Lesson 7: Probabilistic Turing machines

Theme: The notion of probabilistic/randomized Turing machines and some classical results.

Probabilistic Turing machines. A *probabilistic Turing machine* (PTM) is system $\mathcal{M} = \langle \Sigma, \Gamma, Q, q_0, q_{\mathsf{acc}}, q_{\mathsf{rej}}, \delta \rangle$ defined like the NTM, with the difference that $\delta \subseteq (Q - \{q_{\mathsf{acc}}, q_{\mathsf{rej}}\}) \times \Gamma \times Q \times \Gamma \times \{\texttt{Left}, \texttt{Right}\}$ is now a relation such that for every $(p, \sigma) \in (Q - \{q_{\mathsf{acc}}, q_{\mathsf{rej}}\}) \times \Gamma$, there are exactly two transitions that can be applied:

$$(p, \sigma) \to (q_1, \sigma_1, \texttt{Move}_1) \quad \text{and} \quad (p, \sigma) \to (q_2, \sigma_2, \texttt{Move}_2)$$

and the probability that each transition is applied is $1/2$. Intuitively, when it is in state p reading symbol σ, $\mathcal{M}$ tosses an unbiased coin to decide whether to apply $(q_1, \sigma_1, \texttt{Move}_1)$ or $(q_2, \sigma_2, \texttt{Move}_2)$. On an input word w, the probability that $\mathcal{M}$ accepts/rejects w is defined over all possible coin tossing.

Similar to DTM/NTM, we say that $\mathcal{M}$ *runs in time* $f(n)$, if for every word w, every run of $\mathcal{M}$ on w has length $\leqslant f(|w|)$. We say that $\mathcal{M}$ *runs in polynomial time*, if there is a polynomial $p(n) = \mathsf{poly}(n)$ such that $\mathcal{M}$ runs in time $p(n)$. In this case we also say that $\mathcal{M}$ is a *polynomial time PTM.*

The class **BPP** is defined as follows. A language L is in the class **BPP**, if there a polynomial time PTM $\mathcal{M}$ such that for every input word x, the following holds.

$$\mathbf{Pr}[\ \mathcal{M}(x) = L(x)\ ] \quad \geqslant \quad 2/3$$

Here we treat a language L as a function $L : \{0,1\}^* \to \{0,1\}$, where $L(x) = 1$, if $x \in L$, and $L(x) = 0$, if $x \notin L$. Similarly, we treat TM $\mathcal{M}$ as a function $\mathcal{M} : \{0,1\}^* \to \{0,1\}$, where $\mathcal{M}(x) = 1$, if $\mathcal{M}$ accepts x, and $\mathcal{M}(x) = 0$, if $\mathcal{M}$ rejects x.

Note that **BPP** is closed under complement, union and intersection.

Remark 7.1 Alternatively, we can define the class **BPP** as follows. A language L is in the class **BPP**, if there is a polynomial $q(n)$ and a polynomial time DTM $\mathcal{M}$ such that for every $x \in \{0,1\}^*$, the following holds.

$$\mathbf{Pr}_{r \in \{0,1\}^{q(|x|)}}[\ \mathcal{M}(x, r) = L(x)\ ] \quad \geqslant \quad 2/3$$

Note that the DTM $\mathcal{M}$ takes as input (x, r). Intuitively, it can be viewed as a PTM that on input x, first randomly choose a string r of length $q(|x|)$, then run DTM $\mathcal{M}$ on (x, r).

Note the similarity with the alternative definition of **NP** (Def. 0.4), where an NTM first guesses a certificate string r, and then runs a DTM for verification.

Theorem 7.2 (Error reduction) *Let $L \in \mathbf{BPP}$. Then, for every $d \geqslant 1$, there is a polynomial time PTM $\mathcal{M}$ such that for every input word x:*

$$\mathbf{Pr}[\ \mathcal{M}(x) = L(x)\ ] \quad \geqslant \quad 1 - 2^{-\alpha |x|^d} \qquad \textit{(for some fixed } \alpha > 0)$$

Theorem 7.3 (Adleman 1978) $\mathbf{BPP} \subseteq \mathbf{P}_{/\mathsf{poly}}$.

Theorem 7.3 and Theorem 6.4 imply that if $\mathsf{SAT} \in \mathbf{BPP}$, then **PH** collapses to $\mathbf{\Sigma}_2^p$.

Theorem 7.4 (Sipser, Gács, Lautemann 1983) $\mathbf{BPP} \subseteq \mathbf{\Sigma}_2^p \cap \mathbf{\Pi}_2^p$.

One-sided error PTM. The class **RP** is defined as follows. A language L is in the class **RP**, if there a polynomial time PTM $\mathcal{M}$ such that for every input word x, the following holds.

- If $x \in L$, then $\mathbf{Pr}[\ \mathcal{M}(x) = 1\ ] \geqslant 2/3$.
- If $x \notin L$, then $\mathbf{Pr}[\ \mathcal{M}(x) = 0\ ] = 1$.

Note that $\mathcal{M}$ is never wrong when the input $x \notin L$, hence, the name *one-sided*. The class **coRP** is defined as $\mathbf{coRP} \stackrel{\mathsf{def}}{=} \{L : \{0,1\}^* \setminus L \in \mathbf{RP}\}$.

Zero error PTM. A PTM $\mathcal{M}$ for a language L is a zero error PTM, if it never errs, i.e., for every input word x, $\mathbf{Pr}[\ \mathcal{M}(x) = L(x)\ ] = 1$. Now for a PTM $\mathcal{M}$ and input word x, we can define a random variable $T_{\mathcal{M},x}$ to denote the run time of $\mathcal{M}$ on x, where the probability distribution is $\mathbf{Pr}[\ T_{\mathcal{M},x} = t\ ] = p$, if with probability p over the random strings of $\mathcal{M}$ on input x, it halts in t steps .

The class **ZPP** is defined as follows. A language L is in **ZPP**, if there is a polynomial $q(n) = \mathsf{poly}(n)$ and a zero error PTM $\mathcal{M}$ for L such that for every input word x, $\mathbf{Exp}[T_{\mathcal{M},x}] \leqslant q(|x|)$.

The algorithms for languages in **BPP**/**RP**/**coRP** are also called *Monte Carlo* algorithms, and those for languages in **ZPP** are called *Las Vegas* algorithms.

Appendix

A Useful inequalities

Inclusion-exclusion principle: Let $\mathcal{E}_1, \ldots, \mathcal{E}_m$ be some m events. Then, the following holds.

$$\mathbf{Pr}\Big[\bigcup_{i=1}^{m} \mathcal{E}_i\Big] \;=\; \sum_{i=1}^{m} \mathbf{Pr}[\ \mathcal{E}_i\ ] - \sum_{1 \leqslant i_1 < i_2 \leqslant m} \mathbf{Pr}[\ \mathcal{E}_{i_1} \cap \mathcal{E}_{i_2}\ ] + \sum_{1 \leqslant i_1 < i_2 < i_3 \leqslant m} \mathbf{Pr}[\ \mathcal{E}_{i_1} \cap \mathcal{E}_{i_2} \cap \mathcal{E}_{i_3}\ ] - \cdots$$

From here, we also obtain the so called *union bound*:

$$\mathbf{Pr}\Big[\bigcup_{i=1}^{m} \mathcal{E}_i\Big] \;\leqslant\; \sum_{i=1}^{m} \mathbf{Pr}[\ \mathcal{E}_i\ ]$$

Markov inequality: Let X be a non-negative random variable with expectation μ. Then, for every real $c > 0$, the following holds.

$$\mathbf{Pr}[\ X \geqslant c\mu\ ] \;\leqslant\; 1/c$$

Markov inequality is often also called *averaging argument*.

Chebyshev inequality: Let X be a random variable with expectation μ and variance σ^2. Then, for every real $c > 0$, the following holds.

$$\mathbf{Pr}\Big[\ |X - \mu| \geqslant c\sigma\ \Big] \;\leqslant\; 1/c^2$$

Chernoff inequality: Let $X_1, \ldots, X_m$ be (independent) 0,1 random variables. Suppose for every $1 \leqslant i \leqslant m$, $\mathbf{Pr}[X_i = 1] = p$, for some $p > 1/2$. Let $X \stackrel{\mathsf{def}}{=} \sum_{i=1}^{m} X_i$. Then, the following holds.

$$\mathbf{Pr}\Big[\ X > \lfloor m/2 \rfloor\ \Big] \;\geqslant\; 1 - 2^{-\alpha m} \qquad\qquad \text{where } \alpha = \frac{\log_2 e}{2p}\Big(p - \frac{1}{2}\Big)^2$$