RandAlg Formula Sheet

Xiaoshuang Wang, Yuhao Mao

1 Randomized Algorithm and Probabilistic method

Definition 1.1: A randomized algorithm $\mathcal{A}$ is called a Las Vegas algorithm if: i) For each input I and for each call the algorithm computes either the correct answer OPT(I) or returns the answer 'I don't know'. ii) There exists $\varepsilon > 0$ such that for each input I we have $\Pr[\mathcal{A}$ returns the answer $\mathrm{OPT}(I)] \geq \varepsilon$. We say that $\mathcal{A}$ has success probability at least ε. (Note: The probability is computed with respect to the distributions of all random variables that are used by the algorithm.) Moreover, a Las Vegas algorithm is called efficient if on any input I its worst-case running time is polynomial in $|I|$.

Lemma 1.2: Let $\mathcal{A}$ be a Las Vegas algorithm with success probability at least $\varepsilon > 0$, and let $\delta > 0$ some arbitrarily small constant. Then for every input I we have $\Pr[$ At least one of $\left[\frac{\log \delta}{\log(1-\varepsilon)}\right]$ calls of $\mathcal{A}$ returns the answer $\mathrm{OPT}(I)] \geq 1 - \delta$.

Definition 1.3: A randomized algorithm $\mathcal{A}$ is called a Monte Carlo algorithm if there exists $\varepsilon > 0$ such that for each input I we have $\Pr[\mathcal{A}$ returns the answer $\mathrm{OPT}(I)] \geq \frac{1}{2} + \varepsilon$. We say that $\mathcal{A}$ has success probability at least $1/2 + \varepsilon$. Moreover, we call a Monte Carlo algorithm efficient if on any Input I its worst-case running time is polynomial in $|I|$.

Lemma 1.4: Let $\mathcal{A}$ be a Monte Carlo algorithm with success probability at least $1/2 + \varepsilon$ for some $\varepsilon > 0$, and let $\delta > 0$ some arbitrarily small constant. Then for every input I we have

$$\Pr\left[\text{ MAJORITY } \left(\mathcal{A}, I, \left[\frac{2\log\delta}{\log\left(1-4\varepsilon^2\right)}\right]\right) = \mathrm{OPT}(I)\right] \geq 1 - \delta,$$

where MAJORITY $(\mathcal{A}, I, t)$ denotes the value which appears most often in t independent calls of $\mathcal{A}$. (If this value is not unique, we just pick one at random.)

2 Linearity of Expectation

Theorem 2.3: Let X be a random variable with $W_X \subseteq \mathbb{N}_0$ ($W_X := X(\Omega) = \{x \in \mathbb{R} \mid \exists \omega \in \Omega \text{ with } X(\omega) = x\}$). Then

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \Pr[X \geq i]$$

Theorem 2.4 (Linearity of Expectation): For random variables $X_1, \ldots, X_n$ and $X := a_1 X_1 + \ldots + a_n X_n$ with $a_1, \ldots, a_n \in \mathbb{R}$ we have

$$\mathbb{E}[X] = a_1 \mathbb{E}\left[X_1\right] + \ldots + a_n \mathbb{E}\left[X_n\right]$$

Prove maximum/minimum exceeds some threshold: Show the expectation equals to the threshold. Then by the definition of expectation, the maximum/minimum must be at least the expectation.

3 The inequalities of Markov & Chebyshev

Theorem 3.1 **(Markov Inequality)**: Let X be a non-negative random variable. For all $t > 0$ we have

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

Theorem 3.2 (Chebyshev Inequality): Let X be a random variable. For all $t > 0$ we have

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2}$$

4 First & Second Moment Method

Lemma 4.1 (First Moment Method): Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables which take nonnegative integer values. Then $\mathbb{E}\left[X_n\right] = o(1)$ implies $\Pr\left[X_n = 0\right] = 1 - o(1)$.

Lemma 4.2 (Second Moment Method): Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables. Then $\mathbb{E}\left[X_n\right] \neq 0$ (for n large enough) and $\mathrm{Var}\left[X_n\right] = o\left(\mathbb{E}\left[X_n\right]^2\right)$ implies $\Pr\left[X_n = 0\right] = o(1)$.

Definition 4.3: Let Q be a set of graphs. A function $t = t(n)$ is called sharp threshold for Q if for every $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \Pr\left[G_{n,p} \in Q\right] = \begin{cases} 0, & \text{if } p \leq (1-\varepsilon)t(n) \\ 1, & \text{if } p \geq (1+\varepsilon)t(n) \end{cases}$$

Definition 4.4: Let Q be a set of graphs. A function $t = t(n)$ is called weak threshold for Q if

$$\lim_{n\to\infty} \Pr\left[G_{n,p} \in Q\right] = \begin{cases} 0, & \text{if } p \ll t(n) \\ 1, & \text{if } p \gg t(n) \end{cases}$$

5 Chernoff Bounds

Theorem 5.1: Let $X_1, \ldots, X_n$ be independent Bernoulli-distributed random variables with $\Pr\left[X_i = 1\right] = p_i$ and $\Pr\left[X_i = 0\right] = 1 - p_i$. Then for $X := \sum_{i=1}^n X_i$ and $\mu := \mathbb{E}[X] = \sum_{i=1}^n p_i$, and every $\delta > 0$ we have

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

Theorem 5.3: Let $X_1, \ldots, X_n$ be independent Bernoulli-distributed random variables with $\Pr\left[X_i = 1\right] = p_i$ and $\Pr\left[X_i = 0\right] = 1 - p_i$. Then for

$X := \sum_{i=1}^{n} X_i$ and $\mu := \mathbb{E}[X] = \sum_{i=1}^{n} p_i$, and every $0 < \delta < 1$ we have

$$\Pr[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}$$

Corollary 5.4: Let $X_1, \ldots, X_n$ be independent Bernoulli-distributed random variables with $\Pr[X_i = 1] = p_i$ and $\Pr[X_i = 0] = 1 - p_i$. Then the following inequalities hold for $X := \sum_{i=1}^{n} X_i$ and $\mu := \mathbb{E}[X] = \sum_{i=1}^{n} p_i$:

(i) $\Pr[X \geq (1+\delta)\mu] \leq e^{-\mu\delta^2/3}$ for all $0 < \delta \leq 1$,
(ii) $\Pr[X \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}$ for all $0 < \delta \leq 1$,
(iii) $\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}$ for all $0 < \delta \leq 1$,
(iv) $\Pr[X \geq t] \leq 2^{-t}$ for $t \geq 2e\mu$.

6 Negative Correlation

Lemma 6.1: Let $X_1, \ldots, X_n$ be Bernoulli random variables that are pairwise negatively correlated. Then $X = \sum_{i=1}^{n} X_i$ satisfies $\mathrm{Var}[X] \leq \mathbb{E}[X]$

Corollary 6.2: Let $(Z_n)_{n\in\mathbb{N}}$ be a sequence of random variables such that each Z_i is the sum of pairwise negatively correlated Bernoulli random variables. Then

$$\begin{array}{lll} \mathbb{E}[Z_n] \to 0 & \text{implies} & \Pr[Z_n = 0] \to 1, \text{ and} \\ \mathbb{E}[Z_n] \to \infty & \text{implies} & \Pr[Z_n = 0] \to 0 \end{array}$$

Definition 6.3: Random variables $X_1, \ldots, X_n$ are said to be negatively associated iff for every two disjoint subsets $I, J \subseteq \{1, \ldots, n\}$ and every two functions $f : \mathbb{R}^{|I|} \to \mathbb{R}$ and $g : \mathbb{R}^{|J|} \to \mathbb{R}$ that are either both componentwise increasing or both componentwise decreasing (not necessarily strictly) we have

$$\mathbb{E}[f(X_i, i \in I) \cdot g(X_j, j \in J)] \leq \mathbb{E}[f(X_i, i \in I)] \cdot \mathbb{E}[g(X_j, j \in J)].$$

Theorem 6.4: Let $X_1, \ldots, X_n$ be Bernoulli-distributed random variables with $\Pr[X_i = 1] = p_i$ and $\Pr[X_i = 0] = 1 - p_i$ that are negatively associated, and let $X := \sum_{i=1}^{n} X_i$ and $\mu := \mathbb{E}[X] = \sum_{i=1}^{n} p_i$. Then the bounds in Theorems 5.1 and 5.3 and Corollary 5.4 hold for X as well.

Lemma 6.5: Let $X_1, \ldots, X_n$ be Bernoulli random variables such that $\sum_{i=1}^{n} X_i \equiv 1$. Then the variables $X_1, \ldots, X_n$ are negatively associated.

Lemma 6.6: If $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_m)$ are mutually independent and both are negatively associated, then $(\mathbf{X}, \mathbf{Y}) = (X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ is also negatively associated.

Lemma 6.7: If $X_1, \ldots, X_n$ are negatively associated random variables, $I_1, \ldots, I_k \subseteq \{1, \ldots, n\}$ are pairwise disjoint sets, and $h_s : \mathbb{R}^{|I_s|} \to \mathbb{R}, 1 \leq s \leq k$, are functions that are all componentwise increasing or all componentwise decreasing (not necessarily strictly), then the variables $Y_1, \ldots, Y_k$ where $Y_s := h_s(X_i, i \in I_s), 1 \leq s \leq k$, are also negatively associated.

7 The Inequalities of Azuma and Janson

Theorem 7.1 (Azuma's Inequality) Let $(\Omega, \Pr)$ be the product of N probability spaces $(\Omega_1, \Pr_1), \ldots, (\Omega_N, \Pr_N)$, and let $X : \Omega \to \mathbb{R}$ be a random variable with the property that the effect of the i-th coordinate is at most c_i. Then for all $t \geq 0$ we have

$$\Pr[X \geq \mathbb{E}[X]+t] \leq e^{-\frac{t^2}{2\sum_{i=1}^N c_i^2}} \quad \text{and} \quad \Pr[X \leq \mathbb{E}[X]-t] \leq e^{-\frac{t^2}{2\sum_{i=1}^N c_i^2}}$$

Theorem 7.2 (Janson's Inequality) Let $(\Omega, \Pr)$ be the product of N discrete probability spaces wit $\Omega_i = \{0,1\}$ for all $i = 1,\dots,N$. For sets $A_1,\dots,A_m \subseteq [N]$ and $1 \leq i \leq m$ let X_i denote the indicator variable for the event that all coordinates from A_i are equal to one, that is $X_i(\omega_1,\dots,\omega_n) = 1$ if an only if $\omega_j = 1$ for all $j \in A_i$. Furthermore, let $X := \sum_{i=1}^m X_i$. Then we have for

$$\lambda := \mathbb{E}[X] = \sum_{i=1}^m \Pr[X_i = 1] \quad \text{and} \quad \Delta := \sum_{\substack{i \neq j \\ A_i \cap A_j \neq \emptyset}} \Pr[X_i = 1 \wedge X_j = 1],$$

and for all $0 \leq t \leq \mathbb{E}[X]$ that

$$\Pr[X \leq \mathbb{E}[X]-t] \leq e^{-\frac{t^2}{2(\lambda+\Delta)}}$$

In particular we get (setting $t = \mathbb{E}[X]$ in the above inequality),

$$\Pr[X=0] \leq e^{-\frac{\lambda^2}{2(\lambda+\Delta)}} \leq e^{-\min\{\lambda,\lambda^2/\Delta\}/4}.$$

8 Talagrand's Inequality

Theorem 8.1 (Talagrand's Inequality) Let $(\Omega, \Pr)$ be the product of N probability spaces $(\Omega_1, \Pr_1)$, $\dots$, $(\Omega_N, \Pr_N)$, and let $X : \Omega \to \mathbb{R}$ be a random variable with the property that the effect of the i-th coordinate is at most c_i. Moreover, let $\psi : \mathbb{R} \to \mathbb{R}$ be a function such that for every $\omega \in \Omega$ and $r \in \mathbb{R}$ with $X(\omega) \geq r$ there exists a set $J \subseteq \{1,\dots,N\}$ such that for all $\omega' \in \Omega$ with $\omega_i = \omega'_i$ when $i \in J$, we have $X(\omega') \geq r$ and such that $\sum_{i \in J} c_i^2 \leq \psi(r)$. Then for every median $m \in \mathbb{R}$ of X and every $t > 0$ we have

$$\Pr[|X-m| \geq t] \leq 4e^{-\frac{t^2}{4\psi(m+t)}}.$$

Lemma 8.2 If X assumes non-negative real values and there exist constants C_0, C_1 such that $\psi(r) \leq C_0 + C_1 r$ for every $r \geq 0$ satisfies the conditions of Talagrand's inequality, then

$$\Pr[|X-\mathbb{E}[X]| \geq t] \leq 4e^{-\Omega\left(\frac{t^2}{\mathbb{E}[X]+t}\right)}.$$

Proof. At first note that we can without loss of generality assume that $t = \omega(\sqrt{\mathbb{E}[X]})$ and $t = \omega(1)$. If either of these is not true, then we can find a constant D such that $\frac{t^2}{\mathbb{E}[X]+t} \leq D$ which implies that we aim to prove

$$\Pr[|X-\mathbb{E}[X]| \geq t] \leq 4e^{-\Omega(D)}.$$

But as $4e^{-\Omega(D)}$ can be larger than 1 (e.g. if the constant hidden in the Ω-notation is less than $1/D$) this inequality becomes trivial. Also, still without loss of generality, we can assume that $C_0 > 0$. Under the given conditions Theorem 8.1 yields for every median m of X that

$$\Pr[|X-m| \geq t] \leq 4e^{-\frac{t^2}{4C_0+4C_1(m+t)}} \leq \begin{cases} 4e^{-t^2/\left[4(C_0+C_1)\left(1+\frac{C_1}{C_0}m\right)\right]} & \text{if } 0 \leq t \leq 1+\frac{C_1}{C_0}m \\ 4e^{-t/4(C_0+C_1)} & \text{if } t > 1+\frac{C_1}{C_0}m \end{cases}$$

In particular, it follows that

$$\begin{aligned}
|\mathbb{E}[X] - m| &= \left|\sum_{\omega\in\Omega} \Pr[\omega](X(\omega) - m)\right| \\
&\leq \sum_{\omega\in\Omega} \Pr[\omega]|X(\omega) - m| \\
&= \mathbb{E}[|X - m|] \\
&= \int_0^\infty \Pr[|X - m| > t]dt \\
&\leq \int_0^{1+\frac{C_1}{C_0}m} 4e^{-t^2/\left[4(C_0+C_1)\left(1+\frac{C_1}{C_0}m\right)\right]}dt + \int_{1+\frac{C_1}{C_0}m}^\infty 4e^{-t/4(C_0+C_1)}dt \\
&\leq 4\sqrt{\pi\left(C_0 + C_1\right)\left(1 + \frac{C_1}{C_0}m\right)} + 16\left(C_0 + C_1\right)
\end{aligned}$$

By the definition of the median and by Markov's inequality we have $m \leq 2\Pr[X \geq m]m \leq 2\mathbb{E}[X]$, so that

$$|\mathbb{E}[X]-m| \leq 4\sqrt{\pi\left(C_0 + C_1\right)\left(1 + \frac{C_1}{C_0}2\mathbb{E}[X]\right)}+16\left(C_0 + C_1\right) = O(\sqrt{\mathbb{E}[X]}+1) = o(t).$$

With this estimate we can use Theorem 8.1 to obtain concentration around the expectation as follows.

$$\begin{aligned}
\Pr[|X - \mathbb{E}[X]| \geq t] &\leq \Pr[|X - m + m - \mathbb{E}[X]| \geq t] \\
&\leq \Pr[|X - m| \geq t - |m - \mathbb{E}[X]|] \\
&\leq 4e^{-\frac{\left(t-|m-\mathbb{E}[X]|\right)^2}{4C_0+C_1(m+t-|m-\mathbb{E}[X]|)}} \\
&= 4e^{-\frac{t^2}{16C_0+16C_1(\mathbb{E}[X]+o(t)+t)}} \\
&= 4e^{-\Omega\left(\frac{t^2}{\mathbb{E}[X]+t}\right)}
\end{aligned}$$

9 Markov Chains I: Definition and Examples

Definition 9.1 A Markov Chain (with discrete time) over the finite (or countably infinite) state set S consists of an infinite sequence of random variables $(X_t)_{t\in\mathbb{N}_0}$ with domain S and a starting distribution q_0, a vector of $|S|$ non-negative values summing up to one. For each state $i \in S$ we have

$$\Pr\left[X_0 = i\right] = \left(q_0\right)_i$$

Moreover, for $t \geq 1, k \in \{0, 1, \ldots, t\}$ and arbitrary states $j, s_k \in S$ we have

$$\Pr\left[X_{t+1} = j \mid X_t = s_t, X_{t-1} = s_{t-1}, \ldots, X_0 = s_0\right] = \Pr\left[X_{t+1} = j \mid X_t = s_t\right]$$

if both conditional probabilities are well defined, i.e. if $\Pr\left[X_t = s_t, X_{t-1} = s_{t-1}, \ldots, X_0 = s_0\right] > 0$. If the values

$$p_{ij} := \Pr\left[X_{t+1} = j \mid X_t = i\right]$$

are independent of t, we call the Markov chain (time-)homogeneous. In this case its transition matrix is defined by $P = (p_{ij})_{i,j \in S}$.

Definition 9.3 The random variable

$$T_{ij} := \min\left\{n \geq 1 \mid X_n = j, \text{ if } X_0 = i\right\}$$

counts the number of steps needed by the Markov chain to get from i to $j.T_{ij}$ is called hitting time from state i to state j. If j is never reached, we set $T_{ij} = \infty$. Furthermore we define $h_{ij} := \mathbb{E}\left[T_{ij}\right]$. The probability to get from state i to state j after arbitrarily many steps is called visit probability f_{ij}. Formally we define

$$f_{ij} := \Pr\left[T_{ij} < \infty\right].$$

Lemma 9.5 For the expected hitting times we have

$$h_{ij} = 1 + \sum_{k \neq j} p_{ik} h_{kj} \quad \text{for all } i, j \in S$$

provided that the expectations h_{ij} and h_{kj} exist. For the visit probabilities we have similarly

$$f_{ij} = p_{ij} + \sum_{k \neq j} p_{ik} f_{kj} \quad \text{for all } i, j \in S.$$

Theorem 9.9 (Additive Drift Theorem) Let $(X_t)_{t \in \mathbb{N}_0}$ be a Markov chain with state space $\mathcal{S} \subseteq [0, \infty)$ and assume $X_0 = n$. If there exists $c > 0$ such that for all $x \in S, x > 0$ we have

$$\mathbb{E}\left[X_{t+1} \mid X_t = x\right] \leq x - c,$$

then

$$\mathbb{E}[T] \leq \frac{n}{c}$$

Theorem 9.10 Let $(X_t)_{t \in \mathbb{N}_0}$ be a Makov chain with state space $\mathcal{S} = \mathbb{N}_0$ and let T and X_0 be as in Theorem 9.9. Furthermore let $C \in \mathbb{N}$ be some positive constant and denote by T_C the earliest point in time t such that $X_t \leq C$. Assume that there exist constants $p_0 > 0$ and $B > 0$ such that for all $c \in \{1, \ldots, C\} \cap S$:

$$\Pr\left[X_{t+1} = 0 \mid X_t = c\right] \geq p_0$$

and

$$\sum_{n > C} \Pr\left[X_{t+1} = n \mid X_t = c\right] \cdot \mathbb{E}\left[T_C \mid X_0 = n\right] \leq B$$

Then $\mathbb{E}[T] = \mathbb{E}\left[T_C\right] + O(1)$.

Theorem 9.11 Let $(X_t)_{t \in \mathbb{N}_0}$ be a Markov chain with state space $\mathcal{S}$ and with $X_0 = n$. Moreover, assume that $g : \mathcal{S} \to [0, \infty)$ is a function that is injective when restricted to the domain $g^{-1}((0,\infty))$. Let T be the earliest point in time t such that $g(X_t)=0$. Assume furthermore that there exists a constant $c>0$ such that for all

$$\mathbb{E}[g(X_{t+1}) \mid X_t = x] \le g(x) - c \quad \text{for all } x \in \mathcal{S} \text{ with } g(x) > 0.$$

Then we have $\mathbb{E}[T] \le \frac{g(n)}{c}$.

Theorem 9.12 **(Multiplicative Drift Theorem)** Let $(X_t)_{t\in\mathbb{N}_0}$ be a Markov chain with state space $\mathcal{S} \subset \{0\} \cup [1,\infty)$ and assume $X_0 = n$. Let T be the random variable that denotes the earliest point in time $t \ge 0$ such that $X_t = 0$. If there exists a constant $\varepsilon > 0$ such that

$$\mathbb{E}[X_{t+1} \mid X_t = x] \le (1-\varepsilon)x \quad \text{for all } x > 0,$$

then

$$\mathbb{E}[T] \le \frac{1+\log n}{\varepsilon}$$

and

$$\Pr\left[T \ge \frac{x+\log n}{\varepsilon} + 1\right] \le e^{-x} \quad \text{for all } x \ge 1$$

10 Markov Chains II: Stationary Distribution

Definition 10.1 Let P be the transition matrix of a Markov chain. A stationary distribution of the Markov chain is a state vector π satisfying $\pi = \pi \cdot P$.

Definition 10.2 A state i is called absorbing if it has no outgoing transition, i.e., if $p_{ij} = 0$ for all $j \ne i$ and thus $p_{ii} = 1$.

Definition 10.3 A Markov chain is called irreducible if for all pairs of states $i, j \in S$ there exists a number $n \in \mathbb{N}$ such that $p_{ij}^{(n)} > 0$.

Lemma 10.4 For irreducible finite Markov chains we have $f_{ij} = \Pr[T_{ij} < \infty] = 1$ for all states $i, j \in S$. Moreover all expectations $h_{ij} = \mathbb{E}[T_{ij}]$ exist.

Theorem 10.5 An irreducible finite Markov chain has a unique stationary distribution π, namely

$$\pi_j = \frac{1}{h_{jj}}, \quad j \in S.$$

Definition 10.6 The periodicity of a state j is the greatest common divisor of

$$\left\{n \in \mathbb{N}_0 \mid p_{jj}^{(n)} > 0\right\}.$$

A state with periodicity 1 is called aperiodic. A Markov chain is called aperiodic if all states are aperiodic.

Definition 10.7 Irreducible, aperiodic Markov chains are called ergodic.

Theorem 10.8 (Fundamental theorem for ergodic Markov chains) For every ergodic finite Markov chain $(X_t)_{t\in\mathbb{N}_0}$ we have independently of the initial distribution that

$$\lim_{t\to\infty} q_t = \pi,$$

where π denotes the chain's unique stationary distribution.

We close this section with some observations which may help to find the stationary distribution of a Markov chain. Recall that the transition matrix of a Markov chain has the property that all row sums are equal to one, the column sums, however, need not to be all equal to one. If all row and column sums are equal to one, then the transition matrix is said to be doubly stochastic. For such matrices we have the following lemma.

Lemma 10.9 Let $(X_t)_{t\in\mathbb{N}_0}$ be a finite ergodic Markov chains with state space S. Then the uniform distribution $\pi \equiv \frac{1}{|S|}$ is the unique stationary distribution if and only if the transition matrix P is doubly stochastic.

Lemma 10.10 Let $(X_t)_{t\in\mathbb{N}_0}$ be a finite ergodic Markov chains with state space S and transition matrix P. Then if a probability distribution π satisfies

$$\pi_x p_{xy} = \pi_y p_{yx} \quad \text{for all } x, y \in S,$$

then π is the unique stationary distribution.

Definition 10.11 An ergodic Markov chain is called reversible if for all states x, y and the stationary distribution π we have

$$\pi_x p_{xy} = \pi_y p_{yx}$$

Definition 10.12 For a Markov chain $(M_t)_{t\in\mathbb{N}_0}$ with transition matrix P, the ε-convergence time with respect to the total variation distance is given by

$$\tau_{TV}(\varepsilon) := \min\left\{t : \frac{1}{2}\sum_{y\in S}\left|\left(q_0 P^t\right)_y - \pi_y\right| \leq \varepsilon \text{ for all initial distributions } q_0\right\}$$

Definition 10.13 Let $\mathcal{I}$ be a class of problems and $\mathcal{MC}(\mathcal{I})$ the corresponding family of Markov chains. We call $\mathcal{MC}(\mathcal{I})$ rapidly mixing if

$$\tau_{TV}(\varepsilon) = \mathcal{O}\left(\text{ poly }\left(|I|, \log\varepsilon^{-1}\right)\right) \text{ for all } I \in \mathcal{I},$$

where $|I|$ denotes the input length of the instance I of the problem.

10.1 Flow

For this we relate every ergodic Markov chain $M = (M_t)_{t\in\mathbb{N}_0}$ with stationary distribution π to a directed graph $G_M = (V, A)$ with vertex set $V = S$ and edge set

$$A = \{(x, y) \mid x, y \in S, x \neq y \text{ and } p_{xy} > 0\}$$

(note that so far this is simply the transition diagram without loops). We define a weight function c on A by

$$c(x, y) := \pi_x p_{xy},$$

where as usual we write $c(x, y)$ instead of $c((x, y))$. The intuition behind this definition is that for $e = (x, y) \in A, c(e)$ is the probability that a Markov chain obeying its stationary distribution takes the edge e in the next step.

We can now define the following transporting problem on G_M. For every (ordered) pair (x, y) of states $x \neq y$ we are to transport an amount of $\pi_x\pi_y$ of some commodity g_{xy} from x to y. Equivalently, we search for a flow $f_{xy} : A \to \mathbb{R}_{\geq 0}$ such that

$$\sum_u f_{xy}(u, v) = \sum_u f_{xy}(v, u) \quad \text{for all } v \neq x, y$$

and such that the net flow out of x (which is equal to the net flow into y) is equal to $\pi_x \pi_y$, that is

$$\sum_u f_{xy}(x,u) - \sum_u f_{xy}(u,x) = \sum_u f_{xy}(u,y) - \sum_u f_{xy}(y,u) = \pi_x \pi_y.$$

For given flows $(f_{xy})_{x \neq y \in S}$, we define the total flow f as

$$f(e) := \sum_{x \neq y} f_{xy}(e)$$

and denote by

$$\rho(f) := \max_{e \in A} \frac{f(e)}{c(e)}$$

the maximum relative edge load. We now can state the following theorem (without proof), which characterizes rapidly mixing Markov chains with the help of this construction:

Theorem 10.14 Let $\mathcal{MC}(\mathcal{I})$ be a family of ergodic and reversible Markov chains defined by a class of problems $\mathcal{I}$. We denote by $\pi_{\min}(I)$ the smallest state probability that appears in the stationary distribution π of the Markov chain $M(I)$. If $\log \pi_{\min}(I)^{-1}$ is polynomially bounded in the input size $|I|$, then the family $\mathcal{MC}(\mathcal{I})$ is rapidly mixing if and only if for each Markov chain $M(I)$ there exist flows $(f_{xy})_{x \neq y \in S}$ such that $\rho(f)$ is polynomially bounded in $|I|$.

Definition 10.15 Let $(M_t)_{t \in \mathbb{N}_0}$ be a Markov chain with state space S and transition matrix $P = (p_{xy})_{x,y \in S}$ .A coupling of M_t is a Markov chain $Z_t = (X_t, Y_t)$ with state space $S \times S$ such that

$$\begin{aligned} &\Pr\left[X_{t+1} = x' \mid Z_t = (x,y)\right] = p_{xx'} && \forall x, x', y \in S, t \in \mathbb{N}_0 \\ \text{and } &\Pr\left[Y_{t+1} = y' \mid Z_t = (x,y)\right] = p_{yy'} && \forall x, y, y' \in S, t \in \mathbb{N}_0. \end{aligned}$$

Lemma 10.16 Let $(M_t)_{t \in \mathbb{N}_0}$ be a finite ergodic Markov chain with state space S and $Z_t = (X_t, Y_t)$ *a* coupling of M_t. Moreover, let $\epsilon > 0$ and t_0 be such that

$$\Pr\left[X_{t_0} \neq Y_{t_0} \mid X_0 = x, Y_0 = y\right] \leq \epsilon$$

is satisfied for all $x, y \in S$. Then we have

$$\tau_{TV}(\epsilon) \leq t_0$$

where τ_{TV} is defined as in Definition 10.12.

Lemma 10.17 Let $(M_t)_{t \in \mathbb{N}_0}$ be as in Lemma 10.16 and assume that $X_t = Y_t$ implies $X_{t+1} = Y_{t+1}$ and let t_0 be such that

$$\Pr\left[X_{t_0} \neq Y_{t_0} \mid X_0 = x, Y_0 = y\right] \leq 1/2$$

is satisfied for all $x, y \in S$. Then

$$\tau_{TV}(\epsilon) \leq \log_2\left(\varepsilon^{-1}\right) \cdot t_0 \quad \forall \varepsilon > 0$$

11 Generating Functions

Definition 11.1 The (probability) generating function of a nonnegative integer-valued random variable X is defined by

$$G_X(s) := \sum_{k=0}^{\infty} \Pr[X = k] \cdot s^k = \mathbb{E}\left[s^X\right]$$

Theorem 11.2 (Uniqueness of the probability generating function) The density and the distribution of a nonnegative integer-valued random variable X are uniquely determined by its probability generating function.

Some examples of probability generating functions are easily obtained: Binomial distribution. For $X \sim \mathrm{Bin}(n, p)$ we obtain by applying the binomial formula that

$$G_X(s) = \mathbb{E}\left[s^X\right] = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \cdot s^k = (1 - p + ps)^n.$$

Geometric distribution. Let X be a geometrically distributed random variable with success probability p. Then we have

$$G_X(s) = \mathbb{E}\left[s^X\right] = \sum_{k=1}^{\infty} p(1-p)^{k-1} \cdot s^k = ps \cdot \sum_{k=1}^{\infty} ((1-p)s)^{k-1} = \frac{ps}{1-(1-p)s}.$$

Poisson distribution. For $X \sim \mathrm{Po}(\lambda)$ we have

$$G_X(s) = \mathbb{E}\left[s^X\right] = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \cdot s^k = e^{-\lambda + \lambda s} = e^{\lambda(s-1)}$$

Example 11.3 Let X be binomially distributed with $X \sim \mathrm{Bin}(n, \lambda/n)$. For $n \to \infty$ we obtain

$$G_X(s) = \left(1 - \frac{\lambda}{n} + \frac{\lambda s}{n}\right)^n \to e^{\lambda(s-1)}$$

$G_X(s)$ thus converges to the probability generating function of a Poisson distributed random variable with parameter λ. This is in line with the observation that the Poisson distribution appears as limit of the binomial distribution.

Lemma 11.6 If X and Y are two independent nonnegative integer-valued random variables and $Z = X + Y$ denotes their sum, then

$$\Pr[Z = z] = \sum_{x=0}^{z} \Pr[X = x] \cdot \Pr[Y = z - x]$$

Theorem 11.7 (Generating Function of a Sum) Let $X_1, \dots, X_n$ be independent nonnegative integervalued random variables, and let $Z := X_1 + \dots + X_n$. Then

$$G_Z(s) = G_{X_1}(s) \cdot \ldots \cdot G_{X_n}(s)$$

Theorem 11.10 Let $X_1, X_2, \dots$ be independent and identically distributed random variables that are non-negative and integer-valued with probability generating function $G_X(s)$. Let N be another independent nonnegative integer-valued random variable with probability generating function $G_N(s)$. Then the random variable $Z := X_1 + \ldots + X_N$ has the probability generating function $G_Z(s) = G_N(G_X(s))$. Furthermore, we have

$$\mathbb{E}[Z] = \mathbb{E}[N] \cdot \mathbb{E}[X]$$

Definition: recurrent events Formally, the events $H_1, H_2, \ldots$ are called recurrent if

$$\Pr\left[H_i \mid \bar{H}_1 \cap \ldots \cap \bar{H}_{j-1} \cap H_j\right] = \Pr\left[H_{i-j}\right]$$

holds for all $i, j \in \mathbb{N}$ with $i > j$.

Given a sequence of $H_1, H_2, \ldots$ of recurrent events, we let the random variable Z denote the waiting time until one of the H_i occurs. Then

$$\Pr[Z < \infty] = \sum_{k \geq 1} \Pr\left[\bar{H}_1 \cap \ldots \cap \bar{H}_{k-1} \cap H_k\right]$$

The following theorem establishes a necessary and sufficient condition for the waiting time to be finite.

Theorem 11.13 Let $H_1, H_2, \ldots$ be recurrent events. Then for the waiting time Z we have $\Pr[Z < \infty] = 1$ if and only if the sum $\sum_{k=1}^{\infty} \Pr\left[H_k\right]$ diverges. Furthermore, we have

$$\Pr[Z < \infty] = 1 - \frac{1}{1 + \sum_{k=1}^{\infty} \Pr\left[H_k\right]}$$

if the sum converges.