Complete Distributional Problems, Hard Languages, and Resource-Bounded Measure[*]

A. Pavan[†]
Alan L. Selman[‡]
Department of Computer Science
University at Buffalo
Buffalo, NY 14260

October 21, 1997

Abstract

We say that a distribution μ is *reasonable* if there exists a constant $s \geq 0$ such that $\mu(\{x \mid |x| \geq n\}) = \Omega(\frac{1}{n^s})$. We prove the following result, which suggests that all DistNP-complete problems have reasonable distributions.

> If NP contains a DTIME(2^n)-bi-immune set, then every DistNP-complete set has a reasonable distribution.

It follows from work of Mayordomo [May94] that the consequent holds if the p-measure of NP is not zero.

Cai and Selman [CS96] defined a modification and extension of Levin's notion of average polynomial time to arbitrary time-bounds and proved that if L is P-bi-immune, then L is *distributionally hard*, meaning, that for every polynomial-time computable distribution μ, the distributional problem (L, μ) is not polynomial on the μ-average. We prove the following results, which suggest that distributional hardness is closely related to more traditional notions of hardness.

1. If NP contains a distributionally hard set, then NP contains a P-immune set.
2. There exists a language L that is distributionally hard but not P-bi-immune if and only if P contains a set that is immune to all P-printable sets.

The following corollaries follow readily

1. If the p-measure of NP is not zero, then there exists a language L that is distributionally hard but not P-bi-immune.
2. If the p_2-measure of NP is not zero, then there exists a language L in NP that is distributionally hard but not P-bi-immune.

[*] Research supported in part by NSF grant CCR-9400229

[†] aduri@cs.buffalo.edu

[‡] selman@cs.buffalo.edu

1 Introduction

A *distributional problem* is a pair (L,μ), where L is a language over a finite alphabet Σ and μ is a distribution defined on Σ^*. Given a distributional problem, it is an important issue either to find an expected polynomial-time algorithm that solves the problem or to prove that such an algorithm does not exist. Levin [Lev86] provided two central notions for studying this issue. One is analogous to the class P, and provides an *easiness* notion; the other is analogous to the class of NP-complete sets, and provides a *hardness* notion. For the first, Levin defined a robust notion of what it means for an algorithm that accepts L to be polynomial on the μ-average. Using this notion, Average-P denotes the set of all distributional problems (L,μ) such that μ is computable in polynomial time and some algorithm for L is polynomial on the μ-average. Let DistNP denote the collection of all distributional problems (L,μ) such that μ is computable in polynomial time and L belongs to NP. For the second central notion, that of hardness, Levin defined reductions between distributional problems. Using these reducibilities, in the usual manner, we define a distributional problem (L,μ) to be complete for DistNP if (L,μ) belongs to DistNP and every distributional problem in DistNP is reducible to (L,μ). It is not known whether DistNP $\subseteq$ Average-P. If P $=$ NP, then DistNP $\subseteq$ Average-P, and if DistNP $\subseteq$ Average-P, then E $=$ NE [BDCGL92]. Levin showed that distributional tiling with a simple distribution is complete for DistNP, and since then, several additional DistNP-complete problems have been found [BG95, Gur91, VL88, VR92, WB95, Wan95]. However, we do not possess a catalog of natural DistNP-complete problems that is in any way similar to the flood-tide of NP-complete problems. This distinction is one reason that it is important to analyze distributional problems for their potential completeness.

The *standard uniform distribution* on Σ^* is given by $\mu'(x) = \frac{6}{\pi^2}|x|^{-2}2^{-|x|}$. (Given a distribution μ, we let μ' denote the density function on individual strings.) In general, a polynomial-time computable distribution is *uniform* if $\mu'(x) = \rho(|x|)2^{-|x|}$, where $\sum_n \rho(n) = 1$ and there is a polynomial p such that for all n, $\rho(n) \geq 1/p(n)$. Gurevich [Gur91] defined a distribution to be *flat* if there exists a real number $\varepsilon > 0$ such that for all but finitely many x, $\mu'(x) \leq 2^{-|x|^{\varepsilon}}$. Some commonly used distributions on graphs are flat and indeed all uniform distributions are flat. Gurevich proved that no distributional problem with a flat distribution is DistNP-complete unless NEXP $=$ EXP. Assuming that NEXP and EXP are distinct classes, this result asserts that certain natural distributions do not yield complete problems. Thus, one might ask whether the reason that we know only a handful of complete distributional problems is because problems can only be complete when their distributions are unatural.[1] The answer is no. Define a distribution to be *reasonable* if there exists a constant $s > 0$ such that $\mu(\{x \mid |x| \geq n\}) = \Omega(\frac{1}{n^s})$. The reason of course is that distributions that decrease too quickly give too much weight to small instances, and for this reason are unreasonable. The distributions of known DistNP-complete problems, while not uniform, are all reasonable. From our results we will learn that, under generally-accepted complexity-theoretic hypotheses, all DistNP-complete problems have reasonable distributions.

We prove that if NP contains sets that are DTIME(2^n)-bi-immune, then all DistNP-complete problems have reasonable distributions. Therefore, by work of Mayordomo [May94], the consequent follows from the hypothesis that the p-measure of NP is not 0. Thus, we add our results to a growing list of consequences of this hypothesis [May94, LM96].

[1] As a consequence of a result of Wang and Belanger [WB95], for many NP-complete problems A, there is some polynomial-time computable distribution μ so that (A,μ) is DistNP-complete, but the distribution μ in general is not considered to be a natural one *for the problem A*.

Now we will explain another reason for wanting to know that all DistNP-complete problems have reasonable distributions. Cai and Selman [CS96] observed that Levin's definition has limitations when applied to distributional problems with unreasonable distributions and, as extended by Ben David et al. [BDCGL92], when applied to exponential time-bounds. They modified Levin's definition to remove these limitations and, as a consequence of their definition, they obtained a hierarchy theorem for arbitrary average-case time-bounds that is as tight as the Hartmanis-Stearns [HS65] hierarchy theorem for worst-case deterministic time. Consider the class AVP of all distributional problems (L,μ) that are polynomial on the μ-average according to the definition of Cai and Selman, and recall that Average-P denotes Levin's class of distributional problems that are polynomial on the μ-average. (We will provide all formal definitions in the next section.) It is obvious from the definitions that AVP $\subseteq$ Average-P. Cai and Selman showed that $(L,\mu) \in AVP$ if and only if $(L,\mu) \in$ Average-P, for all reasonable distributions μ, but the two definitions differ when applied to distributional problems that have unreasonable distributions. If (L_1,μ_1) is reducible to (L_2,μ_2), both μ_1 and μ_2 are reasonable, and (L_2,μ_2) belongs to AVP, then (L_1,μ_1) belongs to Average-P and so, by the equivalence theorem of Cai and Selman, (L_1,μ_1) belongs to AVP also. (We assume throughout that all distributions are polynomial-time computable.) However, Belanger, Pavan, and Wang [BPW96] have proved that AVP is not in general closed under reductions. They constructed a language L and distributions μ_1 and μ_2 such that μ_2 is reasonable, (L,μ_1) is reducible to (L,μ_2) (by the identity function), $(L,\mu_2) \in$ AVP, and $(L,\mu_1) \notin$ AVP. (Observe as a consequence that μ_1 is not reasonable.) One simple solution is to restrict one's attention to reasonable distributions only. This paper helps to justify this approach, for if μ is a reasonable distribution for every DistNP-complete distributional problem (L,μ), then, for any DistNP-complete problem (L,μ), $(L,\mu) \in$ AVP if and only if DistNP $\subseteq$ AVP. Clearly, this property is important for a meaningful theory of average-case completeness.

Consider now the fundamental question of what it means for a language L to be difficult to recognize. A language that is not in P may still be easy to recognize on many input strings. In contrast, a language that is a.e. complex, or equivalently, P-bi-immune, is difficult to recognize on all but finitely many input strings. Let us say that a language is *distributionally-hard to recognize* if for every polynomial-time computable distribution μ, the distributional problem $(L,\mu) \notin$ AVP; i.e., for every μ, no Turing machine that accepts L has a running-time that is polynomial on the μ-average. Cai and Selman [CS96] proved, as a consequence of their hierarchy theorem, that every P-bi-immune language is distributionally-hard to recognize. Here we prove that there exist languages that are distributionally hard but not P-bi-immune if and only if P contains a set that is immune to all P-printable sets. Also, we show that if NP contains a distributionally hard set, then NP contains a P-immune set. It follows from results of Mayordomo [May94] that if the p-measure of NP is not zero, then there exists a language L that is distributionally hard but not P-bi-immune, and if the p_2-measure of NP is not zero, then there exists a language L in NP that is distributionally hard but not P-bi-immune.

2 Preliminaries

We assume that all languages are subsets of $\Sigma^* = \{0,1\}^*$ and we assume that Σ^* is ordered by standard lexicographic ordering.

A *distribution function* $\mu : \{0,1\}^* \to [0,1]$ is a nondecreasing function from strings to the closed interval $[0,1]$ that converges to one. The corresponding *density function* μ' is defined by $\mu'(0) = \mu(0)$ and $\mu'(x) = \mu(x) - \mu(x-1)$. Clearly, $\mu(x) = \sum_{y \leq x} \mu'(y)$. For any subset of strings S, we will denote by $\mu(S) = \sum_{x \in S} \mu'(x)$, the probability of the event S. Define $u_n = \mu(\{x \mid |x| = n\})$. For each n, let

$\mu'_n(x)$ be the conditional probability of x in $\{x \mid |x| = n\}$. That is, $\mu'_n(x) = \mu'(x)/u_n$, if $u_n > 0$, and $\mu'_n(x) = 0$ for $x \in \{x \mid |x| = n\}$, if $u_n = 0$.

A function μ from Σ^* to $[0,1]$ is *computable in polynomial time* [Ko83] if there is a polynomial time-bounded transducer M such that for every string x and every positive integer n, $|\mu(x) - M(x, 1^n)| < \frac{1}{2^n}$. Consistent with Levin's hypothesis that natural distributions are computable in polynomial time, we restrict our attention *entirely* to such distributions. If μ is computable in polynomial time, then the density function μ' is computable in polynomial time. (The converse is false unless $\mathrm{P} = \mathrm{NP}$ [Gur91].) Also, we explicitly exclude from consideration distributions μ for which $\mu'(x) = 0$ for all but a finite number of strings x. Consideration of such distributions would allow every problem to be an essentially finite problem.

Levin [Lev86] defines a function f from Σ^* to nonnegative reals to be *polynomial on* μ*-average* if there is an integer $k > 0$ such that

$$\sum_{|x| \geq 1} \mu'(x) \frac{(f(x))^{1/k}}{|x|} < \infty. \tag{1}$$

Average-P is the class of distributional problems (L, μ), where L is a language and μ is a polynomial-time computable distribution, such that L can be decided by some Turing machine M whose running time T_M is polynomial on μ-average.

For any time-constructible function T that is monotonically increasing, and hence invertible, Cai and Selman [CS96] define T on the μ-average as follows[2]: Let μ be a distribution on Σ^*, and let $W_n = \mu(\{x : |x| \geq n\})$. A function f is T on the μ-average if for all $n \geq 1$,

$$\sum_{|x| \geq n} \mu'(x) \cdot \frac{T^{-1}(f(x))}{|x|} \leq W_n. \tag{2}$$

Then, $\mathrm{AVTIME}(T(n))$ denotes the class of distributional problems (L, μ), where L is a language and μ is a polynomial-time computable distribution, such that L can be decided by some Turing machine M whose running time T_M is T on the μ-average.

Define $\mathrm{AVP} = \bigcup_{k \geq 1} \mathrm{AVTIME}(n^k)$. Clearly, $\mathrm{AVP} \subseteq$ Average-P.

A distribution μ is *reasonable* if there exists $s > 0$ such that $W_n = \Omega\left(\frac{1}{n^s}\right)$. We will require the following results of Cai and Selman [CS96] and Gurevich [Gur91].

Proposition 1 1. *If* μ *is a reasonable distribution, then* (L, μ) *belongs to* Average-P *(Levin's definition) if and only if* (L, μ) *belongs to* AVP *(Cai and Selman's definition).*

2. *If* μ *satisfies the stronger condition that there exists* $s > 0$ *such that* $u_n = \Omega\left(\frac{1}{n^s}\right)$*, then all of the following are equivalent:*

   (i) (L, μ) *belongs to* Average-P*;*

   (ii) (L, μ) *belongs to* AVP*;*

   (iii) *There is an integer* $k > 0$ *such that for all* $n \geq 1$*,*

$$\sum_{|x| = n} \mu'(x) \frac{(f(x))^{1/k}}{|x|} \leq u_n. \tag{3}$$

[2] Cai and Selman restricted their attention to functions that belong to Hardy's [Har24] class of logarithmico-exponential functions. We do not need to concern ourselves with this for the purpose of this paper.

Now consider reductions. Levin [Lev86] was the first to define polynomial-time many-one reductions on distributional problems; we will use the following form given by Gurevich [Gur91].

Let μ and ν be two distributions. Then, μ is *dominated* by ν, denoted by $\mu \preceq \nu$, if there is a polynomial p such that for all x, $\mu'(x) \leq p(|x|)\nu'(x)$. Let μ_A and μ_B be two distributions and let $f : \Sigma^* \to \Sigma^*$. Recall, for every distribution ν on Σ^*, that f induces a distribution $f(\nu)$ on Σ^* that is defined by $f(\nu)'(y) = \sum_{f(x)=y} \nu'(x)$, for all $y \in range(f)$. Then, we say that μ_A is *dominated by* μ_B *with respect to* f, denoted by $\mu_A \preceq_f \mu_B$, if there exists a distribution ν such that $\mu_A \preceq \nu$ and for all $y \in range(f)$, $\mu'_B(y) = f(\nu)'(y)$.

Let (A,μ_A) and (B,μ_B) be two distributional problems. Then (A,μ_A) is *many–one reducible to* (B,μ_B) *in polynomial time*, denoted by $(A,\mu_A) \leq^p_m (B,\mu_B)$, if there exists a polynomial-time computable function $f : \Sigma^* \to \Sigma^*$ such that A is many–one reducible to B via f and $\mu_A \preceq_f \mu_B$.

Gurevich [Gur91] and Wang [Wan97] provide proofs of the following properties.

Proposition 2

1. *Let (A,μ_A) and (B,μ_B) be two distributional problems such both μ_A and μ_B are polynomial-time computable and such that $(A,\mu_A) \leq^p_m (B,\mu_B)$. If $(B,\mu_B) \in$ Average-P, then $(A,\mu_A) \in$ Average-P.*

2. *Polynomial-time many-one reductions are transitive.*

It is possible to require only that the reduction be computable in polynomial time on the average [Lev86, Gur91]: μ is *weakly dominated* by ν if there is a function g that is polynomial on the μ-average (by Levin's definition) such that for all x, $\mu'(x) \leq g(x)\nu'(x)$. (A,μ_A) is *many–one reducible to* (B,μ_B) *in average polynomial time*, denoted by $(A,\mu_A) \leq^{ap}_m (B,\mu_B)$, if there is a function f that is computable in time a polynomial on the μ_A-average (again, by Levin's definition) such that A is many–one reducible to B via f and μ_A is weakly dominated by some distribution ν such that for all x, $\mu'_B(f(x)) = f(\nu)'(f(x))$.

The analogue of Proposition 2 holds for $\leq^{ap}_m$-reductions.

Once again, if (L_1,μ_1) is reducible to (L_2,μ_2), both μ_1 and μ_2 are reasonable, and (L_2,μ_2) belongs to AVP, then (L_1,μ_1) belongs to Average-P and so, by Proposition 1, (L_1,μ_1) belongs to AVP also. However, Belanger, Pavan, and Wang [BPW96] have proved that AVP is *not* in general closed under reductions.

Given any reducibility $\leq_r$, a distributional problem (L,μ) is $\leq_r$-complete for DistNP if $(L,\mu) \in$ DistNP (i.e., $L \in$ NP and μ is computable in polynomial time) and every distributional problem that belongs to DistNP is $\leq_r$ reducible to (L,μ).

Here we have given only the definitions and properties that we need for this paper; we refer the reader to the recent expositions by Impagliazzo [Imp95] and Wang [Wan97] for deeper understanding of average-case complexity.

2.1 Resource-bounded measure

Let the classes $p_1 = p$ and p_2, both consisting of functions $f : \Sigma^* \to \Sigma^*$, be the classes

$$\begin{aligned} p_1 &= \{f \mid f \text{ is computable in polynomial time}\} \\ p_2 &= \{f \mid f \text{ is computable in } n^{\log n^{O(1)}} \text{ time}\}. \end{aligned}$$

We refer the reader to the papers of Lutz [Lut92, Lut97] for a general introduction to resource-bounded measure theory. Measures are defined in terms of certain capital-preserving betting strategies called *martingales*. Informally, a martingale *succeeds* on a language L if the betting strategy succeeds in winning infinite capital on L. We will not construct martingales in this paper, so we will not define them here. Resource-bounded measures are defined in terms of resource-bounded martingales.

The following definitions are based on these notions: A set X of languages has p_i-*measure* 0 ($i = 1,2$) if there is a p_i-computable martingale that succeeds on every language in X. A set X of languages has p_i-*measure* 1 if the complement of X has p_i-measure 0. A set X has *measure* 0 *in* E if the p-measure of $X \cap \mathrm{E}$ is 0. A set X has *measure* 1 *in* E if the p-measure of the complement of X in E is 0.

We caution that not all sets are measurable. We assume the reader is familiar with standard set-theoretic closure properties of measure theory.

If the p-measure of a class X is 0, then the p_2-measure of X is 0. If the p-measure of X is 0, then the measure of X in E is 0. Lutz has hypothesized that neither the p-measure nor the p_2-measure of NP is 0, and from these strong hypotheses he and others have derived several consequences that do not seem to follow from weaker hypotheses [May94, LM96]. The measure of E in E is 1. The p-measure of P is 0, and we expect that NP is quantitatively different from P. Thus, results of the form “If A, then the p_i-measure of NP is 0” provide evidence that A is false.

A language L is *immune* to a complexity class $\mathcal{C}$, or $\mathcal{C}$-*immune,* if L is infinite and no infinite subset of L belongs to $\mathcal{C}$. A language L is *bi-immune* to a complexity class $\mathcal{C}$, or $\mathcal{C}$-*bi-immune*, if L is infinite, no infinite subset of L belongs to $\mathcal{C}$, and no infinite subset of $\overline{L}$ belongs to $\mathcal{C}$. A language is DTIME($T(n)$)-*complex* if L does not belong to DTIME($T(n)$) almost everywhere; that is, every Turing machine M that accepts L runs in time greater than $T(|x|)$, for all but finitely many words x. Balcázar and Schöning [BS85] proved that for every time-constructible function T, L is DTIME($T(n)$)-complex if and only if L is bi-immune to DTIME($T(n)$).

Mayordomo [May94] proved that the p-measure of the class of DTIME(2^n)-bi-immune sets is not 0, and therefore, if the p-measure of NP is not 0, then NP contains a DTIME(2^n)-bi-immune set. Cai and Selman [CS96] proved, for all P-bi-immune sets L and for all polynomial-time computable distributions μ, that $(L,\mu) \notin$ AVP. Thus, if NP does not have p-measure 0, then there is a language L such that for every polynomial-time computable distribution μ, the distributional problem (L,μ) belongs to DistNP but does not belong to AVP. (Independently, Schuler and Yamakami [SY95] obtained a similar result.)

The set $\{L \mid \exists\mu,(L,\mu) \in \mathrm{AVP}\}$ has has p-measure 0 because it excludes all P-bi-immune sets. However, the set $\{L \mid \exists\mu,(L,\mu) \in \text{Average-P}\}$ has has measure 1 in E because E is a subset. (This is easy to see; for $L \in \mathrm{E}$, take $\mu'(x) = 4^{-|x|}$.) Since the p-measure of P is 0, in terms of resource-bounded measure, AVP is more like a feasible class than Average-P.

3 Complete Distributional Problems

In this section we show that complete distributional problems have reasonable distributions. The Appendix contains proofs that are not given here. We begin with the following lemma.

Lemma 1 *Let μ_1 be the standard uniform distribution, so that $\mu_1(\{x \mid |x| = n\}) = n^{-2}$. Let f be a polynomial-time computable reduction from (A,μ_1) to (B,μ_2), where μ_2 is not reasonable. Then, for* *all $k \geq 1$, there exist infinitely many strings x, such that $|f(x)|^k \leq |x|$.*

Theorem 1 *If $(A,\mu_1) \leq_m^p (B,\mu_2)$, where $B \in \mathrm{NP}$, μ_1 is the standard uniform distribution, and μ_2 is not reasonable, then A is not* DTIME(2^n)*-bi-immune.*

Proof. Let f be a polynomial-time reduction from (A,μ_1) to (B,μ_2) and choose $l \geq 1$ such that $B \in \mathrm{DTIME}(2^{n^l})$.

For all strings x, $x \in A$ if and only if $f(x) \in B$. Membership of $f(x)$ in B can be decided in $2^{|f(x)|^l}$ steps. By Lemma 1, for infinitely many strings x, $|f(x)|^l \leq |x|$. Thus, membership in A of these infinitely many strings can be decided in $2^{|x|}$ steps. Hence, A is not DTIME(2^n)-bi-immune. □

The following corollaries follow immediately.

Corollary 1 *If* NP *contains a* DTIME(2^n)*-bi-immune set, then every $\leq_m^p$-complete distributional problem for* DistNP *has a reasonable distribution.*

Corollary 2 *If the p-measure of* NP *is not* 0*, then every $\leq_m^p$-complete distributional problem for* DistNP *has a reasonable distribution.*

We also obtain these results for $\leq_m^{ap}$-reducibility.

Theorem 2 *If* NP *contains a* DTIME(2^n)*-bi-immune set, then every $\leq_m^{ap}$-complete distributional problem for* DistNP *has a reasonable distribution.*

Corollary 3 *If the p-measure of* NP *is not* 0*, then every $\leq_m^{ap}$-complete distributional problem for* DistNP *has a reasonable distribution.*

Since $\leq_m^p$ is stronger than $\leq_m^{ap}$, Corollaries 1 and 2 also follow from Theorem 2, but Theorem 1 is of independent interest.

4 Distributional Hardness

We define a language L to be *distributionally-hard to recognize* if for all polynomial-time computable distributions μ, $(L,\mu) \notin \mathrm{AVP}$. As we have noted, every P-bi-immune language is distributionally-hard to recognize. We can completely characterize the question of whether there exist any other languages that are distributionally-hard. Recall that set L is P-*printable* if there exists $k \geq 1$ such that all the elements of L up to size n can be printed by a deterministic Turing machine in time $n^k + k$ [HY84, HIS85]. A set A is P-*printable-immune* if no infinite subset of A is P-printable.

Theorem 3 *If* NP *contains a distributionally-hard set, then* NP *contains a P-immune set.*

Theorem 4 *There exist distributionally-hard sets that are not* P*-bi-immune if and only if* P *contains a* P*-printable-immune set.*

Consider the following assertions:

1. NP contains a P-bi-immune set.
2. NP contains a distributionally-hard set.
3. NP contains a P-immune set.
4. P contains a P-printable immune set.
5. There exist distribtionally-hard sets that are not P-bi-immune.

The following corollary summarizes all known relationships among these assertions.

Corollary 4 *Each of the following implications holds:*

$$\begin{aligned} \textit{Assertion 1} &\Rightarrow \textit{Assertion 2} \\ &\Rightarrow \textit{Assertion 3} \\ &\Rightarrow \textit{Assertion 4} \\ &\Leftrightarrow \textit{Assertion 5}\ . \end{aligned}$$

The first implication is due to Cai and Selman [CS96]. For the third implication, let A be an immune set in NP. Since every P-printable set belongs to P, no infinite subset of A is P-printable. Thus, by a result that Allender and Rubinstein [AR88] attribute D. Russo, there exists a set in P with the same property. The remaining implications follow from Theorems 3 and 4.

Corollary 5 *If the p-measure of* NP *is not* 0*, then there is a language L that is distributionally-hard to recognize but not* P*-bi-immune.*

From the presumably stronger hypothesis that the p_2-measure of NP is not 0, we obtain the stronger result that L belongs to NP:

Corollary 6 *If the p_2-measure of* NP *is not* 0*, then there is a language $L \in$* NP *that is distributionally-hard to recognize but not* P*-bi-immune.*

The proof of Theorem 4 from right to left proceeds as follows: Let B be a P-printable-immune set that belongs to P. Let A be any set that is $\mathrm{DTIME}(2^{n^3})$-complex. We define $L = A \cup B$. Note that A and B are not disjoint since A is $\mathrm{DTIME}(2^{n^3})$-bi-immune. Since $B \in \mathrm{P}$, clearly, L is not P-bi-immune. In the Appendix we show that L is distributionally-hard to recognize. The general idea is to suppose that (L, μ) belongs to AVP, for some polynomial-time computable distribution μ, and, from this supposition, demonstrate an infinite P-printable subset of B.

For the proof of Corollary 6, from results of Mayordomo [May94], we know that if the p_2-measure of NP is not 0, then there is a set A in NP that is $\mathrm{DTIME}(2^{n^3})$-bi-immune. The same hypothesis implies that the p-measure of NP is not 0, from which we know that NP contains P-immune sets, so by Corollary 4, there exists P-printable-immune set B that belongs to P. Thus, in this case the set $L = A \cup B$ belongs to NP.

Finally, let us note that Schuler and Yamakami [SY95] considered a notion that in a sense is the opposite of the one we studied here. They examined languages that for all polynomial-time computable distributions are polynomial on the μ-average, and showed that such languages exist that are not in P.

5 Acknowledgments

The proof of Lemma 1 is due to Jin-Yi Cai. Our original assertion of this result was more restrictive and the proof was more complicated. We thank Jin-Yi and the other members of the University at Buffalo's complexity theory research group for their helpful comments. Also, the authors benefited from correspondence with Jack Lutz and D. Sivakumar.

References

[AR88] E. Allender and R. Rubinstein. P-printable sets. *SIAM Journal on Computing*, 17(6):1193–1202, 1988.

[BDCGL92] S. Ben-David, B. Chor, O. Goldreich, and M. Luby. On the theory of average case complexity. *Journal of Computer and System Sciences*, 44(2):193–219, 1992.

[BG95] A. Blass and Y. Gurevich. Matrix transformation is complete for the average case. *SIAM Journal on Computing*, 24:3–29, 1995.

[BPW96] J. Belanger, A. Pavan, and J. Wang. Reductions do not preserve fast convergence rates in average time. *Algorithmica*, to appear.

[BS85] J. Balcázar and U. Schöning. Bi-immune sets for complexity classes. *Mathematical Systems Theory*, 18(1):1–18, June 1985.

[CS96] J-Y. Cai and A. Selman. Fine separation of average time complexity classes. In *Proceedings of the Thirteenth Symposium on Theoretical Aspects of Computer Science*, volume 1046 of *Lecture Notes in Computer Science*, pages 307–318. Springer, Berlin, 1996.

[GHS91] J. Geske, D. Huynh, and J. Seiferas. A note on almost-everywhere-complex sets and separating deterministic-time-complexity classes. *Information and Computing*, 92(1):97–104, 1991.

[Gur91] Y. Gurevich. Average case completeness. *Journal of Computer and System Sciences*, 42:346–398, 1991.

[Har24] G. Hardy. *Orders of Infinity, The 'infinitärcalcül' of Paul du Bois-Reymond*, volume 12 of *Cambridge Tracts in Mathematics and Mathematical Physics.* Cambridge University Press, London, 2nd edition, 1924.

[HIS85] J. Hartmanis, N. Immerman, and V. Sewelson. Sparse sets in NP-P: EXPTIME versus NEXPTIME. *Information and Control*, 65:158–181, 1985.

[HS65] J. Hartmanis and R. Stearns. On the computational complexity of algorithms. *Transactions of the American Mathematical Society*, 117:285–306, 1965.

[HY84] J. Hartmanis and Y. Yesha. Computation times of NP sets of different densities. *Theoretical Computer Science*, 34:17–32, 1984.

[Imp95] R. Impagliazzo. A personal view of average-case complexity. In *Proceedings of the Tenth Annual IEEE Conference on Structure in Complexity Theory*, pages 134–147, 1995.

[Ko83] K. Ko. On the definition of some complexity classes of real numbers *Mathematical Systems Theory*, 16:95–109, 1983.

[Lev86] L. Levin. Average case complete problems. *SIAM Journal on Computing*, 15:285–286, 1986.

[LM96] J. Lutz and E. Mayordomo. Cook versus Karp-Levin: Separating completeness notions if NP is not small. *Theoretical Computer Science*, 164(1–2):141–163, 1996.

[Lut92] J. Lutz. Almost everywhere high nonuniform complexity. *Journal of Computer and System Sciences*, 44:220–258, 1992.

[Lut97] J. Lutz. The quantitative structure of exponential time. In L. Hemaspaandra and A. Selman, editors, *Complexity Theory Retrospective II*, pages 225–254, Springer, New York, 1997.

[May94] E. Mayordomo. Almost every set in exponential time is P-bi-immune. *Theoretical Computer Science*, 136:487–506, 1994.

[SY95] R. Schuler and T. Yamakami. Sets computable in polynomial time on the average. In *Proceedings of the First Annual International Computing and Combinatorics Conference*, volume 959 of *Lecture Notes in Computer Science*, pages 650–661. Springer, Berlin, 1995.

[VL88] R. Venkatesan and L. Levin. Random instances of a graph coloring problem are hard. In *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*, pages 217–222, 1988.

[VR92] R. Venkatesan and S. Rajagopalan. Average case intractability of diophantine and matrix problems. In *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing*, pages 632–642, 1992.

[Wan97] J. Wang. Average-case computational complexity theory. In L. Hemaspaandra and A. Selman, editors, *Complexity Theory Retrospective II*, Springer, New York, 1997.

[Wan95] J. Wang. Average-case completeness of a word problem for groups. In *Proceedings of the Twenty-Seventh ACM Symposium on Theory of Computing*, pages 325–334, 1995.

[WB95] J. Wang and J. Belanger. On the NP-isomorphism problem with respect to random instances. *Journal of Computer and System Sciences*, 50:151–164, 1995.

Appendix

Here we give proofs of results that are not provided in the main body.

Results in Section 3

Lemma 1 *Let μ_1 be the standard uniform distribution, so that $\mu_1(\{x \mid |x| = n\}) = n^{-2}$. Let f be a polynomial-time computable reduction from (A, μ_1) to (B, μ_2), where μ_2 is not reasonable. Then, for all $k \geq 1$, there exist infinitely many strings x, such that $|f(x)|^k \leq |x|$.*

Proof. The function f many-one reduces A to B and $\mu_1 \leq_f \mu_2$. Thus, there exists a distribution ν such that $\mu_1 \preceq \nu$ and for all $y \in range(f)$, $\mu_2'(y) = f(\nu)'(y)$. It is easy to see that ν is reasonable also.

We prove the claim by contradiction. Assume there exist positive integers k and N so that for all strings x, $|x| > N$, $|f(x)|^k > |x|$. We will prove from this assumption that μ_2 is reasonable.

Let $n > N$. Choose s such that $\nu(\{x \mid |x| \geq m\}) = \Omega(m^{-s})$, Consider the following inequalities:

$$\begin{aligned}
\sum_{|z| \geq n^{1/k}} \mu_2'(z) &\geq \sum_{\substack{|z| \geq n^{1/k} \\ z \in f(\Sigma^*)}} \mu_2'(z) \\
&\geq \sum_{\substack{|z| \geq n^{1/k} \\ z \in f(\Sigma^*)}} \sum_{\substack{f(y)=z \\ |y| \geq n}} \nu'(y) \\
&\geq \sum_{|y| \geq n} \nu'(y) \\
&\geq 1/n^s.
\end{aligned}$$

Thus, for all $m \geq N^{1/k}$,

$$\sum_{|z| \geq m} \mu_2'(z) \geq 1/m^{ks},$$

which proves that μ_2 is reasonable. □

Theorem 2 *If* NP *contains a* DTIME(2^n)*-bi-immune set, then every $\leq_m^{ap}$-complete distributional problem for* DistNP *has a reasonable distribution.*

Proof. Let (L, μ) be an $\leq_m^{ap}$-complete distributional problem for DistNP. Define the distribution μ_1 by $\mu_1'(0^n) = n^{-2}$, for all $n \geq 1$, and $\mu_1'(x) = 0$, for all $x \notin \{0\}^*$. For all n, $\mu_1(x \mid |x| = n\}) = n^{-2}$, so, by definition, μ_1 is a reasonable distribution. Let $S \in$ NP; choose $l \geq 1$ such that $S \in$ DTIME$(2^{(n^l)})$. Let f be a function that is computable in time a polynomial on the μ_1-average and that $\leq_m^{ap}$-reduces (S, μ_1) to (L, μ). By Proposition 1, there is a Turing machine M that computes f whose running-time T_M, for some integer $j \geq 1$, satisfies the following inequality, for all $n \geq 1$:

$$\sum_{|x|=n} \frac{T_M(x)^{1/j}}{n} \mu_1'(x) \leq n^{-2}.$$

Thus,

$$\frac{T_M(0^n)^{1/j}}{n} n^{-2} \leq n^{-2},$$

from which it follows that $T_M(0^n) \leq n^j$, for all n. Thus, the restriction of f to $\{0\}^*$ is polynomial-time computable.

Similar to Lemma 1, our first task is to demonstrate that for all $s \geq 1$, there exist infinitely many $n \geq 1$ such that $|f(0^n)|^s \leq n$. Let ν weakly dominate μ_1 so that for all strings $y \in range(f)$, $\mu'(y) = f(\nu)'(y)$. There is a function g that is polynomial on the μ_1-average so that for all x, $\mu_1'(x) \leq g(x)\nu'(x)$. As in the previous paragraph, since μ_1 is reasonable, there exists $j \geq 1$ such that for all $n \geq 1$,

$$\sum_{|x|=n} \frac{g(x)^{1/j}}{n} \mu_1'(x) \leq n^{-2},$$

from which, as above, $g(0^n) \leq n^j$. Then,

$$\begin{aligned} \nu(\{x \mid |x| = n\}) &= \sum_{|x|=n} \nu'(x) \\ &\geq \sum_{|x|=n} \mu_1'/g(x) \\ &\geq (n^{-2})(n^{-j}). \end{aligned}$$

It follows readily that ν is reasonable also. Now the proof of our task proceeds exactly as does the proof of Theorem 1 and we conclude that S is not DTIME(2^n)-bi-immune.

□

Results in Section 4

Theorem 3 *If* NP *contains a distributionally-hard set, then* NP *contains a* P*-immune set.*

Proof. Let $L \in$ NP be distributionally hard. We will show that $L \cap \{0\}^*$ is P-immune.

First we argue that $L \cap \{0\}^*$ is an infinite set. Let us suppose otherwise. Then, $\overline{L}$ contains an infinite subset S of $\{0\}^*$ that belongs to P. For each string x in S, let $r(x)$ be the number of strings in S that are lexicographically less than x. Define a distribution μ on Σ^* as follows: $\mu^{'}(x) = (r(x)+1)^{-2}$, for $x \in S$, and $\mu^{'}(x) = 0$, otherwise. A Turing machine that, on input x, first determines whether $x \in S$, accepts if so, and otherwise simulates an acceptor for L, runs in time a polynomial on the μ-average and accepts L. Thus, (L, μ) belongs to AVP, which contradicts our hypothesis. Thus, $L \cap \{0\}^*$ is an infinite set.

Similarly, $L \cap \{0\}^*$ is P-immune, and, of course, $L \cap \{0\}^* \in$ NP. □

Theorem 4 *There exist distributionally-hard sets that are not* P*-bi-immune if and only if* P *contains a* P*-printable-immune set.*

Proof. Let L be a distributionally-hard set that is not P-bi-immune. Since L is not P-bi-immune, some infinite set S in P either is a subset of L or of $\overline{L}$. Consider the case that $S \subseteq L$. Supposing that S is not P-printable-immune, let S' be an infinite P-printable subset of S. Define μ as follows: For each length n for which S' contains strings of length n, determine the strings $x_1, \ldots, x_{k(n)}$ of length n that belong to S', and define

$$\mu'(x_1) = \cdots = \mu'(x_{k(n)} = \frac{1}{k(n)} \frac{1}{n^2}.$$

All other strings have weight 0. (It follows that $\mu(S') = 1$ and $\mu(\overline{S'}) = 0$.) Define M to be a Turing machine that first behaves like a P-acceptor for S' and then, on words that the P-acceptor does not accept, behaves like a Turing machine that accepts L. Since $S' \subseteq L$, M accepts L, and it is easy to see that T_M is polynomial on the μ-average.

To prove the converse, let $B \in$ P be P-printable-bi-immune. Let A be any set that is DTIME(2^{n^3})-complex. We define $L = A \cup B$. Note that A and B are not disjoint since A is DTIME(2^{n^3})-bi-immune. Since $B \in$ P, clearly, L is not P-bi-immune. Now our goal is to prove that L is distributionally-hard to recognize. Observe that every Turing machine that recognizes L takes more than 2^{n^3} time on all but finitely many strings of $\overline{B}$. Also, recall, for any distribution μ, that $u_n = \mu(\{x \mid |x| = n\})$. We require the following lemma.

Lemma 2 *Suppose that μ is a distribution such that (L,μ) is in* AVP. *Then, there exist infinitely many n such that $u_n \neq 0$ and*

$$\mu(\{x \mid x \in \overline{B}, |x| = n\}) \leq \frac{nu_n}{2^{n^2}}.$$

Proof. We prove the claim by contradiction. Let $X_n = \{x \mid x \in \overline{B}, |x| = n\}$. Let N be a positive integer such that for all $n > N$, $u_n \neq 0$ and

$$\mu(X_n) > \frac{nu_n}{2^{n^2}}.$$

We will prove that (L,μ) is not in AVP. Let M be any Turing machine that accepts L, let T_M denote the running time of M, and assume that N is sufficiently large so that $T_M(x) > 2^{|x|^3}$ for all strings $x \in \overline{B}$, $|x| \geq N$. Let $k \geq 1$ be any positive integer.

The following inequalities demonstrate that (L,μ) does not belong to AVP.

$$\begin{aligned}
\sum_{|x|>N} \frac{T_M^{1/k}(x)\mu'(x)}{|x|} &\geq \sum_{\substack{|x|>N \\ x\in\overline{B}}} \frac{T_M^{1/k}(x)\mu'(x)}{|x|} \\
&\geq \sum_{\substack{m>N \\ u_m\neq 0}} \sum_{\substack{|x|=m \\ x\in\overline{B}}} \frac{T_M^{1/k}(x)\mu'(x)}{|x|} \\
&> \sum_{\substack{m>N \\ u_m\neq 0}} \sum_{\substack{|x|=m \\ x\in\overline{B}}} \frac{(2^{m^3})^{1/k}\mu'(x)}{m} \\
&\geq \sum_{\substack{m>N \\ u_m\neq 0}} \frac{(2^{m^3})^{1/k}\mu(X_m)}{m} \\
&> \sum_{\substack{m>N \\ u_m\neq 0}} \frac{(2^{m^3})^{1/k}}{m} \frac{mu_m}{2^{m^2}} \\
&> \sum_{\substack{m>N \\ u_m\neq 0}} u_m = \sum_{m>N} u_m
\end{aligned}$$

□

Continuing with the proof of Theorem 4, next we show that $(L,\mu) \notin$ AVP, for every polynomial-time computable distribution μ. Again, by contradiction, suppose that μ is a polynomial-time computable distribution such that $(L,\mu) \in$ AVP.

Define an *interval* $[x_1,x_2]$ to be a finite sequence of strings in increasing order that begins with the string x_1 and ends with the string x_2. (If we identify every string with the number it represents in dyadic notation, then lexicographic order of strings and the natural ordering of the positive integers coincide.) For example, the set of all strings of length n is the interval $[0^n,1^n]$. Given strings x_1 and x_2 such that x_1 precedes x_2, let $mid(x_1,x_2) = (x_1+x_2)/2$. Then, $[x_1,mid(x_1,x_2)]$ contains the first $(x_2-x_1+1)/2$ strings in $[x_1,x_2]$, and $[mid(x_1,x_2)+1,x_2]$ contains the last $(x_2-x_1+1)/2$ strings in $[x_1,x_2]$. We will use the following programming variables to simplify notation: Given an interval $I=[x_1,x_2]$, "Left$_I$" denotes the interval $[x_1,mid(x_1,x_2)]$, and "Right$_I$" denotes the interval $[mid(x_1,x_2)+1,x_2]$.

We define a set T to contain at most one string of length n by the following algorithm:

```
Current := $[0^n,1^n]$;
**For** $i=1$ **to** $n$ **do**
    **if** $\mu(\text{Left}_{\text{Current}}) \geq \mu(\text{Right}_{\text{Current}})$
        **then** Current := Left$_{\text{Current}}$ **else** Current := Right$_{\text{Current}}$.
```

The final value of Current contains exactly one string x. Put x into T if and only if $x \in B$.

Next we will prove that T is an infinite P-printable subset of B, which will complete the proof of Theorem 4. Obviously, T is a subset of B. Since μ is computable in polynomial time, $\mu(\text{Left}_{\text{Current}})$ and $\mu(\text{Right}_{\text{Current}})$ can be computed in polynomial time. Thus, T is P-printable.

We need only to show that T is an infinite set. If x is the final value of Current, $|x| = n$, then by the construction, $\mu'(x) \geq u_n/2^n$. However, by Lemma 2, there exist infinitely many n such that $u_n \neq 0$ and $\mu(X_n) \leq \frac{u_n n}{2^{n^2}}$. Thus, for all such n, $\mu'(x)$ is greater than $\mu(X_n)$. Hence, for all such n, the final value of Current belongs to B. Thus, T is an infinite set.

This completes the proof. □