1 Probability theory

1.1 Basics

We first introduce some basic probability theory, and introduce some probabilistic inequalities that will give us some foundations for studying randomised algorithms.

Definition 1.1 (Finite discrete probability space)**.** A *finite discrete probability space* is a pair $(\Omega, \Pr)$ where

(i) Ω is a finite set; and

(ii) $\Pr : \Omega \to [0, 1]$

such that $\sum_{\omega \in \Omega} \Pr[\omega] = 1$.

Remark. As Pr is a function, it is instinctive to use the notation $\Pr(w)$; however, it is commonplace to use square brackets instead: $\Pr[w]$.

Lemma 1.2 (Law of total probability)**.** *Let* $(\Omega, \Pr)$ *be a finite discrete probability space,* $A \subset \Omega$ *be an event, and* $\mathcal{B} = \{B_i\}_{i=1}^n$ *be a disjoint partition of* Ω*. Then*

$$\Pr[A] = \sum_{B \in \mathcal{B}} \Pr[A \cap B].$$

Definition 1.3 (Events)**.** Let $(\Omega, \Pr)$ be a finite discrete probability space.

(i) An *event* A is a subset of Ω; that is, $A \subset \Omega$. Define

$$\Pr[A] := \sum_{\omega \in A} \Pr[w].$$

(ii) Let $A \subset \Omega$ be an event. The *complement* of A, denoted by $\overline{A}$, is the event $\Omega \setminus A$.

(iii) Let $A, B \subset \Omega$ be two events. A and B are *independent* if

$$\Pr[A \cap B] := \Pr[A] \cdot \Pr[B].$$

(iv) Let $A, B \subset \Omega$ be two events. The *conditional probability of A given B*, denoted by $\Pr[A \mid B]$, is

$$\Pr[A \mid B] := \frac{\Pr[A \cap B]}{\Pr[B]}$$

given $\Pr(B) > 0$.

Proposition 1.4. *Let* $(\Omega, \Pr)$ *be a finite discrete probability space and* $A, B \subset \Omega$ *be independent events with* $\Pr[B] > 0$*. Then*

$$\Pr[A \mid B] = \Pr[A].$$

Definition 1.5 (Random variable)**.** Let $(\Omega, \Pr)$ be a finite discrete probability space. A *random variable* X on $(\Omega, \Pr)$ is a function $X : \Omega \to \mathbb{R}$.

Remark. We will abuse notation here, and define

$$(X = x) := \{w \in \Omega : X(\omega) = x\}.$$

For $\geq$, we use similar notation:

$$(X \geq x) := \{w \in \Omega : X(\omega) \geq x\}$$

and similar for $>$, $\leq$, and $<$.

Definition 1.6 (Independent random variables)**.** Let $(\Omega, \Pr)$ be a finite discrete probability space with random variables X and Y. X and Y are *independent* if for all $x, y \in \mathbb{R}$, the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent.

Definition 1.7 (Bernoulli process)**.** Let $(\Omega, \Pr)$ be a finite discrete probability space. A *Bernoulli process* is a (possibly finite) sequence of independent random variables $\{X_i\}_{i=1}^{\infty}$ such that

(i) for all $i \in \mathbb{N}$, $X_i \in \{0, 1\}$; and

(ii) $\Pr[X_i = 1] = \Pr[X_j = 1]$ for all $i, j \in \mathbb{N}$.

Definition 1.8 (Expectation)**.** Let $(\Omega, \Pr)$ be a finite discrete probability space with random variable X. The *expected value* (or *expectation*) of X is

$$\mathbb{E}[X] := \sum_{x \in \operatorname{im} X} x \cdot \Pr(X = x).$$

Remark. We may get lazy and drop the $\operatorname{im} X$ from formulas; for example, we may write

$$\mathbb{E}[X] = \sum_{x} x \cdot \Pr(X = x).$$

Lemma 1.9. *Expectation is linear.*

Definition 1.10 (Conditional expectance on an event)**.** Let $(\Omega, \Pr)$ be a finite discrete probability space with random variable X and event $A \subset \Omega$. The *conditional expected value* of X given A is

$$\mathbb{E}[X \mid A] := \sum_{x} x \cdot \Pr[X = x \mid A].$$

Proposition 1.11. *Let* $(\Omega, \Pr)$ *be a finite discrete probability space with random variables* X *and* Y*. Then*

$$\mathbb{E}[X] = \sum_y \mathbb{E}[X \mid Y = y] \cdot \Pr[Y = y].$$

Definition 1.12 (Conditional expectance)**.** Let $(\Omega, \Pr)$ be a finite discrete probability space with random variables X and Y. The *conditional expected value* of X given Y, denoted by $\mathbb{E}[X \mid Y]$, is the function

$$\begin{aligned} f : \operatorname{im} Y &\mapsto \mathbb{R}, \\ y &\mapsto \mathbb{E}[X \mid Y = y]. \end{aligned}$$

Proposition 1.13 (Law of iterated expectation)**.** *Let* $(\Omega, \Pr)$ *be a finite discrete probability space with random variables* X *and* Y*. Then*

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]].$$

Definition 1.14 (Variance)**.** Let $(\Omega, \Pr)$ be a finite discrete probability space with random variable X. The *variance* of X is

$$\operatorname{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Proposition 1.15. *Let* $(\Omega, \Pr)$ *be a finite discrete probability space with random variable* X*. Then*

$$\operatorname{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Definition 1.16 (Covariance)**.** Let $(\Omega, \Pr)$ be a finite discrete probability space with random variables X and Y. The *covariance* of X and Y is

$$\operatorname{Cov}[X] := \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Proposition 1.17. *Let* $(\Omega, \Pr)$ *be a finite discrete probability space with independent random variables* X *and* Y*. Then* $\operatorname{Cov}[X, Y] = 0$.

Proposition 1.18. *Let* $(\Omega, \Pr)$ *be a finite discrete probability space with random variables* X *and* Y *and* $a, b \in \mathbb{R}$*. Then*

$$\operatorname{Var}[aX + bY] = a^2 \operatorname{Var}[X] + b^2 \operatorname{Var}[Y] + 2ab \operatorname{Cov}[X, Y].$$

This is all the basic theory that we need.

1.2 Inequalities

Theorem 1.19 (Union bound)**.** *Let* $(\Omega, \Pr)$ *be a finite discrete probability space and* $\{A_n\}_{n=1}^{\infty}$ *be a collection of events. Then*

$$\Pr\left[\bigcup_{n=1}^{\infty} A_n\right] \leq \sum_{i=1}^{\infty} \Pr[A_n].$$

Theorem 1.20 (Markov's inequality)**.** *Let* $(\Omega, \Pr)$ *be a finite discrete probability space with random variable* $X : \Omega \to \mathbb{R}_{\geq 0}$ *and* $\alpha \in \mathbb{R}_{\geq 0}$*. Then*

$$\Pr[X \geq \alpha] \leq \frac{\mathbb{E}[X]}{\alpha}.$$

Theorem 1.21 (Jensen's inequality)**.** *Let* $(\Omega, \Pr)$ *be a finite discrete probability space with random variable* X *and* $f : \mathbb{R} \to \mathbb{R}$ *be a function. Then*

(i) if f *is convex, then* $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$*; and*

(ii) if f *is concave, then* $f(\mathbb{E}[X]) \geq \mathbb{E}[f(X)]$*.*

Theorem 1.22 (Chebyshev's inequality)**.** *Let* $(\Omega, \Pr)$ *be a finite discrete probability space with random variable* X *and* $\alpha \in \mathbb{R}_{\geq 0}$*. Then*

$$\Pr\left[|X - \mathbb{E}[X]| \geq \alpha\right] \leq \frac{1}{\alpha^2} \operatorname{Var}[X].$$

Theorem 1.23 (Generic Chernoff bound (multiplicative form))**.** *Let* $(\Omega, \Pr)$ *be a finite discrete probability space with independent random variables* $\{X_i\}_{i=1}^{n}$ *taking values in* $\{0, 1\}$*. Let* $X = \sum_{i=1}^{n} X_i$ *be their sum and* $\delta \geq 0$*. Then*

$$\Pr[X \geq (1 + \delta)\mathbb{E}[X]] \leq \left(\frac{e^{\delta}}{(1 + \delta)^{1+\delta}}\right)^{\mathbb{E}[X]}.$$

The following is a similar bound to the above for random variables that do not take variables in $\{0, 1\}$.

Theorem 1.24 (Hoeffding's inequality)**.** *Let* $(\Omega, \Pr)$ *be a finite discrete probability space with independent random variables* $\{X_i\}_{i=1}^{n}$ *such that for all* $i \in \{1, \ldots, n\}$*,* $|X_i| \leq c_i$ *for some* $c_i \in \mathbb{R}_{\geq 0}$*. Then for all* $t > 0$*,*

$$\Pr\left[\sum_{i=1}^{n} X_i \geq t\right] \leq \exp\left(\frac{-t^2}{2 \sum_{i=1}^{n} c_i^2}\right).$$

The following is again a similar bound to the above for dependent random variables.

Theorem 1.25. *Let* $(\Omega, \Pr)$ *be a finite discrete probability space with independent random variables* $\{X_i\}_{i=1}^n$ *where there exists* $a, b \in \mathbb{R}$ *such that* $a \leq X_i \leq b$ *for all* $i \in \{1, \ldots, n\}$*. Then for all* $\delta > 0$,

$$\Pr[X \geq (1+\delta)\mathbb{E}[X]] \leq \exp\left(\frac{-2\delta^2 \mathbb{E}[X]^2}{n(b-a)^2}\right).$$

2 Martingales

In the previous section, we made sure to mention that we are working over a finite discrete probability space $(\Omega, \Pr)$, but we will now omit this and assume that we are whenever we refer to random variables.

Definition 2.1 (Martingale)**.** A sequence of random variables $\{Z_i\}_{i \in \mathbb{N}_0}$ is a *martingale* with respect to another sequence of random variables $\{X_i\}_{i \in \mathbb{N}_0}$ if for all $n \in \mathbb{N}_0$,

(i) Z_n is a function of $\{X_i\}_{i=0}^n$;

(ii) $\mathbb{E}[|Z_n|] < \infty$; and

(iii) $Z_n = \mathbb{E}[Z_{n+1} \mid X_0, \ldots, X_n]$.

$\{Z_i\}_{i \in \mathbb{N}_0}$ is s *martingale* if it is a martingale with respect to itself.

Lemma 2.2. *Let* $n \in \mathbb{N}_0$*. If* $\{Z_i\}_{i=0}^n$ *is a martingale with respect to the sequence of random variables* $\{X_i\}_{i=0}^n$*, then* $\mathbb{E}[Z_n] = \mathbb{E}[Z_0]$.

Definition 2.3 (Stopping time)**.** Let $\mathcal{Z} = \{Z_i\}_{i \in \mathbb{N}_0}$ be a sequence of random variables. A non-negative, integer-valued random variable T is a *stopping time* for $\mathcal{Z}$ if the event $[T = n]$ depends only on $\{Z_i\}_{i=0}^n$ for all $n \in \mathbb{N}_0$.

Theorem 2.4 (Martingale stopping theorem)**.** *Let* $\{Z_i\}_{i \in \mathbb{N}_0}$ *be a martingale with respect to the sequence of random variables* $\mathcal{X} = \{X_i\}_{i \in \mathbb{N}_0}$ *and let* T *be a stopping time for* $\mathcal{X}$*. Then*

$$\mathbb{E}[Z_T] = \mathbb{E}[Z_0]$$

whenever one of the following holds:

(i) there is $c \in \mathbb{R}_{\geq 0}$ *such that* $|Z_i| \leq c$ *for all* $i \in \mathbb{N}_0$*;*

(ii) T *is bounded; or*

(iii) $\mathbb{E}[T] < \infty$ *and there is* $c \in \mathbb{R}_{\geq 0}$ *such that*

$$\mathbb{E}\left[|Z_{i+1} - Z_i| \mid X_0, \ldots, X_i\right] \leq c$$

for all $i \in \mathbb{N}_0$.

Theorem 2.5 (Wald's equation)**.** *Let $\{X_i\}_{i\in\mathbb{N}}$ be nonnegative, independent, identically distributed random variables with distribution X. Let T be a stopping time for this sequence. If T and X have bounded expectation, then*

$$\mathbb{E}\left[\sum_{i=1}^{T} X_i\right] = \mathbb{E}[T]\cdot\mathbb{E}[X].$$

Theorem 2.6 (Azuma-Hoeffding)**.** *Let $\{Z_i\}_{i=0}^n$ be a martingale such that for all $i \in \mathbb{N}$, there exists $c_i \in \mathbb{R}_{\geq 0}$ such that $|Z_i - Z_{i-1}| \leq c_i$. Then for all $t \in \mathbb{R}_{\geq 1}$ and $\lambda \in \mathbb{R}_{>0}$,*

$$\Pr[|Z_t - Z_0| \geq \lambda] \leq 2\exp\left(\frac{-\lambda^2}{2\sum_{i=1}^{t} c_i^2}\right).$$

Corollary 2.7. *Let $\{Z_i\}_{i=0}^n$ be a martingale and suppose there exists $c \in \mathbb{R}_{\geq 0}$ such that $|Z_i - Z_{i-1}| \leq c$ for all $i \in \mathbb{N}$. Then for all $t \in \mathbb{R}_{\geq 1}$ and $\lambda \in \mathbb{R}_{>0}$,*

$$\Pr[|Z_t - Z_0| \geq \lambda c\sqrt{t}] \leq 2\exp\left(\frac{-\lambda^2}{2}\right).$$

3 Markov chains

Definition 3.1 (Markov chain)**.** A (discrete-time) *Markov chain* on a countable set S is a sequence of random variables $(X_i)_{i\in\mathbb{N}_0}$ such that for all $j, i_0, i_1, \ldots, i_n \in S$ and $n \in \mathbb{N}_0$,

$$\Pr[X_{n+1} = j \mid X_0 = i_0, \ldots, X_n = i_n] = \Pr[X_{n+1} = j \mid X_n = i_n]$$

given that both conditional probabilities are well-defined.

Remark. We are only working with discrete-time Markov chain, so assume all Markov chains are discrete-time Markov chains.

Markov chains are also called *memoryless*; the probability of each event depends only on the state attained in the previous event.

Definition 3.2. Let $\mathcal{X} = (X_i)_{i\in\mathbb{N}_0}$ be a Markov chain on S.

(i) X_0 is the *starting state.*

(ii) S is the *state space.*

(iii) $\mathcal{X}$ is *finite* if it has a finite state space.

(iv) $\mathcal{X}$ is *countable* if it has a countable state space.

Definition 3.3 (Homogenity)**.** A Markov chain $(X_i)_{i\in\mathbb{N}_0}$ is *time-homogeneous* (or just *homogeneous*) if for all $n \in \mathbb{N}$,

$$\Pr[X_{n+1} = j \mid X_n = i] = \Pr[X_n = j \mid X_{n-1} = i].$$

All Markov chains we will look at will be homogenous.

For a homogenous countable Markov chain $\mathcal{X} = (X_i)_{i\in\mathbb{N}_0}$ on S, we can describe its behaviour using *transition probabilities*: for $i, j \in S$ and $n \in \mathbb{N}_0$, define

$$p_{ij} = \Pr[X_{n+1} = j \mid X_n = i].$$

This is the *one-step transition probabilities* of $\mathcal{X}$. For $k \in \mathbb{N}$, we may similarly define the *k-step transition probabilities* by

$$p_{ij}^{(k)} = \Pr[X_{n+k} = j \mid X_n = i].$$

As S is countable, we may pick some enumeration of S and represent p_{ij} as a matrix P such that $P[i, j] = p_{ij}$ called the *transition matrix* of $\mathcal{X}$.

Remark. When S is finite, we pick a enumeration. That is, a bijective function $s : S \to \{1, \ldots, |S|\}$. Thus above, we should write $P[s(i), s(j)] = p_{ij}$, but we will keep with this abuse of notation. Alternatively, we may just assume that $S = \{1, \ldots, l\}$ for some $l \in \mathbb{N}$ from now on.

Proposition 3.4. *Let $(X_i)_{i\in\mathbb{N}_0}$ be a homogeneous countable Markov chain over S with transition probabilities $p_{i,j}$ for all $i, j \in S$ and transition matrix P. Then for all $k \in \mathbb{N}$,*

$$p_{ij}^{(k)} = P^k[i, j].$$

Definition 3.5. Let $n \in \mathbb{N}$ and $A \in M_n(\mathbb{R})$. A is *stochastic* if for all $i \in \{1, \ldots, n\}$,

$$\sum_{j=1}^{n} A[i, j] = 1.$$

Proposition 3.6. *Every stochastic matrix is the transition matrix of some Markov chain.*

Lemma 3.7. *The largest eigenvalue of a stochastic matrix is* 1*.*

Definition 3.8 (Distribution)**.** Let $\mathcal{X} = (X_i)_{i=0}^n$ be a finite Markov chain on S. At some time $t \in \mathbb{N}_0$, the *distribution over states* (or *distribution*) $\boldsymbol{x}^{(t)}$ of $\mathcal{X}$ is given by

$$\boldsymbol{x}^{(t)} = (\Pr[X_t = 1], \Pr[X_t = 2], \ldots).$$

Proposition 3.9. *Let $(X_i)_{i\in\mathbb{N}_0}$ be a homogeneous countable Markov chain over S with transition matrix P. Let $t \in \mathbb{N}_0$ and $\boldsymbol{x}^{(t)}$ be the distribution at that time. Then for all $k \in \mathbb{N}$,*

$$\boldsymbol{x}^{(t+k)} = \boldsymbol{x}^{(t)} P^k.$$

Remark. Again, expect some abuse of notation. Sometimes we may just write $\boldsymbol{x}$ for a distribution if the context is clear.

Definition 3.10 (Stationary distribution)**.** Let $\mathcal{X}$ be a Markov chain with transition matrix P. A distribution $\boldsymbol{\pi}$ of $\mathcal{X}$ is *stationary* if $\boldsymbol{\pi}P = \boldsymbol{\pi}$.

Definition 3.11 (Reversibility)**.** A Markov chain with transition matrix $P \in M_n(\mathbb{R})$ is *reversible* if there is a distribution $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$ such that for all $i, j \in \{1, \ldots, n\}$,

$$\pi_i P[i,j] = \pi_j P[j,i].$$

At this distribution, running the Markov chain *backwards* "looks the same" as running it forward.

Definition 3.12 (Total variation distance)**.** Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two distribution of some Markov chain. The *total variation distance* between $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$d_{\mathrm{TV}}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_1.$$

Total variation distance is a metric, thus we can use it to prove convergence.

Lemma 3.13. *Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two distributions of some discrete random variable Z. Let $\mathbb{E}_{\boldsymbol{x}}[Z]$ and $\mathbb{E}_{\boldsymbol{y}}[Z]$ be the expectations of Z with respect to each of these distributions, and suppose that $|Z| \leq M$ for some $M \in \mathbb{R}$. Then*

$$|\mathbb{E}_{\boldsymbol{x}}[Z] - \mathbb{E}_{\boldsymbol{y}}[Z]| \leq \frac{M}{2} d_{TV}(\boldsymbol{x}, \boldsymbol{y}).$$

Thus if two distributions converge, so does their expected value.

Definition 3.14. Let P be the transition matrix of some Markov chain, $\boldsymbol{\pi}$ be some stationary distribution, and $\boldsymbol{x}$ be an initial distribution. A *mixing time* t_{mix} of $\boldsymbol{x}$ is a function $t_{\mathrm{mix}} : \mathbb{R}_{>0} \to \mathbb{N}_0$ such that for all $\varepsilon > 0$ and $t \geq t_{\mathrm{mix}}(\varepsilon)$ we have

$$d_{\mathrm{TV}}(xP^t, \pi) \leq \varepsilon.$$

Definition 3.15 (Coupling)**.** A *coupling* of two random variables X and Y is a joint distribution on X and Y.

We can use couplings to prove convergence $X \to Y$, such as by creating a dependence between them as to maximise $\Pr[X = Y]$ or to minimise total variation distance.

Lemma 3.16 (Coupling lemma)**.** *For any discrete random variables X and Y,*

$$d_{TV}(X, Y) \leq \Pr[X \neq Y].$$

Not all Markov chains converge, thus we need some constructions to deal with this.

Definition 3.17 (Irredubility)**.** Let $\mathcal{X}$ be a Markov chain with transition matrix $P \in M_n(\mathbb{R})$. $\mathcal{X}$ is *irreducible* if for all $i, j \in \{1, \ldots, n\}$, there exists some $t \in \mathbb{N}_0$ such that $P^t[i, j] \neq 0$.

That is, we can reach any state from any other state (even if it is a long wait). Alternatively, if we represent a Markov chain as a directed graph where the states are vertices and each edge represents a transition that occurs with non-zero probability, then the Markov chain is irreducible if and only if the graph is strongly connected.

Definition 3.18 (Period)**.** The *period* of a state i of a Markov chain with transition matrix P is $\gcd\{t > 0 : P^t[i, i] \neq 0\}$. If the period of i is 1, then i is *aperiodic*. A Markov chain is *aperiodic* if all of its states are aperiodic.

Lemma 3.19. *The period of any state of a reversible Markov chain is at most 2.*

Definition 3.20 (Ergodic)**.** A Markov chain is *ergodic* if it's irreducible and aperiodic.

We will study the stationary distributions of ergodic Markov chains, as they are well behaved.

Lemma 3.21. *Consider a Markov chain $\mathcal{X}$ with transition matrix of P. Then there is an aperiodic Markov chain $\mathcal{X}'$ with transition matrix $\frac{1}{2}(P + I)$ and if $\boldsymbol{\pi}$ is a stationary distribution of $\mathcal{X}$, it is a stationary distribution of $\mathcal{X}'$.*

We may call $\mathcal{X}'$ the *lazy version* of $\mathcal{X}$, thus if we are studying the stationary distributions of a Markov chain, we have methods to get around it being periodic. However, we do not have a method for getting around reducible Markov chains.

Theorem 3.22. *Any finite ergodic Markov chain converges to a unique stationary distribution.*

4 Probabilistic method

The *probabilistic method* is a non-constructive proof for the existence of an object, by showing that some process generates the object with non-zero probability.

Definition 4.1 (Tournament)**.** A *tournament* is a digraph obtained by assigning a direction for each edge in an undirected complete graph.

Theorem 4.2. *For $n \in \mathbb{N}_{\geq 3}$, there exists a tournament with at least $2^{-n}(n-1)!$ directed cycles.*

Lemma 4.3 (Variant of Chernoff's inequality)**.** *Let $(X_i)_{i=1}^n$ independent random variables such that $\Pr[X_i = 1] = \Pr[X_i = -1] = \frac{1}{2}$ for all $i \in \{1, \ldots, n\}$. Let $X = \sum_{i=1}^n X_i$ and $a \in \mathbb{R}_{>0}$. Then*

$$\Pr[|X| > a] \leq 2 \exp\left(\frac{-2a^2}{n}\right).$$

Lemma 4.4 (Lovász local lemma)**.** *Let $\mathcal{A} = (A_i)_{i=1}^m$ be a finite collection of events on some probability space and for each $A \in \mathcal{A}$ let $\Gamma(A)$ be a set of events such that A is independent of all events not in $\Gamma(A) \cup \{A\}$. If there exists a real number $x_A \in (0, 1)$ such that for all $A \in \mathcal{A}$,*

$$\Pr[A] \leq x_A \prod_{B \in \Gamma(A)} (1 - x_B)$$

then

$$\Pr\left[\bigcap_{A \in \mathcal{A}} \overline{A}\right] \geq \prod_{A \in \mathcal{A}} (1 - x_A).$$

Remark. We may think of $\Gamma(A)$ as the neighbours of A in the *dependency graph* of A. A dependency graph, each vertex represents an event, and there is an edge between two events if they are not independent.

Corollary 4.5. *Let $\mathcal{A}$ and Γ as before, and further suppose there is $p \in [0, 1]$ and $d \in \mathbb{R}_{\geq 0}$ such that for all $A \in \mathcal{A}$, $\Pr[A] \leq p$ and $|\Gamma(A)| \leq d$. Then if $ep(d + 1) < 1$ we have*

$$\Pr\left[\bigcap_{A \in \mathcal{A}} \overline{A}\right] > 0.$$

5 The power of two random choices

5.1 Single choice protocol

Theorem 5.1. *If n balls are allocated independently and uniformly at random into n bins, then the maximally loaded bin contains $O\left(\frac{\log n}{\log \log n}\right)$ many balls, with high probability.*

By *high probability*, we mean a probability of at least $1 - \frac{1}{n^c}$ for some $c \in \mathbb{R}_{>0}$.

We can similarly have a asymptotically matching lower bound.

Theorem 5.2. *If n balls are allocated independently and uniformly at random into n bins, then with probability at least $\frac{1}{2}$, there will be a bin receiving $\Omega\left(\frac{\log n}{\log\log n}\right)$ many balls.*

5.2 Multiple choice protocol

Theorem 5.3. *Suppose that n balls are sequentially placed into n bins. Each ball is placed in a least full bin at the time of the placement, among d bins, $d \geq 2$, chosen independently and uniformly at random. Then after all the balls are placed, with high probability, the number of balls in the fullest bin is at most $\frac{\log\log n}{\log d} + O(1)$.*

6 Witness trees

6.1 Delay sequences