TOWARDS A THEORY OF AVERAGE CASE COMPLEXITY

(A SURVEY)

Oded GOLDREICH
CS Dept., TECHNION, ISRAEL
VISITING ICSI, BERKELEY

MAIN CREDIT

- LEVIN'S PAPER ("AVERAGE CASE COMPLETE PROBLEMS")
- EXPOSITIONS DUE TO JOHNSON
  GUREVICH
  GUREVICH & McCAULEY

MOTIVATION

"down to earth"

AVERAGE COMPLEXITY - PREVIOUS WORK

SPECIFIC PROBLEM

WITH SPECIFIC INPUT DISTRIBUTION

IS EASY ON THE AVERAGE

→ SEE SURVEY BY JOHNSON.

INTRODUCTION TO ANY THEORY OF AVERAGE CASE COMPLEXITY

THE OBJECTS: "DISTRIBUTIONAL PROBLEMS"

i.e. (PROBLEM, DISTRIBUTION)

PROBLEM: IN NP

DISTRIBUTION: "SIMPLE"

REQUIRED:

(1) DEF' OF EASY PROBLEMS.

(2) DEF' OF BROADER CLASS OF INTERESTING PROBLEMS.

(3) NOTION OF REDUCTION which "PRESERVES EASINESS".

(4) RESULTS

- EXISTENCE OF COMPLETE PROBLEMS
- STRUCTURE
- ,

(PART I)

LEVIN'S PAPER

Problems, complete in "average" instance*

Leonid A. Levin

BU, MIT

Many interesting combinatorial problems were found to be NP-complete. Since there is little hope to solve them fast in the worst case, researchers look for algorithms which are fast just "on average". This matter is sensitive to the choice of a particular NP-complete problem and a probability distribution of its instances. Some of these tasks were easy and some not. But one needs a way to distinguish the "difficult on average" problems. Such negative results could not only save "positive" efforts but may also be used in areas (like cryptography) where hardness of some problems is a frequent assumption. A concept of "NP-complete random problems" proposed below may serve this purpose.

Conventions: *A random problem* is a pair (μ, R), where $R \subset N^2$ is an "instance-witness" (or "input-output") relation, and $\mu: N \to [0,1]$ is a probability *distribution function* on inputs (i.e. $\mu(x)$ is the probability of all instances not exceeding x). Its *density* $\mu'(x) = \mu(x) - \mu(x-1)$ is the probability of a particular input. A problem is NP, if both R and μ are computable in time polynomial in length $|x| = \lceil \log x \rceil$ of input. A problem is *polynomial on average* if $\bar{R}(x) = \exists y R(x,y)$ is computable in time $t^k(x)$ where $\sum \mu'(x)t(x)/|x| < \infty$. *Domination* $\mu \lesssim \mu_1$ means $\exists k \forall x\ \mu'(x)/\mu_1(x) < |x|^k$. A polynomial time algorithm f *reduces* a problem (μ_1, R_1) to $(f(\mu_1), R_2)$, if $\mu_1 \lesssim \mu_2$ and $\bar{R}_1(x) = \bar{R}_2(f(x))$. Here $f(\mu)$ is the distribution of outputs of f and maps x to $\sum\{\mu'(y): f(y) \le x\}$. So, if $A(x)$ is a "fast on average" algorithm for $(f(\mu_2), R_1)$ then $A(f(x))$ works at most polynomially slower for (μ_1, R_1). The polynomials $|x|^k$ in domination and in reduction time may be replaced by a "polynomial on average" $t^k(x)$ to get "weak reducibility". The definitions can also be modified for a more elegant "inverting" formulation of NP problems: to actually find y for which $x = r(y)$.

*Supported by NSF grants # MCS-8104211, 8304488.

Definition: *A random NP problem is complete, if any random NP problem is reducible to it.*

Example: Tiling. A tile is a square with a latin letter in every side. Tiles with matching letters can be joined. An instance (u, n, s) of the Tiling Problem, has a list u of tile types considered "legal", a string $n = 0^n$ of 0's, and a string s of matching legal tiles. The problem is to extend s to a square of n^2 matching legal tiles. The joint probability, of u, n and $|s| < n$ is, say, $o(n^{-3})$. Then every tile in s is chosen sequentially with equal probability for all "legal" tiles matching the previous one.

Proposition: *Tiling is an NP-complete random problem.*

Proof: Wlog, we may assume $\mu_1(x) > 1/2x^2$ and (μ_1, R_1) to be computable in time, say, $|x|^3$. If $\mu \gtrsim \mu_1$ then the identity transformation reduces (μ_1, R) to (μ, R). Thus one can round μ_1 to $\mu \gtrsim \mu_1$ so that $\mu(x)$ be the shortest binary rational within $(\mu(x-1), \mu(x+1))$. Then $\log_2 \mu'(x)$ and $m(x) = \mu(x)/\mu'(x)$ are *integers*. This m is computable (and invertible by binary search) in polynomial time and the resulting probability of $x = m(x)$ is $\mu'(m^{-1}(x)) = \mu(x)/m(x) < 1/x$.

Thus $f(x) = 0^{|x|}10^x1 \circ m(x)$ reduces (μ, R) to (λ, U), where U is a universal Turing machine with time bound $|x|^3$; p is its program for which $U(0^{|x|}10^x1x, y) = R(m^{-1}(x), y)$ and $\lambda'(0^n1s) = o(n^{-1})/s$ for $|s| < n$. So, $\mu'(x) < 1/m(x) \lesssim \lambda'(f(x))$. Finally, (λ, U) is reducible to the tiling problem in a standard way: the tiled square corresponds to the space-time history of the Turing computation accepting $U(w, y)$, where w is chosen randomly and y is guessed non-deterministically. A node stores a tape (or head) symbol and the direction to the active center. Q.E.D.

Random NP problems look like "fair games" between suppliers of questions and answers (if both are restricted to a polynomial-time probabilistic machine). So, their "average hardness" seems to be a more "balanced" question then "P = NP?".

The author is grateful to R. Rivest for encouragement and discussion and to D. Johnson for finding an error in a previous version.

DISTRIBUTIONAL PROBLEMS

$$NP \triangleq \left\{ D : \{0,1\}^* \to \{0,1\} \;\middle|\; \begin{array}{l} D(x) = 1 \iff \\ \exists y \; R(x,y) = 1 \end{array} \right\}$$

R: POLY-TIME

DISTRIBUTIONS

$$\mu : \{0,1\}^* \to [0,1] \qquad \text{MONOTONE (INCR)}$$

$$\mu'(x) \triangleq \mu(x) - \mu(x-1) = \text{PROB}(x)$$

$$\mu(x) = \sum_{y \le x} \mu'(y)$$

EXAMPLE: UNIFORM DISTRIBUTION ON STRINGS

$$\mu'(x) = \frac{1}{|x|^2} \cdot 2^{-|x|}$$

DIST-PROBLEM = (PROBLEM, DISTRIBUTION)

EASY ON THE AVERAGE

DEF': (D, μ) IS **POLYNOMIAL-TIME ON THE AVERAGE**

IF $\exists A \; \exists \varepsilon > 0$

$$\sum_x \mu'(x) \cdot \frac{t_A(x)^\varepsilon}{|x|} < \infty$$

Why?

LEVINS' ANSWERS

(1) NATURAL

(2) DOESN'T REALLY MATTER

(3) ANY OTHER DEF'-WHICH-MAKES-SENSE IS EQUIVALENT (OR IMPLIED).

FAILURE OF A STRAIGHTFORWARD DEFINITION OF AVERAGE POLY-TIME.

DEF: (D, μ) IS AVERAGE POLY-TIME IF

$$\exists A \exists c \forall n \quad \sum_{x \in \{0,1\}^n} \mu'_n(x) \cdot t_A(x) \leq n^c$$

$\mu'_n(x)$: $\mathrm{Prob}(X = x \mid X \in \{0,1\}^n)$

REMARK ON DERIVATION OF AVERAGE DEF'

$$\exists c \forall x \quad t_A(x) < |x|^c$$

APPLY AVERAGING OPERATOR

$$E(t_A(x) \mid |x| = n) < n^c$$

- BUT WHY NOT WRITE FIRST

$$\exists c \forall x \quad \frac{t_A(x)^{1/c}}{|x|} < 1$$

FAILURE OF A STRAIGHTFORWARD DEFINITION OF AVERAGE POLY-TIME.

DEF: (D, μ) IS **AVERAGE POLY-TIME** IF

$$\exists A \exists c \forall n \quad \sum_{x \in \{0,1\}^n} \mu_n'(x) \cdot t_A(x) \leq n^c$$

$\mu_n'(x)$: $\text{Prob}(X = x \mid X \in \{0,1\}^n)$

PROBLEMS:

(1) NOT CLOSED UNDER REDUCTIONS. } $t(x) \downarrow t(x)^2$

(2) MACHINE DEPENDENT. }

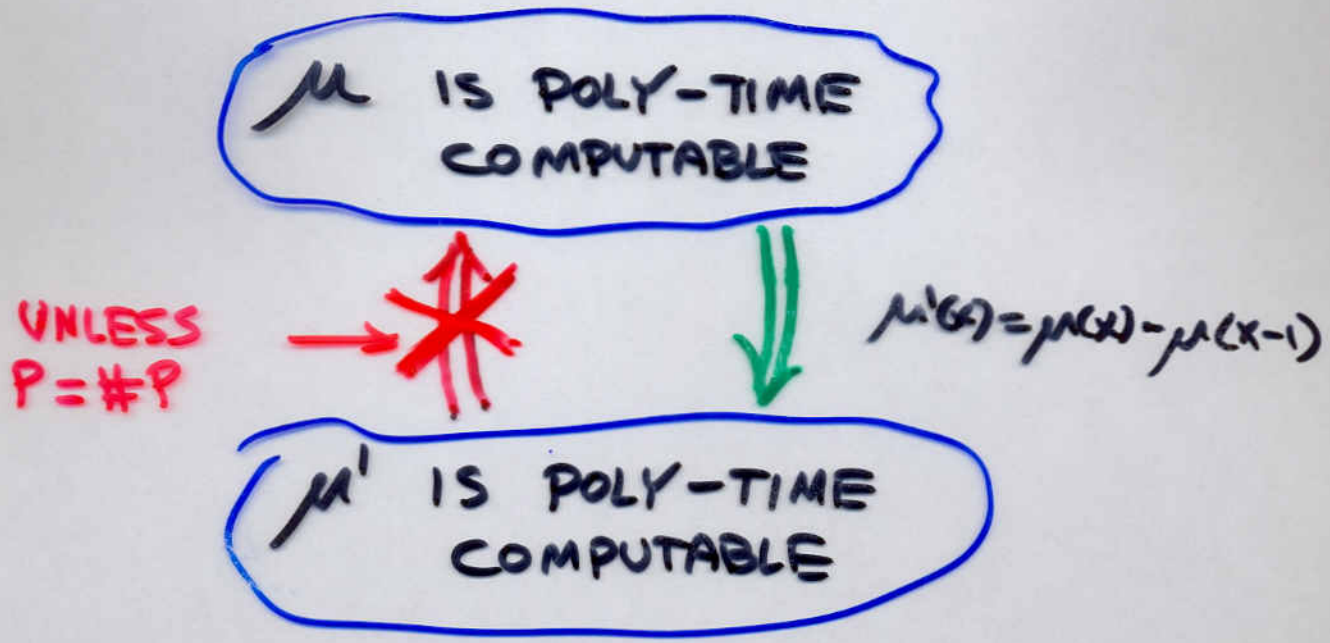
(3) ENCODING DEPENDENT. } (D, μ)

(4) "EQUAL" TREATMENT OF ALL LENGTH. }

A CLASS OF INTERESTING DIST-PROB'

$$\text{DIST-NP} \triangleq \{(D,\mu): \begin{matrix} D \in NP \\ \mu \text{ poly-time computable} \end{matrix}\}$$

REMARK



Reduction between Distributional Problems

Def: (D_1, μ_1) **reduces to** (D_2, μ_2) if there exists f satisfying

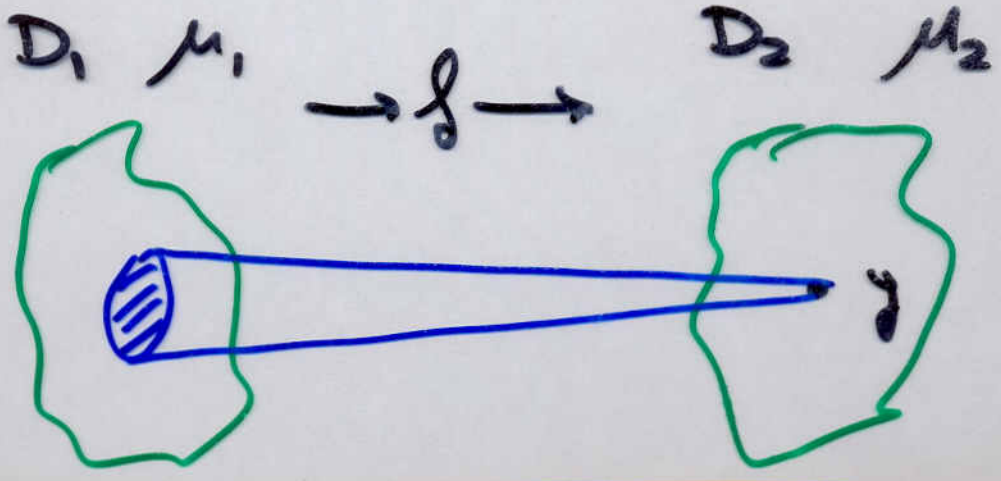
(1) f **is easy:**

f is poly-time computable (on the average w.r.t μ_1).

(2) f **is valid:**

$$\forall x \quad D_2(f(x)) = D_1(x).$$

(3) f **is dominating:**



$$\exists c \forall y \quad \mu_1'(f^{-1}(y)) < |y|^c \cdot \mu_2'(y)$$

Existence of Complete Problems – Comparison to NPCompleteness

Thm: $\exists L \in NPC$

Pf: Use Bounded Halting !

$$BH \triangleq \{(M, x, 1^n) : \exists \text{ computation of } M \text{ on } x \text{ halting in} \leq n \text{ steps}\}$$

M: non-det' TM; 1^n: unary!

$BH \in NP$

$\forall L \in NP, \exists f_L$ reducing L to BH.

$$f_L(x) = (M_L, x, 1^{P_L(|x|)})$$

M_L: non-det' poly TM accepting L

$P_L(|x|)$: the (poly) time bound of M_L

- f_L is poly-time computable.
- $x \in L \iff (M_L, x, 1^{P_L(|x|)}) \in BH$

However $f_L : (L, \mu) \to (BH, ?)$

A COMPLETE PROBLEM FOR DIST-NP

DIST-NP

- DECISION: IS $(M, x, 1^n) \in BH$?
- DISTRIBUTION:

$$\mu_0'(M, x, 1^n) =$$

  - CHOOSE M UNIFORMLY, PROB $= \frac{1}{|M|^2} \cdot 2^{-|M|}$
  - CHOOSE n UNIFORMLY, PROB $= 1/n^2$
  - CHOOSE x IN TWO STEPS

$$k \in_R \{1, 2, \ldots, n\}$$

$$x \in_R \{0,1\}^k$$

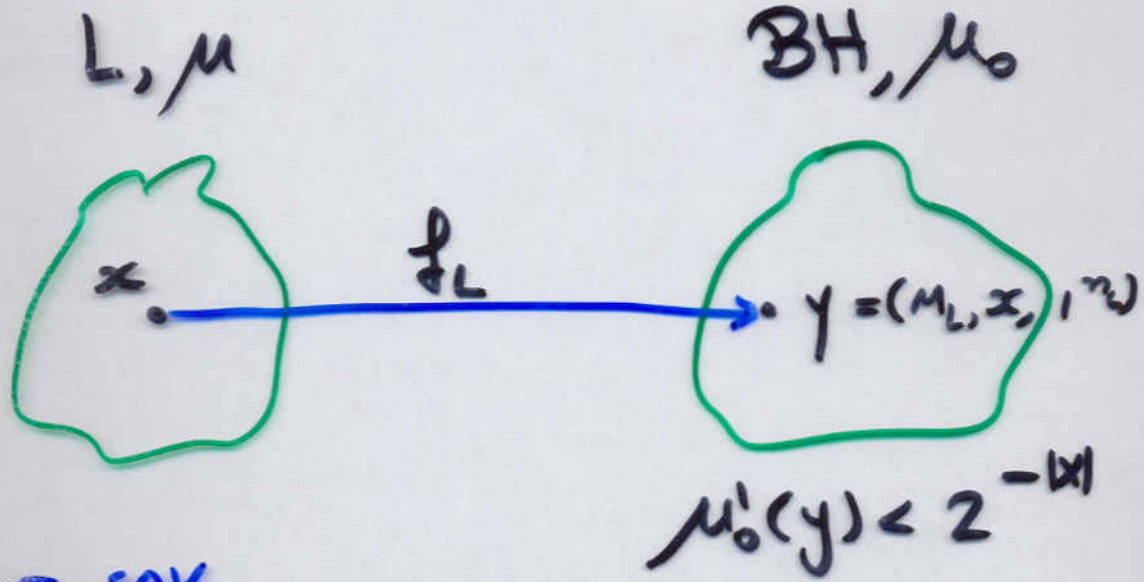
i.e. $\text{PROB}(x) = \frac{1}{n} \cdot 2^{-|x|}$

REMARK: μ_0 IS "NATURAL" BUT NOT UNIFORM

(1^n IS UNIFORM OVER UNARY NOT BINARY !)

COMPLETENESS PROOF – MAIN IDEA

USING f_L (OF NPC PROOF) IS **BAD**!



BUT, SAY

$$\mu'(x) >> 2^{-|x|}$$

THEREFORE, INSTEAD

THE CODING LEMMA

LEMMA: $\forall$ POLY-TIME COMPUTABLE μ

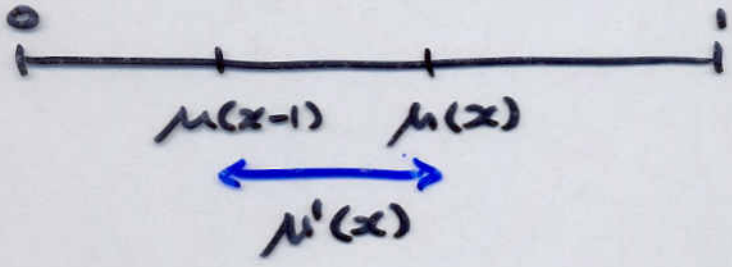
$$\exists\, C_\mu : \{0,1\}^* \to \{0,1\}^* \quad \text{(A CODING)}$$

(1) $\forall x \quad |C_\mu(x)| < \alpha(1) + \log_2 1/\mu'(x)$

(2) C_μ IS POLY-TIME COMPUTABLE

(3) C_μ IS ~~POLY-TIME INVERTABLE~~ 1-1.

PROOF: CONSIDER μ



$\exists$ BINARY FRACTION $\mu(x-1) \le \zeta \le \mu(x)$

WITH (SHORT EXPANSION) $|\zeta| = 1 + \log_2 1/\mu'(x)$

NOW $C_\mu(x) \triangleq \begin{cases} 0x & \text{if } \mu'(x) \le 2^{-|x|} \\ 1\zeta & \text{OTHERWISE} \end{cases}$

(ζ AS ABOVE)

$|C_\mu(x)| \le 2 + \log_2 1/\mu(x)$

THE REDUCTION

$\forall (L,\mu) \in \text{DIST-NP} \quad \exists f_{L,\mu}$

$$f_{L,\mu}(x) = (M'_{L,\mu}, C_\mu(x), 1^{\overbrace{p_L(|x|)+|x|^{O(1)}}^{|x|^{O(1)}}})$$

$$M'_{L,\mu}(y) \triangleq M_L(C_\mu^{-1}(y))$$

VERIFY

(1) $f_{L,\mu}$ IS POLY-TIME COMPUTABLE.

(2) $x \in L$ IFF $(M'_{L,\mu}, C_\mu(x), 1^{|x|^{O(1)}}) \in BH$

(3) DOMINATION

$$\mu'_0(M'_{L,\mu}, C_\mu(x), 1^{|x|^{O(1)}})$$

$$= \underbrace{\text{PROB}(M'_{L,\mu})}_{\text{A CONSTANT}} \cdot \frac{1}{|C_\mu(x)|^2} \cdot \underbrace{2^{-|C_\mu(x)|}}_{\Theta(\mu'(x))} \cdot \underbrace{\frac{1}{|x|^{O(1)}}}_{1/\text{POLY}}$$

$1/\text{POLY}$

Conclusion

Proposition: Dist-BH is complete in Dist-NP (w.r.t. Deter' many-1 reductions).

Other Complete Problems

- "Natural" distributional versions of Bounded Tiling/PCP/Grammer-Problems
- "Unnatural" distributional versions of many/most known NP-Complete problems.

* Open problem:

Show completeness for a "natural" distributional version of a "famous" NPC problem.

PART II

FURTHER DEVELOPMENTS

EVIDENCE TO A GAP BETWEEN RANDOMIZED AND DETERMINISTIC REDUCTIONS

DEF: μ IS CALLED **FLAT** IF $\exists \varepsilon > 0 \; \forall x$

$$\mu'(x) < 2^{-|x|^{\varepsilon}}$$

THM [GUREVICH]: IF $\mathrm{DEXPT} \subsetneq \mathrm{NEXPT}$ THEN

NO DIST-NP WITH FLAT DISTRIBUTION

IS COMPLETE W.R.T DETERMINISTIC REDUCTIONS.

THM [LEVIN]: BH^* WITH UNIFORM DISTRIBUTION

IS COMPLETE W.R.T RANDOMIZED REDUCTIONS.

DOMINATION FOR A RANDOMIZED REDUCTION R

$(R : \pi_1 \to \pi_2)$

$$\exists c \; \forall y \quad \mu_2'(y) \geq \frac{1}{|y|^c} \cdot \sum_x \mu_1'(x) \cdot \mathrm{Prob}(R(x) = y)$$

$BH^* \triangleq \{(M, x, y) : M \text{ halts on } y \text{ in} \leq |y| \text{ steps}\}$

$R(M, x, 1^n) = (M', x', y')$

$R : (BH^* \to \{0,1\}^*)$

EVIDENCE TO A GAP BETWEEN RANDOMIZED AND DETERMINISTIC REDUCTIONS

DEF: μ IS CALLED **FLAT** IF $\exists \varepsilon > 0 \; \forall x$

$$\mu'(x) < 2^{-|x|^{\varepsilon}}$$

EXAMPLE: THE UNIFORM DISTRIBUTION, μ_U, IS FLAT.

RECALL $\mu_U'(x) = \frac{1}{|x|^2} \cdot 2^{-|x|}$

"NEGATIVE-EXAMPLE": μ_0 (OF DIST-BH) IS NOT FLAT.

RECALL $\mu_0'(M,x,1^n) \approx 2^{-|x|} \cdot 1/n^2$

e.g. $n = 2^{|x|}$

$BH^* \triangleq \{(M,x,y):$ M halts on $|y|$ steps $\}$

$R(M,x,1^n) = (M,x,$

$R: (BH, \mu_0) \to (BH^*$

PREVIEW: "ON THE THEORY OF AVERAGE CASE COMPLEXITY"

by BEN-DAVID, CHOR, GOLDREICH, LUBY.

(1) SEARCH vs. DECISION

(2) AVERAGE CASE vs. WORST CASE:

$DEXPT \neq NEXPT \Rightarrow DISTNP \not\subseteq$ EASY ON AVERAGE.

(3) STRUCTURE OF DISTNP UNDER REDUCTIONS:

Eg. $\exists \Pi$ NEITHER COMPLETE NOR EASY.

(4) LACK OF COMPLETENESS RESULTS (?)

(5) NP WITH POLY-TIME SAMPLEABLE DISTRIBUTIONS.

(6) ANALOGOUS RESULTS FOR $\log$SPACE/P-COMPLETE.