# Complexity of distributions and average-case hardness

**Authors:**

Dmitry Itsykson, Alexander Knop, Dmitry Sokolov

**Institute:**

St. Petersburg Department of V.A. Steklov Institute of Mathematics of the Russian Academy of Sciences

# What computer scientists want

### SAT IS EASY ($P = NP$)

There is a plynomial-time algorithm such that for all Boolean formulas it decides correctly if they are satisfiable or not.

# What computer scientists want

### SAT IS EASY ($P = NP$)

There is a plynomial-time algorithm such that for all Boolean formulas it decides correctly if they are satisfiable or not.

### SAT IS HARD ($P \neq NP$)

For any plynomial-time algorithm, there is a Boolean formula such that this algorithm decides if it is satisfiable or not incorrectly.

# What computer scientists really want

**SAT IS EASY**

There is a plynomial-time algorithm such that for almost all Boolean formulas it decides correctly if they are satisfiable or not.

# What computer scientists really want

## SAT IS EASY

There is a plynomial-time algorithm such that for almost all Boolean formulas it decides correctly if they are satisfiable or not.

## SAT IS HARD

For any plynomial-time algorithm there are many Boolean formulas such that this algorithm decides if they are satisfiable or not incorrectly.

# What distribution?

### SAT IS EASY

There is a plynomial-time algorithm such that for almost all Boolean formulas it decides correctly if they are satisfiable or not.

### SAT IS HARD

For any plynomial-time algorithm there are many formulas such that this algorithm decides if they are satisfiable or not incorrectly.

# What distribution?

## SAT IS EASY

There is a plynomial-time algorithm such that for almost all Boolean formulas it decides correctly if they are satisfiable or not.

## SAT IS HARD

For any plynomial-time algorithm there are many formulas such that this algorithm decides if they are satisfiable or not incorrectly.

## LI AND VITANYI, 1992

There is a distribution on strings such that for any language $L$, if $1 - \frac{1}{n^3}$ fraction of $L$ is decidable in polynomial time, then $L$ is decidable in polynomial time.

# Natural classes of distributions

## SAMPLABLE DISTRIBUTIONS

An ensamble of distributions $D$ is samplable in time $f(n)$ iff there is a $f(n)$-time algorithm such that for any $n$ distributions $D_n$ and $A(1^n)$ are equally distributed.

# Natural classes of distributions

## SAMPLABLE DISTRIBUTIONS

An ensamble of distributions $D$ is samplable in time $f(n)$ iff there is a $f(n)$-time algorithm such that for any $n$ distributions $D_n$ and $A(1^n)$ are equally distributed.

## COMPUTABLE DISTRIBUTIONS

An ensamble of distributions $D$ is samplable in time $f(n)$ iff there is a $f(n)$-time algorithm such that for any $n$, function $x \to A(1^n, x)$ is a cumulative distribution function of $D_n$.

# Classes of computations

## HEURISTICALLY DECIDABLE IN POLYNOMIAL TIME

Let $L$ is a language and $D$ is an ensemble of distributions. We call distributional problem $(L, D)$ heuristically decidable in polynomial time with error $\epsilon(n)$ ($(L, D) \in \mathrm{Heur}_{\epsilon(n)}\mathbf{P}$) iff there is a polynomial time algorithm $A$ such that $\Pr_{x \leftarrow D_n}[A(x) \neq L(x)] \leq \epsilon(n)$.

# Complexity of distribution

## FOLKLORE

For any $k > 0$ and $\delta$ there is a language $L$ such that $(L, U) \in \mathrm{Heur}_{\delta}\mathbf{P}$ and for any $R$ holds $(L, R) \notin \mathrm{Heur}_{1-\delta}\mathbf{DTime}(n^k)$.

# Complexity of distribution

### FOLKLORE

For any $k > 0$ and $\delta$ there is a language $L$ such that $(L, U) \in \mathrm{Heur}_\delta\mathbf{P}$ and for any $R$ holds $(L, R) \notin \mathrm{Heur}_{1-\delta}\mathbf{DTime}(n^k)$.

### ITSYKSON, K, AND SOKOLOV, 2015

For any $k > 0$ and $\delta$ there is a language $L$ such that $(L, U) \in \mathrm{Heur}_\delta\mathbf{BPP}$ and for any $R \in \mathbf{DSamp}(n^k)$ holds $(L, R) \notin \mathrm{Heur}_{\frac{1}{2}-\delta}\mathbf{BPTime}(n^k)$.

# Complexity of distribution

## FOLKLORE

For any $k > 0$ and $\delta$ there is a language $L$ such that $(L, U) \in \mathrm{Heur}_\delta\mathbf{P}$ and for any $R$ holds $(L, R) \notin \mathrm{Heur}_{1-\delta}\mathbf{DTime}(n^k)$.

## ITSYKSON, K, AND SOKOLOV, 2015

For any $k > 0$ and $\delta$ there is a language $L$ such that $(L, U) \in \mathrm{Heur}_\delta\mathbf{BPP}$ and for any $R \in \mathbf{DSamp}(n^k)$ holds $(L, R) \notin \mathrm{Heur}_{\frac{1}{2}-\delta}\mathbf{BPTime}(n^k)$.

## DUAL QUESTION?

Is there a distribution $D \in \mathbf{PSamp}$ and a language $L$ such that $(L, D) \notin \mathrm{Heur}_{1-\delta}\mathbf{P}$ and for any $R \in \mathbf{DSamp}(n^k)$ holds $(L, R) \in \mathrm{Heur}_\delta\mathbf{P}$.

# Known results

## DUAL QUESTION?

Is there a distribution $D \in \mathbf{PSamp}$ and a language $L$ such that $(L, D) \notin \mathrm{Heur}_{1-\delta}\mathbf{P}$ and for any $R \in \mathbf{DSamp}(n^k)$ holds $(L, R) \in \mathrm{Heur}_{\delta}\mathbf{P}$.

# Known results

### DUAL QUESTION?

Is there a distribution $D \in \mathbf{PSamp}$ and a language $L$ such that $(L, D) \notin \mathrm{Heur}_{1-\delta}\mathbf{P}$ and for any $R \in \mathbf{DSamp}(n^k)$ holds $(L, R) \in \mathrm{Heur}_{\delta}\mathbf{P}$.

### GUREVICH AND SHELAH, 1987

Let HP denote the language of Hamiltonian graphs. Then $(\mathrm{HP}, U) \in \mathrm{Heur}_{\frac{1}{2^{O(\sqrt{n})}}}\mathbf{DTime}(n)$.

# Known results

## DUAL QUESTION?

Is there a distribution $D \in \mathbf{PSamp}$ and a language $L$ such that $(L, D) \notin \mathrm{Heur}_{1-\delta}\mathbf{P}$ and for any $R \in \mathbf{DSamp}(n^k)$ holds $(L, R) \in \mathrm{Heur}_{\delta}\mathbf{P}$.

## GUREVICH AND SHELAH, 1987

Let HP denote the language of Hamiltonian graphs. Then $(\mathrm{HP}, U) \in \mathrm{Heur}_{\frac{1}{2^{O(\sqrt{n})}}}\mathbf{DTime}(n)$.

## BABAI, ERDOS, AND SELKOW, 1980

Let GI denote the language of pairs of isomorphic graphs. Then $(\mathrm{GI}, U) \in \mathrm{Heur}_{\frac{1}{\sqrt[7]{n}}}\mathbf{DTime}(n)$.

# Equal quesition

## STATISTICAL DISTANCE

A statistical distance between $D_n$ and $R_n$ is
$\Delta(D_n, R_n) = \max_{S \subseteq \{0,1\}^n} |D_n(S) - R_n(S)|$.

# Equal quesition

## STATISTICAL DISTANCE

A statistical distance between $D_n$ and $R_n$ is
$\Delta(D_n, R_n) = \max_{S \subseteq \{0,1\}^n} |D_n(S) - R_n(S)|$.

## EQUIVALENT RESTATEMENT

For any $k$ the following two statements are equal:

- There is a function $\delta(n) \to 0$, a distribution $D \in \mathbf{PSamp}$, and a language $L$ such that $(L, D) \notin \mathrm{Heur}_{1-\delta}\mathbf{P}$ and $(L, R) \in \mathrm{Heur}_{\delta}\mathbf{P}$ for any $R \in \mathbf{DSamp}(n^k)$.
- There is a function $\delta(n) \to 0$, a distribution $D \in \mathbf{PSamp}$ such that for any $R \in \mathbf{DSamp}(n^k)$ the statistical distance between $R$ and $D$ is at least $1 - \delta(n)$.

# Hierarchies for distributions

**WATSON, 2013**

For any constant $k$ and $\epsilon > 0$ there is a distribution $D \in \textbf{PSamp}$ such that $\Delta(D, R) \geq 1 - \frac{1}{k} + \epsilon$ for any $R \in \textbf{DSamp}(n^k)$.

# Hierarchies for distributions

### WATSON, 2013

For any constant $k$ and $\epsilon > 0$ there is a distribution $D \in \mathbf{PSamp}$ such that $\Delta(D, R) \geq 1 - \frac{1}{k} + \epsilon$ for any $R \in \mathbf{DSamp}(n^k)$.

### ITSYKSON, K, SOKOLOV, 2016

For any constant $c$ and $\epsilon > 0$ there is a distribution $D \in \mathbf{DSamp}(n^{\log^c(n)})$ such that $\Delta(D, R) \geq 1 - \lambda(n)$ for any $R \in \mathbf{PSamp}$ where $\lambda(n) \to 0$.

# Hierarchies for distributions

### ITSYKSON, K, SOKOLOV, 2016

For any constant $c$ and $\epsilon > 0$ there is a distribution $D \in \textbf{DSamp}(n^{\log^c(n)})$ such that $\Delta(D, R) \geq 1 - \lambda(n)$ for any $R \in \textbf{PSamp}$ where $\lambda(n) \to 0$.

### ITSYKSON, K, SOKOLOV, 2016

There is a function $\delta(n) \to 0$, a distribution $D \in \textbf{DSamp}(n^{\log^c(n)})$, and a language $L$ such that $(L, D) \notin \mathrm{Heur}_{1-\delta}\mathbf{P}$ and $(L, R) \in \mathrm{Heur}_{1-\delta}\mathbf{P}$ for any $R \in \textbf{PSamp}$ .

# Weak hardness

## OPEN QUESTION

Is there a distribution $D \in \mathbf{PSamp}$ and a language $L$ such that $(L, D) \notin \mathrm{Heur}_{1-\delta}\mathbf{P}$ and $(L, R) \in \mathrm{Heur}_{\delta}\mathbf{P}$ for any $R \in \mathbf{DSamp}(n^k)$?

# Weak hardness

## OPEN QUESTION

Is there a distribution $D \in \mathbf{PSamp}$ and a language $L$ such that $(L, D) \notin \mathrm{Heur}_{1-\delta}\mathbf{P}$ and $(L, R) \in \mathrm{Heur}_{\delta}\mathbf{P}$ for any $R \in \mathbf{DSamp}(n^k)$?

## ITSYKSON, K, SOKOLOV, 2016

For all $a$ and $b$ there is a distribution $D \in \mathbf{PSamp}$ and a language $L$ such that $(L, D) \notin \mathrm{Heur}_{\frac{1}{n^a}}\mathbf{P}$ and for any $R \in \mathbf{DSamp}(n^b)$ there is a constant $c > 0$ such that $(L, R) \in \mathrm{Heur}_{\frac{c}{n^a}}\mathbf{P}$.

# Computable distributions

**ITSYKSON, K, SOKOLOV, 2016**

For all $a$ there is a distribution $D \in \mathbf{PComp}$ and a language $L$ such that $(L, D) \notin \mathrm{Heur}_{1-\frac{1}{2^{n-1}}}\mathbf{P}$ $(L, R) \in \mathrm{Heur}_{O(\frac{1}{2^n})}\mathbf{P}$ for any $R \in \mathbf{DComp}(n^k)$ holds.

## Summary

1. For all $a$ there is a distribution $D \in \mathbf{DSamp}(n^{\log^a(n)})$, a language $L$, and a monotone function $\lambda(n)$ such that $(L, D) \notin \mathrm{Heur}_{1-\lambda(n)}\mathbf{P}$, $(L, R) \in \mathrm{Heur}_{\lambda(n)}\mathbf{P}$ for any $R \in \mathbf{PSamp}$, and $\lambda(n) \to 0$.

## Summary

1. For all $a$ there is a distribution $D \in \mathbf{DSamp}(n^{\log^a(n)})$, a language $L$, and a monotone function $\lambda(n)$ such that $(L, D) \notin \mathrm{Heur}_{1-\lambda(n)}\mathbf{P}$, $(L, R) \in \mathrm{Heur}_{\lambda(n)}\mathbf{P}$ for any $R \in \mathbf{PSamp}$, and $\lambda(n) \to 0$.
2. For all $a$ and $b$ there is a distribution $D \in \mathbf{PSamp}$ and a language $L$ such that $(L, D) \notin \mathrm{Heur}_{\frac{1}{n^a}}\mathbf{P}$ and for any $R \in \mathbf{DSamp}(n^b)$ there is a constant $c > 0$ such that $(L, R) \in \mathrm{Heur}_{\frac{c}{n^a}}\mathbf{P}$.

## Summary

1. For all $a$ there is a distribution $D \in \mathbf{DSamp}(n^{\log^a(n)})$, a language $L$, and a monotone function $\lambda(n)$ such that $(L, D) \notin \mathrm{Heur}_{1-\lambda(n)}\mathbf{P}$, $(L, R) \in \mathrm{Heur}_{\lambda(n)}\mathbf{P}$ for any $R \in \mathbf{PSamp}$, and $\lambda(n) \to 0$.
2. For all $a$ and $b$ there is a distribution $D \in \mathbf{PSamp}$ and a language $L$ such that $(L, D) \notin \mathrm{Heur}_{\frac{1}{n^a}}\mathbf{P}$ and for any $R \in \mathbf{DSamp}(n^b)$ there is a constant $c > 0$ such that $(L, R) \in \mathrm{Heur}_{\frac{c}{n^a}}\mathbf{P}$.
3. For all $a$ there is a distribution $D \in \mathbf{PComp}$ and a language $L$ such that $(L, D) \notin \mathrm{Heur}_{1-\frac{1}{2^{n-1}}}\mathbf{P}$ and for any $R \in \mathbf{DComp}(n^a)$ there is a constant $c > 0$ such that $(L, R) \in \mathrm{Heur}_{\frac{c}{2^n}}\mathbf{P}$.