mathbb{E}(X|\mathcal{F}) = \frac{\mathbb{E}(X(\omega)\mathbf{1}_{\omega \in \Omega_i})}{\mathbb{P}(\Omega_i)} \quad \text{on } \Omega_i. \tag{1.8}$$

2. Martingales

Definition 2.1. *A* **filtration** $\mathcal{F}_n$ *is an increasing sequence of* σ*-fields. A sequence* X_n *is said to be* **adapted** *to* $\mathcal{F}_n$ *if* X_n *is measurable with respect to* $\mathcal{F}_n$ *for all* n*. A sequence* H_n *is said to be* **predictable** *if* H_n *is measurable with respect to* $\mathcal{F}_{n-1}$ *for all* $n \geq 1$*.*

Definition 2.2. *If* X_n *is a sequence such that for all* n, $\mathbb{E}|X_n| < \infty$, X_n *is adapted to* $\mathcal{F}_n$*, and* $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$*, then* X_n *is a* **martingale** *(with respect to* $\mathcal{F}_n$*). If we replace equality in the last condition with* $\leq$ *or* $\geq$*, then* X_n *is a* **supermartingale** *or* **submartingale***, respectively.*

Theorem 2.3. *If* X_n *is a supermartingale, then* $\mathbb{E}(X_n|\mathcal{F}_m) \leq X_m$ *for all* $n \geq m$*.*

Proof. Let $n = m + k$. The statement is trivial for $k < 2$. For $k \geq 2$, since the smaller σ-field always wins, $\mathbb{E}(X_{m+k}|\mathcal{F}_m) = \mathbb{E}(\mathbb{E}(X_{m+k}|\mathcal{F}_{m+k-1})|\mathcal{F}_m) \leq \mathbb{E}(X_{m+k-1}|\mathcal{F}_m)$, and the desired result follows by induction. □

Remark. Often, when we have a result for either supermartingales or submartingales, we can obtain an analogous result for the other as well as for martingales. Here, if X_n is a submartingale, then $-X_n$ is a supermartingale, so by linearity $\mathbb{E}(X_n|\mathcal{F}_m) \geq X_m$ for all $n \geq m$. If X_n is a martingale, then it is both a supermartingale and a submartingale, and hence $\mathbb{E}(X_n|\mathcal{F}_m) = X_m$ for all $n \geq m$. We could also have verified these corollaries directly from the definitions.

Theorem 2.4. *If* X_n *is a martingale and* φ *is a convex function with* $\mathbb{E}|\varphi(X_n)| < \infty$ *for all* n*, then* $\varphi(X_n)$ *is a submartingale. If* X_n *is a submartingale, then we can draw the same conclusion with the added condition that* φ *is non-decreasing.*

Theorem 2.5. *Let* X_n *be a supermartingale. If* $H_n \geq 0$ *is predictable and each* H_n *is bounded, then* $(H\cdot X)_n = \sum_{m=1}^{n} H_m(X_m - X_{m-1})$ *is a supermartingale. Analogous results hold for submartingales and, without the restriction* $H_n \geq 0$*, for martingales.*

Remark. We can think of H_n as a gambling system, namely the amount of money a gambler will bet at time n. Then $(H \cdot X)_n$ represents the gambler's winnings at time n, and the previous theorem says that there is no strategy for beating an unfavorable game.

Corollary 2.6. *If* τ *is a stopping time and* X_n *is a supermartingale, then* $X_{\tau\wedge n}$ *is a supermartingale.*

Proof. Let $H_n = \mathbf{1}_{\tau \geq n}$, then H_n is a predictable sequence and $(H \cdot X)_n = X_{\tau\wedge n} - X_0$ is a supermartingale. Since the constant X_0 is trivially a supermartingale, $X_{\tau\wedge n}$ is a supermartingale. □

Suppose X_n is a submartingale. Let $a < b$, and define stopping times

$$N_k = \begin{cases} -1 & \text{if } k = 0, \\ \inf\{m > N_{k-1} : X_m \leq a\} & \text{if } k \text{ is odd}, \\ \inf\{m > N_{k-1} : X_m \geq b\} & \text{if } k \text{ is even}. \end{cases}$$

Define H_n to be 1 if $N_{2j-1} < n \leq N_{2j}$ for some j and 0 otherwise, so that H_n is a predictable sequence. We can interpret H_n as a gambling system that tries to take advantage of upcrossings from below a to above b. Also let $U_n = \sup\{j : N_{2j} \leq n\}$ be the number of upcrossings completed by time n.

Theorem 2.7 (Upcrossing inequality)**.** *If X_n is a submartingale, then*

$$(b-a)\mathbb{E}U_n \le \mathbb{E}(X_n-a)^+ - \mathbb{E}(X_0-a)^+. \tag{2.1}$$

Proof. Let $Y_n = a+(X_n-a)^+$, then Y_n is a submartingale by Theorem 2.4, and it is clear that Y_n upcrosses $[a,b]$ the same number of times as X_n. Observe that $(b-a)U_n \le (H\cdot Y)_n$, since each upcrossing results in a profit of at least $b-a$ and the final incomplete crossing, if there is one, makes a non-negative contribution to the RHS. Setting $K_n = 1-H_n$, we have that

$$(H\cdot Y)_n + (K\cdot Y)_n = \sum_{m=1}^{n} H_m(Y_m - Y_{m-1}) + \sum_{m=1}^{n}(1-H_m)(Y_m-Y_{m-1}) = Y_n - Y_0. \tag{2.2}$$

Taking expectations, since $\mathbb{E}(K\cdot Y)_n \ge \mathbb{E}(K\cdot Y)_0 = 0$ by Theorem 2.5, we see that

$$(b-a)\mathbb{E}U_n = \mathbb{E}((b-a)U_n) \le \mathbb{E}(H\cdot Y)_n \le \mathbb{E}(Y_n - Y_0), \tag{2.3}$$

which completes the proof. □

Letting a, b range over $\mathbb{Q}$, the upcrossing inequality is used to prove the following:

Theorem 2.8 (Martingale convergence)**.** *If X_n is a submartingale with $\sup \mathbb{E}X_n^+ < \infty$, then as $n \to \infty$, X_n converges a.s. to a limit X with $\mathbb{E}|X| < \infty$.*

Corollary 2.9. *If $X_n \ge 0$ is a supermartingale, then as $n\to\infty$, X_n converges a.s. to a limit X with $\mathbb{E}X \le \mathbb{E}X_0$.*

Proof. Since $Y_n = -X_n \le 0$ is a submartingale with $\mathbb{E}Y_n^+ = 0$, Y_n (and hence X_n) converges a.s. Moreover, $\mathbb{E}(\lim_{n\to\infty} X_n) \le \lim_{n\to\infty} \mathbb{E}X_n \le \mathbb{E}X_0$ by Fatou's lemma. □

Theorem 2.10 (Doob's decomposition)**.** *Any submartingale X_n can be uniquely decomposed as $X_n = M_n + A_n$, where M_n is a martingale and A_n is a predictable increasing sequence with $A_0 = 0$.*

Proof. We want $X_n = M_n + A_n$, where $\mathbb{E}(M_n|\mathcal{F}_{n-1}) = M_{n-1}$ and A_n is measurable with respect to $\mathcal{F}_{n-1}$. Thus we need

$$\mathbb{E}(X_n|\mathcal{F}_{n-1}) = \mathbb{E}(M_n|\mathcal{F}_{n-1}) + \mathbb{E}(A_n|\mathcal{F}_{n-1}) = M_{n-1} + A_n = X_{n-1} - A_{n-1} + A_n. \tag{2.4}$$

With $A_0 = 0$ and $M_0 = X_0$, we can uniquely define A_n and M_n by specifying that

$$A_n - A_{n-1} = \mathbb{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1} \text{ and } M_n = X_n - A_n. \tag{2.5}$$

Since $\mathbb{E}(X_n|\mathcal{F}_{n-1})$ is measurable with respect to $\mathcal{F}_{n-1}$, we see by induction that A_n is $\mathcal{F}_{n-1}$-measurable, so that A_n is a predictable sequence. With X_n being a submartingale, it is clear that $A_n \ge A_{n-1}$. Finally, as

$$\mathbb{E}(M_n|\mathcal{F}_{n-1}) = \mathbb{E}(X_n - A_n|\mathcal{F}_{n-1}) = \mathbb{E}(X_n|\mathcal{F}_{n-1}) - A_n = X_{n-1} - A_{n-1} = M_{n-1}, \tag{2.6}$$

M_n is a martingale. □

Theorem 2.11. *Let X_n be a martingale with $|X_{n-1} - X_n| \le M < \infty$. Let C be the event that $\lim X_n$ exists and is finite, and let D be the event that $\limsup X_n = \infty$ and $\liminf X_n = -\infty$. Then $\mathbb{P}(C \cup D) = 1$.*

Theorem 2.12 (Borel-Cantelli II, conditional version 1)**.** *Let $\mathcal{F}_n$ be a filtration with $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and let A_n be a sequence of events with $A_n \in \mathcal{F}_n$. Then A_n occurs infinitely often a.s. on the event $\sum_{n=1}^{\infty} \mathbb{P}(A_n|\mathcal{F}_{n-1}) = \infty$.*

Proof. Setting $X_0 = 0$ and $X_n = \sum_{m=1}^n \mathbf{1}_{A_m} - \mathbb{P}(A_m|\mathcal{F}_{m-1})$ for $n \geq 1$, X_n is a martingale with $|X_n - X_{n-1}| \leq 1$. Let C and D be as defined in Theorem 2.11. On C, $\sum \mathbf{1}_{A_n} = \infty$ if and only if $\sum \mathbb{P}(A_n|\mathcal{F}_{n-1}) = \infty$; on D, $\sum \mathbf{1}_{A_n} = \infty$ as well as $\sum \mathbb{P}(A_n|\mathcal{F}_{n-1}) = \infty$. Since $\mathbb{P}(C \cup D) = 1$, the result follows at once. □

Theorem 2.13. *If X_n is a submartingale and τ is a stopping time with $\mathbb{P}(\tau \leq k) = 1$, then $\mathbb{E}X_0 \leq \mathbb{E}X_\tau \leq \mathbb{E}X_k$.*

Proof. By Corollary 2.6, $X_{\tau\wedge n}$ is a submartingale, so $\mathbb{E}X_0 = \mathbb{E}X_{\tau\wedge 0} \leq \mathbb{E}X_{\tau\wedge k} = \mathbb{E}X_\tau$. For the second inequality, let $H_n = \mathbf{1}_{\tau \leq n-1}$. Then H_n is predictable and $(H\cdot X)_n = X_n - X_{\tau\wedge n}$ is a submartingale, so $\mathbb{E}X_k - \mathbb{E}X_\tau = \mathbb{E}(H\cdot X)_k \geq \mathbb{E}(H\cdot X)_0 = 0$. □

Let μ and ν be measures on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ that make the coordinates $\xi_n(\omega) = \omega_n$ independent. Suppose $F_n \ll G_n$, where $F_n(x) = \mu(\xi_n \leq x)$ and $G_n(x) = \nu(\xi_n \leq x)$, and define $q_n = dF_n/dG_n$.

Theorem 2.14 (Kakutani dichotomy)**.** *Either $\mu \ll \nu$ or $\mu \perp \nu$, according as*

$$\prod_{n=1}^{\infty} \int \sqrt{q_n}\, dG_n > 0 \quad \text{or} \quad \prod_{n=1}^{\infty} \int \sqrt{q_n}\, dG_n = 0. \tag{2.7}$$

For $i, n \geq 0$, let γ_i^n be IID random variables taking values in the non-negative integers. Define $Z_0 = 1$ and

$$Z_{n+1} = \begin{cases} \gamma_1^{n+1} + \cdots + \gamma_{Z_n}^{n+1} & \text{if } Z_n > 0, \\ 0 & \text{if } Z_n = 0. \end{cases}$$

The idea here is that Z_n is the population in the nth generation, and each member of the nth generation gives birth independently to an identically distributed number of children. The sequence Z_n is called a **Galton-Watson process**.

Lemma 2.15. *Let $\mathcal{F}_n = \sigma(\gamma_i^m : 1 \leq i \leq Z_{n-1}, 1 \leq m \leq n)$ and $\mu = \mathbb{E}\gamma_i^m \in (0,\infty)$. Then Z_n/μ^n is a martingale w.r.t. $\mathcal{F}_n$.*

Theorem 2.16. *If $\mu < 1$, or if $\mu = 1$ and $\mathbb{P}(\gamma_i^m = 1) < 1$, then we have $Z_n = 0$ for all sufficiently large n.*

This makes sense: if each individual has on average no more than one child, then (excluding the trivial case of everyone having exactly one child) the species will die out.

Theorem 2.17. *If $\mu > 1$, then $\mathbb{P}(Z_n > 0$ for all $n) > 0$.*

Proof. (sketch) Let $p_k = \mathbb{P}(\gamma_i^m = k)$. For $s \in [0,1]$, define $\varphi(s) = \sum_{k\geq 0} p_k s^k$, so that φ is the generating function for the offspring distribution p_k. From the theory of generating functions we know that φ is increasing and convex, with $\lim_{s\uparrow 1} \varphi'(s) = \sum_{k=1}^\infty k p_k = \mu$. This is used to prove Facts 2 and 3 below.
Fact 1: If $\theta_n = \mathbb{P}(Z_n = 0)$, then $\theta_n = \sum_{k=1}^\infty p_k \theta_{n-1}^k$.
Fact 2: If $\varphi'(1) = \mu > 1$, then there is a unique $\rho \in [0,1)$ such that $\varphi(\rho) = \rho$.
Fact 3: As $n \uparrow \infty$, $\theta_n \uparrow \rho$.
Thus $\mathbb{P}(Z_n = 0$ for some $n) = \lim_{n\to\infty} \theta_n = \rho < 1$, as desired. □

Theorem 2.18 (Doob's inequality)**.** *Let X_n be a submartingale, $\bar{X}_m = \max_{0\leq n\leq m} X_n^+$, and $\lambda > 0$. Let A be the event $\bar{X}_m \geq \lambda$. Then $\lambda\mathbb{P}(A) \leq \mathbb{E}(X_m \mathbf{1}_A) \leq \mathbb{E}X_m^+$.*

Remark. Kolmogorov's maximal inequality can be easily obtained via Doob's inequality.

Proof. Let $\tau = \inf\{n : X_n \geq \lambda\} \wedge m$. Since $X_\tau \geq \lambda$ on A, $\lambda\mathbb{P}(A) = \mathbb{E}(\lambda \mathbf{1}_A) \leq \mathbb{E}(X_\tau \mathbf{1}_A)$. Furthermore, since $\mathbb{E}X_\tau \leq \mathbb{E}X_m$ by Theorem 2.13 and $X_\tau = X_m$ on A^c,

$$\mathbb{E}(X_\tau \mathbf{1}_A) = \int_A X_\tau d\mathbb{P} = \int_\Omega X_\tau d\mathbb{P} - \int_{A^c} X_\tau d\mathbb{P} \leq \int_\Omega X_m d\mathbb{P} - \int_{A^c} X_m d\mathbb{P} = \mathbb{E}(X_m \mathbf{1}_A), \tag{2.8}$$

which proves the first inequality. The second inequality is obvious. □

Theorem 2.19 (L^p martingale convergence)**.** *If X_n is a martingale with $\sup \mathbb{E}|X_n|^p < \infty$ for some $p > 1$, then $X_n \to X$ a.s. and in L^p.*

Remark. If $X_n \to X$ in L^p, then convergence must also occur in L^1. It is not difficult to see that L^1 convergence implies $\mathbb{E}X = \mathbb{E}X_0$. In fact, by Jensen's inequality and the reverse triangle inequality,

$$|\mathbb{E}|X_n| - \mathbb{E}|X|| \leq \mathbb{E}||X_n| - |X|| \leq \mathbb{E}|X_n - X|, \tag{2.9}$$

so L^1 convergence implies convergence in mean.

Proposition 2.20 (Conditional variance formula)**.** *If X_n is a martingale with $\mathbb{E}X_n^2 < \infty$ for all n, then $\mathbb{E}((X_n - X_m)^2|\mathcal{F}_m) = \mathbb{E}(X_n^2|\mathcal{F}_m) - X_m^2$.*

Suppose X_n is a martingale with $X_0 = 0$ and $\mathbb{E}X_n^2 < \infty$ for all n. Then X_n^2 is a submartingale, and it follows from Doob's decomposition and the conditional variance formula that we can write $X_n^2 = M_n + A_n$, where M_n is a martingale and

$$A_n = \sum_{m=1}^{n} (\mathbb{E}(X_m^2|\mathcal{F}_{m-1}) - X_{m-1}^2) = \sum_{m=1}^{n} \mathbb{E}((X_m - X_{m-1})^2|\mathcal{F}_{m-1}). \tag{2.10}$$

A_n is called the **action process**, or **increasing process**, associated with X_n and can be interpreted as a path by path measurement of the variance at time n.

Theorem 2.21. *On the event $A_\infty < \infty$, $\lim_{n\to\infty} X_n$ exists and is finite.*

Theorem 2.22. *Let $f \geq 1$ be an increasing function with $\int_0^\infty (f(t))^{-2}\, dt < \infty$. Then $X_n/f(A_n) \to 0$ a.s. on the event $A_\infty = \infty$.*

Theorem 2.23 (Borel-Cantelli II, conditional version 2)**.** *Suppose B_n is adapted to $\mathcal{F}_n$ and let $p_n = \mathbb{P}(B_n|\mathcal{F}_{n-1})$. Then*

$$\frac{\sum_{m=1}^n \mathbf{1}_{B_m}}{\sum_{m=1}^n p_m} \to 1 \text{ a.s. on the event } \sum_{m=1}^{\infty} p_m = \infty. \tag{2.11}$$

Proof. Define a martingale X_n by setting $X_0 = 0$ and $X_n - X_{n-1} = \mathbf{1}_{B_n} - p_n$, so that

$$\frac{\sum_{m=1}^n \mathbf{1}_{B_m}}{\sum_{m=1}^n p_m} - \frac{X_n}{\sum_{m=1}^n p_m} = 1. \tag{2.12}$$

If $A_\infty < \infty$, then X_n converges to a finite limit by Theorem 2.21, so $X_n / \sum_{m=1}^n p_m \to 0$ on $\{A_\infty < \infty\} \cap \{\sum_m p_m = \infty\}$. On $\{A_\infty = \infty\}$, if we use $f(t) = \max\{t, 1\}$, then $X_n/f(A_n) \to 0$ a.s. by 2.22, and thus $X_n/A_n \to 0$ a.s. Since

$$A_n - A_{n-1} = \mathbb{E}((X_n - X_{n-1})^2|\mathcal{F}_{n-1}) = \mathbb{E}((\mathbf{1}_{B_n} - p_n)^2|\mathcal{F}_{n-1}) = p_n - p_n^2 \leq p_n, \tag{2.13}$$

$A_n \leq \sum_{m=1}^n p_m$, and $X_n / \sum_{m=1}^n p_m \to 0$ a.s. (Note: $\{A_\infty = \infty\} \subset \{\sum_m p_m = \infty\}$.) □

Definition 2.24. *A set of random variables* X_α, $\alpha \in A$, *is* **uniformly integrable** *if*

$$\lim_{M\to\infty}\left(\sup_{\alpha\in A}\mathbb{E}(|X_\alpha|\mathbf{1}_{|X_\alpha|>M})\right)=0. \tag{2.14}$$

Example. Given $(\Omega,\mathcal{F}_0,\mathbb{P})$ and $X\in L^1(\mathbb{P})$, $\{\mathbb{E}(X|\mathcal{F}):\mathcal{F}\subseteq\mathcal{F}_0\}$ is uniformly integrable.

Note that if $\{X_\alpha\}$ is uniformly integrable, then we can pick M large enough so that $\sup_{\alpha\in A}\mathbb{E}(|X_\alpha|\mathbf{1}_{|X_\alpha|>M})<1$, from which it follows that $\sup_{\alpha\in A}\mathbb{E}|X_\alpha|\le M+1<\infty$. Thus uniform integrability implies uniform L^1-boundedness (which in turn implies tightness). To see that uniform integrability is a strictly stronger condition, let $X_n=n$ with probability $1/n$ and $X_n=0$ otherwise.

Theorem 2.25. *For a martingale, the following are equivalent:*

(1) *It is uniformly integrable.*
(2) *It converges a.s. and in* L^1.
(3) *It converges in* L^1.
(4) *There exists an integrable random variable* X *such that* $X_n=\mathbb{E}(X|\mathcal{F}_n)$.

Remark. For a submartingale, the first three are equivalent.

Remark. If X_n is a uniformly bounded martingale, it is uniformly integrable and, as such, converges in L^1 to some X. It follows that $\mathbb{E}X=\mathbb{E}X_0$.

Theorem 2.26. *If* X_n *is a uniformly integrable submartingale, then for any stopping time* τ, $X_{\tau\wedge n}$ *is uniformly integrable.*

Theorem 2.27. *If* X_n *is a uniformly integrable submartingale (so in particular if* X_n *is uniformly bounded), then for any stopping time* τ, $\mathbb{E}X_0\le\mathbb{E}X_\tau\le\mathbb{E}X_\infty$.

Proof. By Theorem 2.13, $\mathbb{E}X_0\le\mathbb{E}X_{\tau\wedge n}\le\mathbb{E}X_n$. Letting $n\to\infty$, Theorem 2.26 and the second remark following Theorem 2.25 indicate that $X_{\tau\wedge n}\to X_\tau$ and $X_n\to X_\infty$ in L^1, so that $\mathbb{E}X_{\tau\wedge n}\to\mathbb{E}X_\tau$ and $\mathbb{E}X_n\to\mathbb{E}X_\infty$. □

Theorem 2.28 (Optional stopping)**.** *If* $\tau\le\sigma$ *are stopping times and* $X_{\sigma\wedge n}$ *is a uniformly integrable submartingale, then* $X_\tau\le\mathbb{E}(X_\sigma|\mathcal{F}_\tau)$ *and hence* $\mathbb{E}X_\tau\le\mathbb{E}X_\sigma$.

Theorem 2.29. *Suppose* X_n *is a submartingale with* $\mathbb{E}(|X_{n+1}-X_n||\mathcal{F}_n)\le B$ *a.s. If* $\mathbb{E}\tau<\infty$, *then* $X_{\tau\wedge n}$ *is uniformly integrable, and the proof of Theorem 2.27 shows that* $\mathbb{E}X_\tau\ge\mathbb{E}X_0$.

Remark. If X_n is a martingale, then $\mathbb{E}X_\tau=\mathbb{E}X_0$. Wald's equation can be recovered by letting S_n be a random walk, $\mu=\mathbb{E}(S_n-S_{n-1})$, and applying the result to $X_n=S_n-n\mu$.

Theorem 2.30. *If* $X_n\ge 0$ *supermartingale and* τ *is a stopping time, then* $\mathbb{E}X_\tau\le\mathbb{E}X_0$.

Theorem 2.31 (Azuma's inequality)**.** *Let* X_n *be a martingale with* $|X_n-X_{n-1}|\le 1$ *for all* n *and* $X_0=0$. *Then for any constant* $\alpha>0$, $\mathbb{P}(X_k\ge\alpha)\le e^{-\alpha^2/2k}$.

Theorem 2.32 (Azuma with conditional bounds)**.** *Let* X_n *be a martingale with respect to the filtration* $\mathcal{F}_n$, *where* $X_0=0$. *Define* D_n *to be the conditional essential supremum of* $|X_n-X_{n-1}|$ *given* $\mathcal{F}_{n-1}$, *and let* $Q_n=\sum_{k=1}^n D_n^2$. *Then for any constants* $\alpha,\beta>0$,

$$\mathbb{P}(X_k\ge\alpha)\le e^{-\alpha^2/2\beta}+\mathbb{P}(Q_k>\beta). \tag{2.15}$$

3. Finite Markov Chains

Definition 3.1. *Let $(S, \mathcal{S})$ be a measurable space. A sequence X_n taking values in S is said to be a* **Markov chain** *with respect to a filtration $\mathcal{F}_n$ if X_n is $\mathcal{F}_n$-measurable and*

$$\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = \mathbb{P}(X_{n+1} \in A | X_n) \tag{3.1}$$

for every $A \in \mathcal{S}$.

When $(S, \mathcal{S})$ is nice, we can identify a Markov chain with a sequence of **transition kernels**, which is a sequence $p_n(\cdot, \cdot)$ such that

(1) For each $x \in S$, $A \to p_n(x, A)$ is a probability measure on $(S, \mathcal{S})$, and
(2) For each $A \in \mathcal{S}$, $x \to p_n(x, A)$ is a measurable function.

Conversely, given $\{p_n\}$ and an initial measure μ, we can construct a probability measure $\mathbb{P}_\mu$ on the canonical space $(\Omega, \mathcal{F}) = (S, \mathcal{S})^\infty$. We say that a Markov chain is **time-homogeneous** if $p_n = p$ for all n. When $\mu = \delta_x$, we use $\mathbb{P}_x$ as abbreviation for $\mathbb{P}_{\delta_x}$. Similar notations are used to denote expectations.

Theorem 3.2 (Markov property)**.** *Let $\theta_n : \Omega \to \Omega$ send ω_m to ω_{m+n} for each m. Suppose that $Y : \Omega \to \mathbb{R}$ is bounded and measurable. Then for a Markov chain X_n,*

$$\mathbb{E}_\mu(Y \circ \theta_n | \mathcal{F}_n) = \mathbb{E}_{X_n} Y. \tag{3.2}$$

Theorem 3.3 (Strong Markov property)**.** *Suppose that $Y_n : \Omega \to \mathbb{R}$ is measurable for all n and $\{Y_n\}$ is uniformly bounded. Then for a Markov Chain X_n and stopping time τ,*

$$\mathbb{E}_\mu(Y_\tau \circ \theta_\tau | \mathcal{F}_\tau) = \mathbb{E}_{X_\tau} Y_\tau \quad \text{on } \{\tau < \infty\}. \tag{3.3}$$

Intuitively, the Markov property says that given the present, the future does not depend on the past; the strong Markov property defines "present" in terms of a stopping time.

Definition 3.4. *Let $\tau_x = \inf\{n > 0 : X_n = x\}$ and $\rho_{xy} = \mathbb{P}_x(\tau_y < \infty)$. A state x is said to be* **recurrent** *if $\rho_{xx} = 1$ and* **transient** *if $\rho_{xx} < 1$.*

Definition 3.5. *A set $C \subseteq S$ is* **closed** *if $x \in C$ and $\rho_{xy} > 0$ implies that $y \in C$. C is* **irreducible** *if $x, y \in C$ implies that $\rho_{xy} > 0$.*

Remark. If a closed irreducible class contains one recurrent state, then it must contains all recurrrent states and is said to be a recurrent class. Indeed, recurrent states can be partitioned into recurrent classes.

For the remainder of this section we assume S to be finite. In this case there must always be at least one recurrent class. It is useful to represent the transition kernel of a time-homogeneous Markov chain with a stochastic matrix M, for then M^n gives us the n-step transition kernel. A recurrent class C is periodic if the GCD of $\{n : M^n_{xx} > 0\}$ is some integer $l \geq 2$ for any (equivalently all) $x \in C$, and the least such l is called the **period**. Aperiodicity says that for sufficiently large n, $M^n_{xy} > 0$ for all $x, y \in C$.

Theorem 3.6 (Perron-Frobenius I)**.** *Let M be a stochastic matrix. If M (or rather, the corresponding Markov chain) is irreducible and aperiodic, then:*

(1) The eigenvalue of maximum modulus is 1.
(2) This eigenvalue is simple.
(3) The corresponding left eigenvector (the π for which $\pi M = \pi$) is non-negative.
(4) All rows of M^n converge to π as $n \to \infty$.

Theorem 3.7 (Perron-Frobenius II). *Let M be a stochastic matrix. Then:*

(1) *1 is an eigenvalue of maximum modulus.*
(2) *The multiplicity of this eigenvalue is the number of irreducible classes.*
(3) *For each class C, the eigenvector associated to the eigenvalue 1 is the principal eigenvector for M_C, the chain restricted to C.*
(4) *If C is irreducible with period l, then M_C has l eigenvalues of modulus 1, namely $e^{2\pi ij/l}$ for $0 \le j \le l-1$.*
(5) *$n^{-1}\sum_{j=1}^{n}(M_C)^j$ converges.*

In the irreducible case, π, the left eigenvector corresponding to the eigenvalue 1, is the stationary distribution of the Markov chain. Let $\lambda_1, \dots, \lambda_k$ be the eigenvalues of M in the order of decreasing modulus. If M is aperiodic, then $|\lambda_2| < \lambda_1 = 1$, and if all eigenvalues are simple, then we have an eigenbasis $\{v_j\}$ with $v_1 = \pi$. For any initial state i we can write $e_i = \sum_{j=1}^{k} c_{ij} v_j$, where $c_{i1} = 1$. Using $||\pi - v||$ to denote the (half) total variation distance $(1/2)\sum_{m=1}^{k}|\pi(m) - v(m)|$, we obtain an exponential rate of convergence:

$$2||\pi - e_i M^n|| = 2\left\|\pi - \sum_{j=1}^{k} c_{ij}\lambda_j^n v_j\right\| = \sum_{j=2}^{k}|c_{ij}\lambda_j^n v_j| \le \max_{i,j}|c_{ij}|\sum_{j=2}^{k}|\lambda_j|^n \le C|\lambda_2|^n, \tag{3.4}$$

where C is independent of the initial state i. Furthermore, spectral theory shows that when M is symmetric, $2||\pi - \mu M^n|| \le \sqrt{k}|\lambda_2|^n$.

Example (Random walk on the circle). The transition matrix of any random walk on $\mathbb{Z}/n\mathbb{Z}$ is **circulant**, meaning that the rows are successive cyclic permutations of each other. The eigenvectors of an n by n circulant matrix are always

$$\begin{gathered}(1,1,1,\dots,1), (1,\zeta,\zeta^2,\dots,\zeta^{n-1}), (1,\zeta^2,\zeta^4,\dots,\zeta^{2(n-1)}),\\ \dots,(1,\zeta^{n-1},\zeta^{2(n-1)},\dots,\zeta^{(n-1)(n-1)}),\end{gathered} \tag{3.5}$$

where ζ is a primitive n-th root of unity. The eigenvalues of a circulant matrix whose first row is $(a_0, \dots, a_{n-1})$ are $\sum_{j=0}^{n-1}\zeta^{jk}a_j$, $0 \le k \le n-1$. For a SRW, $a_1 = a_{-1} = 1/2$ are the only non-zero values in the first row, and hence the eigenvalues are

$$\frac{1}{2}\left(\zeta^k + \zeta^{-k}\right) = \cos\left(\frac{2\pi k}{n}\right), \ 0 \le k \le n-1. \tag{3.6}$$

When n is odd, the cyclic group of units are not quite symmetric about the imaginary axis, but nevertheless λ_2 is very close to $\cos(2\pi/n)$ for large n. In particular, $1-\lambda_2 \sim 2\pi^2/n^2$, which gives one explanation for why the mixing time for the random walk is of order n^2, i.e. $|\lambda_2|^{n^2} \sim (1-c/n^2)^{n^2} \to e^{-c}$.

The **Metropolis-Hastings algorithm** is a Markov chain Monte Carlo method for obtaining a sequence of random states from a probability distribution for which direct sampling is difficult. Pick any D-regular graph on S. If $X_n = x$, choose y uniformly from among the neighbors of x. If $w(y) \ge w(x)$ (where w assigns an easily computable weight), let $X_{n+1} = y$. If $w(y) < w(x)$, let $X_{n+1} = y$ with probability $w(y)/w(x)$ and $X_{n+1} = x$ otherwise. Note that

$$w(x)p(x,y) = w(y)p(y,x) = \frac{\min\{w(x), w(y)\}}{D}. \tag{3.7}$$

4. Countable Markov Chains

Let $\tau_x^0 = 0$, $\tau_x^k = \inf\{n > \tau_x^{k-1} : X_n = x\}$, $\tau_x = \tau_x^1$, and as before $\rho_{xy} = \mathbb{P}_x(\tau_y < \infty)$. Let $N(x) = \sum_{n=0}^{\infty} \mathbf{1}_{X_n=x}$ be the number of visits to x. By the Markov property, $\rho_{xx} = 1$ implies $\mathbb{P}_x(N(x) = \infty) = 1$, and by the strong Markov property, $\mathbb{P}_x(\tau_y^k < \infty) = \rho_{xy}\rho_{yy}^{k-1}$. The latter is intuitive because in order to make k visits to y from x, we have to first go from x to y and then return $k-1$ times to y.

When the state space S is countably infinite, there may be no recurrent states, in which case the chain is said to be transient. Since $\mathbb{E}_x N(y) = \rho_{xy}/(1-\rho_{yy})$, x is recurrent if and only if $\mathbb{E}_x N(x) = \infty$. A recurrent state x is said to be **positive recurrent** if $\mathbb{E}_x\tau_x < \infty$ and **null recurrent** if $\mathbb{E}_x\tau_x = \infty$.

Definition 4.1. *A measure μ is said to be a* **stationary measure** *if*

$$\sum_{x\in S} \mu(x)p(x,y) = \mu(y) \tag{4.1}$$

for all $y \in S$. If μ is a probability measure, we call μ a **stationary distribution**.

Remark. If μ is a stationary measure, then by (4.1) $\mathbb{P}_\mu(X_1 = y) = \mu(y)$, so by the Markov property and induction $\mathbb{P}_\mu(X_n = y) = \mu(y)$ for all $n \geq 1$. In addition, if μ is a stationary distribution, then it represents a possible equilibrium for the chain.

Theorem 4.2 (Markov chain convergence)**.** *For an irreducible, aperiodic Markov chain with stationary distribution π, $p^n(x,y) \to \pi(y)$ as $n \to \infty$ (here p^n denotes the n-step transition probability).*

Definition 4.3. *A function $h : S \to \mathbb{R}$ is* **harmonic** *at x if $h(x) = \sum_{y\in S} p(x,y)h(y)$. We say that h is harmonic if it is harmonic at every x. Similarly, we say that h is* **dual harmonic** *at y if $h(y) = \sum_{x\in S} h(x)p(x,y)$ and that h is dual harmonic if it is dual harmonic at every y.*

Remark. Harmonicity in this context is essentially a discrete analogue of the classic notion of harmonicity; in both cases, harmonic functions satisfy a mean value property.

The Greens function is defined by $G(x,y) = \mathbb{E}_x N(y)$. It is not difficult to see that $G(\cdot, y)$ is harmonic at x. For any transient Markov chain with starting point α, we define the **Martin kernel** by $K(x,y) = G(x,y)/G(\alpha,y)$. Also define the capacity by

$$\mathrm{Cap}_K(A) = \frac{1}{\inf_\mu \mathcal{E}_K(\mu)}. \tag{4.2}$$

For the rest of this section, a stopping time τ is allowed to be 0 unless otherwise specified.

Theorem 4.4. *Let X_n be a transient Markov chain with $X_0 = \alpha$. For any set Λ,*

$$\frac{1}{2}\mathrm{Cap}_K(\Lambda) \leq \mathbb{P}_\alpha(\tau_\Lambda < \infty) \leq \mathrm{Cap}_K(\Lambda). \tag{4.3}$$

Proof. (sketch) Let ν be the harmonic subprobability measure on Λ. Since energy is given by $\mathcal{E}(\nu) = \sum_y \nu(y)V_\nu(y)$, $\mathcal{E}_K(\nu) = ||\nu||$. As energy is a homogeneous quadratic, letting $\mu = \nu/||\nu||$ yields $\mathcal{E}_K(\mu) = 1/||\nu||$. Thus $\mathrm{Cap}(\Lambda) \geq 1/\mathcal{E}_K(\mu) = ||\nu||$, which is equivalent to the second inequality.

For the first inequality, let μ be any probability measure supported on Λ, and define $Z = \sum_{y\in\Lambda} \mu(y)N(y)/G(\alpha,y)$. By the second moment method, $\mathbb{P}(Z \neq 0) \geq \mathbb{E}Z^2/(\mathbb{E}Z)^2$

for any non-negative Z. Computations show that $\mathbb{E}_\alpha Z = 1$ and $\mathbb{E}_\alpha Z^2 \leq 2\mathcal{E}_K(\mu)$, from which it follows that

$$\mathbb{P}_\alpha(\tau_\Lambda < \infty) = \mathbb{P}_\alpha(Z \neq 0) \geq \frac{1}{2\mathcal{E}_K(\mu)}. \tag{4.4}$$

The RHS is equal to $\mathrm{Cap}_K(\Lambda)/2$ when μ is the energy minimizing measure. □

Corollary 4.5. *Define* $\mathrm{Cap}_K^\infty(\Lambda) = \inf\{\mathrm{Cap}_K(\Lambda \setminus A) : |A| < \infty\}$. *Then*

$$\frac{1}{2}\mathrm{Cap}_K^\infty(\Lambda) \leq \mathbb{P}_\alpha(X_n \in \Lambda \ i.o.) \leq \mathrm{Cap}_K(\Lambda). \tag{4.5}$$

Let H be any event for which $\mathbb{P}(H|\mathcal{F}_n) = \mathbb{P}(H|X_n)$, and let $h(x)$ be the function $\mathbb{P}_x(H)$. On the event that $X_n = x$,

$$\begin{aligned} \mu(X_{n+1} = y|\mathcal{F}_n) &= \frac{\mathbb{P}(X_{n+1} = y, H|\mathcal{F}_n)}{\mathbb{P}(H|\mathcal{F}_n)} = \frac{\mathbb{P}(X_{n+1} = y, H|X_n)}{\mathbb{P}(H|X_n)} \\ &= p(x,y)\frac{\mathbb{P}(H|X_{n+1} = y)}{\mathbb{P}(H|X_n = x)} = p(x,y)\frac{h(y)}{h(x)}. \end{aligned} \tag{4.6}$$

We call the Markov transition kernel $p_h(x,y) = p(x,y)h(y)/h(x)$ the h**-transform** of p. By the harmonicity of h, it is easy to verify that $\sum_{y\in S} p_h(x,y) = 1$.

A Markov chain on the graph $G = (V,E)$ satisfying

$$\sum_{j=1}^{n} p(x_j, x_{j+1}) = \sum_{j=1}^{n} p(x_{j+1}, x_j) \tag{4.7}$$

for all sets of vertices $x_1, \ldots, x_n$, where $j+1$ is evaluated mod n, is said to be **reversible**. This is equivalent to the existence of non-negative weights $\{c(e) : e \in E\}$, unique up to a constant multiple, such that $p(x,y) = c(x,y)/\sum_{z\in E} c(x,z)$, where the denominator is denoted $c(x)$. As the measure c satisfies $c(x)p(x,y) = c(y)p(y,x)$, it is in fact stronger than a stationary measure.

We can define an electrical network on G. For disjoint subsets A and Z of V, let a battery hold voltages constant at 1 and 0 respectively on A and Z, and let voltages $h(x)$ at the other vertices satisfy the following:

(1) There is a signed current flow $I : E \to \mathbb{R}$ such that

$$I(x,y) = -I(y,x) = c(x,y)[h(x) - h(y)]. \tag{4.8}$$

(2) Current flow is conserved: total current into x = total current out.

The voltage function can be shown to be the map $x \mapsto \mathbb{P}_x(\tau_A < \tau_Z)$.

Theorem 4.6. *Let A be the singleton set $\{a\}$, and define $\tau_a^+ = \inf\{k \geq 1 : X_k = a\}$.*

(1) Let c_{eff} be the effective conductance of the circuit. Then $\mathbb{P}_a(\tau_Z < \tau_a^+) = c(a)c_{\text{eff}}$.

(2) Let $\{G_n\}$ be a sequence of connected subgraphs converging upward to the infinite graph G, and let Z_n be the boundary of G_n in G. Let $\mathbb{P}^{(n)}$ denote the law of the Markov chain induced on G_n. Then

$$\mathbb{P}_a(\tau_a^+ = \infty) = \lim_{n\to\infty} \mathbb{P}_a^{(n)}(\tau_Z < \tau_a^+) = \frac{\lim_{n\to\infty} c_{\text{eff}}^{(n)}}{c(a)}. \tag{4.9}$$

Theorem 4.7 (Nash-Williams criterion). *Let $\{V_n\}$ be a partition of the vertices of an infinite graph according to distance from a. Let E_n be the set of edges between V_n and V_{n+1}, and define $C_n = \sum_{e \in E_n} c(e)$. Then $\sum_n 1/C_n = \infty$ implies that a is recurrent.*

5. Special Topic: Erdös-Kac Theorem

Define $\delta_p(n) = 1$ if $p \,|\, n$ and 0 otherwise. Then $\omega(n) = \sum_{p \le n} \delta_p(n)$ counts the number of distinct prime divisors of n. The $\delta_p(n)$ behave like independent random variables X_p with $\mathbb{P}(X_p = 1) = 1/p$ and $\mathbb{P}(X_p = 0) = 1 - 1/p$. It follows that $\omega(n)$ behaves like $\sum_{p \le n} X_p$, which has mean $\sum_{p \le n} 1/p$ and variance $\sum_{p \le n}(1/p)(1 - 1/p)$. Since

$$\sum_{p \le n} 1/p = \log\log n + O(1) \text{ and } \sum_p 1/p^2 < \infty, \tag{5.1}$$

the mean and variance of $\sum_{p \le n} X_p$ are both about $\log\log n$.

Recall that in the classical version of Central Limit Theorem, we take a sum, subtract off the mean, divide by the standard deviation, and end up with the standard normal as the limiting distribution. Heuristically, then, it would make sense for the number of distinct prime factors of a randomly chosen large integer n to be approximately normal with mean and variance $\log\log n$. This is formalized in the Erdös-Kac theorem.

Let P_n denote the uniform distribution on $\{1, \ldots, n\}$. For $A \subset \mathbb{N}$,

$$P_\infty(A) := \lim_{n \to \infty} P_n(m \le n : m \in A), \tag{5.2}$$

if it exists, is the **asymptotic density** of A.

Theorem 5.1 (Erdös-Kac). *As $n \to \infty$,*

$$P_n\left(m \le n : \frac{\omega(m) - \log\log n}{\sqrt{\log\log n}} \le x\right) \to \mathbb{P}(\chi \le x). \tag{5.3}$$

Example. Numbers near 10^{25} have, on average, 4 distinct prime factors. About 95% have no more than 8 distinct prime factors. This sheds some light on why factoring is so hard!

Proof. (sketch) Let $\alpha_n = n^{1/\log\log n}$. Define $\omega_n(m) = \sum_{p \le \alpha_n} \delta_p(m)$ and let E_n denote expected value w.r.t. P_n. Then

$$E_n(\omega - \omega_n) = E_n\left(\sum_{\alpha_n < p \le n} \delta_p\right) = \sum_{\alpha_n < p \le n} P_n(m : \delta_p(m) = 1) \le \sum_{\alpha_n < p \le n} 1/p. \tag{5.4}$$

It is not difficult to show that

$$\frac{\sum_{\alpha_n < p \le n} 1/p}{\sqrt{\log\log n}} \to 0. \tag{5.5}$$

This indicates that we can ignore the primes "near" n, i.e. it suffices to show that

$$P_n\left(m \le n : \frac{\omega_n(m) - \log\log n}{\sqrt{\log\log n}} \le x\right) \to \mathbb{P}(\chi \le x). \tag{5.6}$$

The strategy is to replace $\omega_n = \sum_{p \le \alpha_n} \delta_p$ with $S_n = \sum_{p \le \alpha_n} X_p$. Define

$$a_n = \mathbb{E}S_n = \sum_{p \le \alpha_n} 1/p \text{ and } b_n^2 = \mathrm{Var}(S_n) = \sum_{p \le \alpha_n} (1/p)(1 - 1/p). \tag{5.7}$$

Both are $\log\log n + o(\sqrt{\log\log n})$, so now it suffices to show that

$$P_n\left(m \le n : \frac{\omega_n(m) - a_n}{b_n} \le x\right) \to \mathbb{P}(\chi \le x). \tag{5.8}$$

By a well-known result on the moment problem, (5.8) will follow if we can demonstrate that $E_n((\omega_n - a_n)/b_n)^r \to \mathbb{E}\chi^r$ for all r.

An application of the Lindeberg-Feller theorem shows that $(S_n - a_n)/b_n \Rightarrow \chi$; in fact, it can be shown that $\mathbb{E}((S_n - a_n)/b_n)^r \to \mathbb{E}\chi^r$ for all r. Thus it remains to verify that $|\mathbb{E}(S_n - a_n)^r - E_n(\omega_n - a_n)^r| \to 0$ for all r. This follows from straightforward but tedious computations which make use of the following:

Fact 1: For any $\epsilon > 0$, $\alpha_n \le n^\epsilon$ for sufficiently large n. Hence $\alpha_n^r/n \to 0$ for all $r < \infty$.

Fact 2: If $p_1, \ldots, p_k$ are primes $\le n$, then $\mathbb{E}(X_{p_1} \cdots X_{p_k}) = 1/(p_1 \cdots p_k)$ (obvious) and $E_n(\delta_{p_1} \cdots \delta_{p_k}) = \lfloor n/(p_1 \cdots p_k) \rfloor / n$ (not difficult to see either). Thus their difference is bounded by $1/n$. □

E-mail address: shanshand@math.upenn.edu