Logical Methods in Computer Science
Vol. 7 (3:22) 2011, pp. 1–31
www.lmcs-online.org

Submitted Mar. 8, 2011
Published Sep. 29, 2011

POWER OF RANDOMIZATION IN AUTOMATA ON INFINITE STRINGS *

ROHIT CHADHA [a], A. PRASAD SISTLA [b], AND MAHESH VISWANATHAN [c]

[a] LSV, ENS Cachan & CNRS & INRIA, France
e-mail address: chadha.rohit@gmail.com

[b] Univ. of Illinois at Chicago, U.S.A.
e-mail address: sistla@cs.uic.edu

[c] Univ. of Illinois at Urbana-Champaign, U.S.A
e-mail address: vmahesh@cs.uiuc.edu

Abstract. Probabilistic Büchi Automata (PBA) are randomized, finite state automata that process input strings of infinite length. Based on the threshold chosen for the acceptance probability, different classes of languages can be defined. In this paper, we present a number of results that clarify the power of such machines and properties of the languages they define. The broad themes we focus on are as follows. We present results on the decidability and precise complexity of the emptiness, universality and language containment problems for such machines, thus answering questions central to the use of these models in formal verification. Next, we characterize the languages recognized by PBAs topologically, demonstrating that though general PBAs can recognize languages that are not regular, topologically the languages are as simple as ω-regular languages. Finally, we introduce Hierarchical PBAs, which are syntactically restricted forms of PBAs that are tractable and capture exactly the class of ω-regular languages.

1. Introduction

Automata on infinite (length) strings have played a central role in the specification, modeling and verification of non-terminating, reactive and concurrent systems [VW86, Kur94, VWS83, HP96, Sis83]. However, there are classes of systems whose behavior is probabilistic in nature; the probabilistic behavior being either due to the employment of randomization in the algorithms executed by the system or due to other uncertainties in the system, such as failures, that are modeled probabilistically. While Markov Chains and Markov

1998 ACM Subject Classification: F.4.3,D.2.4,F.1.1,F.1.2.
Key words and phrases: Automata on infinite strings, Randomization, Omega-regular languages, Expressiveness, Decidability, Probabilistic Monitors.

* An extended abstract of the paper appeared in [CSV09b].

[a] Supported in part by NSF grants CCF04-29639 and NSF CCF04-48178.
[b] Supported in part by NSF CCF-0742686.
[c] Supported in part by NSF CCF04-48178 and NSF CCF05-09321.

DOI:10.2168/LMCS-7 (3:22) 2011

Decision Processes have been used to model such behavior in the formal verification community [RKNP04], both these models do not adequately capture *open*, *reactive* probabilistic systems that continuously accept inputs from an environment. The most appropriate model for such systems are probabilistic automata on infinite strings, which are the focus of study in this paper.

Probabilistic Büchi Automata (PBA) have been introduced in [BG05] to capture such computational devices. These automata generalize probabilistic finite automata (PFA) [Rab63, Sal73, Paz71] from finite length inputs to infinite length inputs. Informally, PBAs are like finite-state automata except that they differ in two respects. First, from each state and on each input symbol, the PBA may roll a dice to determine the next state. Second, the notion of acceptance is different because PBAs are probabilistic in nature and have infinite length input strings. The behavior of a PBA on a given infinite input string can be captured by an infinite Markov chain that defines a probability measure on the space of runs/executions of the machine on the given input. Like Büchi automata, a run is considered to be accepting if some accepting state occurs infinitely often, and therefore, the probability of acceptance of the input is defined to be the measure of all accepting runs on the given input. There are two possible languages that one can associate with a PBA $\mathcal{B}$ [BG05, BBG08] — $\mathcal{L}_{>0}(\mathcal{B})$ (called *probable semantics*) consisting of all strings whose probability of acceptance is non-zero, and $\mathcal{L}_{=1}(\mathcal{B})$ (called *almost sure semantics*) consisting of all strings whose probability of acceptance is 1. Based on these two languages, one can define two classes of languages — $\mathbb{L}(\text{PBA}^{>0})$, and $\mathbb{L}(\text{PBA}^{=1})$ which are the collection of all languages (of infinite length strings) that can be accepted by some PBA with respect to probable, and almost sure semantics, respectively. In this paper we study the expressive power of, and decision problems for these classes of languages.

We present a number of new results that highlight three broad themes. First, we establish results on decidability and precise complexity of the canonical decision problems in verification, namely, emptiness, universality, and language containment, for the classes $\mathbb{L}(\text{PBA}^{>0})$ and $\mathbb{L}(\text{PBA}^{=1})$. For the decision problems, we focus our attention on RatPBAs which are PBAs in which all transition probabilities are rational. For RatPBAs $\mathcal{B}$ and $\mathcal{B}'$, our results are as follows.

(A) Checking if $\mathcal{L}_{=1}(\mathcal{B}) = \emptyset$ and $\mathcal{L}_{=1}(\mathcal{B}) = \Sigma^*$ are **PSPACE**-complete.
(B) The problems of checking if $\mathcal{L}_{>0}(\mathcal{B}) = \emptyset$, and if $\mathcal{L}_{>0}(\mathcal{B}) = \Sigma^*$ are $\mathbf{\Sigma}^0_2$-complete.
(C) The problems of checking if $\mathcal{L}_{=1}(\mathcal{B}) \subseteq \mathcal{L}_{=1}(\mathcal{B}')$ and if $\mathcal{L}_{>0}(\mathcal{B}) \subseteq \mathcal{L}_{>0}(\mathcal{B}')$ are $\mathbf{\Sigma}^0_2$-complete.

The decidability of the universality checking of $\mathcal{L}_{=1}(\mathcal{B})$ (bullet (A) above) is a new result. The result establishing the **PSPACE**-completeness of emptiness checking of $\mathcal{L}_{=1}(\mathcal{B})$ (bullet (A) above) substantially improves the result of [BBG08] where it was shown to be decidable in **EXPTIME** and conjectured to be **EXPTIME**-hard. The improved upper bound for emptiness checking is established by observing that the complement of the language $\mathcal{L}_{=1}(\mathcal{B})$ is recognized by a special PBA $\mathcal{M}$ (with probable semantics) called a *finite state probabilistic monitor* (FPM) [CSV08, CSV09a] and then exploiting a result in [CSV09a] that shows that the language of an FPM is universal if and only if there is an *ultimately periodic word* in the complement of the language recognized by a FPM. This observation of the existence of ultimately periodic words does not carry over to the class $\mathbb{L}(\text{PBA}^{>0})$. However, we show that $\mathcal{L}_{>0}(\mathcal{B})$ is non-empty iff it contains a *strongly asymptotic word*, which is a generalization of ultimately periodic word. This allows us to show that the emptiness problem for $\mathbb{L}(\text{PBA}^{>0})$, though undecidable as originally shown in [BBG08], is $\mathbf{\Sigma}^0_2$-complete (bullet (B) above), where $\mathbf{\Sigma}^0_2$ is a set in the second level of the arithmetic hierarchy. This result is noteworthy because typically problems of automata on infinite words that are undecidable tend to lie way beyond the arithmetic hierarchy in the analytical hierarchy. Finally, given that the emptiness and universality problems for $\mathbb{L}(\text{PBA}^{=1})$ are in **PSPACE** (bullet (A)), one would expect language containment under almost sure semantics to be at least decidable. However, surprisingly, we show that it is, in fact, $\mathbf{\Sigma}^0_2$-complete (bullet (C) above).

The second theme brings to sharper focus the correspondence between nondeterminism and probable semantics, and between determinism and almost sure semantics, in the context of automata on infinite words. This correspondence was hinted at in [BBG08]. There it was observed that $\mathbb{L}(\text{PBA}^{=1})$ is a strict subset of $\mathbb{L}(\text{PBA}^{>0})$ and that while Büchi, Rabin and Streett acceptance conditions all yield the same class of languages under the probable semantics, they yield different classes of languages under the almost sure semantics. These observations mirror the situation in non-probabilistic automata — languages recognized by deterministic Büchi automata are a strict subset of the class of languages recognized by nondeterministic Büchi automata, and while Büchi, Rabin and Streett acceptances are equivalent for nondeterministic machines, Büchi acceptance is strictly weaker than Rabin and Streett for deterministic machines. In this paper we further strengthen this correspondence through a number of results on the closure properties as well as the topological structure of $\mathbb{L}(\text{PBA}^{>0})$ and $\mathbb{L}(\text{PBA}^{=1})$.

First we consider closure properties. It was shown in [BBG08] that the class $\mathbb{L}(\text{PBA}^{>0})$ is closed under all the Boolean operations (like the class of languages recognized by nondeterministic Büchi automata) and that $\mathbb{L}(\text{PBA}^{=1})$ is not closed under complementation. We extend these observations as follows.

(D) $\mathbb{L}(\text{PBA}^{=1})$ is closed under intersection and union.
(E) Every language in $\mathbb{L}(\text{PBA}^{>0})$ can be expressed as the Boolean combination of languages in $\mathbb{L}(\text{PBA}^{=1})$.

These results mimic similar observations about Büchi automata — the class of languages recognized by deterministic Büchi automata is closed under union and intersection, but not complementation; and, any ω-regular language (or languages recognized by nondeterministic Büchi machines) can be expressed as the Boolean combination of languages recognized by deterministic Büchi automata.

Next, we characterize the classes topologically. There is a natural topological space on infinite length strings called the *Cantor topology* [Tho90]. We show that, like ω-regular languages, all the classes of languages defined by PBAs lie in very low levels of this Borel hierarchy. We show that–

(F) $\mathbb{L}(\text{PBA}^{=1})$ is strictly contained in $\mathcal{G}_\delta$, just like the class of languages recognized by deterministic Büchi is strictly contained in $\mathcal{G}_\delta$.
(G) $\mathbb{L}(\text{PBA}^{>0})$ is strictly contained in the Boolean closure of $\mathcal{G}_\delta$ much like the case for ω-regular languages.

The last theme identifies syntactic restrictions on PBAs that captures regularity. Much like PFAs for finite word languages, PBAs, though finite state, allow one to recognize non-regular languages. It has been shown [BG05, BBG08] that both $\mathbb{L}(\text{PBA}^{>0})$ and $\mathbb{L}(\text{PBA}^{=1})$ contain non-ω-regular languages. A question initiated in [BG05] was to identify restrictions on PBAs that ensure that PBAs have the same expressive power as finite-state (non-probabilistic) machines. One such restriction was identified in [BG05], where it was shown that *uniform* PBAs with respect to the probable semantics capture exactly the class of ω-regular languages. However, the uniformity condition identified by Baier et. al. was semantic in nature. In this paper, we identify one simple syntactic restriction (i.e., one that is based only on the local transition structure of the machine, and can be efficiently checked) that captures regularity both for probable semantics and almost sure semantics. Not only, the restricted PBAs capture the notion of regularity, they are also very tractable.

The restriction we consider is that of a hierarchical structure. A *Hierarchical* PBA *(*HPBA*)* is a PBA whose states are partitioned into different levels such that, from any state q, on an input symbol a, at most one transition with non-zero probability goes to a state at the same level as q and all others go to states at higher level. We show that –

(H) HPBAs with respect to probable semantics define exactly the class of ω-regular languages.
(I) HPBAs with respect to almost sure semantics define exactly the class of ω-regular languages in $\mathbb{L}(\mathrm{PBA}^{=1})$, namely, those recognized by deterministic Büchi automata.
(J) Emptiness and universality problems for probable semantics for HPBAs with rational transition probabilities are **NL**-complete and **PSPACE**-complete, respectively.
(K) Emptiness and universality problems for almost sure semantics for HPBAs with rational transition probabilities **PSPACE**-complete and **NL**-complete, respectively.

The complexity of decision problems for HPBAs under probable semantics is interesting because this is the exact same complexity as that for (non-probabilistic) Büchi automata. In contrast, the emptiness problem for uniform PBA has been shown to be in **EXPTIME** and co-**NP**-hard [BG05]; thus, they seem to be less tractable than HPBA.

The rest of the paper is organized as follows. After discussing closely related work, we start with some preliminaries (in Section 2) before introducing PBAs. We present our results about the probable semantics in Section 3, and almost sure semantics in Section 4. Hierarchical PBAs are introduced in Section 5, and conclusions are presented in Section 6.

Related Work. Probabilistic Büchi automata (PBA), introduced in [BG05], generalize the model of Probabilistic Finite Automata [Rab63, Sal73, Paz71] to consider inputs of infinite length. In [BG05], Baier and Größer only considered the probable semantics for PBA. They also introduced the model of uniform PBAs to capture ω-regular languages and showed that the emptiness problem for such machines is in **EXPTIME** and co-**NP**-hard. The almost sure semantics for PBA was first considered in [BBG08] where a number of results were established. It was shown that $\mathbb{L}(\mathrm{PBA}^{>0})$ are closed under all Boolean operations, $\mathbb{L}(\mathrm{PBA}^{=1})$ is strictly contained in $\mathbb{L}(\mathrm{PBA}^{>0})$, the emptiness problem for $\mathbb{L}(\mathrm{PBA}^{>0})$ is undecidable, and the emptiness problem of $\mathbb{L}(\mathrm{PBA}^{=1})$ is in **EXPTIME**. We extend and sharpen the results of this paper. In a series of previous papers [CSV08, CSV09a], we considered a special class of PBAs called FPMs (Finite state Probabilistic Monitors) whose accepting set of states consists of all states excepting a rejecting state which is also absorbing. There we proved a number of results on the expressiveness and decidability/complexity of problems for FPMs. We draw on many of these observations to establish new results for the more general model of PBAs.

An extended abstract of this paper appeared in [CSV09b]. Several proofs were omitted in [CSV09b] for lack of space, and the current version includes all of these proofs.

2. Preliminaries

The set of natural numbers will be denoted by $\mathbb{N}$, the closed unit interval by $[0,1]$ and the open unit interval by $(0,1)$. The power-set of a set X will be denoted by 2^X.

Sequences. Given a finite set S, $|S|$ denotes the cardinality of S. Given a sequence (finite or infinite) $\kappa = s_0, s_1, \ldots$ over S, $|\kappa|$ will denote the length of the sequence (for infinite sequence $|\kappa|$ will be ω), and $\kappa[i]$ will denote the ith element s_i of the sequence. As usual S^* will denote the set of all finite sequences/strings/words over S, S^+ will denote the set of all finite non-empty sequences/strings/words over S and S^ω will denote the set of all infinite sequences/strings/words over S. Given $\eta \in S^*$ and $\kappa \in S^* \cup S^\omega$, $\eta\kappa$ is the sequence obtained by concatenating the two sequences in order. Given $\mathsf{L}_1 \subseteq \Sigma^*$ and $\mathsf{L}_2 \subseteq \Sigma^\omega$, the set $\mathsf{L}_1\mathsf{L}_2$ is defined to be $\{\eta\kappa \mid \eta \in \mathsf{L}_1 \text{ and } \kappa \in \mathsf{L}_2\}$. Given natural numbers $i, j \leq |\kappa|$, $\kappa[i:j]$ is the finite sequence $s_i, \ldots s_j$, where $s_k = \kappa[k]$. The set of *finite prefixes* of κ is the set $\mathit{Pref}(\kappa) = \{\kappa[0,j] \mid j \in \mathbb{N}, j \leq |\kappa|\}$.

Arithmetical Hierarchy. Let Γ be a finite alphabet. A language L over Γ is a set of finite strings over Γ. Arithmetical hierarchy consists of classes of languages $\mathbf{\Sigma}^0_n$, $\mathbf{\Pi}^0_n$ for each integer $n > 0$. Fix an $n > 0$. A language $\mathsf{L} \in \mathbf{\Sigma}^0_n$ iff there exists a recursive predicate $\phi(u, \vec{x}_1, ..., \vec{x}_n)$ where u is a variable ranging over Γ^*, and for each i,$0 < i \leq n$, $\vec{x}_i$ is a finite sequence of variables ranging over integers such that

$$L \;=\; \{u \in \Gamma^* \mid \exists \vec{x}_1, \forall \vec{x}_2, \ldots, Q_n \vec{x}_n \; \phi(u, \vec{x}_1, ..., \vec{x}_n)\}$$

where Q_n is an existential quantifier if n is odd, else it is a universal quantifier. Note that the quantifiers in the above equation are alternating starting with an existential quantifier. The class $\mathbf{\Pi}^0_n$ is exactly the class of languages that are complements of languages in $\mathbf{\Sigma}^0_n$. $\mathbf{\Sigma}^0_1$, $\mathbf{\Pi}^0_1$ are exactly the class of **R.E.**-sets and **co-R.E.**-sets. A canonical $\mathbf{\Sigma}^0_1$-complete [1] language is the set of deterministic Turing machine encodings that halt on some input string. A well known $\mathbf{\Sigma}^0_2$-complete language is the set of deterministic Turing machine encodings that halt on finitely many inputs.

Languages of infinite words. A language L of infinite words over a finite alphabet Σ is a subset of Σ^ω. (Please note we restrict only to finite alphabets). A set of languages of infinite words over Σ is said to be a class of languages of infinite words over Σ. Given a class $\mathcal{L}$, the Boolean closure of $\mathcal{L}$, denoted $\mathsf{BCl}(\mathcal{L})$, is the smallest class containing $\mathcal{L}$ that is closed under the Boolean operations of complementation, union and intersection.

Automata and ω-regular Languages. A *finite automaton on infinite words*, $\mathcal{A}$, over a (finite) alphabet Σ is a tuple (Q, q_0, F, Δ), where Q is a finite set of states, $\Delta \subseteq Q \times \Sigma \times Q$ is the transition relation, $q_0 \in Q$ is the initial state, and F defines the accepting condition. The nature of F depends on the type of automaton we are considering; for a *Büchi automaton* $F \subseteq Q$, while for a *Rabin automaton* F is a finite subset of $2^Q \times 2^Q$. If for every $q \in Q$ and $a \in \Sigma$, there is exactly one q' such that $(q, a, q') \in \Delta$ then $\mathcal{A}$ is called a *deterministic* automaton. Let $\alpha = a_0, a_1, \ldots$ be an infinite string over Σ. A *run* r *of* $\mathcal{A}$ on α is an infinite sequence $s_0, s_1, \ldots$ over Q such that $s_0 = q_0$ and for every $i \geq 0$, $(s_i, a_i, s_{i+1}) \in \Delta$. The notion of an *accepting run* depends on the type of automaton we consider. For a Büchi automaton, r is accepting if some state in F appears infinitely often in r. On the other hand for a Rabin automaton, r is accepting if it satisfies the *Rabin acceptance condition* — there

[1] Let $\mathcal{C}$ be a class in the arithmetic hierarchy. $L \in \mathcal{C}$ is said to be $\mathcal{C}$-complete if $L \in C$, and for every $L' \in \mathcal{C}$ there is a computable function f such that $x \in L'$ iff $f(x) \in L$.

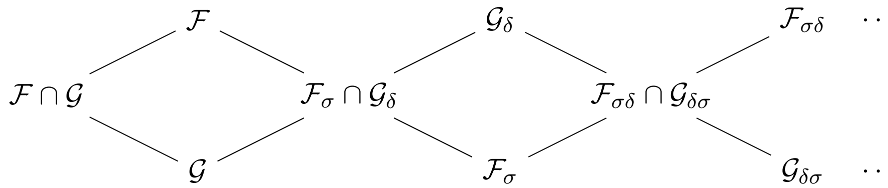

Figure 1: The Borel Hierarchy. Inclusions from left to right are strict.

is some pair $(B_i, G_i) \in F$ such that all the states in B_i appear only finitely many times in r, while at least one state in G_i appears infinitely many times. The automaton $\mathcal{A}$ *accepts* the string α if it has an accepting run on α. The *language accepted (recognized) by* $\mathcal{A}$, denoted by $\mathcal{L}(\mathcal{A})$, is the set of strings that $\mathcal{A}$ accepts. A language $\mathsf{L} \subseteq \Sigma^\omega$ is called *ω-regular* iff there is some Büchi automata $\mathcal{A}$ such that $\mathcal{L}(\mathcal{A}) = \mathsf{L}$. In this paper, given a fixed alphabet Σ, we will denote the class of ω-regular languages by $\mathsf{Regular}$. It is well-known that unlike the case of finite automata on finite strings, deterministic Büchi automata are less powerful than nondeterministic Büchi automata. On the other hand, nondeterministic Rabin automata and deterministic Rabin automata have the expressive power and they recognize exactly the class $\mathsf{Regular}$. Finally, we will sometimes find it convenient to consider automata $\mathcal{A}$ that do not have finitely many states. We will say that a language L is *deterministic* iff it can be accepted by a deterministic Büchi automaton that does not necessarily have finitely many states. We denote by Deterministic the collection of all deterministic languages. Please note that the class Deterministic strictly contains the class of languages recognized by finite state deterministic Büchi automata. The following are well-known results [PP04, Tho90].

Proposition 2.1. $\mathsf{L} \in \mathsf{Regular} \cap \mathrm{Deterministic}$ iff there is a finite state deterministic Büchi automaton $\mathcal{A}$ such that $\mathcal{L}(\mathcal{A}) = \mathsf{L}$. Furthermore, $\mathsf{Regular} \cap \mathrm{Deterministic} \subsetneq \mathsf{Regular}$ and $\mathsf{Regular} = \mathsf{BCl}(\mathsf{Regular} \cap \mathrm{Deterministic})$.

Topology on infinite strings. The set Σ^ω comes equipped with a natural topology called the *Cantor topology*. The collection of open sets is the collection $\mathcal{G} = \{\mathsf{L}\Sigma^\omega \mid \mathsf{L} \subseteq \Sigma^+\}$.[2] The collection of closed sets, $\mathcal{F}$, is the collection of *prefix-closed sets* — L is prefix-closed if for every infinite string α, if every prefix of α is a prefix of some string in L, then α itself is in L. In the context of verification of reactive systems, closed sets are also called *safety languages* [Lam85, AS85].

Borel Hierarchy on the Cantor space. For a class $\mathcal{L}$ of languages, we define $\mathcal{L}_\delta = \{\cap_{i\in\mathbb{N}} \mathsf{L}_i \mid \mathsf{L}_i \in \mathcal{L}\}$ and $\mathcal{L}_\sigma = \{\cup_{i\in\mathbb{N}} \mathsf{L}_i \mid \mathsf{L}_i \in \mathcal{L}\}$. The set of open sets of the Cantor space is closed under arbitrary unions but only finite intersections. Similarly the set of closed sets of the Cantor union is closed arbitrary intersections but only finite unions. The Borel hierarchy of the Cantor space is obtained by the means of countable unions, intersections and complementation, and is shown in Figure 1. This yields a transfinite hierarchy, but we will restrict our attention to the first few levels. At the lowest level of this hierarchy is the collection $\mathcal{G} \cap \mathcal{F}$ which is strictly contained in both $\mathcal{G}$ and $\mathcal{F}$ which form the next level of the hierarchy. Both $\mathcal{G}$ and $\mathcal{F}$ are strictly contained in the collection $\mathcal{G}_\delta \cap \mathcal{F}_\sigma$ which forms the next level. The collection $\mathcal{G}_\delta \cap \mathcal{F}_\sigma$ is strictly contained in $\mathcal{G}_\delta$ and $\mathcal{F}_\sigma$ which is at the next

[2] This topology is also generated by the metric $d : \Sigma^\omega \times \Sigma^\omega \to [0, 1]$ where $d(\alpha, \beta)$ is 0 iff $\alpha = \beta$; otherwise it is $\frac{1}{2^i}$ where i is the smallest integer such that $\alpha[i] \neq \beta[i]$.

level. $\mathcal{G}_\delta$ and $\mathcal{F}_\sigma$ are strictly contained in $\mathcal{G}_{\delta\sigma} \cap \mathcal{F}_{\sigma\delta}$ which itself is strictly contained in both $\mathcal{G}_{\delta\sigma}$ and $\mathcal{F}_{\sigma\delta}$. One remarkable result in automata theory is that the class of languages $\mathcal{G}_\delta$ coincides exactly with the class of languages recognized by infinite-state deterministic Büchi automata [Lan69, PP04, Tho90]. This combined with the fact that the class of ω-regular languages is the Boolean closure of ω-regular deterministic Büchi automata yields that the class of ω-regular languages is strictly contained in $\mathsf{BCl}(\mathcal{G}_\delta)$ which itself is strictly contained in $\mathcal{G}_{\delta\sigma} \cap \mathcal{F}_{\sigma\delta}$ [PP04, Tho90].

Proposition 2.2. $\mathcal{G}_\delta =$ Deterministic, and $\mathsf{Regular} \subsetneq \mathsf{BCl}(\mathcal{G}_\delta) \subsetneq \mathcal{G}_{\delta\sigma} \cap \mathcal{F}_{\sigma\delta}$.

2.1. **Probabilistic Büchi automata.** We shall now recall the definition of probabilistic Büchi automata given in [BG05]. Informally, PBAs are like finite-state deterministic Büchi automata except that the transition function from a state on a given input is described as a probability distribution that determines the probability of the next state. PBAs generalize the probabilistic finite automata (PFA) [Rab63, Sal73, Paz71] on finite input strings to infinite input strings. Formally,

Definition 2.3. A *finite state probabilistic Büchi automata (*PBA*)* over a finite alphabet Σ is a tuple $\mathcal{B} = (Q, q_s, Q_f, \delta)$ where Q is a finite set of *states*, $q_s \in Q$ is the *initial state*, $Q_f \subseteq Q$ is the set of *accepting/final states*, and $\delta : Q \times \Sigma \times Q \to [0, 1]$ is the *transition relation* such that for all $q \in Q$ and $a \in \Sigma$, $\sum_{q' \in Q} \delta(q, a, q') = 1$. In addition, if $\delta(q, a, q')$ is a rational number for all $q, q' \in Q, a \in \Sigma$, then we say that $\mathcal{M}$ is a rational probabilistic Büchi automata (RatPBA).

Notation: The transition function δ of PBA $\mathcal{B}$ on input a can be seen as a square matrix δ_a of order $|Q|$ with the rows labeled by "current" state, columns labeled by "next state" and the entry $\delta_a(q, q')$ equal to $\delta(q, a, q')$. Given a word $u = a_0 a_1 \dots a_n \in \Sigma^+$, δ_u is the matrix product $\delta_{a_0}\delta_{a_1} \dots \delta_{a_n}$. For an empty word $\epsilon \in \Sigma^*$ we take δ_ϵ to be the identity matrix. Finally for any $Q_0 \subseteq Q$, we define $\delta_u(q, Q_0) = \sum_{q' \in Q_0} \delta_u(q, q')$. Given a state $q \in Q$ and a word $u \in \Sigma^+$, $\mathsf{post}(q, u) = \{q' \mid \delta_u(q, q') > 0\}$.

Intuitively, the PBA starts in the initial state q_s and if after reading $a_0, a_1 \dots, a_i$ results in state q, then it moves to state q' with probability $\delta_{a_{i+1}}(q, q')$ on symbol a_{i+1}. Given a word $\alpha \in \Sigma^\omega$, the PBA $\mathcal{B}$ can be thought of as a infinite state Markov chain which gives rise to the standard σ-algebra defined using cylinders and the standard probability measure on Markov chains [Var85, KS76] as follows. Given a word $\alpha \in \Sigma^\omega$, the *probability space generated* by $\mathcal{B}$ and α is the probability space $(Q^\omega, \mathcal{F}_{\mathcal{B},\alpha}, \mu_{\mathcal{B},\alpha})$ where

- $\mathcal{F}_{\mathcal{B},\alpha}$ is the smallest σ-algebra on Q^ω generated by the collection $\{\mathsf{C}_\eta \mid \eta \in Q^+\}$ where $\mathsf{C}_\eta = \{\rho \in Q^\omega \mid \eta \text{ is a prefix of } \rho\}$.
- $\mu_{\mathcal{B},\alpha}$ is the unique probability measure on $(Q^\omega, \mathcal{F}_{\mathcal{B},\alpha})$ such that $\mu_{\mathcal{B},\alpha}(\mathsf{C}_{q_0 \dots q_n})$ is
  - 0 if $q_0 \neq q_s$,
  - 1 if $n = 0$ and $q_0 = q_s$, and
  - $\delta(q_0, \alpha(0), q_1) \dots \delta(q_{n-1}, \alpha(n-1), q_n)$ otherwise.

A *run* of the PBA $\mathcal{B}$ is an infinite sequence $\rho \in Q^\omega$. A run ρ is *accepting* if $\rho[i] \in Q_f$ for infinitely many i. A run ρ is said to be *rejecting* if it is not accepting. The set of accepting runs and the set of rejecting runs are measurable [Var85]. Given a word α, the measure of the set of accepting runs is said to be the *probability of accepting* α and is henceforth denoted by $\mu^{acc}_{\mathcal{B},\alpha}$; and the measure of the set of rejecting runs is said to be the *probability* *of rejecting* α and is henceforth denoted by $\mu^{rej}_{\mathcal{B},\alpha}$. Clearly $\mu^{acc}_{\mathcal{B},\alpha} + \mu^{rej}_{\mathcal{B},\alpha} = 1$. Following, [BG05, BBG08], a PBA $\mathcal{B}$ on alphabet Σ defines two *semantics*:

- $\mathcal{L}_{>0}(\mathcal{B}) = \{\alpha \in \Sigma^\omega \mid \mu^{acc}_{\mathcal{B},\alpha} > 0\}$, henceforth referred to as the *probable semantics* of $\mathcal{B}$, and
- $\mathcal{L}_{=1}(\mathcal{B}) = \{\alpha \in \Sigma^\omega \mid \mu^{acc}_{\mathcal{B},\alpha} = 1\}$, henceforth referred to as the *almost-sure semantics* of $\mathcal{B}$.

This gives rise to the following classes of languages of infinite words.

Definition 2.4. Given a finite alphabet Σ,

- $\mathbb{L}(\mathrm{PBA}^{>0}) = \{\mathsf{L} \subseteq \Sigma^\omega \mid \exists \mathrm{PBA}\ \mathcal{B}.\ \mathsf{L} = \mathcal{L}_{>0}(\mathcal{B})\}$;
- $\mathbb{L}(\mathrm{PBA}^{=1}) = \{\mathsf{L} \subseteq \Sigma^\omega \mid \exists \mathrm{PBA}\ \mathcal{B}.\ \mathsf{L} = \mathcal{L}_{=1}(\mathcal{B})\}$.

Probabilistic Rabin automaton. Analogous to the definition of a PBA and RatPBA, one can define a Probabilistic Rabin automaton PRA and RatPRA [BBG08, Grö08]; where instead of using a set of final states, a set of pairs of subsets of states is used. A run in that case is said to be accepting if it satisfies the Rabin acceptance condition. It is shown in [BBG08, Grö08] that PRAs have the same expressive power under both probable and almost-sure semantics. Furthermore, it is shown in [BBG08, Grö08] that for any PBA $\mathcal{B}$, there is PRA $\mathcal{R}$ such that a word α is accepted by $\mathcal{R}$ with probability 1 iff α is accepted by $\mathcal{B}$ with probability > 0. All other words are accepted with probability 0 by $\mathcal{R}$.

Proposition 2.5 ([BBG08])**.** For any PBA $\mathcal{B}$ there is a PRA $\mathcal{R}$ such that $\mathcal{L}_{>0}(\mathcal{B}) = \mathcal{L}_{>0}(\mathcal{R}) = \mathcal{L}_{=1}(\mathcal{R})$ and $\mathcal{L}_{=0}(\mathcal{B}) = \mathcal{L}_{=0}(\mathcal{R})$. Furthermore, if $\mathcal{B}$ is a RatPBA then $\mathcal{R}$ is a RatPRA and the construction of $\mathcal{R}$ is recursive.

Finite probabilistic monitors (FPM**)s.** We identify one useful syntactic restriction of PBAs, called *finite probabilistic monitors* (FPM)s. In an FPM, all the states are accepting except a special absorbing *reject* state. We studied them extensively in [CSV08, CSV09a].

Definition 2.6. A PBA $\mathcal{M} = (Q, q_s, Q_f, \delta)$ on Σ is said to be an FPM if there is a state $q_r \in Q$ such that $q_r \neq q_s$, $Q_f = Q \setminus \{q_r\}$ and $\delta(q_r, a, q_r) = 1$ for each $a \in \Sigma$. The state q_r said to be the *reject* state of $\mathcal{M}$. If in addition $\mathcal{M}$ is a RatPBA, we say that $\mathcal{M}$ is a rational finite probabilistic monitor (RatFPM).

3. Probable semantics

In this section, we shall study the expressiveness of the languages contained in $\mathbb{L}(\mathrm{PBA}^{>0})$ as well as the complexity of deciding emptiness and universality of $\mathcal{L}_{>0}(\mathcal{B})$ for a given RatPBA $\mathcal{B}$. We assume that the alphabet Σ is fixed and contains at least two letters.

3.1. Expressiveness.

We shall establish new expressiveness results for the class $\mathbb{L}(\mathrm{PBA}^{>0})$–

- We show that although the class $\mathbb{L}(\mathrm{PBA}^{>0})$ strictly contains ω-regular languages [BG05], it is not topologically harder. More precisely, we will show that for any PBA $\mathcal{B}$, $\mathcal{L}_{>0}(\mathcal{B})$ is a $\mathsf{BCl}(\mathcal{G}_\delta)$-set. This will be a consequence of following facts.
  (a) $\mathbb{L}(\mathrm{PBA}^{>0}) = \mathsf{BCl}(\mathbb{L}(\mathrm{PBA}^{=1}))$ (see Theorem 3.1).
  (b) $\mathbb{L}(\mathrm{PBA}^{=1}) \subseteq \mathcal{G}_\delta$ (see Lemma 3.2).
- However, there are $\mathsf{BCl}(\mathcal{G}_\delta)$ sets that are not in $\mathbb{L}(\mathrm{PBA}^{>0})$ (see Lemma 3.3).

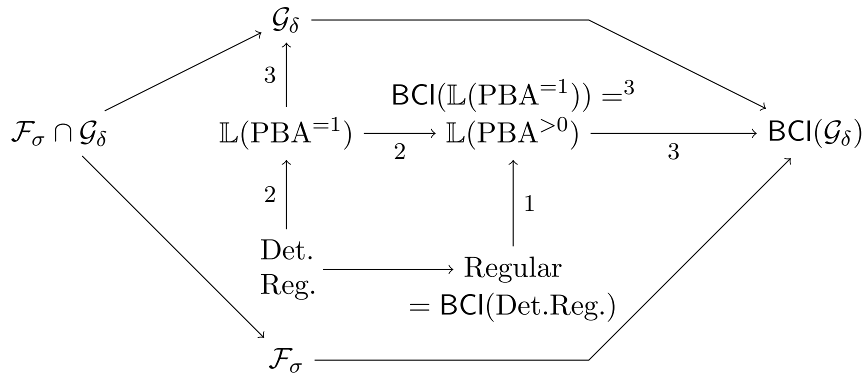

Figure 2: Relationship between languages recognized by PBAs and sets in the Borel hierarchy defined by the Cantor topology. Arrows indicate strict containment. 'Det. Reg.' refers to the class of languages recognized by deterministic Büchi automata, while 'Regular' refers to the class of ω-regular languages. Containment arrows with label 1 were proved in [BG05] and those labelled 2 were proved in [BBG08]. Results relating the classes Regular and Det. Reg. are classical results; see survey [PP04, Tho90]. Containment arrows with label 3 and the equality $\mathsf{BCl}(\mathbb{L}(\text{PBA}^{=1})) = \mathbb{L}(\text{PBA}^{>0})$ are proved in this paper.

Our expressiveness results are summarized in Figure 2.

We first show that just as the class of ω-regular languages is the Boolean closure of the class of ω-regular recognized by deterministic Büchi automata, the class $\mathbb{L}(\text{PBA}^{>0})$ coincides with the Boolean closure of the class $\mathbb{L}(\text{PBA}^{=1})$. This is the content of the following theorem whose proof is of independent interest and shall be used later in establishing that the containment of languages of two PBAs under almost-sure semantics is undecidable (see Theorem 4.5).

Theorem 3.1. $\mathbb{L}(\text{PBA}^{>0}) = \mathsf{BCl}(\mathbb{L}(\text{PBA}^{=1}))$.

Proof. First observe that it was already shown in [BBG08] that $\mathbb{L}(\text{PBA}^{=1}) \subseteq \mathbb{L}(\text{PBA}^{>0})$. Since $\mathbb{L}(\text{PBA}^{>0})$ is closed under Boolean operations, we get $\mathsf{BCl}(\mathbb{L}(\text{PBA}^{=1})) \subseteq \mathbb{L}(\text{PBA}^{>0})$. We have to show the reverse inclusion.

It suffices to show that given a PBA $\mathcal{B}$, the language $\mathcal{L}_{>0}(\mathcal{B}) \in \mathsf{BCl}(\mathbb{L}(\text{PBA}^{=1}))$. Fix $\mathcal{B}$. Recall that results of [BBG08, Grö08] (see Proposition 2.5) imply that there is a probabilistic Rabin automaton (PRA) $\mathcal{R}$ such that 1) $\mathcal{L}_{>0}(\mathcal{B}) = \mathcal{L}_{=1}(\mathcal{R}) = \mathcal{L}_{>0}(\mathcal{R})$ and 2) $\mathcal{L}_{=0}(\mathcal{B}) = \mathcal{L}_{=0}(\mathcal{R})$. Let $\mathcal{R} = (Q, q_s, F, \delta)$ where $F \subseteq 2^Q \times 2^Q$ is the set of the Rabin pairs. Assuming that F consists of n-pairs, let $F = \{(B_1, G_1), \ldots, (B_n, G_n)\}$.

Given an index set $\mathcal{I} \subseteq \{1, \ldots, n\}$, let $\mathsf{Good}_{\mathcal{I}} = \cup_{r \in \mathcal{I}} G_r$. Let $\mathcal{R}_{\mathcal{I}}$ be the PBA obtained from $\mathcal{R}$ by taking the set of final states to be $\mathsf{Good}_{\mathcal{I}}$. In other words, $\mathcal{R}_{\mathcal{I}} = (Q, q_s, \mathsf{Good}_{\mathcal{I}}, \delta)$. Given $\mathcal{I} \subseteq \{1, \ldots, n\}$ and an index $j \in \mathcal{I}$, let $\mathsf{Bad}_{\mathcal{I},j} = B_j \cup (\cup_{r \in \mathcal{I}, r \neq j} G_r)$. Let $\mathcal{R}_{\mathcal{I}}^j$ be the PBA obtained from $\mathcal{R}$ by taking the set of final states to be $\mathsf{Bad}_{\mathcal{I},j}$, *i.e.*, $\mathcal{R}_{\mathcal{I}}^j = (Q, q_s, \mathsf{Bad}_{\mathcal{I},j}, \delta)$. The result follows from the following claim.

Claim:

$$\mathcal{L}_{>0}(\mathcal{B}) = \bigcup_{\mathcal{I} \subseteq \{1, \ldots, n\}, j \in \mathcal{I}} \mathcal{L}_{=1}(\mathcal{R}_{\mathcal{I}}) \cap (\Sigma^\omega \setminus \mathcal{L}_{=1}(\mathcal{R}_{\mathcal{I}}^j)).$$

Proof of the claim: Given $\mathcal{I} \subseteq \{1, \ldots, n\}, j \in \mathcal{I}$, let $\mathsf{L}_{\mathcal{I},j} = \mathcal{L}_{=1}(\mathcal{R}_{\mathcal{I}}) \cap (\Sigma^\omega \setminus \mathcal{L}_{=1}(\mathcal{R}^j_{\mathcal{I}}))$. We will say that a run ρ of PRA $\mathcal{R}$ satisfies the Rabin pair (B_r, G_r) if all states in B_r occur only finitely many times in ρ and at least one state in G_r occurs infinitely often in ρ.

We first show that $\mathsf{L}_{\mathcal{I},j} \subseteq \mathcal{L}_{>0}(\mathcal{R})$. Fix any $\alpha \in \mathsf{L}_{\mathcal{I},j}$. Since $\mathsf{L}_{\mathcal{I},j} \subseteq \mathcal{L}_{=1}(\mathcal{R}_{\mathcal{I}})$, it follows that on input α the measure of runs that visit the set $\mathsf{Good}_{\mathcal{I}} = \cup_{i \in \mathcal{I}} G_i$ infinitely often must be 1. On the other hand, as $\mathsf{L}_{\mathcal{I},j} \cap \mathcal{L}_{=1}(\mathcal{R}^j_{\mathcal{I}}) = \emptyset$, it follows that on input α the measure of runs that visit $\mathsf{Bad}_{\mathcal{I},j} = B_j \cup (\cup_{i \in \mathcal{I}, i \neq j} G_i)$ only finitely many times has strictly positive measure. Since $\mathsf{Good}_{\mathcal{I}} \setminus \mathsf{Bad}_{\mathcal{I},j} \subseteq G_j$, it now follows from the previous two observations that the measure of runs that visit G_j infinitely often but visit $\mathsf{Bad}_{\mathcal{I},j}$ only finitely many times is strictly positive. Since $B_j \subseteq \mathsf{Bad}_{\mathcal{I},j}$, we get that the set of runs that satisfy the Rabin pair (B_j, G_j) has non-zero measure on input α. Therefore, we have that $\mathsf{L}_{\mathcal{I},j} \subseteq \mathcal{L}_{>0}(\mathcal{R})$. But, we have that $\mathcal{L}_{>0}(\mathcal{R}) = \mathcal{L}_{=1}(\mathcal{R}) = \mathcal{L}_{>0}(\mathcal{B})$. Hence, we get

$$\bigcup_{\mathcal{I} \subseteq \{1,\ldots,n\}, j \in \mathcal{I}} \mathsf{L}_{\mathcal{I},j} \subseteq \mathcal{L}_{>0}(\mathcal{B}).$$

We will be done if we can show the reverse inclusion. Thus, given word α in $\mathcal{L}_{>0}(\mathcal{B})$, we have to construct $\mathcal{I}$ and j such that $\alpha \in \mathsf{L}_{I,j}$. We construct them as follows. First, let $\widetilde{\mathcal{I}}$ be the set of all indices r such that the measure of all runs that satisfy the Rabin pair (B_r, G_r) on input α is > 0. $\widetilde{\mathcal{I}}$ is non-empty (since $\alpha \in \mathcal{L}_{=1}(\mathcal{R})$). Clearly, we have that on input α, the measure of runs such that $\mathsf{Good}_{\widetilde{\mathcal{I}}}$ is visited infinitely often is 1 (again, since $\alpha \in \mathcal{L}_{=1}(\mathcal{R})$). In other words, $\alpha \in \mathcal{L}_{=1}(\mathcal{R}_{\widetilde{\mathcal{I}}})$. Required $\mathcal{I}$ will be a subset of $\widetilde{\mathcal{I}}$ and will be constructed by induction as follows.

At step 1 of the induction, we pick an arbitrary index r in $\widetilde{\mathcal{I}}$. Then we check if it is the case that on α, the probability of visiting G_r infinitely often in $\mathcal{R}$ is 1. Note that it is the case that the probability that B_r is visited infinitely often in $\mathcal{R}$ is < 1 (as α satisfies (B_r, G_r) with non-zero probability). Note that this implies that $\alpha \in \mathcal{L}_{=1}(\mathcal{R}_{\{r\}}) \cap (\Sigma^\omega \setminus \mathcal{L}_{=1}(\mathcal{R}^r_{\{r\}})$ and the induction stops at this point. If it is not the case, then let $\mathcal{I}_1 = \{r\}$.

Proceed by induction. At step m, we would have produced an index set $\mathcal{I}_m \subseteq \widetilde{\mathcal{I}}$ such that on α, we have that $\alpha \notin \mathcal{L}_{=1}(\mathcal{R}_{\mathcal{I}_m})$ (meaning the set of runs which visit $\mathsf{Good}_{\mathcal{I}_m}$ infinitely often have probability < 1). Now since α is accepted by PRA $\mathcal{R}$ with probability 1, there must be some index r in $\widetilde{\mathcal{I}} \setminus \mathcal{I}_m$ such that the set of runs that satisfy (B_r, G_r) and visit Good_{I_m} only finitely many times is > 0. Fix one such r. Now, there are two cases.

(1) On the input α, the set of runs that visit $\mathsf{Good}_{\mathcal{I}_m} \cup G_r$ infinitely often has measure 1. In that case, by construction, we also have that $\alpha \in \mathsf{L}_{\mathcal{I}_m \cup \{r\}, r}$ and induction stops.
(2) Otherwise, we let $\mathcal{I}_{m+1} = \mathcal{I}_m \cup \{r\}$ and proceed.

The induction must stop at a finite point at which we will satisfy the required condition (since $\alpha \in \mathcal{L}_{=1}(\mathcal{R}_{\widetilde{\mathcal{I}}})$). □

The second component needed for showing that $\mathbb{L}(\mathrm{PBA}^{>0}) \subseteq \mathsf{BCl}(\mathcal{G}_\delta)$ is the fact that for any PBA $\mathcal{B}$ and $x \in [0,1]$, the language $\mathcal{L}_{\geq x}(\mathcal{B})$ is a $\mathcal{G}_\delta$-set; which we prove next.

Lemma 3.2. For any PBA $\mathcal{B}$ and $x \in [0,1]$, $\mathcal{L}_{\geq x}(\mathcal{B})$ is a $\mathcal{G}_\delta$ set.

Proof. Let $\mathcal{B} = (Q, q_s, Q_f, \delta)$. Now given $k > 0$, let $\mathsf{Paths}^k \subseteq \Sigma^\omega$ be the set of all infinite runs which start at the state q_s and visit the set of final states at least k-times. Let Paths^ω be the set of all infinite runs which start at the state q_s and visit the final states infinitely often. Formally, $\mathsf{Paths}^k = \{\rho \in Q^\omega \mid \rho[0] = q_s \text{ and } |\{i \in \mathbb{N} \mid \rho[i] \in Q_f\}| \geq k\}$ and

$\mathsf{Paths}^\omega = \{\rho \in Q^\omega \mid \rho[0] = q_s \text{ and } |\{i \in \mathbb{N} \mid \rho[i] \in Q_f\}| = \omega\}$. We have that $Path^k, k > 0$ forms a decreasing sequence and

$$\cap_{k\in\mathbb{N},k>0}\mathsf{Paths}^k = \mathsf{Paths}^\omega.$$

From standard probability theory, we get that for any word α,

$$\lim_{k\to\infty} \mu_{\mathcal{B},\alpha}(\mathsf{Paths}^k) = \mu_{\mathcal{B},\alpha}(\mathsf{Paths}^\omega)$$

where $\mu_{\mathcal{B},\alpha}$ is the probability measure generated by the infinite word α and PBA $\mathcal{B}$. From this, we immediately see that an infinite word α is accepted with probability at least x iff for all $k > 0$ the probability of visiting the set of final states on input α at least k-times $\geq x$. In other words,

$$\{\alpha \in \Sigma^\omega \mid \mu^{acc}_{\mathcal{B},\alpha} \geq x\} = \cap_{k\in\mathbb{N},k>0}\{\alpha \in \Sigma^\omega \mid \mu_{\mathcal{B},\alpha}(\mathsf{Paths}^k) \geq x\}.$$

Hence, it suffices to show that for each $k \in \mathbb{N}, k > 0$ the set $\{\alpha \in \Sigma^\omega \mid \mu_{\mathcal{B},\alpha}(\mathsf{Paths}^k) \geq x\}$ is a $\mathcal{G}_\delta$ set. Note that for each $k > 0$,

$$\{\alpha \in \Sigma^\omega \mid \mu_{\mathcal{B},\alpha}(\mathsf{Paths}^k) \geq x\} = \cap_{n\in\mathbb{N}}\{\alpha \in \Sigma^\omega \mid \mu_{\mathcal{B},\alpha}(\mathsf{Paths}^k) > x - \frac{1}{n}\}.$$

Hence, it suffices to show that for each $k \in \mathbb{N}, n \in \mathbb{N}, k > 0$ the set $\{\alpha \in \Sigma^\omega \mid \mu_{\mathcal{B},\alpha}(\mathsf{Paths}^k) > x - \frac{1}{n}\}$ is a $\mathcal{G}$-set. In order to see that this is the case, given $k > 0$ and $\ell > 0$, let $Path^{k,\ell} \subseteq \Sigma^\omega$ be the set of infinite runs that start at the initial state and visit Q_f at least k times in the first ℓ steps. Formally, $\mathsf{Paths}^{k,\ell} = \{\rho \in Q^\omega \mid \rho[0] = q_s \text{ and } |\{i \in \mathbb{N}, i < \ell \mid \rho[i] \in Q_f\}| \geq k\}$.

Now, the result follows immediately from the observation that

$$\{\alpha \in \Sigma^\omega \mid \mu_{\mathcal{B},\alpha}(\mathsf{Paths}^k) > x - \frac{1}{n}\} = \cup_{\ell\in\mathbb{N}}\{\alpha \in \Sigma^\omega \mid \mu_{\mathcal{B},\alpha}(\mathsf{Paths}^{k,\ell}) > x - \frac{1}{n}\}$$

and the observation that each of the set $\{\alpha \in \Sigma^\omega \mid \mu_{\mathcal{B},\alpha}(\mathsf{Paths}^{k,\ell}) > x - \frac{1}{n}\}$ is a $\mathcal{G}$-set. □

Using Lemma 3.2, one immediately gets that $\mathbb{L}(\text{PBA}^{>0}) \subseteq \mathsf{BCl}(\mathcal{G}_\delta)$. Even though PBAs accept non-ω-regular languages, they cannot accept all the languages in $\mathsf{BCl}(\mathcal{G}_\delta)$.

Lemma 3.3. $\mathsf{Regular} \subsetneq \mathbb{L}(\text{PBA}^{>0}) \subsetneq \mathsf{BCl}(\mathcal{G}_\delta)$.

Proof. Note that $\mathsf{Regular} \subsetneq \mathbb{L}(\text{PBA}^{>0})$ follows immediately from results of [BG05]. Thanks to Lemma 3.2, we also have that $\mathbb{L}(\text{PBA}^{=1}) \subseteq \mathcal{G}_\delta$. Since $\mathbb{L}(\text{PBA}^{>0}) = \mathsf{BCl}(\mathbb{L}(\text{PBA}^{=1}))$ (see Theorem 3.1), we get that $\mathbb{L}(\text{PBA}^{>0}) \subseteq \mathsf{BCl}(\mathcal{G}_\delta)$. We only have to show that this containment is strict. The proof of this fact utilizes the following result which shows that for any $\mathsf{L} \in \mathbb{L}(\text{PBA}^{>0})$, the smallest safety language containing L is guaranteed to be ω-regular even if L is not.[3]

Claim: For any PBA $\mathcal{B}$, let $\mathsf{cl}(\mathsf{L})$ be the smallest safety language containing $\mathsf{L} = \mathcal{L}_{>0}(\mathcal{B})$. Then $\mathsf{cl}(\mathsf{L})$ is ω-regular.

Proof of the claim: Without loss of generality, we can assume that $\mathsf{L} \neq \emptyset$. Let $\mathcal{B} = (Q, q_s, Q_f, \delta)$. Given $q \in Q$, let $\mathcal{B}_q$ be the PBA which is exactly like $\mathcal{B}$, except that the initial state is q. That is $\mathcal{B}_q = (Q, q, Q_f, \delta)$. Let $Q_{>0} \subseteq Q$ be the set of states $\{q \mid \exists\alpha.\mu^{acc}_{\mathcal{B}_q,\alpha} > 0\}$. Consider the finite state Büchi automata $\mathcal{A} = (Q_{>0}, q_s, Q_{>0}, \Delta)$ where $(q_1, a, q_2) \in \Delta$ iff

[3]As arbtitrary intersection of safety languages is also a safety language, for every language L, there is a smallest safety language containing L. Topologically, this is the *closure* of L.

	Emptiness	Universality	Containment
$\mathbb{L}(\text{PBA}^{>0})$	$\boldsymbol{\Sigma}_2^0$-complete$^{(\dagger)}$	$\boldsymbol{\Sigma}_2^0$-complete$^{(\dagger)}$	$\boldsymbol{\Sigma}_2^0$-complete$^{(\dagger)}$
$\mathbb{L}(\text{PBA}^{=1})$	**PSPACE**-complete$^{(\dagger\dagger)}$	**PSPACE**-complete	$\boldsymbol{\Sigma}_2^0$-complete

Figure 3: Hardness of decision problems for RatPBAs. $^{(\dagger)}$The problems of checking emptiness, universality and containment for probable semantics was shown to be **R.E.**-hard in [BBG08]. $^{(\dagger\dagger)}$the problem of checking emptiness of almost sure semantics was shown to decidable in **EXPTIME** in [BBG08].

$\delta(q_1, a, q_2) > 0$. It is easy to see that $\mathsf{cl}(\mathsf{L})$ is exactly the language recognized by $\mathcal{A}$. This implies that $\mathsf{cl}(\mathsf{L})$ is ω-regular.(**End of the claim**) □

We proceed as follows. Fix two letters a, b of the alphabet Σ and consider the language L consisting of exactly one word $\alpha = abaabb \ldots a^i b^i a^{i+1} b^{i+1} \ldots$. Now, $\mathsf{cl}(\mathsf{L}) = \mathsf{L}$ (every single element set in a metric space is a closed set) and L is not ω-regular (since L does not contain any periodic word). Therefore, the closed set L is not in the class $\mathbb{L}(\text{PBA}^{>0})$ (note that $\mathsf{L} \in \mathcal{G}_\delta$ as $\mathcal{F} \subseteq \mathcal{G}_\delta$). □

3.2. **Decision problems.** For the rest of this section, we shall focus our attention on decision problems for probable semantics for RatPBAs. Results of this section are summarized in the first row of Figure 3 and stated in Theorem 3.9. Given a RatPBA $\mathcal{B}$, the problems of emptiness and universality of $\mathcal{L}_{>0}(\mathcal{B})$ are known to be undecidable [BBG08]. We sharpen this result by showing that these problems are $\boldsymbol{\Sigma}_2^0$-complete. This is interesting in the light of the fact that problems on infinite string automata that are undecidable tend to typically lie in the analytical hierarchy, and not in the arithmetic hierarchy.

Before we proceed with the proof of the upper bound, let us recall an important property of finite-state Büchi automata [Tho90, PP04]. The language recognized by a finite-state Büchi automaton $\mathcal{A}$ is non-empty iff there is a final state q_f of $\mathcal{A}$, and finite words u and v such that q_f is reachable from the initial state on input u, and q_f is reachable from the state q_f on input v. This implies that any non-empty ω-regular language contains an ultimately periodic word. We had extended this observation to FPMs in [CSV08, CSV09a]. In particular, we had shown that the language $\mathcal{L}_{>x}(\mathcal{M})$ is non-empty for a given $\mathcal{M}$ iff there exists a set of final states C of $\mathcal{M}$ and words u and v such that the probability of reaching C from the initial state on input u is $> x$ and for each state $q \in C$ the probability of reaching C from q on input v is 1. This immediately implies that if $\mathcal{L}_{>x}(\mathcal{M})$ is non-empty then $\mathcal{L}_{>x}(\mathcal{M})$ must contain an ultimately periodic word. In contrast, this fact does not hold for non-empty languages in $\mathbb{L}(\text{PBA}^{>0})$. In fact, Baier and Größer [BG05], construct a PBA $\mathcal{B}$ such that $\mathcal{L}_{>0}(\mathcal{B})$ does not contain any ultimately periodic word.

However, we will show that even though the probable semantics of a PBA may not contain an ultimately periodic, they nevertheless are restrained in the sense that they must contain a *strongly asymptotic* word. In order to define strongly asymptotic words formally, we introduce the following notation–

Notation: Let $\mathcal{B} = (Q, q_s, Q_f, \delta)$. Given $C \subseteq Q$, $q \in C$ and a finite word $u = a_0 a_1 \ldots a_k \in \Sigma^+$, let $\delta_u^{Q_f}(q, C) = \sum_{q' \in C} \delta_u^{Q_f}(q, q')$ where

$$\delta_u^{Q_f}(q, q') = \sum_{(\{q\} \cup \{q'\} \cup \{q_i \mid 1 \leq i \leq k\}) \cap Q_f \neq \emptyset} \delta_{a_0}(q, q_1)\delta_{a_1}(q_1, q_2) \ldots \delta_{a_k}(q_k, q').$$

Informally, in the above notation $\delta_u^{Q_f}(q, C)$ is the probability that the PBA $\mathcal{B}$, when started in state q, on the input string u, is in some state in C at the end of u after passing through a final state. We can now define strongly asymptotic words.

Definition 3.4. Given a PBA $\mathcal{B} = (Q, q_s, Q_f, \delta)$ and a set C of states of $\mathcal{B}$, a word $\alpha \in \Sigma^\omega$ is said to be *strongly asymptotic with respect to $\mathcal{B}$ and C* if there is an infinite sequence $i_1 < i_2 < ....$ such that
(1) $\delta_{\alpha[0:i_1]}(q_s, C) > 0$ and
(2) all $j > 0$ and for all $q \in C$, $\delta^{Q_f}_{\alpha[i_j+1,i_{j+1}]}(q, C) > 1 - \frac{1}{2^j}$.
A word α is said to be *strongly asymptotic with respect to $\mathcal{B}$* if there is some C such that α is strongly asymptotic with respect to $\mathcal{B}$ and C.

We will now show that if the probable semantics of a PBA is non-empty then it must contain a strongly asymptotic word. We need one more notation.
Notation: $\mathit{Reach}(\mathcal{B}, C, x)$ denotes the predicate $\exists u \in \Sigma^+ . \delta_u(q_s, C) > x$.
Intuitively, the predicate $\mathit{Reach}(\mathcal{B}, C, x)$ is true iff there is some finite non-empty string u, such that the probability of being in C having started from the initial state q_s and after having read u is $> x$. The existence of strongly asymptotic word in probable semantics is an immediate consequence of the following Lemma.

Lemma 3.5. Let $\mathcal{B} = (Q, q_s, Q_f, \delta)$. For any $x \in [0, 1)$, $\mathcal{L}_{>x}(\mathcal{B}) \neq \emptyset$ iff $\exists C \subseteq Q$ such that $\mathit{Reach}(\mathcal{B}, C, x)$ is true and for all $j > 0$ there is a finite non-empty word u_j such that for all $q \in C$. $\delta^{Q_f}_{u_j}(q, C) > (1 - \frac{1}{2^j})$.

Proof. ($\Leftarrow$) Note that it is a well-known fact that the product $\prod_{j=1}^{\infty}(1 - \frac{1}{2^j})$ converges and is > 0. Assume now that $\exists C \subseteq Q$ such that $\mathit{Reach}(\mathcal{B}, C, x)$ is true and for all $j > 0$ there is a finite word u_j such that $\forall q \in C\ .\delta^{Q_f}_{u_j}(q, C) > (1 - \frac{1}{2^j})$. Since $\mathit{Reach}(\mathcal{B}, C, x)$ is true, there is a finite word u such that $\delta_u(q_s, C) > x$. Fix u. Also for each $j > 0$, fix u_j such that $\forall q \in C$, and $\delta^{Q_f}_{u_j}(q, C) > (1 - \frac{1}{2^j})$.

Let $z = \delta_u(q_s, C)$. Let $y = \frac{x}{z}$. We have that $y < 1$. Since $\prod_{j=1}^{\infty}(1 - \frac{1}{2^j}) > 0$ and $y < 1$, there is a $j_0 > 0$ such that $\prod_{j=j_0}^{\infty}(1 - \frac{1}{2^j}) > y$. Now it is easy to see that the word $\alpha = u u_{j_0} u_{j_0+1} \dots$ is accepted by $\mathcal{B}$ with probability $> zy$. But zy is x and the result follows.

($\Rightarrow$) Assume that $\mathcal{L}_{>x}(\mathcal{B}) \neq \emptyset$. Fix an infinite input string $\gamma \in \mathcal{L}_{>x}(\mathcal{B})$. Recall that the probability measure generated by γ and $\mathcal{B}$ is denoted by $\mu_{\mathcal{B},\gamma}$. For the rest of this proof we will just write μ for $\mu_{\mathcal{B},\gamma}$.

We will call a non-empty set of states C *good* if there is an $\epsilon > 0$, a measurable set $\mathsf{Paths} \subseteq Q^\omega$ of runs, and an infinite sequence of natural numbers $i_1 < i_2 < i_3 < \dots$ such that following conditions hold.

- $\mu(\mathsf{Paths}) \geq x + \epsilon$;
- For each $j > 0$ and each run ρ in Paths, we have that
  (a) $\rho[0] = q_s$, $\rho[i_j] \in C$ and
  (b) at least one state in the finite sequence $\rho[i_j, i_{j+1}]$ is a final state.

We say that a good set C is *minimal* if C is good but for each $q \in C$, the set $C \setminus \{q\}$ is not good. Clearly if there is a good set of states then there is also a minimal good set of states.
Claim:

- There is a good set of states C.

- Let C be a minimal good set of states. Fix ϵ, Paths and the sequence $i_1 < i_2 < \dots$ which witness the fact that C is a good set of states. For each $q \in C$ and each $j > 0$, let $\mathsf{Paths}_{j,q}$ be the subset of Paths such that each run in $\mathsf{Paths}_{j,q}$ passes through q at point i_j, *i.e.*, $\mathsf{Paths}_{j,q} = \{\rho \in \mathsf{Paths} \mid \rho[i_j] = q\}$. Then there exists a $p > 0$ such that $\mu(\mathsf{Paths}_{j,q}) \geq p$ for each $q \in C$ and each $j > 0$.

We first show how to obtain the Lemma using the above claim. Fix a minimal set of good states C. Fix ϵ, Paths and the sequence $i_1 < i_2 < \dots$ which witness the fact that C is a good set of states. We claim that C is the required set of states. As $\mu(\mathsf{Paths}) \geq x + \epsilon$ and for each $\rho \in \mathsf{Paths}$, $\rho[i_1] \in C$, it follows immediately that $Reach(\mathcal{B}, C, x)$. Assume now, by way of contradiction, that there exists a $j_0 > 0$ such that for each finite word u, there exists a $q \in C$ such that $\delta_u^{Q_f}(q, C) \leq 1 - \frac{1}{2^{j_0}}$. Fix j_0. Also fix $p > 0$ such that $\mu(\mathsf{Paths}_{j,q}) \geq p$ for each j and $q \in C$, where $\mathsf{Paths}_{j,q}$ is the subset of Paths such that each run in $\mathsf{Paths}_{j,q}$ passes through q at point i_j; the existence of p is guaranteed by the above claim.

We first construct a sequence of sets $L_i \subseteq Q^+$ as follows. Let $L_1 \subseteq Q^+$ be the set of finite words on states of Q of length $i_1 + 1$ such that each word in L_1 starts with the state q_s and ends in a state in C. Formally $L_1 = \{\eta \in Q^+ \mid |\eta| = i_1 + 1, \eta[0] = q_s \text{ and } \eta[i_1] \in C\}$. Assume that L_r has been constructed. Let $L_{r+1} \subseteq Q^+$ be the set of finite words on states of Q of length $i_{r+1} + 1$ such that each word in L_{r+1} has a prefix in L_r, passes through a final state in between i_r and i_{r+1}, and ends in a state in C. Formally, $L_{r+1} = \{\eta \in Q^+ \mid |\eta| = i_{r+1} + 1, \eta[0 : i_r] \in L_r, \exists i.(i_r < i < i_{r+1} \ \wedge \ \eta[i] \in Q_f)\}$.

Note that $(L_r\Sigma^\omega)_{r\geq 1}$ is a decreasing sequence of measurable subsets and $\mathsf{Paths} \subseteq \cap_{r>1} L_r\Sigma^\omega$. Now, it is easy to see from the choice of j_0 and p that $\mu(L_{r+1}\Sigma^\omega) \leq \mu(L_r\Sigma^\omega) - \frac{p}{2^{j_0}}$. This, however, implies that there is a r_0 such that $\mu(L_{r_0}\Sigma^\omega) < 0$. A contradiction. Thus, it suffices to show that the claim is correct.

Proof of the claim:

(1) For each $k > 0$, let $C_k = \mathsf{post}(q_s, \gamma[0 : k])$. Since the set of states Q is finite, there must be some C such that $C_k = C$ for infinitely many k's. Fix one such C. We claim that C is a good set of states. We need to show that C satisfies the definition of good set of states. So we need to construct ϵ, Paths and the infinite sequence $i_1 < i_2 < \dots$ as in the definition of good set of states. We will pick $\epsilon > 0$ such that $\mu^{acc}_{\mathcal{B},\gamma} = x + 2\epsilon$. We construct Paths and the sequence $i_1 < i_2 < \dots$ as follows.

First let Paths_0 be the set of all runs starting in q_s and visiting the final states infinitely often. Paths_0 is measurable and $\mu(\mathsf{Paths}_0) = x + 2\epsilon$. Take $i_1 > 0$ to be the smallest integer such that $\mathsf{post}(q_s, \gamma[0 : i_1]) = C$. Inductively, assume that we have constructed a sequence of integers $i_1 < i_2 < \cdots < i_{j+1}$, and a measurable set $\mathsf{Paths}_j \subseteq \mathsf{Paths}_0$ such that

(a) $\mu(\mathsf{Paths}_j) > x + \epsilon + \frac{\epsilon}{2^j}$,
(b) for each $\rho \in \mathsf{Paths}_j$ and $k \leq j + 1$, $\rho[i_k] \in C$, and
(c) for each $\rho \in \mathsf{Paths}_j$ and $k < j + 1$, there is some i between i_k and i_{k+1} such that $\rho[i] \in Q_f$.

Observe that Paths_0 and i_1 satisfy that above conditions as condition (c) holds vaccuously. Now for each $\ell > 0$, $\mathsf{Paths}_j^\ell \subseteq \mathsf{Paths}_j$ be the set of runs that visit a final state at least one time between i_{j+1} and $i_{j+1} + \ell$. Formally, $\mathsf{Paths}_j^\ell = \{\rho \in \mathsf{Paths}_j \mid \exists i.(i_{j+1} < i < i_{j+1} + \ell \wedge \rho[i] \in Q_f)\}$. Clearly Paths_j^ℓ is an increasing sequence of measurable sets and $\cup_{\ell \in \mathbb{N}} \mathsf{Paths}_j^\ell = \mathsf{Paths}_j$ (each run in Paths_j visits the set of final states infinitely often).

Since $\mu(\mathsf{Paths}_j) > x+\epsilon+\frac{\epsilon}{2^j}$, there must exist a ℓ_0 such that $\mu(\mathsf{Paths}_j^{\ell_0}) > x+\epsilon+\frac{1}{2}(\frac{\epsilon}{2^j})$. Fix ℓ_0 and let $i_{j+2} > i_{j+1}+\ell_0$ be the smallest integer such that $\mathsf{post}(q_s,\gamma[0:i_{j+2}]) = C$. Let $\mathsf{Paths}_{j+1} = \mathsf{Paths}_j^{\ell_0}$. It is easy to see that Paths_{j+1} is measurable and that it satisfies the conditions (a), (b) and (c), assumed inductively about Paths_j.

Observe that the above inductive construction ensures that $\mathsf{Paths}_{j+1} \subseteq \mathsf{Paths}_j$. Take $\mathsf{Paths} = \cap_{j\in\mathbb{N}}\mathsf{Paths}_j$. It is easy to see that Paths and the sequence $i_1 < i_2 < \cdots$ constructed inductively, satisfy the claim.

(2) We have that C is minimal good set of states. Note that as C is finite, we only need to show that for each $q \in Q$, $\inf_{j>0}\mu(\mathsf{Paths}_{j,q}) > 0$. We proceed by contradiction. Assume that there is some q such that $\inf_{j>0}\mu(\mathsf{Paths}_{j,q}) = 0$. Fix one such q. We will obtain a contradiction to minimality if we can show that $C\setminus\{q\}$ is also a good set of states.

In order to show that $C\setminus\{q\}$ is a good set of states, we have to satisfy the definition of a good set of states.

Now, since $\inf_{j>0}\mu(\mathsf{Paths}_{j,q}) = 0$, there is some j_1 such that $\mu(\mathsf{Paths}_{j_1,q}) < \frac{\epsilon}{4}$. Let $\mathsf{Paths}^1 = \mathsf{Paths}\setminus\mathsf{Paths}_{j_1,q}$. We have that $\mathsf{Paths}^1 \subseteq \mathsf{Paths}$, $\mu(\mathsf{Paths}^1) \geq x+\frac{\epsilon}{2}+\frac{\epsilon}{4}$ and for each $\rho \in \mathsf{Paths}^1$, $\rho[i_{j_1}] \in C\setminus\{q\}$.

Now, again as $\inf_{j>0}\mu(\mathsf{Paths}_{j,q}) = 0$, there is some $j_2 > j_1$ such that $\mu(\mathsf{Paths}_{j_2,q}) < \frac{\epsilon}{8}$. Let $\mathsf{Paths}^2 = \mathsf{Paths}^1\setminus\mathsf{Paths}_{j_2,q}$. We have that $\mathsf{Paths}^2 \subseteq \mathsf{Paths}^1$, $\mu(\mathsf{Paths}^2) \geq x+\frac{\epsilon}{2}+\frac{\epsilon}{8}$ and for each $\rho \in \mathsf{Paths}^2$, $\rho[i_{j_2}] \in C\setminus\{q\}$. Note also that as $j_2 > j_1$ and $\mathsf{Paths}^1 \subseteq \mathsf{Paths}$, we have that for each each $\rho \in \mathsf{Paths}^2$ there is some i such that $i_{j_1} < i < i_{j_2}$ and $\rho[i] \in Q_f$.

We can continue and obtain a sequence $\mathsf{Paths}^{j_1} \supseteq \mathsf{Paths}^{j_2} \supseteq \ldots$ of measurable sets, and sequence $i_{j_1} < i_{j_2} < \ldots$ such that for each $l > 0$, $\mu(\mathsf{Paths}^{j_l}) \geq x+\frac{\epsilon}{2}+\frac{\epsilon}{2^l}$ and for each $\rho \in \mathsf{Paths}^{j_l}$, $\rho[i_{j_l}] \in C\setminus\{q\}$. Furthermore for each $l > 1$ and each $\rho \in \mathsf{Paths}^l$ there is some i such that $i_{j_{l-1}} < i < i_{j_l}$ and $\rho[i] \in Q_f$.

Let $\mathsf{Paths}' = \cap_{l>0}\mathsf{Paths}^l$. We have that $\mu(\mathsf{Paths}') \geq x+\frac{\epsilon}{2}$. Clearly $\frac{\epsilon}{2}$, Paths' and the sequence $i_{j_1} < i_{j_2} < \ldots$ witness the fact that $C\setminus\{q\}$ is a good set of states. ☐

We get immediately that the probable semantics of a PBA, if non-empty, must contain a strongly asymptotic word.

Corollary 3.6. Given a PBA $\mathcal{B}$, $\mathcal{L}_{>0}(\mathcal{B}) \neq \emptyset$ iff $\mathcal{L}_{>0}(\mathcal{B})$ contains a strongly asymptotic word.

Lemma 3.5 also implies that emptiness-checking of $\mathcal{L}_{>0}(\mathcal{B})$ for a given a RatPBA $\mathcal{B}$ is in $\mathbf{\Sigma}_2^0$.

Corollary 3.7. Given a RatPBA, $\mathcal{B}$, the problem of deciding whether $\mathcal{L}_{>0}(\mathcal{B}) = \emptyset$ is in $\mathbf{\Sigma}_2^0$.

Proof. Let us fix a PBA $\mathcal{B} = (Q, q_s, Q_f, \delta)$, and $x \in [0,1)$. Now Lemma 3.5 says that the non-emptiness of $\mathcal{L}_{>x}(\mathcal{B})$ is equivalent to the following property

$$\begin{aligned}\varphi = \exists C\subseteq Q.\ \exists u\in\Sigma^*.\quad &((\delta_u(q_s,C) > x)\wedge\\ &(\forall j.\ \exists u_j\in\Sigma^*.\ (\forall q\in C.\ \delta_{u_j}^{Q_f}(q,C) > (1-\tfrac{1}{2^j})))\end{aligned}$$

which can be rewritten as (by moving quantifiers out)

$$\begin{aligned}\varphi = \exists C\subseteq Q.\ \forall j.\ \exists u\in\Sigma^*.\ \exists u_j\in\Sigma^*.\quad &((\delta_u(q_s,C) > x)\wedge\\ &(\forall q\in C.\ \delta_{u_j}^{Q_f}(q,C) > (1-\tfrac{1}{2^j})))\end{aligned}$$

Now consider the property ψ given as

$$\psi = \forall j.\ \exists C_j \subseteq Q.\ \exists u \in \Sigma^*.\ \exists u_j \in \Sigma^*.\quad ((\delta_u(q_s, C_j) > x) \wedge (\forall q \in C_j.\ \delta_{u_j}^{Q_f}(q, C_j) > (1 - \tfrac{1}{2^j})))$$

Clearly, ψ logically follows from φ. However, in our specific case, it turns out that in fact, ψ is equivalent to φ due to the following observations. First note, that since there are only finitely many subsets of Q, there must be a $C \subseteq Q$ such that $C = C_j$ for infinitely many j (if ψ holds). Further observe that if $\exists u_j.(\forall q \in C.\ \delta_{u_j}^{Q_f}(q, C) > (1 - \frac{1}{2^j}))$ for some j then $\exists u_i.(\forall q \in C.\ \delta_{u_i}^{Q_f}(q, C) > (1 - \frac{1}{2^i}))$ holds for all $i \leq j$. From these it follows that φ logically follows from ψ.

Observe that $(\delta_u(q_s, C) > x)$ and $(\forall q \in C.\ \delta_{u_j}^{Q_f}(q, C) > (1 - \frac{1}{2^j}))$ are recursive predicates. Thus, ψ demonstrates that the non-emptiness problem is in $\mathbf{\Pi}_2^0$, which means that emptiness is in $\mathbf{\Sigma}_2^0$. □

We will now show that the emptiness problem is also $\mathbf{\Sigma}_2^0$-hard.

Lemma 3.8. Given a RatPBA, $\mathcal{B}$, the problem of deciding whether $\mathcal{L}_{>0}(\mathcal{B}) = \emptyset$ is $\mathbf{\Sigma}_2^0$-hard.

Proof. The hardness result will be obtained by significantly modifying the proof in [BBG08], where the emptiness problem was shown to be **R.E.**-hard.

Consider a deterministic two counter machine M with two counters and a one way, read only input tape. We can capture the computation of M, as a sequence of configurations where each configuration is a 4-tuple (q, x, a^i, b^j, m) where q is the state of the finite state control that M changed to, x is the input symbol that is read and i, j are the new counter values and m indicates whether the input head stayed in the same place, or moved right and read a new input symbol. Here $m \in \{\mathit{same}, \mathit{right}\}$. Note that the two counter values are represented in unary having a string of as and bs, respectively. Thus, a computation of M can be described by a string over alphabet Σ' that includes the states of M, the input symbols of M, the symbols $a, b, \mathit{same}, \mathit{right}, (,)$, and ','. In this proof we will restrict our attention to machines M that read all the input symbols; thus, the number of steps in a computation is at least the length of the input. A halting computation is a sequence of configurations ending in a halting state. Define $L(M)$ to be the set of input strings on which M halts. Let $\langle M \rangle$ be a binary encoding of M. Consider $\mathcal{H} = \{\langle M \rangle \mid L(M) \neq \emptyset\}$ and $\mathcal{D} = \{\langle M \rangle \mid L(M) \text{ is finite}\}$. Recall that $\mathcal{H}$ is **R.E.**-complete and $\mathcal{D}$ is $\mathbf{\Sigma}_2^0$-complete.

Our proof of hardness will be as follows: Given a deterministic two counter machine M, we will construct three RatPBAs $\mathcal{P}_1, \mathcal{P}_2$, and $\mathcal{P}_3$ such that $\langle M \rangle \in \mathcal{D}$ iff $\mathcal{L}_{>0}(\mathcal{P}_1) \cap \mathcal{L}_{>0}(\mathcal{P}_2) \cap \mathcal{L}_{>0}(\mathcal{P}_3) = \emptyset$. Since $\mathbb{L}(\mathrm{PBA}^{>0})$ is closed under intersection and the intersection of automata can be effectively constructed [BBG08, Grö08], this will demonstrate a reduction from $\mathcal{D}$ to the emptiness problem and therefore prove the hardness result. Our construction of the RatPBAs $\mathcal{P}_1, \mathcal{P}_2$, and $\mathcal{P}_3$ relies on ideas in [CL89] and [BBG08, Grö08]. Hence, we begin by recalling the key ideas from these papers that we will exploit.

For the rest of this proof, let us fix a deterministic two-counter machine M whose computations can be encoded as strings over Σ'. Consider any rational ϵ such that $0 < \epsilon < \frac{1}{2}$. [CL89] give the construction of a PFA $\mathcal{R}$ (that depends on M and ϵ) over alphabet $\Sigma_{\mathcal{R}} = \Sigma' \cup \{@\}$, where $@ \notin \Sigma'$. We can show that this PFA $\mathcal{R}$ satisfies the following properties.

(1) There exists an (computable) integer constant $d \geq 2$ such that if w is a valid and halting computation of M of length n, then the input string $(w@)^{d^n}$ is accepted by $\mathcal{R}$

with probability $\geq (1-\epsilon)$; that is, the string obtained by concatenating w, d^n number of times, where successive concatenations are separated by @, is accepted with probability at least $1-\epsilon$.

(2) Consider any input $u = w_1@w_2@\cdots@w_m@$, where no w_i is a valid halting computation of M. $\mathcal{R}$ accepts u with probability at most ϵ.

The proof that $\mathcal{R}$ satisfies these properties is defered to the Appendix. Let $\mu^{acc}_{\mathcal{R},w}$ denote the probability with which the input w is accepted by $\mathcal{R}$. Observe that the above construction has the following property: if $L(M) = \emptyset$ then any input string is accepted by $\mathcal{R}$ with probability at most ϵ; on the other hand, if $L(M) \neq \emptyset$ then there is some input that is accepted with probability at least $1-\epsilon$.

Using the above construction of $\mathcal{R}$, [BBG08, Grö08] reduce $\mathcal{H}$ to the emptiness problem, thus demonstrating its **R.E.**-hardness. The main ideas behind this are as follows. Let $\Sigma_{\mathcal{P}} = \Sigma_{\mathcal{R}} \cup \{\sharp, \$\}$, where $\sharp$ and \$ are symbols not in $\Sigma_{\mathcal{R}}$. [BBG08, Grö08] construct two RatPBAs $\mathcal{P}_1$ and $\mathcal{P}_2$ over alphabet $\Sigma_{\mathcal{P}}$ such that $\mathcal{L}_{>0}(\mathcal{P}_1)$ is

$$\{w^1_1\sharp w^1_2\sharp\cdots w^1_{k_1}\$\$w^2_1\sharp w^2_2\cdots w^2_{k_2}\$\$\cdots \mid w^j_i \in \Sigma^*_{\mathcal{R}} \text{ and } \prod_{j\geq 1}(1-(\prod_{i=1}^{k_j-1}(1-\mu^{acc}_{\mathcal{R},w^j_i}))) > 0\}.$$

and $\mathcal{L}_{>0}(\mathcal{P}_2)$ is

$$\{v_1\$\$v_2\$\$\cdots\ :\ v_i \in (\Sigma \cup \{\sharp\})^* \text{ and } \prod_{i\geq 1}(1-(1-\epsilon)^{g(v_i)}) = 0\}$$

where $g(v_i)$ is the number of $\sharp$ symbols in v_i. Let $L_1 = \mathcal{L}_{>0}(\mathcal{P}_1)$ and $L_2 = \mathcal{L}_{>0}(\mathcal{P}_2)$. The following two observations are shown in [BBG08, Grö08].

(1) Consider any input $w = w^1_1\sharp w^1_2\sharp\cdots w^1_{k_1}\$\$w^2_1\sharp w^2_2\cdots w^2_{k_2}\$\$\cdots$, where $w^j_i \in \Sigma^*_{\mathcal{R}}$ and $\mu^{acc}_{\mathcal{R},w^j_i} \leq \epsilon$. If $w \in L_2$ then $w \notin L_1$.

(2) Suppose $w_1, w_1, \ldots$ are (not necessarily distinct) words over $\Sigma_{\mathcal{R}}$ such that $\mu^{acc}_{\mathcal{R},w_i} \geq 1-\epsilon$. For any $\epsilon < \frac{1}{2}$, there are $k_1, k_2, k_3, \ldots$ such that

$$(w_1\sharp)^{k_1-1}w_1\$\$(w_2\sharp)^{k_2-1}w_2\$\$\cdots$$

belongs to $L_1 \cap L_2$.

Observe that the above two observations allow [BBG08, Grö08] to conclude that $L_1 \cap L_2 \neq \emptyset$ iff there is some w such that $\mu^{acc}_{\mathcal{R},w} \geq 1-\epsilon$. Thus, using properties of $\mathcal{R}$, one can see that $\mathcal{H}$ can be reduced to the emptiness problem, therefore demonstrating its undecidability.

In order to prove the tighter lower bound of $\mathbf{\Sigma}^0_2$, we would like to extend the above ideas to obtain a reduction from $\mathcal{D}$, instead of $\mathcal{H}$. We first outline the intuitions behind the extension. Suppose $u_1, u_2, \ldots$ are (not necessarily distinct) halting computations of M. Consider the input word

$$w(k_1, k_2, \ldots) = ((u_1@)^{\ell_1}\sharp)^{k_1-1}(u_1@)^{\ell_1}\$\$((u_2@)^{\ell_2}\sharp)^{k_2-1}(u_2@)^{\ell_2}\$\$\cdots$$

where $\ell_i = d^{|u_i|}$. From the preceding paragraphs it can be seen that there is a choice of $k_1, k_2, \ldots$ such that $w(k_1, k_2, \ldots) \in L_1 \cap L_2$. To obtain a reduction from $\mathcal{D}$, we need to "check" that infinitely many among the computations $u_1, u_2, \ldots$ correspond to distinct inputs. To do this we will construct a third RatPBA $\mathcal{P}_3$ that will check that the computations u_i grow unboundedly. Since M is deterministic, passing the test imposed by $\mathcal{P}_3$ ensures that $L(M)$ is infinite, and conversely, if $L(M)$ is infinite then our assumption that M reads all the input symbols ensures that there will be some string that passes the $\mathcal{P}_3$ test.

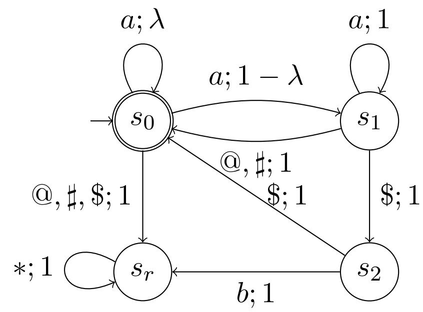

Figure 4: Automata $\mathcal{P}_3$. Here $*$ indicates any input symbol, b any input symbol that is not \$, and a any input symbol that is not $@, \sharp$ or \$.

We now outline the formal details. The RatPBA $\mathcal{P}_3$ has $\Sigma_{\mathcal{P}}$ as input alphabet and is shown in Figure 4. It has four states s_0, s_1, s_2, s_r where s_0 is the initial state and the only final state. s_r is an absorbing state. The transition probabilities depend on a parameter λ that we will fix later. In state s_0, on inputs other than $@, \sharp, \$$, the machine transitions to s_0 with probability λ and s_1 with probability $1-\lambda$; on $@, \sharp, \$$ it goes to s_r with probability 1. In state s_1 it behaves as follows. On input $@, \#$ it goes to state s_0 with probability 1; on input \$ it goes to s_2; on all other inputs, it remains in s_1 with probability 1. In state s_2 it behaves as follows. On input \$ it goes to state s_0 with probability 1; on all other inputs, it goes to s_r with probability 1. We will pick λ to be such that $\lambda \cdot d \ll \frac{1}{2}$; recall that d is the constant associated with $\mathcal{R}$. Let $\mathsf{SeqComp} = \{u_0 x_0 u_1 x_1 \cdots \mid x_i \in \{@, \#, \$\$\}$ and $@, \#, \$$ do not appear in the strings $u_i\}$. It can be easily shown that $\mathcal{L}_{>0}(\mathcal{P}_3) \subseteq \mathsf{SeqComp}$. In addition, consider $\alpha = u_0 x_0 u_1 x_1 \cdots \in \mathsf{SeqComp}$ with $x_i \in \{@, \#, \$\$\}$ and $@, \#, \$$ not in u_i. If there is an ℓ such that for infinitely many i, $|u_i| = \ell$ then $\alpha \notin \mathcal{L}_{>0}(\mathcal{P}_3)$. We conclude the proof by showing the following claim.

Claim: $L(M)$ is a finite set iff $\mathcal{L}_{>0}(\mathcal{P}_1) \cap \mathcal{L}_{>0}(\mathcal{P}_2) \cap \mathcal{L}_{>0}(\mathcal{P}_3) = \emptyset$.

Proof of the claim: Let $\mathcal{C}(M)$ be the set of valid halting computations of M. Since we assume that M is deterministic and reads all the input symbols, we can conclude that $L(M)$ is finite iff $\mathcal{C}(M)$ is finite. Suppose $L(M)$ is finite and $\alpha \in \mathcal{L}_{>0}(\mathcal{P}_2) \cap \mathcal{L}_{>0}(\mathcal{P}_3)$. Since α is accepted by $\mathcal{P}_3$, we know that the computations in α grown unboundedly. However, since $\mathcal{C}(M)$ is finite, we can conclude that there is a suffix β of α such that none of computations of M in β are valid and halting. Thus, $\beta = w_1^1 \sharp w_2^1 \sharp \cdots w_{k_1}^1 \$\$ w_1^2 \sharp w_2^2 \cdots w_{k_2}^2 \$\$ \cdots$ such that $\mu^{acc}_{\mathcal{R}, w_i^j} \leq \epsilon$. Coupled with the fact that $\alpha \in \mathcal{L}_{>0}(\mathcal{P}_2)$, we can argue that $\alpha \notin \mathcal{L}_{>0}(\mathcal{P}_1)$ using a similar reasoning as in [BBG08].

Suppose $L(M)$ is an infinite set. Hence $\mathcal{C}(M)$ is also an infinite set. Let $u_1, u_2, \ldots$ be some *distinct* computations in $\mathcal{C}(M)$; we will describe how to choose u_i later. As before, consider

$$w(k_1, k_2, \ldots) = ((u_1 @)^{\ell_1} \sharp)^{k_1 - 1} (u_1 @)^{\ell_1} \$\$ ((u_2 @)^{\ell_2} \sharp)^{k_2 - 1} (u_2 @)^{\ell_2} \$\$ \cdots$$

where $\ell_i = d^{|u_i|}$. As mentioned before, there are $k_1, k_2, \ldots$ such that $w(k_1, k_2, \ldots)$ is accepted by both $\mathcal{P}_1$ and $\mathcal{P}_2$. In addition, the probability that $\mathcal{P}_3$ accepts $w(k_1, k_2, \ldots)$ is $\prod_{i>0} p_i$ where $p_i = (1 - \lambda^{|u_i|})^{\ell_i \cdot k_i}$. We will choose u_i (or rather its length) to be a computation so that p_i is $> (1 - \frac{1}{2^i})$; this will ensure that $\prod_{i>0} p_i$ is non-zero. Assuming λ to be very small and substituting for $\ell_i = d^{|u_i|}$, it is easily seen that $p_i > (1-(d\cdot\lambda)^{|u_i|})^{k_i}$. Here d, k_i are fixed and λ is a small constant such that $d\cdot\lambda \ll \frac{1}{2}$. Now, it should be easy to see that we can chose a sufficiently long halting computation u_i so that $(1-(d\cdot\lambda)^{|u_i|})^{k_i} > (1-\frac{1}{2^i})$. □

Since the class $\mathbb{L}(\mathrm{PBA}^{>0})$ is closed under complementation and the complementation procedure is recursive [BBG08] for RatPBAs, we can conclude that checking universality of $\mathcal{L}_{>0}(\mathcal{B})$ is also $\mathbf{\Sigma}^0_2$-complete. The same bounds also apply to checking language containment under probable semantics. Note that these problems were already shown to undecidable in [BBG08], but the exact complexity was not computed therein.

Theorem 3.9. Given a RatPBA, $\mathcal{B}$, the problems 1) deciding whether $\mathcal{L}_{>0}(\mathcal{B}) = \emptyset$ and 2) deciding whether $\mathcal{L}_{>0}(\mathcal{B}) = \Sigma^\omega$, are $\mathbf{\Sigma}^0_2$-complete. Given another RatPBA, $\mathcal{B}'$, the problem of deciding whether $\mathcal{L}_{>0}(\mathcal{B}) \subseteq \mathcal{L}_{>0}(\mathcal{B}')$ is also $\mathbf{\Sigma}^0_2$-complete.

Proof. Since $\mathbb{L}(\mathrm{PBA}^{>0})$ is closed under complementation and the complementation is recursive [BBG08, Grö08], Lemma 3.8 immediately implies that the problems of universality, emptiness and set containment are $\mathbf{\Sigma}^0_2$-hard. Also observe that given $\mathcal{B}_1$ and $\mathcal{B}_2$, we have that $\mathcal{L}_{>0}(\mathcal{B}_1) \subseteq \mathcal{L}_{>0}(\mathcal{B}_2)$ iff $\mathcal{L}_{>0}(\mathcal{B}_1) \cap (\Sigma^\omega \setminus \mathcal{L}_{>0}(\mathcal{B}_2)) = \emptyset$. Now, results of [BBG08, Grö08] show that there is a constructible $\mathcal{B}_3$ such that $\mathcal{L}_{>0}(\mathcal{B}_3) = \mathcal{L}_{>0}(\mathcal{B}_1) \cap (\Sigma^\omega \setminus \mathcal{L}_{>0}(\mathcal{B}_2))$. Now, thanks to Corollary 3.7, the problems of universality, emptiness and set containment are in $\mathbf{\Sigma}^0_2$. □

Remarks 3.10. Lemma 3.5 can be used to show that emptiness-checking of $\mathcal{L}_{>\frac{1}{2}}(\mathcal{B})$ for a given RatPBA $\mathcal{B}$ is in $\mathbf{\Sigma}^0_2$. In contrast, we had shown in [CSV09a] that the problem of deciding whether $\mathcal{L}_{>\frac{1}{2}}(\mathcal{M}) = \Sigma^\omega$ for a given FPM $\mathcal{M}$ lies beyond the arithmetical hierarchy.

4. Almost-sure semantics

The class $\mathbb{L}(\mathrm{PBA}^{=1})$ was first studied in [BBG08], although it was not characterized topologically. In this section, we study the expressiveness and complexity of the class $\mathbb{L}(\mathrm{PBA}^{=1})$. We will also demonstrate that the class $\mathbb{L}(\mathrm{PBA}^{=1})$ is closed under finite unions and intersections. As in the case of probable semantics, we assume that the alphabet Σ is fixed and contains at least two letters.

4.1. Expressiveness.

In this section, we shall establish new expressiveness results for the class $\mathbb{L}(\mathrm{PBA}^{=1})$–

- $\mathbb{L}(\mathrm{PBA}^{=1}) \subsetneq \mathcal{G}_\delta$. This is an immediate consequence of Theorem 3.1, Lemma 3.2 and Lemma 3.3.
- $\mathsf{Regular} \cap \mathbb{L}(\mathrm{PBA}^{=1}) = \mathsf{Regular} \cap \mathrm{Deterministic}$ (see Proposition 4.1).
- The class $\mathbb{L}(\mathrm{PBA}^{=1})$ is closed under union and intersection. (see Corollary 4.3).

We start by characterizing the intersection $\mathsf{Regular} \cap \mathbb{L}(\mathrm{PBA}^{=1})$. Note that the fact every language $\mathbb{L}(\mathrm{PBA}^{=1})$ is contained in $\mathcal{G}_\delta$ implies immediately that there are ω-regular languages not in $\mathbb{L}(\mathrm{PBA}^{=1})$. That there are ω-regular languages not in $\mathbb{L}(\mathrm{PBA}^{=1})$ was also proved in [BBG08], although the proof therein is by explicit construction of an ω-regular language which is then shown to be not in $\mathbb{L}(\mathrm{PBA}^{=1})$. Our topological characterization of the class $\mathbb{L}(\mathrm{PBA}^{=1})$ has the advantage that we can characterize the intersection $\mathsf{Regular} \cap \mathbb{L}(\mathrm{PBA}^{=1})$

exactly: $\mathsf{Regular} \cap \mathbb{L}(\mathrm{PBA}^{=1})$ is the class of ω-regular languages that can be recognized by a finite-state deterministic Büchi automaton.

Proposition 4.1. For any PBA $\mathcal{B}$, $\mathcal{L}_{=1}(\mathcal{B})$ is a $\mathcal{G}_\delta$ set. Furthermore, $\mathsf{Regular} \cap \mathbb{L}(\mathrm{PBA}^{=1}) = \mathsf{Regular} \cap \mathrm{Deterministic}$ and $\mathsf{Regular} \cap \mathrm{Deterministic} \subsetneq \mathbb{L}(\mathrm{PBA}^{=1}) \subsetneq \mathcal{G}_\delta = \mathrm{Deterministic}$.

Proof. Lemma 3.2, Theorem 3.1 and Lemma 3.3 already imply that $\mathbb{L}(\mathrm{PBA}^{=1}) \subsetneq \mathcal{G}_\delta = \mathrm{Deterministic}$. We only need to show that $\mathsf{Regular} \cap \mathbb{L}(\mathrm{PBA}^{=1}) = \mathsf{Regular} \cap \mathrm{Deterministic}$. Since every language in $\mathbb{L}(\mathrm{PBA}^{=1})$ is deterministic, we get immediately that $\mathsf{Regular} \cap \mathbb{L}(\mathrm{PBA}^{=1}) \subseteq \mathsf{Regular} \cap \mathrm{Deterministic}$. For the reverse inclusion, note that every ω-regular, deterministic language is recognizable by a finite-state deterministic Büchi automaton. It is easy to see that any language recognized by a deterministic finite-state Büchi automaton is in $\mathbb{L}(\mathrm{PBA}^{=1})$. The result follows. □

A direct consequence of the characterization of the intersection $\mathsf{Regular} \cap \mathrm{Deterministic}$ is that the class $\mathbb{L}(\mathrm{PBA}^{=1})$ is not closed under complementation as the class of ω-regular languages recognized by deterministic Büchi automata is not closed under complementation. That the class $\mathbb{L}(\mathrm{PBA}^{=1})$ is not closed under complementation is also observed in [BBG08], and is proved by constructing an explicit example. However, even though the class $\mathbb{L}(\mathrm{PBA}^{=1})$ is not closed under complementation, we have a "partial" complementation operation— for any PBA $\mathcal{B}$ there is another PBA $\mathcal{B}'$ such that $\mathcal{L}_{>0}(\mathcal{B}')$ is the complement of $\mathcal{L}_{=1}(\mathcal{B})$. This also follows from the following results of [BBG08] as they showed that $\mathbb{L}(\mathrm{PBA}^{=1}) \subseteq \mathbb{L}(\mathrm{PBA}^{>0})$ and $\mathbb{L}(\mathrm{PBA}^{>0})$ is closed under complementation. However our construction has two advantages: 1) it is much simpler than the one obtained by the constructions in [BBG08], and 2) the PBA $\mathcal{B}'$ belongs to the restricted class of finite probabilistic monitors FPMs (see Section 2 for definition of FPMs). This construction plays a critical role in our complexity analysis of decision problems.

Lemma 4.2. For any PBA $\mathcal{B}$, there is an FPM $\mathcal{M}$ such that $\mathcal{L}_{=1}(\mathcal{B}) = \Sigma^\omega \setminus \mathcal{L}_{>0}(\mathcal{M})$.

Proof. Let $\mathcal{B} = (Q, q_s, Q_f, \delta)$. We construct $\mathcal{M}$ as follows. First we pick a new state q_r, which will be the reject state of the FPM $\mathcal{M}$. The set of states of $\mathcal{M}$ would be $Q \cup \{q_r\}$. The initial state of $\mathcal{M}$ will be q_s, the initial state of $\mathcal{B}$. The set of final states of $\mathcal{M}$ will be Q, the set of states of $\mathcal{B}$. The transition relation of $\mathcal{M}$ would be defined as follows. If q is not a final state of $\mathcal{B}$ then the transition function would be the same as for $\mathcal{B}$. If q is an final state of $\mathcal{B}$ then $\mathcal{M}$ will transit to the reject state with probability $\frac{1}{2}$ and with probability $\frac{1}{2}$ continue as in $\mathcal{B}$. Formally, $\mathcal{M} = (Q \cup \{q_r\}, q_s, Q, \delta_\mathcal{M})$ where $\delta_\mathcal{M}$ is defined as follows. For each $a \in \Sigma$, $q, q' \in Q$,

- $\delta_\mathcal{M}(q, a, q_r) = \frac{1}{2}$ and $\delta_\mathcal{M}(q, a, q') = \frac{1}{2}\delta(q, a, q')$ if $q \in Q_f$,
- $\delta_\mathcal{M}(q, a, q_r) = 0$ and $\delta_\mathcal{M}(q, a, q') = \delta(q, a, q')$ if $q \in Q \setminus Q_f$,
- $\delta_\mathcal{M}(q_r, a, q_r) = 1$.

It is easy to see that a word $\alpha \in \Sigma^\omega$ is rejected with probability 1 by $\mathcal{M}$ iff it is accepted with probability 1 by $\mathcal{B}$. The result now follows. □

The "partial" complementation operation has many consequences. One consequence is that the class $\mathbb{L}(\mathrm{PBA}^{=1})$ is closed under union. The class $\mathbb{L}(\mathrm{PBA}^{=1})$ is easily shown to be closed under intersection. Hence for closure properties, $\mathbb{L}(\mathrm{PBA}^{=1})$ behave like deterministic Büchi automata. Please note that closure properties were not studied in [BBG08].

Corollary 4.3. The class $\mathbb{L}(\mathrm{PBA}^{=1})$ is closed under finite union and finite intersection.

Proof. Let $\mathcal{B}_1 = (Q^1, q_s^1, Q_f^1, \delta^1)$ and $\mathcal{B}_2 = (Q^2, q_s^2, Q_f^2, \delta^2)$ be two PBAs, and we assume without loss of generality that $Q^1 \cap Q^2 = \emptyset$. We will present construction of PBAs that recognize the union and intersection of these languages under the almost sure semantics.

We begin by first considering the construction for union. Now by Lemma 4.2, we know that there are FPMs $\mathcal{M}_1$ and $\mathcal{M}_2$ such that $\mathcal{L}_{=1}(\mathcal{B}_i) = \Sigma^\omega \setminus \mathcal{L}_{>0}(\mathcal{M}_i)$. Now, we had shown in [CSV09a] that there is a FPM $\mathcal{M} = (Q, q_s, Q_f, \delta)$ such that for any word α, $\mu^{acc}_{\mathcal{M},\alpha} = \mu^{acc}_{\mathcal{M}_1,\alpha} \times \mu^{acc}_{\mathcal{M}_2,\alpha}$. It is easy to see that $\mathcal{L}_{>0}(\mathcal{M}) = \mathcal{L}_{>0}(\mathcal{M}_1) \cap \mathcal{L}_{>0}(\mathcal{M}_2)$.

Now, the FPM $\mathcal{M}$ can be easily "complemented". If q_r is the reject state of $\mathcal{M}$, then consider the PBA $\overline{\mathcal{M}} = (Q, q_s, \{q_r\}, \delta)$; clearly $\mathcal{L}_{=1}(\overline{\mathcal{M}}) = \Sigma^\omega \setminus \mathcal{L}_{>0}(\mathcal{M})$. Thus, by DeMorgan Laws, $\mathcal{L}_{=1}(\overline{\mathcal{M}}) = \mathcal{L}_{=1}(\mathcal{B}_1) \cup \mathcal{L}_{=1}(\mathcal{B}_2)$.

The PBA recognizing the intersection of the languages recognized by $\mathcal{B}_1$ and $\mathcal{B}_2$ with respect to almost-sure semantics does the following: on an input α, with probability $\frac{1}{2}$ it runs $\mathcal{B}_1$ on α, and with probability $\frac{1}{2}$ it runs $\mathcal{B}_2$. Clearly, such a machine will accept (with respect to almost-sure semantics) iff both $\mathcal{B}_1$ and $\mathcal{B}_2$ accept. Formally, $\mathcal{B} = (Q, q_s, Q_f, \delta)$ is given by

- $Q = Q^1 \cup Q^2 \cup \{q_s\}$ where $q_s \notin Q^1 \cup Q^2$
- $Q_f = Q_f^1 \cup Q_f^2$
- The transition relation δ is defined as follows
  - For $q \in Q^1$, $\delta(q_s, a, q) = \frac{1}{2}\delta^1(q_s^1, a, q)$, and for $q \in Q^2$, $\delta(q_s, a, q) = \frac{1}{2}\delta^2(q_s^2, a, q)$
  - For $q, q' \in Q^1$, $\delta(q, a, q') = \delta^1(q, a, q')$ and for $q, q' \in Q^2$, $\delta(q, a, q') = \delta^2(q, a, q')$. □

4.2. **Decision problems.** For the rest of this section, we shall focus our attention on decision problems for almost sure semantics for RatPBAs. Results of this section are summarized in the second row of Figure 3 on page 12 and proved in Theorem 4.4 and Theorem 4.5. The problem of checking whether $\mathcal{L}_{=1}(\mathcal{B}) = \emptyset$ for a given RatPBA $\mathcal{B}$ was shown to be decidable in **EXPTIME** in [BBG08], where it was also conjectured to be **EXPTIME**-complete. The decidability of the universality problem was left open in [BBG08]. We can leverage our "partial" complementation operation to show that a) the emptiness problem is in fact **PSPACE**-complete, thus tightening the bound in [BBG08] and b) the universality problem is also **PSPACE**-complete.

Theorem 4.4. Given a RatPBA $\mathcal{B}$, the problem of deciding whether $\mathcal{L}_{=1}(\mathcal{B}) = \emptyset$ is **PSPACE**-complete. The problem of deciding whether $\mathcal{L}_{=1}(\mathcal{B}) = \Sigma^\omega$ is also **PSPACE**-complete.

Proof. (**Upper bounds**.) We first show the upper bounds. The proof of Lemma 4.2 shows that for any RatPBA $\mathcal{B}$, there is a RatFPM $\mathcal{M}$ constructed in polynomial time such that $\mathcal{L}_{=1}(\mathcal{B}) = \Sigma^\omega \setminus \mathcal{L}_{>0}(\mathcal{M})$. $\mathcal{L}_{=1}(\mathcal{B})$ is empty (universal) iff $\mathcal{L}_{>0}(\mathcal{M})$ is universal (empty respectively). Now, we had shown in [CSV08, CSV09a] that given a RatFPM $\mathcal{M}$, the problems of checking emptiness and universality of $\mathcal{L}_{>0}(\mathcal{M})$ are in **PSPACE**, thus giving us the desired upper bounds.
(**Lower bounds**.) We had shown in [CSV08, CSV09a] that given a RatFPM $\mathcal{M}$, the problems of deciding the emptiness and universality of $\mathcal{L}_{>0}(\mathcal{M})$ are **PSPACE**-hard respectively. Given a RatFPM $\mathcal{M} = (Q, q_s, Q_0, \delta)$ with q_r as the absorbing reject state, consider the PBA $\overline{\mathcal{M}} = (Q, q_s, \{q_r\}, \delta)$ obtained by considering the unique reject state of $\mathcal{M}$ as the only final state of $\overline{\mathcal{M}}$. Clearly we have that $\mathcal{L}_{>0}(\mathcal{M}) = \Sigma^\omega \setminus \mathcal{L}_{=1}(\overline{\mathcal{M}})$. Thus $\mathcal{L}_{>0}(\mathcal{M})$ is empty (universal) iff $\mathcal{L}_{=1}(\overline{\mathcal{M}})$ is universal (empty respectively). The result now follows. □

	Emptiness	Universality
$\mathbb{L}(\text{HPBA}^{>0})$	**NL**-complete	**PSPACE**-complete
$\mathbb{L}(\text{HPBA}^{=1})$	**PSPACE**-complete	**NL**-complete

Figure 5: Complexity of decision problems for RatHPBAs.

Even though the problems of checking emptiness and universality of almost-sure semantics of a RatPBA are decidable, the problem of deciding language containment under almost-sure semantics turns out to be undecidable, and is indeed as hard as the problem of deciding language containment under probable semantics (or, equivalently, checking emptiness under probable semantics).

Theorem 4.5. Given RatPBAs, $\mathcal{B}_1$ and $\mathcal{B}_2$, the problem of deciding whether $\mathcal{L}_{=1}(\mathcal{B}_1) \subseteq \mathcal{L}_{=1}(\mathcal{B}_2)$ is $\mathbf{\Sigma}^0_2$-complete.

Proof. Observe first that given RatPBAs $\mathcal{B}_1$ and $\mathcal{B}_2$, there are (constructible) RatFPMs $\mathcal{M}_1$ and $\mathcal{M}_2$ such that $\mathcal{L}_{=1}(\mathcal{B}_i) = \Sigma^\omega \setminus \mathcal{L}_{>0}(\mathcal{M}_i)$ for $i = 1, 2$ (see Lemma 4.2). Thus, $\mathcal{L}_{=1}(\mathcal{B}_1) \subseteq \mathcal{L}_{=1}(\mathcal{B}_2)$ iff $\mathcal{L}_{>0}(\mathcal{M}_2) \subseteq \mathcal{L}_{=1}(\mathcal{M}_1)$. The upper bound then follows from the upper bound of the containment of PBAs under probable semantics.

The lower bound is shown by a reduction from emptiness-checking of probable semantics. Recall from the proof of the fact that $\mathbb{L}(\text{PBA}^{>0}) = \mathsf{BCl}(\mathbb{L}(\text{PBA}^{=1}))$ (Theorem 3.1) that given a RatPBA $\mathcal{B}$, there are RatPBAs $\mathcal{B}^+_1, \mathcal{B}^+_2, \dots \mathcal{B}^+_m$ and $\mathcal{B}^-_1, \mathcal{B}^-_2 \dots \mathcal{B}^-_m$ such that

$$\mathcal{L}_{>0}(\mathcal{B}) = \bigcup_{1 \le i \le m} \mathcal{L}_{=1}(\mathcal{B}^+_i) \cap (\Sigma^\omega \setminus \mathcal{L}_{=1}(\mathcal{B}^-_i)).$$

Furthermore, the construction in the proof of Theorem 3.1 and results of [Grö08] (Proposition 2.5 and the fact that complementation of probable semantics is a recursive operation for RatPBAs) implies that $\mathcal{B}^+_i$ and $\mathcal{B}^-_i$ are constructible. Now, $\mathcal{L}_{>0}(\mathcal{B}) = \emptyset$ iff for each i, $\mathcal{L}_{=1}(\mathcal{B}^+_i) \cap (\Sigma^\omega \setminus \mathcal{L}_{=1}(\mathcal{B}^-_i)) = \emptyset$. The lower bound now follows from the observation that $\mathcal{L}_{=1}(\mathcal{B}^+_i) \cap (\Sigma^\omega \setminus \mathcal{L}_{=1}(\mathcal{B}^-_i)) = \emptyset$ iff $\mathcal{L}_{=1}(\mathcal{B}^+_i) \subseteq \mathcal{L}_{=1}(\mathcal{B}^-_i)$. ☐

5. Hierarchical PBAs

We now identify a simple syntactic restriction on PBAs which–

- under probable semantics coincide exactly with ω-regular languages, and
- under almost-sure semantics coincide exactly with ω-regular deterministic languages.

We will also establish complexity of decision problems of emptiness and universality for the case when transition probabilities are given as rational numbers. The complexity results are summarized in Figure 5.

Intuitively, a hierarchical PBA is a PBA such that the set of its states can be stratified into (totally) ordered levels. From a state q, for each letter a, the machine can transition with non-zero probability to at most one state in the same level as q, and all other probabilistic transitions go to states that belong to a higher level. Formally,

Definition 5.1. Given a natural number k, a PBA $\mathcal{B} = (Q, q_s, Q, \delta)$ over an alphabet Σ is said to be a *k-level hierarchical PBA (k-*HPBA) if there is a function $\mathsf{rk} : Q \to \{0, 1, \dots, k\}$ such that the following holds.

Given $j \in \{0, 1, \ldots, k\}$, let $Q_j = \{q \in Q \mid \mathsf{rk}(Q) = j\}$. For every $q \in Q$ and $a \in \Sigma$, if $j_0 = \mathsf{rk}(q)$ then $\mathsf{post}(q, a) \subseteq \cup_{j_0 \leq \ell \leq k} Q_\ell$ and $|\mathsf{post}(q, a) \cap Q_{j_0}| \leq 1$.

The function rk is said to be a *compatible ranking function* of $\mathcal{B}$ and for $q \in Q$ the natural number $\mathsf{rk}(q)$ is said to be the *rank* or *level* of q. $\mathcal{B}$ is said to be a *hierarchical* PBA *(*HPBA*)* if $\mathcal{B}$ is k-hierarchical for some k. If $\mathcal{B}$ is also a RatPBA, we say that $\mathcal{B}$ is a rational hierarchical PBA (RatHPBA).

We can define classes analogous to $\mathbb{L}(\text{PBA}^{>0})$ and $\mathbb{L}(\text{PBA}^{=1})$; and we shall call them $\mathbb{L}(\text{HPBA}^{>0})$ and $\mathbb{L}(\text{HPBA}^{=1})$ respectively. Before we proceed to discuss the probable and almost-sure semantics for HPBAs, we point out two interesting facts about hierarchical HPBAs. First is that for the class of ω-regular deterministic languages, HPBAs like non-deterministic Büchi automata can be exponentially more succinct.

Proposition 5.2. Let $\Sigma = \{a, b, c\}$. For each $n \in \mathbb{N}$, there is a ω-regular deterministic language $\mathsf{L}_n \subseteq \Sigma^\omega$ such that i) any deterministic Büchi automata for L_n has at least $O(2^n)$ number of states, and ii) there are HPBAs $\mathcal{B}_n$ s.t. $\mathcal{B}_n$ has $O(n)$ number of states and $\mathsf{L}_n = \mathcal{L}_{=1}(\mathcal{B}_n)$.

Proof. Given $n \in \mathbb{N}$, let L_n be the safety language in which for every a there is a c after exactly n-steps. In other words, $\mathsf{L}_n = \Sigma^\omega \setminus (\Sigma^* a \Sigma^n \{a, b\} \Sigma^\omega)$. This could model, for instance, the property "every request a is answered after exactly n-steps". We can build a deterministic Büchi automaton for L_n and the number of states of such a automaton is $O(2^n)$. We could build a HPBA $\mathcal{B}_n$ with $O(n)$ state such that $\mathsf{L}_n = \mathcal{L}_{=1}(\mathcal{B}_n)$. The HPBA $\mathcal{B}_n$ will be an FPM also. The construction of $\mathcal{B}_n$ is as follows— $\mathcal{B}_n$ scans the input and upon encountering a, $\mathcal{B}_n$ decides with probability $\frac{1}{2}$ to check if there is a c after n steps and with probability $\frac{1}{2}$, $\mathcal{B}_n$ decides to continue scanning the rest of the input. In the former case, if the check $\mathcal{B}_n$ reveals an error then $\mathcal{B}_n$ rejects the input; otherwise $\mathcal{B}_n$ accepts the input. □

The second thing is that even though HPBAs yield only ω-regular languages under both almost-sure semantics and probable semantics, we can recognize non-ω-regular languages with cutpoints.

Proposition 5.3. There is a HPBA $\mathcal{B}$ such that both $\mathcal{L}_{\geq \frac{1}{2}}(\mathcal{B})$ and $\mathcal{L}_{> \frac{1}{2}}(\mathcal{B})$ are not ω-regular.

Proof. The HPBA we will construct will actually an FPM. The following construction is given in [CSV08]. Let $\Sigma = \{\mathbf{0}, \mathbf{1}\}$. Let $Q = \{q_0, q_1, q_r\}$ and $\delta : Q \times \Sigma \times Q \to [0, 1]$ be defined as follows. The states q_r and q_1 are absorbing, *i.e.*, $\delta(q_r, \mathbf{0}, q_r) = \delta(q_r, \mathbf{1}, q_r) = \delta(q_1, \mathbf{0}, q_1) = \delta(q_1, \mathbf{1}, q_1) = 1$. Transitions out of q_0 satisfy $\delta(q_0, \mathbf{0}, q_0) = \delta(q_0, \mathbf{0}, q_r) = \delta(q_0, \mathbf{1}, q_0) = \delta(q_0, \mathbf{1}, q_1) = \frac{1}{2}$. Consider the FPM $\mathcal{M}_{\mathsf{Id}} = (Q, q_0, \{q_0, q_1\}, \delta)$. $\mathcal{M}_{\mathsf{Id}}$ can be seen to be 2-hierarchical with $\mathsf{rk}(q_0) = 0, \mathsf{rk}(q_1) = 1$ and $\mathsf{rk}(q_r) = 2$. Given $\alpha = a_0 a_1 \ldots$, it can be shown that $\mu^{acc}_{\mathcal{M}_{\mathsf{Id}}, \alpha} = \mathsf{bin}(\alpha)$ where $\mathsf{bin}(\alpha)$ is the real number: $\sum_i \frac{\mathsf{num}(a_i)}{2^{i+1}}$ where $\mathsf{num}(\mathbf{0})$ is the integer 0 and $\mathsf{num}(\mathbf{1})$ is the integer 1.

Now, consider the FPM $\mathcal{M}_{\mathsf{Id}} \circ \mathcal{M}_{\mathsf{Id}}$ constructed as follows. The states of this FPM are $\{q_0, q_1\} \times \{q_0, q_1\} \cup q_{r_{\mathsf{new}}}$. The initial state is (q_0, q_0) and the reject state is $q_{r_{\mathsf{new}}}$. The transition probabilities, from the state (q_{i_1}, q_{j_1}) on input $a \in \{\mathbf{0}, \mathbf{1}\}$ is defined as follows— to state (q_{i_2}, q_{j_2}) the transition probability is $\delta(q_{i_1}, a, q_{i_2}) \times \delta(q_{j_1}, a, q_{j_2})$ and to state $q_{r_{\mathsf{new}}}$ the transition probability is $1 - \sum_{i_2, j_2 \in \{0,1\}} \delta(q_{i_1}, a, q_{i_2}) \times \delta(q_{j_1}, a, q_{j_2})$. The state $q_{r_{\mathsf{new}}}$ is absorbing.

The FPM $\mathcal{M}_{\mathsf{Id}} \circ \mathcal{M}_{\mathsf{Id}}$ can be seen to be hierarchical with $\mathsf{rk}(q_{i_1}, q_{i_2}) = i_1 + i_2$. Furthermore, it can be shown that on word α, $\mu^{acc}_{\mathcal{M}_{\mathsf{Id}} \circ \mathcal{M}_{\mathsf{Id}}, \alpha} = (\mathsf{bin}(\alpha))^2$. Thus, $\mathcal{L}_{>\frac{1}{2}}(\mathcal{M}_{\mathsf{Id}} \circ \mathcal{M}_{\mathsf{Id}}) = \{\alpha \mid \mathsf{bin}(\alpha) > \sqrt{\frac{1}{2}}\}$ and $\mathcal{L}_{\geq\frac{1}{2}}(\mathcal{M}_{\mathsf{Id}} \circ \mathcal{M}_{\mathsf{Id}}) = \{\alpha \mid \mathsf{bin}(\alpha) \geq \sqrt{\frac{1}{2}}\}$; both of which are not ω-regular. ☐

Remarks 5.4. We will see shortly that the problems of deciding emptiness and universality for a HPBA turn out to be decidable under both probable and almost-sure semantics. However, with cutpoints, they turn out to be undecidable. The latter observation is out of scope of the paper.

5.1. **Probable semantics.** We shall now show that the class $\mathbb{L}(\mathrm{HPBA}^{>0})$ coincides with the class of ω-regular languages. In [BG05], a restricted class of PBAs called uniform PBAs was identified that also accept exactly the class of ω-regular languages. We make a couple of observations, contrasting our results here with theirs. First the definition of uniform PBA was semantic (i.e., the condition depends on the acceptance probability of infinitely many strings from different states of the automaton), whereas HPBA are a syntactic restriction on PBA. Second, we note that the definitions themselves are incomparable in some sense; in other words, there are HPBAs which are not uniform, and vice versa. Finally, HPBAs appear to be more tractable than uniform PBAs. We show that the emptiness problem for $\mathbb{L}(\mathrm{HPBA}^{>0})$ is **NL**-complete. In contrast, the same problem was demonstrated to be in **EXPTIME** and co-**NP**-hard [BG05] for uniform PBAs.

We first establish that every ω-regular language can be recognized by a hierarchical PBA; this is the content of the next Lemma.

Lemma 5.5. For every ω-regular language L, there is a hierarchical PBA $\mathcal{B}$ such that $\mathsf{L} = \mathcal{L}_{>0}(\mathcal{B})$.

Proof. Let $\mathcal{R} = (Q, q_s, F, \Delta)$ be a deterministic Rabin automaton recognizing L, where $F = \{(B_1, G_1), \ldots (B_k, G_k)\}$. The hierarchical PBA will, intuitively, in the first step choose the pair (B_i, G_i) that will be satisfied in the run, and then ensure that the measure of paths that visit B_i infinitely often is 0. Formally, $\mathcal{B} = (Q', q'_s, Q'_f, \delta')$ is given as follows.

- $Q' = \{q'_s, q'_r\} \cup (\{1, \ldots k\} \times Q)$, where $q'_s, q'_r \notin Q$
- $Q'_f = \bigcup_{i=1}^{k}(\{i\} \times G_i)$
- The transition relation δ' is given by
  - $\delta'(q'_s, a, (i, q)) = \frac{1}{k}$ iff $(q_s, a, q) \in \Delta$
  - For $q \notin B_i$, $\delta'((i, q), a, (i, q')) = 1$ iff $(q, a, q') \in \Delta$
  - For $q \in B_i$, $\delta'((i, q), a, q'_r) = \frac{1}{2}$ for all $a \in \Sigma$, and $\delta'((i, q), a, (i, q')) = \frac{1}{2}$ iff $(q, a, q') \in \Delta$
  - Finally, $\delta'(q'_r, a, q'_r) = 1$ for all $a \in \Sigma$.

It is easy to see that $\mathcal{L}_{>0}(\mathcal{B}) = \mathsf{L}$. Finally, we point out that $\mathcal{B}$ is a $k+1$-level hierarchical PBA. This is witnessed by the ranking function rk defined as follows — $\mathsf{rk}(q'_s) = 0$, $\mathsf{rk}((i, q)) = i$, and $\mathsf{rk}(q'_r) = k + 1$. ☐

Theorem 5.6. $\mathbb{L}(\mathrm{HPBA}^{>0}) = \mathsf{Regular}$.

Proof. Thanks to Lemma 5.5 we need to show that every language in $\mathbb{L}(\mathrm{HPBA}^{>0})$ is ω-regular. The other inclusion follows from the following Claim.

Claim: For any hierarchical PBA $\mathcal{B} = (Q, q_s, Q_f, \delta)$ and any word $\alpha \in \Sigma^\omega$, $\alpha \in \mathcal{L}_{>0}(\mathcal{B})$ iff there is an infinite sequence of states $q_s = q_0, q_1, \ldots$ such that $q_i \in Q_f$ for infinitely many $i \in \mathbb{N}$, $\delta(q_i, \alpha[i], q_{i+1}) > 0$ for all $i \in \mathbb{N}$ and $\exists j \geq 0$ such that $\delta(q_i, \alpha[i], q_{i+1}) = 1$ for all $i \geq j$.

Proof of the claim: Let $\mathcal{B}$ be a k-level hierarchical PBA with compatible ranking function rk. Let $Q_j = \{q \in Q \mid \mathsf{rk}(q) = j\}$. The proof will proceed by induction on the level k.

Base Case: Suppose $k = 0$. Based on the definition of hierarchical PBAs, this means that $\mathcal{B}$ is a deterministic Büchi automaton, i.e., for all $q, q' \in Q$ and $a \in \Sigma$, either $\delta(q, a, q') = 1$ or $\delta(q, a, q') = 0$. Thus, the claim clearly holds in this case.

Induction Step: Let $\alpha \in \Sigma^\omega$ be such that $\alpha \in \mathcal{L}_{>0}(\mathcal{B})$, with $\mu^{acc}_{\mathcal{B},\alpha} = x > 0$. Observe that for every i, $|\mathsf{post}(q_s, \alpha[0, i]) \cap Q_0| \leq 1$. There are two cases to consider.

Case 1: Suppose $|\mathsf{post}(q_s, \alpha[0, i]) \cap Q_0| = 1$ for all i; let us denote the unique state in $\mathsf{post}(q_s, \alpha[0, i]) \cap Q_0$ by q_i. Suppose in addition, there is a j such that for all $\ell > j$, $\delta(q_\ell, \alpha[\ell], q_{\ell+1}) = 1$. Then clearly the sequence $q_0, q_1, \ldots$ satisfies the conditions of the lemma.

Case 2: Suppose Case 1 does not hold. Then there are two possibilities. The first possibility is that there is a i_0 such that $\mathsf{post}(q_s, \alpha[0, i_0]) \cap Q_0 = \emptyset$. The second possibility is that for every j, there is a $\ell > j$ such that $\delta(q_\ell, \alpha[\ell], q_{\ell+1}) < 1$, where once again we are denoting the unique state of Q_0 in $\mathsf{post}(q_s, \alpha[0, \ell])$ by q_ℓ. In this second subcase, there must then exist an i_0 such that $\delta_u(q_s, q_{i_0}) < x$, where $u = \alpha[0, i_0]$.

Now, based on the definition of i_0 given for the two subcases above, it must be the case that for some state $q \in \mathsf{post}(q_s, \alpha[0, i_0]) \setminus Q_0$, the measure of accepting runs from q on the word $\alpha[i_0+1]\alpha[i_0+2]\cdots$ is non-zero. Consider the hierarchical PBA $\mathcal{B}' = (Q', q, Q'_f, \delta')$, where $Q' = Q \setminus Q_0$, $Q'_f = Q_f \setminus Q_0$ and $\delta' = \delta|_{Q' \times \Sigma \times Q'}$. Clearly, $\mathcal{B}'$ is a $k-1$-level hierarchical PBA, and thus by induction hypothesis, the string $\alpha[i_0+1]\alpha[i_0+2]\cdots$ has a run $q = q'_0 q'_1 \ldots$ satisfying the conditions in the claim. The desired run for α (in PBA $\mathcal{B}$) satisfying the conditions in the lemma is obtained by concatenating a run from q_s to q on $\alpha[0, i_0]$ with $q'_0 q'_1 \ldots$. **(End proof of claim).** □

We now proceed with the main theorem. Let $\mathcal{B} = (Q, q_s, Q_f, \delta)$. We will construct a finite-state nondeterministic Büchi automaton $\mathcal{A} = (Q', q'_s, Q'_f, \Delta')$, such that the language recognized by $\mathcal{A}$ is exactly $\mathcal{L}_{>0}(\mathcal{B})$. Intuitively, the set of states Q' will consist of two copies of Q— $Q \times \{0\}$ and $Q \times \{1\}$. In the first copy, we will simulate the possible transitions between pair of states of $\mathcal{B}$ (we ignore the exact transition probabilities of $\mathcal{B}$). In the second copy, we will only simulate *deterministic* transitions of $\mathcal{B}$, *i.e.*, those transitions between pair of states which happen with probability 1. From the first copy, we can transit to the second copy if the probability of transiting between the corresponding states in $\mathcal{B}$ is non-zero. From the second copy, we will never transit to the first state. The set of final states of $\mathcal{A}$ are those states in second level that correspond to the final states of $\mathcal{B}$. Intuitively, the construction ensures that if $\alpha \in \mathsf{L}_{>0}(\mathcal{B})$, and the sequence $q_s = q_0, q_1, \ldots$ and natural number $j \geq 0$ are such that

(1) $q_\ell \in Q_f$ for infinitely many ℓ,
(2) $0 < \delta(q_i, \alpha[i], q_{i+1}) < 1$ for all $i < j$ and $\delta(q_i, \alpha[i], q_{i+1}) = 1$ for all $i \geq j$

then $(q_0, 0), \ldots (q_j, 0), (q_{j+1}, 1), (q_{j+2}, 1) \ldots$ is an accepting run of $\mathcal{A}$ on input α.

Formally, Q' is the set $Q \times \{0, 1\}$, $q'_s = (q_s, 0)$, $Q'_f = \{(q, 1) \mid q \in Q_f\}$, and Δ' is defined as follows. For each $q_1, q_2 \in Q$,

- $((q_1, 0), a, (q_2, 0)) \in \Delta'$ iff $\delta(q_1, a, q_2) > 0$.

- $((q_1, 0), a, (q_2, 1)) \in \Delta'$ iff $\delta(q_1, a, q_2) > 0$.
- $((q_1, 1), a, (q_2, 1)) \in \Delta'$ iff $\delta(q_1, a, q_2) = 1$.
- $((q_1, 1), a, (q_2, 0)) \in \Delta'$ iff never.

The claim above immediately implies that $\mathcal{L}_{>0}(\mathcal{B})$ is the language recognized by $\mathcal{A}$ and hence is ω-regular. □

We will show that the problem of deciding whether $\mathcal{L}_{>0}(\mathcal{B})$ is empty for hierarchical RatPBA's is **NL**-complete while the problem of deciding whether $\mathcal{L}_{>0}(\mathcal{B})$ is universal is **PSPACE**-complete. Thus "algorithmically", hierarchical PBAs are much "simpler" than both PBAs and uniform PBAs. Note that the emptiness and universality problem for finite state Büchi-automata are also **NL**-complete and **PSPACE**-complete respectively.

Theorem 5.7. Given a RatHPBA, $\mathcal{B}$, the problem of deciding whether $\mathcal{L}_{>0}(\mathcal{B}) = \emptyset$ is **NL**-complete. The problem of deciding whether $\mathcal{L}_{>0}(\mathcal{B}) = \Sigma^\omega$ is **PSPACE**-complete.

Proof. **(Upper Bounds).** First note since $\mathcal{B}$ is hierarchical, the language $\mathcal{L}_{>0}(\mathcal{B})$ is ω-regular (see Theorem 5.6). The proof of Theorem 5.6 also allows us to construct a finite-state Büchi automata $\mathcal{A}$ such that a) $\mathcal{L}_{>0}(\mathcal{B})$ is the language recognized by $\mathcal{A}$ and b) the size of the automaton $\mathcal{A}$ is at-most twice the size of the automaton $\mathcal{B}$. Furthermore, the construction can be carried out in **NL**. Since the emptiness problem of finite-state Büchi automata is in **NL** and the universality problem is in **PSPACE**, we immediately get that the desired upper bounds.

(Lower Bounds). Please note that the **NL**-hardness of the emptiness problem can be proved easily from the emptiness problem of deterministic finite state machines. For the universality problem, we make the following claim.

Claim: Given an FPM $\mathcal{M}$ such that the $\mathcal{M}$ is also a hierarchical PBA, the problem of deciding whether $\mathcal{L}_{=1}(\mathcal{M})$ is empty is **PSPACE**-hard.

Before, we proceed to prove the claim, we first show how the lower bound follows from the reduction. Given an FPM $\mathcal{M} = (Q, q_s, Q_f, \delta)$ with reject state q_r, consider the PBA $\overline{\mathcal{M}} = (Q, q_s, \{q_r\}, \delta)$ obtained by taking the reject state of $\mathcal{M}$ as the unique final state of $\overline{\mathcal{M}}$. Clearly,

(1) $\overline{\mathcal{M}}$ is HPBA if $\mathcal{M}$ is.
(2) $\mathcal{L}_{>0}(\overline{\mathcal{M}})$ is universal iff $\mathcal{L}_{=1}(\mathcal{M})$ is empty.

From these two observations the desired result will follow if we can prove the claim. We now prove the claim.

Proof of the claim. We show that there is a polynomial time bounded reduction from every language in **PSPACE** to the language

$$\{(\mathcal{M}, \Sigma) \mid \mathcal{M} \text{ is an FPM on } \Sigma, \mathcal{M} \text{ is a HPBA and } \mathcal{L}_{=1}(\mathcal{M}) = \emptyset\}.$$

Consider a language $L \in$ **PSPACE** and $\mathbb{T}$ be a single tape deterministic Turing machine that accepts L in space $p(n)$ for some polynomial p where n is the length of its input. We assume that $\mathbb{T}$ accepts an input by halting in a specific final state q_f and $\mathbb{T}$ rejects an input by not halting. Let $\mathbb{T}$ be given by the tuple $(Q, \Lambda, \Gamma, \Delta, q_0, q_f)$. Here Q is the set of states of the finite control of $\mathbb{T}$; Λ, Γ are the input and tape alphabets and $\Lambda \subseteq \Gamma$ and the blank symbol # is in $\Gamma \setminus \Lambda$; $\Delta : Q \times \Gamma \to \Gamma \times Q \times \{Left, Right\}$; q_0 is the initial state and q_f is the final state. Each tuple $\Delta(q, a) = (a', q', d)$ indicates that when $\mathbb{T}$ is in state q, scanning a cell containing the symbol a, then $\mathbb{T}$ writes value a' in the current cell, changes to state

q' and moves in the direction d. Without loss of generality, we assume that head position of $\mathbb{T}$ initially is at cell number 0.

Let $\Phi' = \Gamma \times Q$ and $\Phi = \Phi' \cup \Gamma$. We call members of Φ' composite symbols. A configuration of $\mathbb{T}$, on an input of length n, is a string of symbols, of length $p(n)$, drawn from Φ. We can define a valid configuration in the standard way. In each valid configuration there can be only one composite symbol (i.e.,from Φ') and that indicates the head position of $\mathbb{T}$. A computation of $\mathbb{T}$ is a sequence of configurations which is either finite or infinite depending on whether the input is accepted or not. A computation starts in an initial configuration and each succeeding configuration is obtained by one move of $\mathbb{T}$ from the previous configuration. The initial configuration contains the input string and the first symbol in it is from $\Gamma \times Q$ indicating its head position is on the first cell.

For given input σ, we construct a FPM $\mathcal{M}_\sigma$ such that $\mathcal{M}_\sigma$ is a 2-HPBA and $\mathcal{M}_\sigma$ accepts some infinite input with probability 1 iff $\mathbb{T}$ rejects σ, i.e., $\mathbb{T}$ does not halt on σ. Let σ be an input to $\mathbb{T}$ of length n and let $m = p(n)$. A state of the automaton $\mathcal{M}_\sigma$ is a pair of the form (i, s) where $0 \leq i < m$ and $s \in \Phi$, or is in $\{q_s, q_r\}$; here q_s is the initial state and is of rank 0 and q_r is the reject state and is of rank 2. The rank of states $\{(i, s) \mid 0 \leq i < m \text{ and } s \in \Phi\}$ will be 1. Intuitively, if $\mathcal{M}_\sigma$ is in state (i, s) that denotes that i^{th} element of the current configuration of the computation of $\mathbb{T}$ has value s. Note that s is in Φ' or is in Γ. The input alphabet to $\mathcal{M}_\sigma$ is the set $\{0, ..., m-1\} \times \Phi' \times \{left, right\}$ together with an additional input symbol τ; that is each input to the automaton is τ or is of the form $(i, (b, q), d)$.

Let $\sigma = \sigma_0, ..., \sigma_{n-1}$ be the input to $\mathbb{T}$. The transitions of $\mathcal{M}_\sigma$ are defined as follows. From the initial state q_s, on input τ, there are transitions to the states (i, r_i), for each $i \in \{0, ..., m-1\}$ where, $r_i = (\sigma_0, q_0)$ and $r_i = \sigma_i$ for $0 < i < n$, and is the blank symbol otherwise; the probability of each of these transitions is $\frac{1}{m}$. Thus the input τ sets up the initial configuration when $\mathcal{M}_\sigma$ is in the initial state q_r. From every other state on input τ there is a transition to the reject state q_r with probability 1. Also, from the initial state q_s, there is a transition to the reject state q_r with probability 1 for all input symbols other than τ.

From any state of the form $(j, (b, q))$ on input of the form $(i, (a, q'), d)$ the transition is defined as follows: if $i = j$, $q = q'$, $b = a$ and $\Delta(q, a) = (q_1, a_1, d)$, then there is a transition to the automaton state (j, a_1); otherwise, the transition is to q_r; in either case, the probability of the transition is 1. Note that if $\mathbb{T}$ halts then also there is a transition to q_r.

From any state of the form (j, b), where $b \in \Gamma$, on input symbol of the form $(i, (a, q), d)$ the transitions are defined as follows: if either $i = j - 1$, $d = right$ and $\Delta(q, a) = (q', a', d')$, or if $i = j + 1$, $d = left$ and $\Delta(q, a) = (q', a', d')$ then the transition is to the state $(j, (b, q'))$; otherwise the transition is back to (j, b); in both cases the probability of the transition is 1.

Suppose σ is rejected, i.e., $\mathbb{T}$ does not terminate on σ. Furthermore assume that the composite symbols in each successive configuration of the infinite computation of T on input σ are $(a_0, q_0), (a_1, q_1), ...$ and they occur in positions $i_0, ...$ and the direction of the head movement is given by $d_0, ...$ respectively. Then $\mathcal{M}_\sigma$ accepts the infinite string $\tau(i_0, (a_0, q_0), d_0), .., (i_k, (a_k, q_k), d_k), ...$ with probability 1 and accepts all others with probability less than 1. It is not difficult to see that if σ is accepted by $\mathbb{T}$, all input strings are accepted by $\mathcal{M}_\sigma$ with probability less than 1. The above reduction is clearly polynomial time bounded. **(End proof of the claim.)** □

5.2. **Almost-sure semantics.** For a hierarchical PBA, the "partial" complementation operation for almost-sure semantics discussed in Section 4 yields a hierarchical PBA. Therefore using Theorem 5.6, we immediately get that a language $\mathcal{L} \in \mathbb{L}(\text{HPBA}^{=1})$ is ω-regular. Thanks to the topological characterization of $\mathbb{L}(\text{HPBA}^{=1})$ as a sub-collection of deterministic languages, we get that $\mathbb{L}(\text{HPBA}^{=1})$ is exactly the class of languages recognized by deterministic finite-state Büchi automata.

Theorem 5.8. $\mathbb{L}(\text{HPBA}^{=1}) = \mathsf{Regular} \cap$ Deterministic.

Proof. The inclusion $\mathsf{Regular} \cap$ Deterministic $\subseteq \mathbb{L}(\text{HPBA}^{=1})$ follows immediately from the fact that any language in $\mathsf{Regular} \cap$ Deterministic is recognizable by a finite-state deterministic Büchi automaton. For the reverse inclusion $\mathbb{L}(\text{HPBA}^{=1}) \subseteq \mathsf{Regular} \cap$ Deterministic, note that since $\mathbb{L}(\text{PBA}^{=1}) \subseteq$ Deterministic, it suffices to show that $\mathbb{L}(\text{HPBA}^{=1}) \subseteq \mathsf{Regular}$. Now, given $\mathsf{L} \in \mathbb{L}(\text{HPBA}^{=1})$, Lemma 4.2 immediately implies that there is an FPM $\mathcal{M}$ such that $\mathcal{L}_{>0}(\mathcal{M}) = \Sigma^\omega \setminus \mathsf{L}$. Furthermore, it is easy to see from the proof of Lemma 4.2 that we can take $\mathcal{M}$ to be hierarchical given that $\mathsf{L} \in \mathbb{L}(\text{HPBA}^{=1})$. Now, thanks to Theorem 5.6, $\mathcal{L}_{>0}(\mathcal{M})$ is ω-regular which implies that L is also ω-regular. □

The "partial" complementation operation also yields the complexity of emptiness and universality problems.

Theorem 5.9. Given a RatHPBA, $\mathcal{B}$, the problem of deciding whether $\mathcal{L}_{=1}(\mathcal{B}) = \emptyset$ is **PSPACE**-complete. The problem of deciding whether $\mathcal{L}_{=1}(\mathcal{B}) = \Sigma^\omega$ is **NL**-complete.

(Upper Bounds.) The upper bounds are obtained by constructing the FPM $\mathcal{M}$ such that $\mathcal{L}_{=1}(\mathcal{B}) = \Sigma^\omega \setminus \mathcal{L}_{>0}(\mathcal{M})$ as in the proof of Lemma 4.2. Now, $\mathcal{M}$ is hierarchical if $\mathcal{B}$ is hierarchical. The result now follows immediately from Theorem 5.7.

Proof. **(Lower Bounds.)** The **NL**-hardness of checking universality can be shown from **NL**-hardness of checking emptiness of deterministic finite state machines. Please recall that in the proof of Theorem 5.7, we had shown that given an FPM $\mathcal{M}$ such that $\mathcal{M}$ is a HPBA, the problem of checking whether $\mathcal{L}_{=1}(\mathcal{M})$ is empty is **PSPACE**-hard. Thus, it follows immediately that checking emptiness of $\mathcal{L}_{=1}(\mathcal{B})$ for a HPBA is **PSPACE**-hard. □

6. Conclusions

In this paper, we investigated the power of randomization in finite state automata on infinite strings. We presented a number of results on the expressiveness and decidability problems under different notions of acceptance based on the probability of acceptance. In the case of decidability, we gave tight bounds for both the universality and emptiness problems. As part of future work, it will be interesting to investigate the power of randomization in other models of computations on infinite strings such as pushdown automata etc. Since the universality and emptiness problems are PSPACE-complete for almost-sure semantics, their application to practical systems needs further enquiry.

Acknowledgements.

The authors thank anonymous referees whose comments have improved the presentation of the paper.

References

[AS85] B. Alpern and F. Schneider. Defining liveness. *Information Processing Letters*, 21:181–185, 1985.

[BBG08] C. Baier, N. Bertrand, and M. Größer. On decision problems for probabilistic Büchi automata. In *11th International Conference on Foundations of Software Science and Computational Structures, FoSSaCS*, pages 287–301, 2008.

[BG05] C. Baier and M. Größer. Recognizing ω-regular languages with probabilistic automata. In *Proceedings of the IEEE Symposium on Logic in Computer Science*, pages 137–146, 2005.

[CL89] A. Condon and R. J. Lipton. On the complexity of space bounded interactive proofs (extended abstract). In *30th Annual Symposium on Foundations of Computer Science*, pages 462–467, 1989.

[CSV08] R. Chadha, A. P. Sistla, and M. Viswanathan. On the expressiveness and complexity of randomization in finite state monitors. In *LICS 2008- 23rd Annual IEEE Symposium on Logic in Computer Science*, pages 18–29. IEEE Computer Society, 2008.

[CSV09a] R. Chadha, A. P. Sistla, and M. Viswanathan. On the expressiveness and complexity of randomization in finite state monitors. *Journal of the ACM*, 56(5), 2009.

[CSV09b] R Chadha, A. P. Sistla, and M. Viswanathan. Power of randomization in automata on infinite strings. In *Proceedings of the 20th International Conference on Concurrency Theory (CONCUR'09)*, volume 5710 of *Lecture Notes in Computer Science*, pages 229–243. Springer, 2009.

[Grö08] M. Größer. *Reduction Methods for Probabilistic Model Checking*. PhD thesis, TU Dresden, 2008.

[HP96] G. J. Holzmann and D. Peled. The state of spin. In *Proceedings of the 8th International Conference on Computer Aided Verification*, 1996.

[KS76] J. Kemeny and J. Snell. *Denumerable Markov Chains*. Springer-Verlag, 1976.

[Kur94] R. P. Kurshan. *Computer Aided Verification of the Coordinated Processes: The Automata Theoretic Approach*. Princeton University Press, 1994.

[Lam85] L. Lamport. Logical foundation, distributed systems- methods and tools for specification. *Springer-Verlag Lecture Notes in Computer Science*, 190, 1985.

[Lan69] L. H. Landweber. Decision problems for omega-automata. *Mathematical Systems Theory*, 3(4):376–384, 1969.

[Paz71] A. Paz. *Introduction to Probabilistic Automata*. Academic Press, 1971.

[PP04] D. Perrin and J-E. Pin. *Infinite Words: Automata, Semigroups, Logic and Games*. Elsevier, 2004.

[Rab63] M. O. Rabin. Probabilitic automata. *Information and Control*, 6(3):230–245, 1963.

[RKNP04] J. M. Rutten, M. Kwiatkowska, G. Norman, and D. Parker. *Mathematical Techniques for Analyzing Concurrent and Probabilistic Systems*. AMS, 2004.

[Sal73] A. Salomaa. *Formal Languages*. Academic Press, 1973.

[Sis83] A. P. Sistla. *Theoretical Issues in the Design and Verification of Distributed Systems*. PhD thesis, Harvard University, 1983.

[Tho90] W. Thomas. Automata on infinite objects. In *Handbook of Theoretical Computer Science*, volume B, pages 133–192. 1990.

[Var85] M. Vardi. Automatic verification of probabilistic concurrent systems. In *26th annual Symposium on Foundations of Computer Science*, pages 327–338. IEEE Computer Society Press, 1985.

[VW86] M. Vardi and P. Wolper. An automata theoretic approach to automatic program verification. In *Proceedings of the first IEEE Symposium on Logic in Computer Science*, 1986.

[VWS83] M. Vardi, P. Wolper, and A. P. Sistla. Reasoning about infinite computation paths. In *Proceedings of the 24th IEEE Symposium on Foundations of Computer Science*, 1983.

Appendix A. Properties of $\mathcal{R}$ in the proof of Lemma 3.8

Lemma A.1. Let M be a deterministic 2-counter machine with a one way read only input tape whose configurations are encoded over alphabet Σ'. Let ϵ be any rational such that $0 < \epsilon < \frac{1}{2}$. There is a PFA $\mathcal{R}$ over alphabet $\Sigma_{\mathcal{R}} = \Sigma' \cup \{@\}$, where $@ \notin \Sigma'$, such that

(1) There exists an (computable) integer constant $d \geq 2$ such that if w is a valid and halting computation of M of length n, then the input string $(w@)^{d^n}$ is accepted by $\mathcal{R}$ with probability $\geq (1-\epsilon)$, and

(2) Any input $x = w_1@w_2@\cdots@w_m@$, where no w_i is a valid halting computation of M, is accepted by $\mathcal{R}$ with probability at most ϵ.

Proof. We begin by recalling some details of the construction given in [CL89]. Given M and ϵ_0, [CL89][4] give the construction of a PFA $\mathcal{R}^M(\epsilon_0)$ which expects the input to be of the form $x = w_1@w_2@\cdots w_m@$, where $w_i \in (\Sigma')^*$. The automaton $\mathcal{R}^M(\epsilon_0)$ tries to check that each w_i is a valid, halting computation of M. Since $\mathcal{R}^M(\epsilon_0)$ has only finitely many states, it cannot reliably check consistency as it requires maintaining counter values. Instead $\mathcal{R}^M(\epsilon_0)$ plays a game on reading a computation w_i, wherein it tosses $O(n)$ coins; here n is $|w_i|$. The game has four possible outcomes.

(1) *reject*, when $\mathcal{R}^M(\epsilon_0)$ discovers an error in w_i,
(2) *double wins*,
(3) *sum wins*, or
(4) *neither wins*

Details of how this game is played are beyond the scope of this paper, and the interested reader is referred to [CL89]. If w_i is a valid halting computation, the following properties are known to hold: (a) *reject* is never an outcome of the game, (b) the probability of outcome *double wins* is equal to the probability of the outcome *single wins*, which we will denote by p in this proof, and (c) $p \geq 2^{-4n}$. The automaton $\mathcal{R}^M(\epsilon_0)$ maintains two counters D and S that take values between 0 and q — q is a constant integer whose value will be fixed later in the next paragraph. After playing the game on w_i, $\mathcal{R}^M(\epsilon_0)$ takes the following steps depending on the outcome of the game. If the outcome is *reject*, then $\mathcal{R}^M(\epsilon_0)$ moves to a special reject state q_r and ignores the rest of the input. If the outcome is *double wins* then counter D is incremented, and if the outcome is *sum wins* then counter S is incremented. When the outcome is *neither wins*, the counters D and S are left unchanged. The automaton $\mathcal{R}^M(\epsilon_0)$ then checks the value of D and S — if either of them are q then it ignores the rest of the input and does not play anymore games; on the other hand if both S and D are less than q then it processes w_{i+1} by playing the game.

After processing the entire input $x = w_1@w_2@\cdots w_m@$, the automaton $\mathcal{R}^M(\epsilon_0)$ decides to accept or reject x as follows.

(1) If $\mathcal{R}^M(\epsilon_0)$ is in the reject state q_r (i.e., one of the games played had outcome *reject*) then x is rejected.
(2) x is also rejected if either (a) both S and D are $< q$, or (b) $D = q$ and $S = 0$.
(3) In all other cases, x is accepted, i.e., when $S = q$ or when $D = q$ and $S \neq 0$.

The constant q is fixed to ensure that the following property holds: Assuming that after processing x at least one of the counters D or S is q, the probability that x is accepted is (a) $> 1 - \epsilon_0$ if all the w_is are valid, halting computations of M, and (b) $\leq \epsilon_0$ if none of the w_is are valid, halting computations.

In proving this lemma, we will take $\mathcal{R}$ to be the PFA obtained by taking $\epsilon_0 = \frac{\epsilon}{2}$, i.e., $\mathcal{R} = \mathcal{R}^M(\frac{\epsilon}{2})$. We will first show that any input $x = w_1@w_2@\cdots@w_m@$, where no w_i is a valid halting computation of M, is accepted by $\mathcal{R}$ with probability at most ϵ. Now, we know by construction of $\mathcal{R}$ and properties stated above, assuming that one of S or D is q, the probability that x is accepted is at most $\frac{\epsilon}{2}$. Moreover, when both S and D are $< q$, we know (by construction) that $\mathcal{R}$ rejects x. Thus the second condition in the lemma holds.

[4]The construction in [CL89] is actually carried out only for deterministic 2-counter machines without an input tape. However, the construction easily carries over to deterministic 2-counter machines with one-way read only input tape.

We will now prove the first condition. Consider input $x = w_1@w_2@\cdots@w_m@$, where all the w_is are valid, halting computations of M of (equal) length n. Taking p_{SD} to be the probability that either D or S is q after processing x, the probability that x is accepted by $\mathcal{R}$ is at least $p_{SD}(1-\frac{\epsilon}{2})$. Thus, to prove the first condition, all we need to show is that if m is larger than d^n, for some fixed, computable d, then $p_{SD}(1-\frac{\epsilon}{2}) > 1-\epsilon$. In other words, we need to prove that for such large m, $p_{SD} > \delta$, where $\delta = (1-\epsilon)/(1-\frac{\epsilon}{2})$.

Let $\overline{p_{SD}} = 1 - p_{SD}$. So $\overline{p_{SD}}$ is the probability that both S and D are less than q after x is processed. Recall that p is the probability that D is incremented after a single computation w_i is processed by $\mathcal{R}$. Moreover, since all the w_is are assumed to be valid computations, p is also the probability that S is incremented after playing one game. Thus, we can say that

$$\overline{p_{SD}} = \sum_{0\le S<q}\sum_{0\le D<q}\binom{m}{S+D}\binom{S+D}{S}p^{S+D}(1-2p)^{m-(S+D)}$$

where $\binom{k}{\ell}$ is the number of ways of choosing ℓ objects from k objects. We can simplify the above expression as follows.

$$\begin{aligned}\textstyle\sum_{0\le S<q}\sum_{0\le D<q} & \binom{m}{S+D}\binom{S+D}{S}p^{S+D}(1-2p)^{m-(S+D)}\\ &\le \binom{2q-2}{q-1}\textstyle\sum_{0\le k\le 2q-2}\min(k,q-1)\binom{m}{k}p^k(1-2p)^{m-k}\\ &= \binom{2q-2}{q-1}\textstyle\sum_{0\le k\le 2q-2}\frac{\min(k,q-1)}{2^k}\binom{m}{k}(2p)^k(1-2p)^{m-k}\\ &\le \binom{2q-2}{q-1}\textstyle\sum_{0\le k\le 2q-2}\binom{m}{k}(2p)^k(1-2p)^{m-k}\end{aligned}$$

In the above reasoning, the second line follows from the observation that since $S+D \le 2q-2$, $\binom{S+D}{S} \le \binom{2q-2}{q-1}$, and the last line follows from the fact that $\frac{\min(k,q-1)}{2^k} \le 1$. Also observe that $\sum_{0\le k\le 2q-2}\binom{m}{k}(2p)^k(1-2p)^{m-k}$ is nothing but the cumulative distribution function for a binomial distribution with parameters $2p$ and m. Taking m such that $2q-2 < m(2p)$, we upper bound the above expression using Chernoff bounds as follows,

$$\overline{p_{SD}} \le \binom{2q-2}{q-1}\sum_{0\le k<2q-2}\binom{m}{k}(2p)^k(1-2p)^{m-k} \le \binom{2q-2}{q-1}\exp\left(-\frac{(2pm-(2q-2))^2}{4pm}\right)$$

Let $\rho = \binom{2q-2}{q-1}$. Now $\rho \cdot \exp(-\frac{(2pm-(2q-2))^2}{4pm}) < 1-\delta$ when $m > \frac{\theta+(2q-2)}{p}$, where $\theta = \log(\frac{\rho}{1-\delta})$. Finally since $p \ge 2^{-4n}$, we get the desired d for the lemma. □