Notes for my Oral Examination

Owen Mireles

May 24, 2020

I am typing these notes mainly as a means to an active review for my oral examination. They are not meant to be exhaustive, of course, and may contain inaccuracies, of course, and will *almost surely* have plenty of typos, of course. If something seems too simplified, I apologize beforehand: these notes are meant as a summary.

My major topic was Probability; my minor topic was Fractals. I will list bibliography on the fly.

1 Probability: Theory and Examples. 4th Edition. (Rick Durret)

I feel comfortable with the fundamentals of measure theory, so I will skip the context-setting aspect of introductory probability. Basic vocabulary is also omitted.

1.1 Some Probability Distributions

Important distributions: Gaussian, Poisson, Geometric, Exponential, Binomial.

What do they measure? What is a situation described by them?

- **Gaussian.** Describes distributions where one expects values to be concentrated around a certain mean, which is also its mode. Mathematically important because it is the central limit of sums of i. i. d. random variables.
- **Poisson.** A Poisson of parameter λ describes how many occurrences of an event happen in a fixed interval (i.e., in the next minute) if each occurrence is independent of the others, and the average rate of occurrence is known. This can be treated rigorously to see that the Poisson distribution is the limit of Binomial distributions.
- **Geometric.** If a trial is executed successfully with probability p, each trial independent of other, then the number of trials needed for one success is modelled by a geometric distribution of parameter p.
- **Exponential.** The continuous version of the Poisson distribution.
- **Binomial.** A binomial distribution of parameters (n, p) describes the probability of having k successes out of n tries, if you succeed independently, and with probability p, each try.

The following table summarizes the relevant information of these distributions.

Distribution	Params	Density Fn.	Mean	Var	Char. Fn.
Gaussian	μ, σ	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$	μ	σ^2	$e^{it\mu - \frac{1}{2}\sigma^2 t^2}$
Poisson	λ	$f(k) = e^{-\lambda}\frac{\lambda^k}{k!}$	λ	λ	$e^{\lambda(e^{it}-1)}$
Geometric	p	$f(k) = (1-p)^{k-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{p}{e^{-it}-(1-p)}$
Exponential	λ	$f(x) = \lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$(1 - it\lambda^{-1})^{-1}$
Binomial	n, p	$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$	$(1 - p + pe^{it})^n$

1.2 Weak and Strong Law of Large Numbers

The following are fundamental theorems in probability theory.

Theorem 1 (Weak Law of Large Numbers)**.** *Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}(X_1) = \mu$ and $\mathrm{Var}(X_1) = \sigma^2$. If $S_n = X_1 + \ldots + X_n$, then $S_n/n \to \mu$* ***in probability****.*

Proof. By Chebyshev's inequality, for a fixed $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|S_n/n - \mu| > \epsilon) \leq \lim_{n\to\infty} \frac{\mathrm{Var}(S_n)}{n^2\epsilon} = \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon} = 0.$$

□

Theorem 2 (Strong Law of Large Numbers)**.** *Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}(|X_1|) < \infty$ and $\mathbb{E}(X_1) = \mu$. If $S_n = X_1 + \ldots + X_n$, then $S_n/n \to \mu$* ***almost surely****.*

Proof. The more basic proof is a sequence of analysis arguments. More enlightening, and easier to remember, is the proof using (backwards) Martingales.

Let $S_n = X_1 + \ldots + X_n$, and define $Y_{-n} = S_n/n$, for $n \leq 0$. Also define $\mathcal{F}_{-n} = \sigma(S_n, X_{n+1}, X_{n+2}, \ldots)$. We claim that Y_{-n} is a reversed Martingale with respect to the filtration $\mathcal{F}_{-n}$. To see this,

$$\begin{aligned}
\mathbb{E}(Y_{-(n-1)}|\mathcal{F}_{-n}) &= \mathbb{E}((S_{n-1} + X_n - X_n)/(n-1)|\mathcal{F}_{-n}) \\
&= \frac{S_n}{n-1} - \mathbb{E}(X_n/(n-1)|\mathcal{F}_{-n}) \\
&= \frac{S_n}{n-1} - \frac{S_n}{n(n-1)} \text{ , by symmetry.} \\
&= \frac{S_n}{n} = Y_{-n}.
\end{aligned}$$

Since the mean of X_1 is finite, by Martingale convergence theorems, we know there is a limit. It is $\lim_{n\to\infty} S_n/n = \mathbb{E}(X_1|\mathcal{F}_\infty)$ almost surely . Since $\mathcal{F}_\infty$ is clearly made up of "exchangeable events", Hewitt-Savage's 0-1 Law tells us that $\mathbb{E}(X_1|\mathcal{F}_\infty) = \mu$.

□

1.3 Borel-Cantelli Lemmas

If A_n is a sequence of events (measurable sets), we define the event

$$\{A_n \text{ infinitely often}\} = \{A_n \text{ i.o.}\} = \limsup A_n = \bigcap_{n\geq 1} \bigcup_{m\geq n} A_m.$$

Theorem 3 (First Borel-Cantelli Lemma). *If* $\sum_{n\geq 1}\mathbb{P}(A_n)<\infty$*, then* $\mathbb{P}(A_n\ i.o.)=0$.

Proof. Notice that for any k,

$$\mathbb{P}(A_n \text{ i.o.}) \leq \mathbb{P}(\cup_{m\geq k}A_m) \leq \sum_{m\geq k}\mathbb{P}(A_m).$$

This holds for any $k\in\mathbb{N}$, and the series converges: we get the result. □

The complementary result requires independence.

Theorem 4 (Second Borel-Cantelli Lemma). *If the events* A_n *are mutually independent and* $\sum_{n\geq 1}\mathbb{P}(A_n)=\infty$*, then* $\mathbb{P}(A_n\ i.o.)=1$.

Proof. We recall the ever helpful inequality $1-x\leq e^{-x}$, when $x>0$. Being loose with notation, we get

$$\begin{aligned}
1-\mathbb{P}(A_n \text{ i.o.}) &= \mathbb{P}(A_n \text{ finitely o.}) = \mathbb{P}(\cup_{n\geq 1}\cap_{m\geq n}A_m^c) \\
&= \lim_{n\to\infty}\mathbb{P}(\cap_{m\geq n}A_m^c) \text{ , by monotone convergence} \\
&= \lim_{n\to\infty}\prod_{m\geq n}[1-\mathbb{P}(A_m)] \text{ , by independence} \\
&\leq \lim_{n\to\infty}\prod_{m\geq n}e^{-\mathbb{P}(A_m)} = \lim_{n\to\infty}e^{-\sum_{m\geq n}\mathbb{P}(A_m)} = 0.
\end{aligned}$$

□

1.4 Weak Convergence

We say r.v. $X_n\to X$ *in distribution* (or weakly) if $F_n(x)\to F(x)$ at every x which is a point of continuity of F (F_n is the distribution function of X_n). Equivalently, we have this type of convergence if, for every bounded measurable g we have $\mathbb{E}(g(X_n)\to\mathbb{E}(g(X))$ (which explains more clearly the reason behind calling it *weak convergence*).

Three theorems seem particularly helpful.

Theorem 5. *If* $X_n\to X$ *in distribution, there exist* Y_n *and* Y *with* $Y_n\overset{d}{=}X_n$ *and* $Y\overset{d}{=}X$*, such that* $Y_n\to Y$ *almost surely.*

Theorem 6 (Portmanteau's Theorem). *The following are equivalent.*

- $X_n\to X$ *in distribution.*
- *For every open* $A\subset\mathbb{R}$*, it is true that* $\liminf_{n\to\infty}\mathbb{P}(X_n\in A)\geq\mathbb{P}(X\in A)$.
- *For every closed* $B\subset\mathbb{R}$*, it is true that* $\limsup_{n\to\infty}\mathbb{P}(X_n\in B)\leq\mathbb{P}(X\in B)$.
- *For every measurable set* C *with* $\mathbb{P}(X\in\partial C)=0$*, we have* $\lim_{n\to\infty}\mathbb{P}(X_n\in C)=\mathbb{P}(X\in C)$.

Theorem 7 (Continuity Theorem). *Say* ϕ_n *is the characteristic function for* X_n *and* ϕ *is the one of* X*. If* $\phi_n\to\phi$ *as* $n\to\infty$ *point-wise, then* $X_n\to X$ *in distribution, provided that* ϕ *is continuous at 0.*

1.5 Central Limit Theorem

Another quite important theorem is the Central Limit Theorem.

Theorem 8 (Classic Central Limit Theorem)**.** *Let $X_1, X_2, \ldots$ be i.i.d. with mean μ and variance σ^2. Define $S_n = X_1 + \ldots + X_n$. Then, $S_n/\sqrt{n}$ converges in distribution to $\mathcal{N}(\mu, \sigma)$, a normal distribution with mean μ and standard deviation σ.*

Proof. Taking $(X_n - \mu)/\sigma$ instead of X_n, it suffices to show that $S_n/\sqrt{n} \to \mathcal{N}(0,1)$ in distribution when the X_i have mean 0 and standard deviation 1.

Let ϕ_k be the characteristic function of X_k. Since

$$\phi_k(t) = \mathbb{E}(e^{itX_k}) = \sum_{m \geq 0} \mathbb{E}\left(\frac{(itX_k)^m}{m!}\right),$$

we can write $\phi_k(t) = 1 - \frac{t^2}{2} + o(t^2)$ as $t \to 0$. If ϕ is the ch. fn. of $S_n/\sqrt{n}$, then

$$\phi(t) = \prod_{k=1}^{n} \phi_k\left(\frac{t}{\sqrt{n}}\right) = \phi_1\left(\frac{t}{\sqrt{n}}\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \longrightarrow e^{-t^2/2} \qquad \text{as } n \to \infty .$$

The ch. fn. of $\mathcal{N}(0,1)$ is $e^{-t^2/2}$, which completes the proof. □

Slightly more general, and sometimes useful, is the following. Recall that $\mathbb{E}(X; A) = \mathbb{E}(X \cdot \mathbb{1}_A)$

Theorem 9 (Lindeberg-Feller Theorem)**.** *For each n, let $X_{n,1}, X_{n,2}, \ldots, X_{n,n}$ be independent with $\mathbb{E}(X_{n,m}) = 0$ and $\mathrm{Var}(X_{n,m}) = \sigma^2_{n,m}$. If*

- $\lim_{n\to\infty} \sum_{m=1}^{n} \sigma^2_{n,m} = \sigma^2 > 0$, *and*
- *For every $\epsilon > 0$, we have $\sum_{m_1}^{n} \mathbb{E}(X^2_{n,m}; |X_{n,m}| > \epsilon) \longrightarrow 0$ as $n \to \infty$,*

then $S_n = X_{n,1} + \ldots + X_{n,n} \to \sigma^2 \mathcal{N}(0,1)$ in distribution as $n \to \infty$.

1.6 Local Limit Theorem

We say a r.v. X has *lattice distribution* if there exist $b, h \in \mathbb{Z}$ such that $\mathbb{P}(X \in b + h\mathbb{Z}) = 1$. The minimum number h that satisfies this condition is called the *span of* X.

Theorem 10 (Local Limit Theorem)**.** *Let $X_1, \ldots$ be i.i.d. with $\mathbb{E}(X_1) = 0$, $\mathrm{Var}(X_1) = \sigma^2$, and a common lattice distribution of span h. If $\mathbb{P}(X_1 \in b + h\mathbb{Z}) = 1$, define $S_n = X_1 + \ldots + X_n$ and $\mathcal{L}_n = \{(nb + h\mathbb{Z})/\sqrt{n}\}$. Also make*

$$p(x) = \mathbb{P}\left(\frac{S_n}{\sqrt{n}} = x\right), \qquad \text{for } x \in \mathcal{L}_n,$$

and

$$n(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}.$$

Then,

$$\sup_{x \in \mathcal{L}_n} \left| \frac{\sqrt{n}}{h} p(x) - n(x) \right| \longrightarrow 0, \qquad \text{as } n \to \infty.$$

1.7 Martingales

A martingale can be thought of as a *fair game*. The rigorous way of stating this is via conditional expectation.

1.7.1 Conditional Expectation

Let $\mathcal{G} \subset \mathcal{F}$ be two σ-algebras.

If X is $\mathcal{F}$-measurable, there is a $\mathcal{G}$-measurable r.v. Y such that $\int_A Y d\mathbb{P} = \int_A X d\mathbb{P}$ for any $A \in \mathcal{G}$. It is denoted by $\mathbb{E}(X|\mathcal{G})$; it is called *conditional expectation.*

By Radon-Nikodym's Theorem, it always exists. If $\nu(\cdot) = \int_{(\cdot)} X d\mathbb{P}$, then ν is a measure supported $\mathcal{G}$. Is is clear that $\nu \ll \mathbb{P}$ when $\mathbb{P}$ is restricted to $\mathcal{G}$. Hence, there exists $d\nu/d\mathbb{P}$, and we call it $\mathbb{E}(X|\mathcal{G})$, such that

$$\int_A X d\mathbb{P} = \nu(A) = \int_A (d\nu/d\mathbb{P}) d\mathbb{P} = \int_A \mathbb{E}(X|\mathcal{G}) d\mathbb{P}.$$

There are at least four useful properties of the conditional expectation.

- Linearity. $\mathbb{E}(aX + b|\mathcal{G}) = a\mathbb{E}(X|\mathcal{G}) + b$.
- Monotonicity. If $X \leq Y$, then $\mathbb{E}(X|\mathcal{G}) \leq \mathbb{E}(Y|\mathcal{G})$.
- If X is $\mathcal{G}$-measurable and Y is $\mathcal{F}$-measurable, then $\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G})$.
- Jensen's Inequality. If ϕ is convex, and both $\mathbb{E}|X| < \infty$ and $\mathbb{E}|\phi(X)| < \infty$, then $\phi(\mathbb{E}(X|\mathcal{G})) \leq \mathbb{E}(\phi(X)|\mathcal{G})$. In particular, if $p > 1$, then $\mathbb{E}(|\mathbb{E}(X|\mathcal{G})|^p) \leq \mathbb{E}(|X|^p|\mathcal{G})$.

1.7.2 Martingale Convergence

Definition 1 (Martingales and Sub/Supermartingales)**.** *Let $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots$ be a filtration of σ-algebras. Let X_n be $\mathcal{F}_n$-measurable. If $\mathbb{E}(X_n|\mathcal{F}_{n-1}) = X_{n-1}$, we say $\{X_n\}$ is a* ***martingale****. If $\mathbb{E}(X_n|\mathcal{F}_{n-1}) \leq X_{n-1}$, it is a* ***supermartingale****; if $\mathbb{E}(X_n|\mathcal{F}_{n-1}) \geq X_{n-1}$, it is a* ***submartingale****.*

Definition 2 (Stopping Times)**.** *N is a* ***stopping time*** *if it is positive-integer-valued and $\{N = n\} \in \mathcal{F}_n$.*

Theorem 11. *If N is a stopping time and X_n is a (super/sub) martingale, then $X_{N\wedge n}$ is a (super/sub) martingale.*

The following theorem is key. Many posterior results rely on it.

Theorem 12 (Martingale Convergence Theorem)**.** *If X_n is a submartingale with finite $\sup_n \mathbb{E}(X_n^+)$), then there exists a r.v. X, with $E|X| < \infty$, such that $X_n \to X$ a.s.*

Proof. Let U_n denote the "up-crossings" of $[a, b]$ that X_n completes. Geometrically and intuitively, we note that

$$(b - a)\mathbb{E}U_n \leq |a| + \mathbb{E}X_n^+.$$

Hence, $U_n < \infty$ a.s. As this is true for every a, b, we have

$$\mathbb{P}\left\{\bigcup_{a,b\in\mathbb{Q}} \liminf X_n \leq a < b \leq \limsup X_n\right\} = 0.$$

Thus, point-wise, there is a limit X a.s.

□

1.7.3 Doob's Inequality and L^p Convergence

We list four important theorems.

Theorem 13. *If X_n is a submartingale, and N is a stopping time with $\mathbb{P}(N \leq k) = 1$ for some k, we have*

$$\mathbb{E}X_0 \leq \mathbb{E}X_N \leq \mathbb{E}X_k.$$

Theorem 14 (Doob's Inequality)**.** *Let X_m be a submartingale. Define $M_n = \max_{1 \leq m \leq n} X_n^+$; for $\lambda > 0$, define $A = \{M_n \geq \lambda\}$. Then,*

$$\lambda \mathbb{P}(A) \leq \mathbb{E}(X_n \mathbb{1}_A) \leq \mathbb{E}X_n^+.$$

Using Holder's theorem, Fubini's theorem, a bit of analysis, and some cleverness, we can conclude:

Theorem 15 (Doob's L^p Maximum Inequality)**.** *Let X_n be a submartingale. With the previous notation, for $\infty > p > 1$,*

$$\mathbb{E}(M_n^p) \leq \left(\frac{p}{p-1}\right)^p (\mathbb{E}(X_n^+))^p.$$

Martingale convergence and L^p convergence imply:

Theorem 16 (L^p convergence theorem)**.** *If X_n is a martingale with $\sup \mathbb{E}|X_n| < \infty$, where $1 < p < \infty$, there exists a X in L^1 such that $X_n \to X$ a.s. and in L^p.*

1.7.4 Uniform Integrability and L^1 Convergence

Definition 3. *A family of r.v. $\{X_n\}$ is* ***uniformly integrable*** *if, for every $\epsilon > 0$, there exists a $K > 0$ such that, for all n, we have*

$$\mathbb{E}(|X_n| \mathbb{1}_{|X_n| > K}) < \epsilon.$$

Martingales are prime examples of uniformly integrable families.

Theorem 17. *If X is $\mathcal{F}$-measurable and in L^1, then the family of r.v. given by*

$$\{\mathbb{E}(X|\mathcal{G}) : \mathcal{G} \subset \mathcal{F} \text{ is a sub } \sigma\text{-algebra}\}$$

is uniformly integrable.

One useful characterization is the following.

Theorem 18. *If $X_n \to X$ in probability, then the following are equivalent:*

- *$X_n \to X$ in L^1.*
- *The X_n are uniformly integrable.*
- *$\mathbb{E}|X_n| \to \mathbb{E}|X| < \infty$.*

We have three useful theorems.

Theorem 19. *Suppose $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, then $\mathbb{E}(X|\mathcal{F}_n) \to \mathbb{E}(X|\mathcal{F}_\infty)$ a.s. and in L^1.*

A consequence of this is:

Theorem 20 (Lévy's 0-1 Law)**.** *If $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ and $A \in \mathcal{F}_\infty$, then $\mathbb{E}(\mathbb{1}_A|\mathcal{F}_n) \to \mathbb{1}_A$ a.s.*

Theorem 21 (Dominated Convergence for Conditional Expectation)**.** *Suppose $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, $X_n \in \mathcal{F}_n$ with $X_n \to X$ a.s., and that there exists $Y \in L^1$ such that $|X_n| \leq Y$, then*

$$\mathbb{E}(X_n|\mathcal{F}_n) \to \mathbb{E}(X|\mathcal{F}_\infty) \text{ a.s.}$$

1.7.5 Optional Stopping Times

A stopping time can be thought of as an "optionally terminated" process, as you may choose to keep effectuating the stochastic process or not. To be honest, I am not sure why this subsection is called this way other than the *somewhat superficial* reason I just gave. However, there are two important theorems relating to stopping.

Theorem 22. *If N is a stopping time and X_n is a uniformly integrable submartingale, then $X_{N\wedge n}$ is also uniformly integrable.*

To prove the main theorem of this section, this result is useful.

Theorem 23. *Given a uniformly integrable submartingale X_n, if N is a stopping time, we have*

$$\mathbb{E}X_0 \leq \mathbb{E}X_N \leq \mathbb{E}X_\infty,$$

where $X_\infty = \lim_{n\to\infty} X_n$.

Theorem 24 (Optional Stopping Theorem)**.** *Let $L \leq M$ be two stopping times, and $X_{n\wedge M}$ be a uniformly integrable submartingale. Then $\mathbb{E}(X_L) \leq \mathbb{E}(E_M)$. Further,*

$$X_L \leq \mathbb{E}(X_M|\mathcal{F}_L).$$

A recurrent case where this theorem is used is when X_n is actually a martingale and $L = 0$

1.8 Finite Markov Chains

This section was done following the exposition of *Markov Chains and Mixing Times. 1st Edition. (David A. Levin, Yuval Peres, and Elizabeth L. Wilmer).*

Definition 4 (Markov Chain)**.** *A Markov chain is a stochastic process $X_1, X_2, \ldots$ where*

$$\mathbb{P}(X_i = a|X_{i-1} = b_1, \ldots, X_1 = b_{i-1}) = \mathbb{P}(X_i = a|X_{i-1} = b_1).$$

Hence, if X_1 can only take n different values (meaning the "state space" of the Markov chain has n states), the complete information of the Markov chain can be visualized in an $n \times n$ matrix P, a Markov transition matrix. The condition $\sum_{j=1}^{n} P_{ij} = 1$ must be satisfied for every $1 \leq i \leq n$. Such a matrix is called a *stochastic matrix.* Each row of such a matrix defines a probability measure, which we call P_i. It seems to be customary to use x, y, z for naming states, instead of i, j, k.

Something useful to note and remember is that $P^n(x, y)$ represents the probability of starting at state x and being at state y after precisely n steps.

1.8.1 Stationary Distribution

If for every pair of states x, y there is a $n > 0$ such that $P^n(x, y) > 0$, we say the chain is **irreducible**.

A **return time** is a time $n \geq 1$ such that $P^n(x, x) > 0$. The period of the state x is the gcd of all return times. If every period is 1, we say the chain is **aperiodic**.

Definition 5 (Stationary Distribution). *If π is a measure on state space Ω that satisfies $\pi = \pi P$, or, more explicitly,*

$$\pi(y) = \sum_{x \in \Omega} \pi(x) P(x, y),$$

we say π is a ***stationary distribution****.*

The following is a fundamental fact that allows the study of mixing times.

Theorem 25 (Existence and Uniqueness of a Stationary Distribution). *Let P be be the transition matrix for an irreducible Markov chain. Then, there exists π, a probability distribution on Ω, such that*

- *$\pi = \pi P$ and $\pi(x) > 0$ for every $x \in \Omega$.*
- *If τ_x^+ is the first return time to x, then $\pi(x) = 1/(\mathbb{E}_x(\tau_x^+))$.*

Also, π is unique.

1.8.2 Reversibility

Definition 6 (Detailed Balance Equations). *Let π be a probability on Ω. For all $x, y \in \Omega$, the expressions*

$$\pi(x) P(x, y) = \pi(y) P(y, x)$$

are called the detailed balance equations.

If π satisfies the detailed balance equations, it is a stationary measure for P.

We can observe the following, which is suggestive of the title of this subsection,

$$\mathbb{P}_\pi(X_0 = x_0, \ldots, X_n = x_n) = \mathbb{P}_\pi(X_0 = x_n, \ldots, X_n = x_0).$$

An irreducible chain P having a distribution π which satisfies the detailed balance equations is called **reversible**.

Definition 7 (Time Reversal of a Markov Chain). *Let P be an irreducible chain with stationary distribution π. The* ***time reversal*** *of P is the chain with matrix*

$$\widehat{P}(x, y) = \frac{\pi(y) P(y, x)}{\pi(x)}$$

When a Markov chain is reversible, we have $P = \widehat{P}$.

1.8.3 Convergence to Uniformity

The following is a quantitative version of a very important qualitative result.

Theorem 26 (Convergence Theorem). *Let TV be the total variation distance. Suppose that P is irreducible and aperiodic with stationary distribution π. Then there exist constants $0 < \alpha < 1$ and $C > 0$ such that*

$$\max_{x \in \Omega} ||P^n(x, \cdot) - \pi||_{TV} \leq C\alpha^n.$$

In particular, P^n converges to its stationary distribution as $n \to \infty$.

2 The Probabilistic Method. 4th Edition. (Noga Alon and Joel H. Spencer)

Briefly, the probabilistic method is using probability theory to solve problems in combinatorics. The focus of the probabilistic method is learning *how to do*, rather than learning theory. This hands-on aspect is even more emphasized here than in other areas of math. As such, each section will list the main theoretical tools of a given chapter and illustrate one example showing them in action.

2.1 The Basic Method

Key Idea 1. *To assert something exists/is possible, show the probability of it happening is positive.*

Example 1.

Let $R(n, k)$ be the Ramsey numbers.

We want to show that if $\binom{n}{k} \cdot 2^{1-\binom{k}{2}} < 1$, then $R(k, k) > n$.

Let G be a complete graph with n vertices. We color the edges either red or blue independently and with probability $1/2$. Let I be a subset of k vertices. Let A_I denote the event where I is a monochromatic k-clique. Thus, the probability that there is a monochromatic k-clique is

$$\mathbb{P}(\cup_{|I|=k} A_I) \leq \sum_{|I|=k} \mathbb{P}(A_I) = \sum_{|I|=k} \frac{2}{\binom{k}{2}} = \binom{n}{k} \cdot 2^{1-\binom{k}{2}} < 1.$$

Hence, it is possible that there are no monochromatic k-cliques, and $R(k, k) > n$.

From this, it follows that $R(k, k) > \lfloor 2^{k/2} \rfloor$.

2.2 Linearity of Expectation

Key Idea 2. *Expectation is a linear functional. Also, if* $\mathbb{E}(X) = \mu$, *then there are realizations* ω_1, ω_2 *such that* $X(\omega_1) \geq \mu$ *and* $X(\omega_2) \leq \mu$.

Example 2.

Let $v_1, \ldots, v_n \in \mathbb{R}^n$ be unit-length vectors. We claim there exist $\epsilon_1, \ldots, \epsilon_n$ with values ± 1 such that $|\sum \epsilon_i v_i| \leq \sqrt{n}$.

Choose each ϵ_i to be ± 1 with probability $1/2$ and independently. If $X = |\sum \epsilon_i v_i|$, then

$$\mathbb{E}(X^2) = \sum_{i,j} \mathbb{E}(\epsilon_i \epsilon_j) v_i \cdot v_j = \sum_{i=1}^{n} \mathbb{E}(\epsilon_i^2) |v_i|^2 = n.$$

The result follows.

2.3 Alterations

Key Idea 3. *Sometimes a problem is too specific. Un-restrain the conditions a bit, use the probabilistic method, and then tweak the result deterministically.*

Example 3.

For a set S of n points in the unit square U, let $T(S)$ be the minimum area of a triangle whose three vertices are distinct points of S. We claim there is a set S of n points in U such that $T(S) \geq 1/(100n^2)$.

Let P, Q, R be three points uniformly chosen from U. Let μ be the area of the triangle they determine. We wish to bound from above the quantity $\mathbb{P}(\mu < \epsilon)$.

Let x be the distance between P and Q. We have $\mathbb{P}(b \leq x \leq b+\Delta b) \leq \pi((b+\Delta b)^2 - b^2)$, and when $\Delta b \to 0$, we get $\mathbb{P}(b \leq x \leq b + db) \leq \pi b 2 db$. If $x = b$ and $\mu < \epsilon$, then R must be in a strip of width $4\epsilon/b$ and length at most $\sqrt{2}$. Hence, knowing that b is bounded by $\sqrt{2}$,

$$\mathbb{P}(\mu < \epsilon) \leq \int_0^{\sqrt{2}} 2\pi b \left(\frac{4\epsilon}{b}\sqrt{2}\right) db \leq 16\pi\epsilon.$$

Now, let $P_1, \ldots, P_{2n}$ be uniformly and independently chosen points in U; let X be the number of triangles $P_i P_j P_k$ with area less than $1/100n^2$. The probability of this happening is at most $0.6n^{-2}$. Hence,

$$\mathbb{E}(X) \binom{n}{3} (0.6n^{-2}) < n.$$

Thus, there exists a set of $2n$ points in which at most n triangles have area less than $1/100n^2$. *Removing one point from each of these triangles* (this is the **alteration**) leaves us with a set of at least n points in which no triangle has area less than $1/100n^2$.

2.4 Second Moment Method

Key Idea 4. *Sometimes, looking at only expected values and probabilities by themselves is not enough. Chebyshev's inequality is also helpful. Any argument that relies on computing a variance and using Chebyshev's inequality is called the "second moment method".*

Theorem 27 (Chebyshev's Inequality)**.** *For any $\lambda > 0$, we have*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{\mathrm{Var}(X)}{\lambda^2}.$$

Example 4.

Here is a result from number theory. Let $\nu(n)$ be the number of distinct primes p dividing n. Let $\omega(n) \to \infty$ arbitrarily slowly. The the numbers of x in $\{1, \ldots, n\}$ such that

$$|\nu(x) - \ln\ln n| > \omega(n)\sqrt{\ln\ln n}$$

is $o(n)$.

Let x be randomly chosen from $\{1, \ldots, n\}$. For p prime, set $X_p = 1$ if $p|x$ or $X_p = 0$ otherwise. Set $M = n^{1/10}$, and make $X = \sum_{p \leq M} X_p$. At most 10 different numbers less than M can divide $x \leq n$, so that $\nu(x) - 10 \leq X(x) \leq \nu(x)$, so bounds on X translate to bounds on ν.

Observe that

$$\mathbb{E}(X_p) = \frac{\lfloor n/p \rfloor}{n} = \frac{1}{p} + O\left(\frac{1}{n}\right).$$

Hence,

$$\mathbb{E}(X) = \ln\ln n + O(1),$$

where we are using the number theoretic fact that $\sum_{p\leq m} = \ln\ln m + O(1)$.

To compute the variance, we have

$$\mathrm{Var}(X) = \sum_{p\leq M} \mathrm{Var}(X_p) + \sum_{p\neq q} \mathrm{Cov}(X_p, X_q).$$

Further, $X_p^2 = X_p$, so $\mathrm{Var}(X_p) = (1/p)(1-1/p) + O(1/n)$, and

$$\sum_{p\leq M} \mathrm{Var}(X_p) \leq \sum_{p\leq M} \frac{1}{p} + O(1) = \ln\ln n + O(1).$$

For the covariances, $X_pX_q = 1$ only if $pq|x$, hence

$$\mathrm{Cov}(Xp, X_q) = \mathbb{E}(X_pX_q) - \mathbb{E}(X_p)\mathbb{E}(X_q) \leq \frac{1}{n}\left(\frac{1}{p} + \frac{1}{q}\right).$$

Thus,

$$\sum_{p\neq q} \mathrm{Cov}(X_p, X_q) \leq \frac{2M}{n} \sum \frac{1}{p} = o(1).$$

Similarly, $\sum_{p\neq q} \mathrm{Cov}(X_p, X_q) \geq -o(1)$. Thus, for any $\lambda > 0$,

$$\mathbb{P}(|X - \ln\ln n| > \lambda\sqrt{\ln\ln n}) \leq \lambda^{-2} + o(1).$$

Since $|X - \nu| \leq 10$, we get the result.

2.5 Lovász's Local Lemma

Key Idea 5. *When events of positive probability are independent, the probability of all of them happening, perhaps small, is still positive. Independence can be relaxed to "bounded dependence". Lovász's Local Lemma gives a quantitative way of doing this.*

While the Local Lemma is more general, one version that is used often is the following.

Theorem 28 (The Local Lemma; Symmetric Case)**.** *Let $A_1, A_2, \ldots, A_n$ be events in an arbitrary probability space. Suppose that each event A_i is mutually independent of a set of all the other A_j but at most d, and that $\mathbb{P}(A_i) \leq p$ for all $| \leq i \leq n$. If*

$$ep(d+1) \leq 1,$$

then $\mathbb{P}(\cap_{i=1}^n A_i^C) > 0$.

Example 5.

For a k-coloring of $\mathbb{R}$, which is a function $c : \mathbb{R} \to \{1, 2, \ldots, c\}$, we say a set $T \subset \mathbb{R}$ is *multi-colored* if $c|_T$ is surjective. The example is the following result.

Let m and k be two positive integers satisfying

$$e(m(m-1)+1)k\left(1 - \frac{1}{k}\right)^m \leq 1.$$

Then, for any set S of m real numbers, there is a k-coloring so that each translation $x+S$ (for any $x \in \mathbb{R}$) is multi-colored.

The solution is a combination of combinatorics and topology.

First, we fix a finite subset $X \subset \mathbb{R}$ and show the existence of a k-coloring so that each translation $x + S$ (for $x \in X$) is multi-colored.

We use the Local Lemma for this. Call $Y = \cup_{x \in X} x + S$, and the probability space is given by uniformly k-coloring the elements of Y. For each $x \in X$, the event A_x is "$x + S$ is **NOT** multi-colored". Then $\mathbb{P}(A_x) \leq k((k-1)/k)^m$ (use at most $k-1$ colors for the elements of $x + S$). Also, A_x is independent of A_y if $(x + S) \cap (y + S) = \varnothing$, so A_x is dependent on at most $m(m-1)$ events A_y. By the Local Lemma, there is a k-coloring that makes every $x + S$ multi-colored.

To extend this result, we employ a compactness argument. By Tikhonov's Theorem, the space $\{1, \ldots, k\}^{\mathbb{R}}$ is compact in the product topology. The elements of this space are k-colorings. In this space, for every fixed x, the set C_x of all colorings, such that $x + S$ is multi-colored, is closed. The previous paragraph showed that the intersection of finitely many C_x is non-empty. By compactness, and the "nested-compacts argument", the intersection of all C_x is non-empty. Any c in this intersection satisfies the property we were looking for.

2.6 Correlation Inequalities, FKG

Key Idea 6. *Results of the type $\mathbb{E}(XY) \gtrless \mathbb{E}(X)\mathbb{E}(Y)$ are useful.*

Definition 8 (Distributive Lattice)**.** *A **lattice** is a partially ordered set in which any two elements $x, y \in L$ have a unique minimal upper bound, denoted $x \vee y$, and a a unique maximal lower bound, denoted $x \wedge y$. A lattice is **distributive** if, for all $x, y, z \in L$,*

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z).$$

Theorem 29 (KFG Inequality)**.** *A function $\mu : L \to \mathbb{R}^+$, where L is a finite distributive lattice, is called* log-supermodular *if, for all $x, y \in L$,*

$$\mu(x)\mu(y) \leq \mu(x \vee y)\mu(x \wedge y).$$

Let L be a finite distributive lattice, and μ be log-supermodular. Then, for any two increasing functions $f, g : L \to \mathbb{R}^+$, we have

$$\left(\sum_{x \in L} \mu(x) f(x)\right)\left(\sum_{x \in L} \mu(x) g(x)\right) \leq \left(\sum_{x \in L} \mu(x) f(x) g(x)\right)\left(\sum_{x \in L} f(x)\right).$$

If f and g are both decreasing, the result is still true. If one of f and g is increasing and the other decreasing, the inequality is reversed.

If μ is a probability measure, then the KFG inequality says $\mathbb{E}(f)\mathbb{E}(g) \leq \mathbb{E}(fg)$.

Example 6.

Let $G(n, p)$ be the Erdös-Renyi graph. Let H be the event where G is Hamiltonian; and P denotes the event where G is planar. The FKG inequality can be used to prove that $\mathbb{P}(P \cap H) \leq \mathbb{P}(P)\mathbb{P}(H)$.

2.7 Martingale Methods

Key Idea 7. *Use martingales and Lipschitz functions to access the world of concentration inequalities.*

We use finite-time martingales whose domain Ω is $G(n,p)$.

First, some constructions.

The Edge Exposure Martingale. Let the random graph $G(n,p)$ be the underlying probability space. Label the potential edges $\{i,j\} \subset [n]$ by $e_1, \dots, e_m$, setting $m = \binom{n}{2}$, in any specific, but fixed, manner. Let f be any graph-theoretic function. We define a martingale $X_0, \dots, X_m$ by giving values $X_i(H)$, where H is sampled from $G(n,p)$. $X_m(H)$ is simply $f(H)$. $X_0(H)$ is $\mathbb{E}(f(G))$; it is a constant. In general,

$$X_i(H) = \mathbb{E}[f(G) | e_j \in G \iff e_j \in H, 1 \le j \le i].$$

The Vertex Exposure Martingale. Again, let $G(n,p)$ be the underlying probability space and f any graph theoretic function. Define $X_1, \dots, X_n$ by

$$X_i(H) = \mathbb{E}[f(G) | \text{ for } x, y \le i, \{x,y\} \in G \iff \{x,y\} \in H \ ].$$

Martingales give us access to stronger concentration inequalities than the Second Moment Method.

Theorem 30 (Azuma's Inequality)**.** *Let* $0 = X_0, \dots, X_m$ *be a martingale with*

$$|X_i - X_{i+1}| \le 1$$

for all $0 \le i < m$*. Let* $\lambda > 0$ *be arbitrary. Then*

$$\mathbb{P}(X_m > \lambda\sqrt{m}) < e^{-\lambda^2/2}.$$

Proof. Make $a = \lambda/\sqrt{m}$. Set $Y_i = X_i - X_{i-1}$, so that $|Y_i| \le 1$ and $\mathbb{E}(Y_i | X_{i-1}, \dots, X_0) = 0$. Then,

$$\mathbb{E}(e^{aY_i} | X_{i-1}, \dots, X_0) \le \cosh(a) \le e^{a^2/2}.$$

Then,

$$\begin{aligned}
\mathbb{E}(e^{aX_m}) &= \mathbb{E}\left(\prod_{i=1}^{m} e^{aY_i}\right) \\
&= \mathbb{E}\left(\prod_{i=1}^{m-1} e^{aY_i}\mathbb{E}(e^{aY_i} | X_{i-1}, \dots, X_0)\right) \le e^{a^2/2}\mathbb{E}\left(\prod_{i=1}^{m-1} e^{aY_i}\right) \\
&\le e^{ma^2/2}.
\end{aligned}$$

By Markov's inequality,

$$\begin{aligned}
\mathbb{P}(X_m > \lambda\sqrt{m}) &= \mathbb{P}(e^{aX_m} > e^{a\lambda\sqrt{m}}) \le e^{-a\lambda\sqrt{m}}\mathbb{E}(e^{aX_m}) \\
&\le e^{-a\lambda\sqrt{m}} e^{ma^2/2} = e^{-\lambda^2/2}.
\end{aligned}$$

□

We can generalize the setting more.

Fix a filtration

$$\varnothing = B_0 \subset B_1 \subset \cdots \subset B_m = B.$$

Let $L : A^B \to \mathbb{R}$ be a functional. We define a martingale $X_0, \ldots, X_m$ by setting

$$X_i(h) = \mathbb{E}(L(g) : g(b) = h(b) \text{ for all } b \in B_i).$$

We say the functional L satisfies the Lipschitz condition relative to the filtration if, for all $0 \le i \le m$,

$$h, h' \text{ differ only on } B_{i+1} - B_i \qquad \Longrightarrow \qquad |L(h') - L(h)| \le 1.$$

We can say:

Theorem 31. *Let L satisfy the Lipschitx condition. Then the corresponding martingale satisfies*

$$|X_{i+1}(h) - X_i(h)| \le 1$$

for all $0 \le i < m$, with $h \in A^B$.

And we finish with the more general version of Azuma's inequality.

Theorem 32 (General Azuma's Inequality)**.** *Let L satisfy the Lipschitz condition relative to a filtration of length m and let $\mu = \mathbb{E}(L(g))$. Then for all $\lambda > 0$,*

$$\mathbb{P}(L(g) \ge \mu + \lambda\sqrt{m}) < e^{-\lambda^2/2},$$

$$\mathbb{P}(L(g) \le \mu - \lambda\sqrt{m}) < e^{-\lambda^2/2}.$$

2.8 Chernoff Inequalities

Key Idea 8. *Markov's inequality can be exploited. Markov's inequality only works for non-negatives r.v. However, for any X, the key observation is that for any $\lambda > 0$, we have*

$$\mathbb{P}(X > a) \qquad \Longleftrightarrow \qquad \mathbb{P}(e^{\lambda X} > e^{\lambda a}) \le e^{-\lambda a}\mathbb{E}(e^{\lambda X}).$$

After this, we can optimize in λ.

Theorem 33. *Let $\mathbb{E}X = 0$ and $|X| \le 1$, then for any $\lambda > 0$, we have*

$$\mathbb{E}(e^{\lambda X}) \le \cosh \lambda.$$

Proof. Analyze the function

$$h(x) = \frac{e^\lambda + e^{-\lambda}}{2} + \frac{e^\lambda - e^{-\lambda}}{2}x.$$

Notice that, for $x \in [-1, 1]$, $e^{\lambda x} \le h(x)$, by convexity of the function $x \mapsto e^{\lambda x}$, since $(x, h(x))$ is a point on the chord joining $(-1, e^{-\lambda})$ and $(1, e^\lambda)$. Thus,

$$\mathbb{E}(e^{\lambda X}) \le \mathbb{E}(h(X)) = h(\mathbb{E}X) = h(0) = \cosh \lambda$$

□

3 Brownian Motion. 1st Edition. (Peter Mörters and Yuval Peres)

Going into every detail of this book would prove too ambitious. I try to focus on the main ideas.

3.1 Brownian Motion

Definition 9. *Brownian motion is a continuous-time $t \geq 0$ stochastic process $B(t)$ that satisfies the following three properties.*

- *Increments are independent. That is, for any sequence $t_1 < \ldots < t_n$, the r.v. $B(t_n) - B(t_{n-1}), \ldots, B(t_2) - B(t_1)$ are independent.*
- *Increments are stationary. That is, for any $t \geq 0$ and $h > 0$, the r.v. $B(t+h) - B(t)$ has mean 0 and variance h.*
- *Almost surely (meaning, for almost every $\omega \in \Omega$), $B(t)$ is continuous for all $t \geq 0$.*

If $B(0) = 0$, we say it is a standard linear (one dimensional) Brownian motion. Otherwise, we have to specify that $B(0) = x \in \mathbb{R}$ and say that Brownian motion is started at x.

Such an object, while containing many restrictions which a priori may imply contradictions, exists.

Theorem 34. *Brownian motion exists.*

Proof. We construct $B(t)$ for $t \in [0, 1]$, and then concatenate countably many of these. □

Brownian motion has invariance properties. Some basic ones are the following.

Theorem 35 (Scaling Invariance)**.** *Let $B(t)$ be a standard Brownian motion. For any $a > 0$, the process $X(t)$ defined by $X(t) = \frac{1}{a} B(a^2 t)$ is also a standard Brownian motion.*

Theorem 36 (Time Inversion)**.** *Let $B(t)$ be a standard Brownian motion. For any $a > 0$, the process $X(t)$ defined by $X(t) = tB(1/t)$, if $t > 0$, and $X(t) = 0$, if $t = 0$, is also a standard Brownian motion.*

According to the authors of the book, it is useful to think of Brownian motion as a *random fractal.* The following results justify this nomenclature.

Theorem 37 (α-Hölder, with $\alpha < 1/2$)**.** *If $\alpha < 1/2$ then, almost surely, Brownian motion is everywhere locally α-Hölder continuous.*

Theorem 38 (Nowhere Differentiability)**.** *Almost surely, for all $0 < a < b < \infty$, Brownian motion is not monotone on the interval $[a, b]$.*

Almost surely,

$$\limsup_{n \to \infty} \frac{B(n)}{\sqrt{n}} = +\infty, \qquad \textit{and} \qquad \liminf_{n \to \infty} \frac{B(n)}{\sqrt{n}} = -\infty.$$

Fix $t \geq 0$. Then, almost surely, Brownian motion is not differentiable at t. Moreover, $D^* B(t) = +\infty$ and $D_* B(t) = -\infty$, where D^* is the upper right derivative and D_* is the lower right derivative.

3.2 Markov Process

3.3 Dirichlet's Problem and Harmonic Measure

3.4 Dimension and Potential Theory

3.5 Random Walks and Brownian Motion