Appendix A

Some definitions and useful facts

These lemmas will eventually be inserted in the main text in a suitable place.

Binomial coefficients

Recall the following bounds on factorials and binomial coefficients:
\[ e \left( \frac{n}{e} \right)^n \leq n! \leq e \left( \frac{n + 1}{e} \right)^{n + 1} \]
\[ \frac{n^k}{k^k} \leq \binom{n}{k} \leq e^{k^k} \frac{n^k}{k^k} \]
\[ \frac{2^n}{n} = (1 + o(1)) \frac{4^n}{\sqrt{\pi n}} \]

and
\[ \log \binom{n}{k} = (1 + o(1)) n H(k/n) \]

where \( H(p) := -p \log p - (1 - p) \log(1 - p) \).
### A.1.2 Conditional expectation: definition and properties

Recall the definition of the conditional expectation (see e.g. [Wil91, Section 9.2]).

**Theorem A.1** (Conditional expectation). Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ a sub $\sigma$-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 $\mathbb{E}[X \mid \mathcal{G}]$.

In $L^2$ conditional expectation reduces to an orthogonal projection (see e.g. [Wil91, Section 9.4]).

**Theorem A.2** (Conditional expectation: $L^2$ case). Let $\langle X, Y \rangle := \mathbb{E}[XY]$. Let $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ a sub $\sigma$-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, $\langle Z, X - Y \rangle = 0, \ \forall Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$. Such $Y$ is called an orthogonal projection of $X$ on $L^2(\Omega, \mathcal{G}, \mathbb{P})$.

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 of the conditional expectation (see e.g. [Wil91, Section 9.7]).

**Lemma A.3** (Taking out what is known). If $Z \in \mathcal{G}$ is bounded then $\mathbb{E}[ZX \mid \mathcal{G}] = Z \mathbb{E}[X \mid \mathcal{G}]$. This is also true if $X, Z \geq 0$ and $\mathbb{E}[ZX] < +\infty$ or $X \in L^p(\mathcal{F})$ and $Z \in L^q(\mathcal{G})$ with $p^{-1} + q^{-1} = 1$ and $p > 1$.

**Lemma A.4** (Role of independence). If $X$ is independent of $\mathcal{H}$ then $\mathbb{E}[X \mid \mathcal{H}] = \mathbb{E}[X]$. In fact, if $\mathcal{H}$ is independent of $\sigma(\{X\}, \mathcal{G})$, then $\mathbb{E}[X \mid \sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X \mid \mathcal{G}]$.

**Lemma A.5** (Tower property (or law of total probability)). We have $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X]$. In fact, if $\mathcal{H} \subseteq \mathcal{G}$ is a $\sigma$-field

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

That is, the smallest $\sigma$-field wins.

The following fact will also prove useful (see e.g. [Dur10, Example 5.1.5] for a proof).
**Lemma A.6** (Conditioning on an independent RV). Suppose $X$ and $Y$ are independent. Let $\phi$ be a function with $\mathbb{E}|\phi(X,Y)| < +\infty$ and let $g(x) = \mathbb{E}(\phi(x,Y))$. Then,

$$\mathbb{E}(\phi(X,Y)|X) = g(X).$$

**A.1.3 A Taylor expansion**

To be written. See [LL10, Lemmas 12.1.1, 12.1.4].

**A.1.4 Spectral representation of reversible matrices**

Let $P$ be the transition matrix of a finite, irreducible Markov chain on $V$ reversible with respect to $\pi$. Define $n := |V|$. We let $\ell^2(\pi)$ be the vector space of real-valued functions with inner product

$$\langle f, g \rangle_{\pi} := \sum_{x \in V} \pi(x)f(x)g(x).$$

**Lemma A.7** (Spectral representation: reversible matrices). The space $\ell^2(\pi)$ has an orthonormal basis of eigenfunctions $\{f_j\}_{j=1}^n$ with real eigenvalues $\{\lambda_j\}_{j=1}^n$ such that $|\lambda_j| \leq 1$, for all $j$. The eigenfunction $f_1$ corresponding to the eigenvalue 1 can be taken to be the all-1 function. Furthermore, we have the following decomposition

$$\frac{P^t(x,y)}{\pi(y)} = 1 + \sum_{j=2}^n f_j(x)f_j(y)\lambda_j^t.$$

**Proof.** To be written. See [LPW06, Lemma 12.2] ■

**A.1.5 A fact about trees**

**Lemma A.8.** A cycle-free undirected graph with $n$ vertices and $n - 1$ edges is a spanning tree.
A.1.6 A Poincaré inequality

The Dirichlet form is defined as $\mathcal{E}(f, g) := \langle f, (I - P)g \rangle_\pi$. Note that

$$2\langle f, (I - P)f \rangle_\pi = 2\langle f, f \rangle_\pi - 2\langle f, Pf \rangle_\pi$$

$$= \sum_x \pi(x)f(x)^2 + \sum_y \pi(y)f(y)^2 - 2\sum_x \pi(x)f(x)f(y)P(x, y)$$

$$= \sum_{x,y} f(x)^2\pi(x)P(x, y) + \sum_{x,y} f(y)^2\pi(y)P(y, x) - 2\sum_x \pi(x)f(x)f(y)P(x, y)$$

$$= \sum_{x,y} \pi(x)P(x, y)[f(x) - f(y)]^2 = 2\mathcal{E}(f)$$

where

$$\mathcal{E}(f) := \frac{1}{2} \sum_{x,y} c(x, y)[f(x) - f(y)]^2,$$

is the Dirichlet energy encountered previously. We note further that if $\sum_x \pi(x)f(x) = 0$ then

$$\langle f, f \rangle_\pi = \langle f - \langle 1, f \rangle_\pi, f - \langle 1, f \rangle_\pi \rangle_\pi = \text{Var}_\pi[f],$$

where the last expression denotes the variance under $\pi$. So the variational characterization of $\lambda_2$ translates into

$$\text{Var}_\pi[f] \leq \gamma\mathcal{E}(f),$$

for all $f$ such that $\sum_x \pi(x)f(x) = 0$ (in fact for any $f$ by considering $f - \langle 1, f \rangle_\pi$ and noticing that both sides are unaffected by adding a constant), which is known as a Poincaré inequality.

**Lemma A.9** (Poincaré inequality).

$$\text{Var}_\pi[f] \leq \gamma\mathcal{E}(f), \quad \forall f,$$

with equality for $f_2$, the eigenfunction of $P$ corresponding to the second largest eigenvalue $\lambda_2$. 351
Bibliography


Index

$\epsilon$-packing, 64, 93
adapted process, 99
approximate counting, 267
Azuma-Hoeffding inequality, 117, 119, 120, 129, 130, 136, 140, 145, 182, 288
balancing vectors, 22
ballot theorem, 103
balls and bins, 126
Bernstein’s inequality, 59–61
Berry-Esseen theorem, 128
binary classification, 83
binomial variable, 51
birth-and-death chain, 183
bond percolation, 11
trees, 114
Bonferroni inequalities, 95
Boole’s inequality, see union bound
bottleneck ratio, 290
bounded differences inequality, 124
branching number, 44, 45
branching processes
dual branching process, 319
duality principle, 319
exploration process, 317, 328
extinction, 309, 311
Galton-Watson branching process, 344
Galton-Watson process, 309, 317
Galton-Watson tree, 310, 316
infinite line of descent, 345
Poisson offspring, 314, 320, 322
random-walk representation, 317–319
Chebyshev polynomials, 79
Chebyshev’s inequality, 19, 46, 47, 49, 197, 339, 342
Cheeger’s inequality, 290
Chen-Stein
Stein coupling, 249
Chen-Stein method, 246, 261
dissociated case, 249
Stein coupling, 246
Chernoff bound, 52, 127
Chernoff-Cramér bound, 47, 58, 62
Chernoff-Cramér method, 17, 47, 50, 52, 70, 96, 117
chi-square variable, 71
chromatic number, 130
clique number, 35, 261
commute time, 14, 173
commute time identity, 173, 184
compressed sensing, 73
compressen sensing
sensing matrix, 73
concentration inequalities, 98
concentration phenomenon, 117
conditional expectation
definition, 349
connectivity, 39
contour argument, 29
convex duality, 161
Lagrangian, 161
weak duality, 162
correlation inequalities, 187
coupling, 187, 188, 199, 201, 266, 327, 331, 341
coaescence time, 231
coupling inequality, 192, 197
coupling time, see coaescence time
Markovian, 230
maximal coupling, 193
monotone coupling, 201
of Markov chains, 230
path coupling, 239
coupling time, 187
covering number, 64
covering numbers, 92
critical value, 42
cumulant-generating function, 47, 49
Curie-Weiss model, 303
cutset, 44, 45, 164–166
dependency graph, 34
dimensionality reduction, 70
Dirichlet form, 164, 274
Dirichlet problem, 149
Dirichlet’s principle, 164, 184
Dudley’s inequality, 91, 92
dependency graph, 34
dimensionality reduction, 70
Dirichlet form, 164, 274
Dirichlet problem, 149
Dirichlet’s principle, 164, 184
Dudley’s inequality, 91, 92
dependency graph, 34
dimensionality reduction, 70
Dirichlet form, 164, 274
Dirichlet problem, 149
Dirichlet’s principle, 164, 184
Dudley’s inequality, 91, 92
dependency graph, 34
dimensionality reduction, 70
Dirichlet form, 164, 274
Dirichlet problem, 149
Dirichlet’s principle, 164, 184
Dudley’s inequality, 91, 92
dependency graph, 34
dimensionality reduction, 70
Dirichlet form, 164, 274
Dirichlet problem, 149
Dirichlet’s principle, 164, 184
Dudley’s inequality, 91, 92
edge boundary, 289
Efron-Stein inequality, 123
electrical network, 147
definitions, 151
effective conductance, 159, 164
effective resistance, 158, 160, 173, 184
Kirchhoff’s cycle law, 152
Kirchhoff’s node law, 152
Ohm’s law, 152, 160, 176, 183
parallel law, 154
series law, 154
empirical measure, 91, 93
empirical risk minimization, 84
epsilon-net, 63, 70, 74
Erdős-Rényi graph, 35, 39, 61, 129, 213, 218
Erdős-Rényi graphs, 327
cluster, 328
connectivity, 327
degree sequence, 197
evolution, 327
giant component, 327, 328, 339
escape probability, 155, 163
exhaustive sequence, 158, 184
expander graphs
Pinsker’s model, 296
Fenchel-Legendre dual, 49
filtered space, 99
filtration, 99
first moment method, 17, 22–25, 28, 32, 34, 36, 42, 129, 132, 339
first moment principle, 22–24, 85
FKG condition, 214
FKG inequality, 214, 221, 266
FKG measure, 214
flow, 152, 160
energy, 160, 171
finite energy, 167
flow to ∞, 166, 171
flow-conservation constraints, 152, 167
strength, 152
gambler’s ruin, 154, 159
gamma variable, 71
Gaussian variable, 70
generalization error, 84
McDiarmid’s inequality, 125, 130
method of bounded differences, 125, 127, 134
method of moments, 97
method of random paths, 168, 184
mixing time, 14, 78, 229
cutoff, 236, 307
diameter bound, 82
lower bound, 81
pre-cutoff, 307
mixing times
lower bounds, 243
upper bounds, 243
moment-generating function, 18, 47, 52
moments, 17
central moments, 17
Nash-Williams inequality, 164, 184
negative association, 176
network, 13
No Free Lunch Theorem, 84
orthogonality of increments, 111, 120
Polya’s theorem, 163, 168
Polya’s urn, 109
packing number, 64
pairwise uncorrelated, 95
Paley-Zygmund inequality, 33
parity functions, 280
path coupling, 244
pattern matching, 126
Peierls’ argument, see contour argument
percolation, 28, 42, 188
critical exponents, 326
critical value, 28, 42, 221
dual lattice, 29
Galton-Watson tree, 316
Harris’ theorem, 266
Kesten’s theorem, 266
percolation function, 28, 42, 204, 221
percolation on \( \mathbb{L}^2 \), 28, 221
percolation on \( \mathbb{L}^d \), 204
percolation on trees, 42
phase transition, 327
RSW theorem, 266
percolation function, 42
Poincaré inequality, 124, 274
Poisson approximation, 196
Poisson trials, 52
Poisson variable, 50
poset, 203
positive association, 199, 213
strong, 263
positive correlation, 213
predictable process, 99
preferential attachment graph, 133
probabilistic method, 21, 85, 93
probability generating function, 311
pseudo-regret, 141
random
\( b \)-ary tree, 236
random \( k \)-SAT, 25
random permutation
longest increasing subsequence, 26
random projection method, 70
random target lemma, 150
random walk
biased random walk on \( \mathbb{Z} \), 189
cycle, 232, 277
hypercube, 232, 280
lazy, 231
simple random walk on \( \mathbb{Z} \), 78
random walk on network, 14
Rayleigh quotient, 285
Rayleigh’s principle, 166, 176
recurrence, 14, 147, 158, 166
reflection principle, 102, 345
relative entropy, 51
relaxation time
random walk on cycle, 278
random walk on hypercube, 280
restricted isometry property, 73, 74
rough embedding, 169, 171
rough equivalence, 170, 184
rough isometry, 184
RSW theorem, 221
Russo’s formula, 221
satisfiability threshold, 25
Sauer’s lemma, 89, 90, 93
scale-free trees, 133
second moment method, 17, 32, 39, 42, 339
weighted second moment method, 45
separation distance, 269
set balancing, 49
shattering, 89
simple random walk on a graph, 13
slicing method, 145
stochastic bandit
arm, 141
spanning arborescence, 179
sparse signal recovery, 73
sparsity, 73
spectral gap, 276
spectral radius, 286
Spitzer’s combinatorial lemma, 320
stochastic bandit
Upper Confidence Bound, 141
stochastic domination, 187, 199, 204, 213, 332
Markov chain, 208
stochastic matrix, 6
stochastic monotonicity, 208
stopping time, 100
Strassen’s theorem, 266
sub-exponential variable, 57
sub-Gaussian increments, 68
sub-Gaussian variable, 53, 118
submodularity, 266
symmetrization, 54, 87, 92
tail probabilities, 18
Thomson’s principle, 161
threshold phenomenon, 25, 28, 35, 42
threshold function, 35
tilting, 56
total variation distance, 10
transience, see recurrence
trees
branching ratio, 115
Turán graphs, 24
type, see recurrence
uniform spanning tree
weighted uniform spanning tree, 177
Wilson’s method, 175
uniform spanning trees, 175
cycle popping algorithm, 178
uniform uncertainty principle, 74
union bound, 24, 95
Varopoulos-Carne bound, 78, 98
VC dimension, 89, 92
vertex boundary, 289
Wasserstein distance, 265