Stochasticity in Algorithmic Statistics for Polynomial Time

Alexey Milovanov, Nikolay Vereshchagin

National Research University Higher School of Economics

CCC 2017, Riga

Algorithmic Statistics

A black box that samples from an unknown probability distribution $\longrightarrow$

Algorithmic Statistics

A black box that samples from an unknown probability distribution

$$\longrightarrow x = \underbrace{1000010101\ldots1}_{n}$$

Algorithmic Statistics

A black box that samples from an unknown probability distribution $\longrightarrow x = \underbrace{1000010101\ldots1}_{n}$

A general question:

Given the black box's output x and a distribution μ, is it plausible that the black box samples from μ?

Algorithmic Statistics

A black box that samples from an unknown probability distribution $\longrightarrow x = \underbrace{1000010101\ldots1}_{n}$

A general question:

Given the black box's output x and a distribution μ, is it plausible that the black box samples from μ?

Example:

Let $x = 101100101110100101010000101100101110100101010000$ and let μ be the uniform distribution over strings of length $n = |x|$.
Is it plausible that the black box samples from μ?

$$\boxed{\text{A black box that samples from an unknown probability distribution}} \longrightarrow x = \underbrace{1000010101\ldots 1}_{n}$$

A general question:

Given the black box's output x and a distribution μ, is it plausible that the black box samples from μ?

Example:

Let $x = 101100101110100101010000101100101110100101010000$ and let μ be the uniform distribution over strings of length $n = |x|$. Is it plausible that the black box samples from μ?

An answer:

No, since x is a square ($x = uu$) and the probability of being a square is negligible ($2^{-n/2}$).

Algorithmic Statistics with no time bounds

Definition (Kolmogorov)

A probability distribution μ is an acceptable explanation for x if *the randomness deficiency* of x wrt μ,

$$-\log_2 \mu(x) - C(x|\mu)$$

is negligible.

Algorithmic Statistics with no time bounds

Definition (Kolmogorov)

A probability distribution μ is an acceptable explanation for x if *the randomness deficiency* of x wrt μ,

$$-\log_2 \mu(x) - C(x|\mu)$$

is negligible.

Majority Principle: for all μ, if x is sampled from μ, then the probability of having

$$-\log_2 \mu(x) - C(x|\mu) > \beta$$

is less than $2^{-\beta}$.

Algorithmic Statistics with no time bounds

Definition (Kolmogorov)

A probability distribution μ is an acceptable explanation for x if *the randomness deficiency* of x wrt μ,

$$-\log_2 \mu(x) - C(x|\mu)$$

is negligible.

Majority Principle: for all μ, if x is sampled from μ, then the probability of having

$$-\log_2 \mu(x) - C(x|\mu) > \beta$$

is less than $2^{-\beta}$.

Proposition

$-\log \mu(x) - C(x|\mu)$ is large if and only if there is a simple $T \ni x$ (that is, T is enumerated by a short program) with negligible $\mu(T)$.

Algorithmic Statistics with no time bounds

Back to our example:

$x = 101100101110100101010000101100101110100101010000$

Back to our example:

$x = 101100101110100101010000101100101110100101010000$

- If μ is the uniform distribution over n-bit strings, then $-\log \mu(x) - C(x|\mu) \approx n - n/2 = n/2$;

Algorithmic Statistics with no time bounds

Back to our example:

$x = 101100101110100101010000101100101110100101010000$

- If μ is the uniform distribution over n-bit strings, then $-\log \mu(x) - C(x|\mu) \approx n - n/2 = n/2$;
- If μ is the uniform distribution over all n-bit squares, then $-\log \mu(x) - C(x|\mu) \approx n/2 - n/2 = 0$.

Algorithmic Statistics with no time bounds

Back to our example:

$x = 1011001011101001010100001011001011101001010100000$

- If μ is the uniform distribution over n-bit strings, then $-\log \mu(x) - C(x|\mu) \approx n - n/2 = n/2$;
- If μ is the uniform distribution over all n-bit squares, then $-\log \mu(x) - C(x|\mu) \approx n/2 - n/2 = 0$.

Another example:

Let x be an arbitrary n-bit string,
let μ be concentrated on x, i.e., $\mu(x) = 1$.
Then μ is acceptable for x, since $-\log \mu(x) - C(x|\mu) \approx 0 - 0 = 0$.

Back to our example:

$x = 1011001011101001010100001011001011101001010100000$

- If μ is the uniform distribution over n-bit strings, then $-\log \mu(x) - C(x|\mu) \approx n - n/2 = n/2$;
- If μ is the uniform distribution over all n-bit squares, then $-\log \mu(x) - C(x|\mu) \approx n/2 - n/2 = 0$.

Another example:

Let x be an arbitrary n-bit string,
let μ be concentrated on x, i.e., $\mu(x) = 1$.
Then μ is acceptable for x, since $-\log \mu(x) - C(x|\mu) \approx 0 - 0 = 0$.

The goal:

given x, find a simple ($C(\mu) \approx 0$) acceptable explanation μ for x.

Algorithmic Statistics with no time bounds

Back to our example:

$x = 10110010111010010101000010110010111010010101010000$

- If μ is the uniform distribution over n-bit strings, then $-\log \mu(x) - C(x|\mu) \approx n - n/2 = n/2$;
- If μ is the uniform distribution over all n-bit squares, then $-\log \mu(x) - C(x|\mu) \approx n/2 - n/2 = 0$.

Another example:

Let x be an arbitrary n-bit string,
let μ be concentrated on x, i.e., $\mu(x) = 1$.
Then μ is acceptable for x, since $-\log \mu(x) - C(x|\mu) \approx 0 - 0 = 0$.

The goal:

given x, find a simple $(C(\mu) \approx 0)$ acceptable explanation μ for x.

Theorem (A. Shen 1983)

This goal is not always achievable (there are non-stochastic *strings).*

Algorithmic Statistics with time bounds: acceptable explanations

Now we care about computation time!

Algorithmic Statistics with time bounds: acceptable explanations

Now we care about computation time!

Question: How do we define acceptable explanations? Why not say that time-bounded version of Kolmogorov's randomness deficiency $-\log \mu(x) - C^t(x|\mu)$ is small?

Algorithmic Statistics with time bounds: acceptable explanations

Now we care about computation time!

Question: How do we define acceptable explanations? Why not say that time-bounded version of Kolmogorov's randomness deficiency $-\log \mu(x) - C^t(x|\mu)$ is small?

Answer: For polynomial time bounded computations, we cannot prove that randomness deficiency is small if and only if there is no simple refutation set. We will define acceptability using refutation sets.

Algorithmic Statistics with time bounds: acceptable explanations

Now we care about computation time!

Question: How do we define acceptable explanations? Why not say that time-bounded version of Kolmogorov's randomness deficiency $-\log \mu(x) - C^t(x|\mu)$ is small?

Answer: For polynomial time bounded computations, we cannot prove that randomness deficiency is small if and only if there is no simple refutation set. We will define acceptability using refutation sets.

Back to our example:
$x = 1011001011101001010100001011001011101001010100 00$,
μ is the uniform distribution over strings of length $n = |x|$.
We refute μ, since x falls into a simple set $T \ni x$ having negligible $\mu(T)$. Notice that T can be recognized by a short program in a short (polynomial) time.

Algorithmic Statistics with time bounds: acceptable explanations

Definition (informal)

μ is an *acceptable explanation* for x if there is no $T \ni x$ with negligible $\mu(T)$ which is recognizable by a short program in a short time.

Algorithmic Statistics with time bounds: acceptable explanations

Definition (informal)

μ is an *acceptable explanation* for x if there is no $T \ni x$ with negligible $\mu(T)$ which is recognizable by a short program in a short time.

Definition (formal)

μ is a (t, α, ε)-*acceptable explanation* for x if for all $T \ni x$ with $CD^t(T) < \alpha$, we have $\mu(T) \geqslant \varepsilon$.

Algorithmic Statistics with time bounds: acceptable explanations

Definition (informal)

μ is an *acceptable explanation* for x if there is no $T \ni x$ with negligible $\mu(T)$ which is recognizable by a short program in a short time.

Definition (formal)

μ is a (t, α, ε)*-acceptable explanation* for x if for all $T \ni x$ with $CD^t(T) < \alpha$, we have $\mu(T) \geqslant \varepsilon$.

Majority principle: if $\varepsilon \ll 2^{-\alpha}$, then the μ-probability of the event

μ is not (t, α, ε)-acceptable explanation for x

is negligible (the probability of this event is smaller than $\varepsilon 2^{\alpha}$).

Simple explanations

Example

x is an arbitrary string and μ is concentrated on x.
Then μ is $(*, *, 1)$-acceptable for x.

Simple explanations

Example

x is an arbitrary string and μ is concentrated on x.
Then μ is $(*, *, 1)$-acceptable for x.

Goal: Given x find a simple acceptable explanation for x.

Simple explanations

Example

x is an arbitrary string and μ is concentrated on x.
Then μ is $(*, *, 1)$-acceptable for x.

Goal: Given x find a simple acceptable explanation for x.

Definition (informal)

A distribution μ is simple if there is a fast sampler with a short program for μ.

Simple explanations

Example

x is an arbitrary string and μ is concentrated on x.
Then μ is $(*, *, 1)$-acceptable for x.

Goal: Given x find a simple acceptable explanation for x.

Definition (informal)

A distribution μ is simple if there is a fast sampler with a short program for μ.

Definition (formal)

μ is (t', α')-simple if there is a sampler for μ with program of length less than α' and running time less than t'.

Simple explanations

Example

x is an arbitrary string and μ is concentrated on x.
Then μ is $(*, *, 1)$-acceptable for x.

Goal: Given x find a simple acceptable explanation for x.

Definition (informal)

A distribution μ is simple if there is a fast sampler with a short program for μ.

Definition (formal)

μ is (t', α')-simple if there is a sampler for μ with program of length less than α' and running time less than t'.

Remark

For one result we will need that μ be computed rather than sampled in a short time.

The main result

We consider (t', α')-simple (t, α, ε)-acceptable explanations where

The main result

We consider (t', α')-simple (t, α, ε)-acceptable explanations where α', α are $O(\log n)$,

The main result

We consider (t', α')-simple (t, α, ε)-acceptable explanations where
α', α are $O(\log n)$,
$t', t, 1/\varepsilon$ are polynomial in n, and

The main result

We consider (t', α')-simple (t, α, ε)-acceptable explanations where
α', α are $O(\log n)$,
$t', t, 1/\varepsilon$ are polynomial in n, and
$\alpha \gg \alpha'$, $t \gg t'$, and $\varepsilon \ll 2^{-\alpha}$.

The main result

We consider (t', α')-simple (t, α, ε)-acceptable explanations where
α', α are $O(\log n)$,
$t', t, 1/\varepsilon$ are polynomial in n, and
$\alpha \gg \alpha'$, $t \gg t'$, and $\varepsilon \ll 2^{-\alpha}$.

Conjecture (informal)

There are strings x that have no simple acceptable explanations (*non-stochastic* strings).

The main result

We consider (t', α')-simple (t, α, ε)-acceptable explanations where
α', α are $O(\log n)$,
$t', t, 1/\varepsilon$ are polynomial in n, and
$\alpha \gg \alpha'$, $t \gg t'$, and $\varepsilon \ll 2^{-\alpha}$.

Conjecture (informal)

There are strings x that have no simple acceptable explanations (*non-stochastic* strings).

Conjecture (formal)

For all c there is d such that for infinitely many n there is a n-bit string x without $(n^c, c \log n)$-simple $(n^d, d \log n, n^{-c})$-acceptable explanations.

The main result

We consider (t', α')-simple (t, α, ε)-acceptable explanations where
α', α are $O(\log n)$,
$t', t, 1/\varepsilon$ are polynomial in n, and
$\alpha \gg \alpha'$, $t \gg t'$, and $\varepsilon \ll 2^{-\alpha}$.

Conjecture (informal)

There are strings x that have no simple acceptable explanations (*non-stochastic* strings).

Conjecture (formal)

For all c there is d such that for infinitely many n there is a n-bit string x without $(n^c, c \log n)$-simple $(n^d, d \log n, n^{-c})$-acceptable explanations.

Theorem

If $NE \neq RE$, then the Conjecture holds (and, moreover, the Conjecture holds for a constant d, which does not depend on c).

Other results

Conjecture (formal)

For all c there is d such that for infinitely many n there is a n-bit string x without $(n^c, c \log n)$-simple $(n^d, d \log n, n^{-c})$-acceptable explanations.

Theorem

If $NE \neq RE$, then the Conjecture holds (and, moreover, the Conjecture holds for a constant d, which does not depend on c).

Other results

Conjecture (formal)

For all c there is d such that for infinitely many n there is a n-bit string x without $(n^c, c \log n)$-simple $(n^d, d \log n, n^{-c})$-acceptable explanations.

Theorem

If $NE \neq RE$, then the Conjecture holds (and, moreover, the Conjecture holds for a constant d, which does not depend on c).

Theorem

If the Conjecture holds for a constant d (not depending on c), then $P \neq PSPACE$.

Other results

Conjecture (formal)

For all c there is d such that for infinitely many n there is a n-bit string x without $(n^c, c\log n)$-simple $(n^d, d\log n, n^{-c})$-acceptable explanations.

Theorem

If $NE \neq RE$, then the Conjecture holds (and, moreover, the Conjecture holds for a constant d, which does not depend on c).

Theorem

If the Conjecture holds for a constant d (not depending on c), then $P \neq PSPACE$.

Theorem

If $P = PSPACE$, then the Conjecture holds. (Moreover, the Conjecture holds unconditionally for space restrictions in place of time restrictions.)

Definition

A set T of strings is called *elusive* if $T \in \mathrm{P}$, however for all c there are infinitely many n such that $T^{=n} \neq \emptyset$ but for any randomized machine with program of length $c \log n$ running in time n^c we have Prob[M's output falls in $T^{=n}$] $< n^{-c}$

The techniques

Definition

A set T of strings is called *elusive* if $T \in \mathrm{P}$, however for all c there are infinitely many n such that $T^{=n} \neq \emptyset$ but for any randomized machine with program of length $c \log n$ running in time n^c we have $\mathrm{Prob}[M\text{'s output falls in } T^{=n}] < n^{-c}$

The sketch of the proof.

NE $\neq$ RE $\Rightarrow$ There exists an elusive set $\Rightarrow$ The Conjecture $\square$

A gap between Kolmogorov and distinguishing complexities

Theorem (informal)

If there exists an elusive set, then there are strings x with

$$CD^{poly(n)}(x|r) \ll C^{poly(n)}(x|r)$$

for 99% of r's of length $poly(n)$.

A gap between Kolmogorov and distinguishing complexities

Theorem (informal)

If there exists an elusive set, then there are strings x with

$$CD^{poly(n)}(x|r) \ll C^{poly(n)}(x|r)$$

for 99% of r's of length $poly(n)$.

Theorem (formal)

If there exists an elusive set, then for some d for all c there are infinitely many strings x with

$$CD^{|x|^d}(x|r) \ll C^{|x|^c}(x|r) - c \log |x|$$

for 99% of r's of length n^d.

Other approaches do define acceptable explanations: Plausible explanations

Definition

μ is a (t, ε)-*plausible* explanation for x if for all $T \ni x$ we have $\mu(T) \geqslant \varepsilon 2^{-\mathrm{CD}^t(T)}$.

Other approaches do define acceptable explanations: Plausible explanations

Definition

μ is a (t, ε)-*plausible* explanation for x if for all $T \ni x$ we have $\mu(T) \geqslant \varepsilon 2^{-\mathrm{CD}^t(T)}$.

Theorem

Assume that there exists a PRNG $G : \{0,1\}^n \to \{0,1\}^{2n}$. *Then for all* c *for all sufficiently large* n *for* 99% *of strings* s *of length* n *the uniform distribution is an* $(n^c, c \log n, n^{-c}/200)$*-acceptable hypothesis for* $G_n(s)$.

Other approaches do define acceptable explanations: Plausible explanations

Definition

μ is a (t, ε)-*plausible* explanation for x if for all $T \ni x$ we have $\mu(T) \geqslant \varepsilon 2^{-\mathrm{CD}^t(T)}$.

Theorem

Assume that there exists a PRNG $G : \{0,1\}^n \to \{0,1\}^{2n}$. Then for all c for all sufficiently large n for 99% of strings s of length n the uniform distribution is an $(n^c, c\log n, n^{-c}/200)$-acceptable hypothesis for $G_n(s)$.

On the other hand, the set $T = \{x\}$ proves that the uniform distribution is not $(poly(n), 2^{-n})$-*plausible* for x, as the fraction of $T = \{x\}$ among all $2n$-bit strings is 2^{-2n} and $\mathrm{CD}^{\mathrm{poly}(n)}(T) \leqslant n$.

Other approaches do define acceptable explanations: optimal explanations

Definition

μ is a (t, ε)-*optimal* explanation for x if $\mu(x) \geqslant \varepsilon 2^{-C^t(x)}$.

Other approaches do define acceptable explanations: optimal explanations

Definition

μ is a (t, ε)-*optimal* explanation for x if $\mu(x) \geqslant \varepsilon 2^{-C^t(x)}$.

Relations between acceptability, plausibility and optimality

Other approaches do define acceptable explanations: optimal explanations

Definition

μ is a (t, ε)-*optimal* explanation for x if $\mu(x) \geqslant \varepsilon 2^{-C^t(x)}$.

Relations between acceptability, plausibility and optimality

(t, ε)-plausible $\Rightarrow$ $(t, \alpha, \varepsilon 2^{-\alpha})$-acceptable for all α.

Other approaches do define acceptable explanations: optimal explanations

Definition

μ is a (t,ε)-*optimal* explanation for x if $\mu(x) \geqslant \varepsilon 2^{-C^t(x)}$.

Relations between acceptability, plausibility and optimality

(t,ε)-plausible $\Rightarrow$ $(t,\alpha,\varepsilon 2^{-\alpha})$-acceptable for all α.

(t,ε)-plausible $\Rightarrow$ (t,ε)-optimal (let $T = \{x\}$ in the definition of plausibility).

Other approaches do define acceptable explanations: optimal explanations

Definition

μ is a (t, ε)-*optimal* explanation for x if $\mu(x) \geqslant \varepsilon 2^{-C^t(x)}$.

Relations between acceptability, plausibility and optimality

(t, ε)-plausible $\Rightarrow$ $(t, \alpha, \varepsilon 2^{-\alpha})$-acceptable for all α.

(t, ε)-plausible $\Rightarrow$ (t, ε)-optimal (let $T = \{x\}$ in the definition of plausibility).

Theorem (informal)

(1) If $CD^{poly}(x) \ll C^{poly}(x)$, then x has no simple plausible explanations (under the assumption that $Time(2^{O(n)}) \not\subset Space(2^{o(n)})$ for almost all n).

Other approaches do define acceptable explanations: optimal explanations

Definition

μ is a (t, ε)-*optimal* explanation for x if $\mu(x) \geqslant \varepsilon 2^{-C^t(x)}$.

Relations between acceptability, plausibility and optimality

(t, ε)-plausible $\Rightarrow$ $(t, \alpha, \varepsilon 2^{-\alpha})$-acceptable for all α.

(t, ε)-plausible $\Rightarrow$ (t, ε)-optimal (let $T = \{x\}$ in the definition of plausibility).

Theorem (informal)

(1) If $CD^{poly}(x) \ll C^{poly}(x)$, then x has no simple plausible explanations (under the assumption that $Time(2^{O(n)}) \not\subset Space(2^{o(n)})$ for almost all n).
(2) If $CD^{poly}(x) \approx C^{poly}(x)$, then every simple optimal explanation is plausible (under the assumption that $Time(2^{O(n)}) \not\subset Size(2^{o(n)})$ for almost all n).

Open questions

Question

Under which assumptions there exist non-stochastic strings with polynomial bounds for time and linear bounds for program length.

Open questions

Question

Under which assumptions there exist non-stochastic strings with polynomial bounds for time and linear bounds for program length.

Question

Under which assumptions there exist strings that do not possess simple optimal hypotheses?

Open questions

Question

Under which assumptions there exist non-stochastic strings with polynomial bounds for time and linear bounds for program length.

Question

Under which assumptions there exist strings that do not possess simple optimal hypotheses?

Question

How acceptability is related to optimality for strings x with $\mathrm{CD}^{\mathrm{poly}}(x) \ll C^{\mathrm{poly}}(x)$?

Thank you!