Lecture 2:

Complexity Theory Through the Lens of Kolmogorov Complexity

Igor Carboni Oliveira

University of Warwick

CIRM - Randomness, Information & Complexity

February/2024

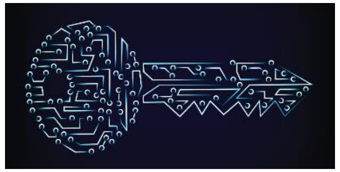
Plan for the Week

Lecture 1 (Monday)

Probabilistic Notions of (Time-Bounded) Kolmogorov Complexity

“**Unconditional** results & applications to average-case complexity”

Lecture 2 (Tuesday)



OWF P vs NP

Connections to Cryptography and Complexity Theory

“Major questions in complexity are **equivalent** to statements about Kolmogorov Complexity”

Lecture 3 (Thursday)

Connections to Algorithms (explicit constructions, generating primes, etc.)

“Existence of large primes with efficient short descriptions”

What this lecture is about

Computational Complexity Theory versus **Kolmogorov Complexity Theory**

NP vs **BPP**

DistNP vs **HeurBPP**

Existence of **OWFs**

Theory of **Time-Bounded** Kolmogorov Complexity

Based on joint work with **Shuichi Hirahara**, **Rahul Ilango**, **Zhenjian Lu**, and **Mikito Nanashima**

Key principles of Kolmogorov complexity

Incompressibility

Symmetry of Information (SoI)

Coding

Language Compression

Q. Do these principles survive in the **time-bounded** setting?

(even predates the P vs NP problem **[Levin'03]**)

Results:

Equivalences to main conjectures of complexity theory

	$\Longleftrightarrow$	$\Longrightarrow$
$\nexists$ i.o. OWFs	• Average-case conditional coding • Average-case conditional language compression • Average-case SoI	
$\mathrm{NP} \subseteq \mathrm{BPP}$	• Worst-case conditional coding • Worst-case conditional language compression	• Worst-case SoI ($\mathrm{NP} \subseteq \mathrm{AvgBPP}$ suffices)
$\mathrm{NP} \subseteq \mathrm{HeurBPP}$	• "Independent average-case" conditional coding • "Independent average-case" conditional language compression	• "Independent average-case" SoI

Almost complete picture, but fully understanding the role of **Symmetry of Information** remains a mystery

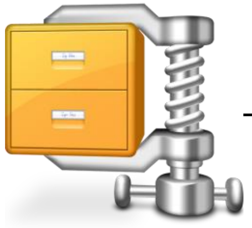
Background and Main Result

(Focus on Symmetry of Information and **OWFs**)

Kolmogorov Complexity

string x

01001101 … 0111010111



short encoding of x

010001110

Kolmogorov Complexity:

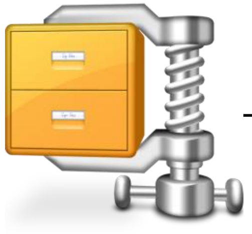
$$\mathrm{K}(x) = \min_{M} \{|M| : U(M) \text{ outputs } x\}$$

"minimum length of a program that recovers x"

Kolmogorov Complexity

string x

01001101 … 0111010111



short encoding of x

010001110

Conditional Kolmogorov Complexity:

$$\mathrm{K}(x \mid y) = \min_{M} \{|M| : U(M, y) \text{ outputs } x\}$$

"minimum length of a program that recovers x given y"

Time-Bounded Kolmogorov Complexity

t-time-bounded Kolmogorov complexity:

$$\boxed{\mathrm{K}^t(x) = \min_M \{|M| : U(M) \text{ outputs } x \text{ within } t(|x|) \text{ steps}\}}$$

Symmetry of Information

$$\boxed{\mathrm{K}(x, y) \lesssim \mathrm{K}(x) + \mathrm{K}(y \mid x)}$$

Symmetry of Information

$$\boxed{\mathrm{K}(x, y) \lesssim \mathrm{K}(x) + \mathrm{K}(y \mid x)}$$

Symmetry of information (SoI) for time-unbounded Kolmogorov complexity:

$$\boxed{\mathrm{K}(x, y) \gtrsim \mathrm{K}(x) + \mathrm{K}(y \mid x)}$$

Symmetry of Information

$$\boxed{\mathrm{K}(x, y) \lesssim \mathrm{K}(x) + \mathrm{K}(y \mid x)}$$

Symmetry of information (SoI) for time-unbounded Kolmogorov complexity:

$$\boxed{\mathrm{K}(x, y) \gtrsim \mathrm{K}(x) + \mathrm{K}(y \mid x)}$$

$$\boxed{\mathrm{K}(x) + \mathrm{K}(y \mid x) \approx \mathrm{K}(x, y) \approx \mathrm{K}(y) + \mathrm{K}(x \mid y)}$$

Symmetry of Information

$$\boxed{\mathrm{K}(x, y) \lesssim \mathrm{K}(x) + \mathrm{K}(y \mid x)}$$

Symmetry of information (SoI) for time-unbounded Kolmogorov complexity:

$$\boxed{\mathrm{K}(x, y) \gtrsim \mathrm{K}(x) + \mathrm{K}(y \mid x)}$$

$$\boxed{\mathrm{K}(x) + \mathrm{K}(y \mid x) \approx \mathrm{K}(x, y) \approx \mathrm{K}(y) + \mathrm{K}(x \mid y)}$$

$$\boxed{\mathrm{K}(x) - \mathrm{K}(x \mid y) \approx \mathrm{K}(y) - \mathrm{K}(y \mid x)}$$

Symmetry of Information

$$\boxed{\mathrm{K}(x,y) \lesssim \mathrm{K}(x) + \mathrm{K}(y \mid x)}$$

Symmetry of information (SoI) for time-unbounded Kolmogorov complexity:

$$\boxed{\mathrm{K}(x,y) \gtrsim \mathrm{K}(x) + \mathrm{K}(y \mid x)}$$

$$\boxed{\mathrm{K}(x) + \mathrm{K}(y \mid x) \approx \mathrm{K}(x,y) \approx \mathrm{K}(y) + \mathrm{K}(x \mid y)}$$

$$\boxed{\mathrm{K}(x) - \mathrm{K}(x \mid y) \approx \mathrm{K}(y) - \mathrm{K}(y \mid x)}$$

$$\boxed{\mathrm{H}(\boldsymbol{X}) - \mathrm{H}(\boldsymbol{X} \mid \boldsymbol{Y}) = \mathrm{H}(\boldsymbol{Y}) - \mathrm{H}(\boldsymbol{Y} \mid \boldsymbol{X})}$$

SoI Principle in Shannon's Information Theory

Symmetry of Information

$$\boxed{\mathrm{K}(x, y) \lesssim \mathrm{K}(x) + \mathrm{K}(y \mid x)}$$

Symmetry of information (SoI) for time-unbounded Kolmogorov complexity:

$$\boxed{\mathrm{K}(x, y) \gtrsim \mathrm{K}(x) + \mathrm{K}(y \mid x)}$$

Does symmetry of information hold in the time-bounded setting, for K^t?

$$\boxed{\mathrm{K}^t(x, y) \geq \mathrm{K}^{\mathrm{poly}(t)}(x) + \mathrm{K}^{\mathrm{poly}(t)}(y \mid x) - O(\log t(|x| + |y|))}$$

Symmetry of Information and One-Way Functions

Definition (One-Way Functions):

An efficiently computable function $f: \{0,1\}^n \to \{0,1\}^n$ is a one-way function if for every probabilistic polynomial-time algorithm A,

$$\Pr_{x \sim \{0,1\}^n}\left[A\big(f(x)\big) \in f^{-1}(f(x))\right] \leq 1/n^{\omega(1)}$$

Symmetry of Information and One-Way Functions

Definition (One-Way Functi

An efficiently comput
function if for every p

Pr
$x \sim \{0,$

OWFs are both **necessary** and **sufficient** for:

- Private-key encryption [GM84, HILL99]
- Pseudorandom generators [HILL99]
- Authentication schemes [FS90]
- Pseudorandom functions [GGM84]
- Digital signatures [Rompel90]
- Commitment schemes [Naor90]
- Coin-tossing [Blum84]

....

Symmetry of Information and One-Way Functions

Definition (One-Way Functions):

An efficiently computable function $f\colon\{0,1\}^n \to \{0,1\}^n$ is a one-way function if for every probabilistic polynomial-time algorithm A,

$$\Pr_{x\sim\{0,1\}^n}\left[A\big(f(x)\big) \in f^{-1}(f(x))\right] \le 1/n^{\omega(1)}$$

Theorem [Longpré-Watanabe'95]:

One-way functions exist $\Longrightarrow$ No symmetry of information for K^t

$\Longleftarrow$?

Symmetry of Information and One-Way Functions

Definition (One-Way Functions):

An efficiently computable function $f\colon\{0,1\}^n \to \{0,1\}^n$ is a one-way function if for every probabilistic polynomial-time algorithm A,

$$\Pr_{x\sim\{0,1\}^n}\left[A\big(f(x)\big) \in f^{-1}(f(x))\right] \le 1/n^{\omega(1)}$$

Two key points:

- Average-case symmetry of information? --- Consider “average” pairs (x, y)
- Probabilistic versions of time-bounded Kolmogorov complexity?

Probabilistic Versions of Kolmogorov Complexity

Randomized t-time-bounded Kolmogorov complexity:

$$\boxed{\mathrm{rK}^t(x) = \min_k \left\{ k : \exists\ t(|x|) \text{ time program } M \in \{0,1\}^k \text{ s.t. } \Pr_{\text{randomness of } M}[M \text{ outputs } x] \geq \tfrac{2}{3} \right\}}$$

There exists a **fixed** small (randomized) program that outputs x **w.h.p over its internal randomness**

Probabilistic Versions of Kolmogorov Complexity

Randomized t-time-bounded Kolmogorov complexity:

$$\mathrm{rK}^t(x) = \min_k \left\{ k : \exists\ t(|x|) \text{ time program } M \in \{0,1\}^k \text{ s.t. } \Pr_{\text{randomness of } M}[M \text{ outputs } x] \geq \frac{2}{3} \right\}$$

There exists a **fixed** small (randomized) program that outputs x **w.h.p over its internal randomness**

Probabilistic t-time-bounded Kolmogorov complexity:

$$\mathrm{pK}^t(x) = \min_k \left\{ k : \Pr_{w \in \{0,1\}^{t(|x|)}}[\exists\ M \in \{0,1\}^k \text{ s.t. } M(w) \text{ outputs } x \text{ within } t(|x|) \text{ steps}] \geq \frac{2}{3} \right\}$$

For **most w**, there exists a small program, **which can depend on w**, that outputs x given w

Symmetry of Information and One-Way Functions

Theorem:

One-way functions do not exist	$\Longleftrightarrow$	"Average-case" SoI for $\mathrm{pK}^{\mathrm{poly}}$ holds

The following are equivalent:

- One-way functions do not exist.
- For every poly-time-samplable distribution $\{D_n\}$ over $\{0,1\}^n \times \{0,1\}^n$, constant $c \geq 0$, and sufficiently large $t \geq \mathrm{poly}(n)$, there are infinitely many n such that

$$\Pr_{(x,y)\sim D_n}\left[\mathrm{pK}^t(x,y) \geq \mathrm{pK}^t(x) + \mathrm{pK}^t(y \mid x) - \log t(n)\right] \geq 1 - 1/n^c$$

Symmetry of Information and One-Way Functions

Theorem:

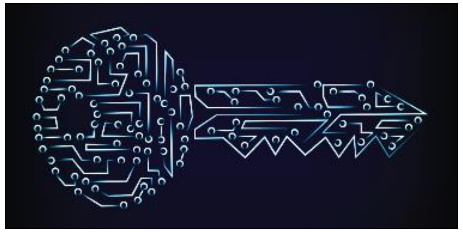
One-way functions do not exist $\Longleftrightarrow$ "Average-case" SoI for $\mathrm{pK}^{\mathrm{poly}}$ holds

The following are equivalent:

- Infinitely-often one-way functions do not exist.
- For every poly-time-samplable distribution $\{D_n\}$ over $\{0,1\}^n \times \{0,1\}^n$, constant $c \geq 0$, and sufficiently large $t \geq \mathrm{poly}(n)$, and for all but finitely many n,

$$\Pr_{(x,y)\sim D_n}\left[\mathrm{pK}^t(x,y) \geq \mathrm{pK}^t(x) + \mathrm{pK}^t(y \mid x) - \log t(n)\right] \geq 1 - 1/n^c$$

Relevance to the foundations of cryptography



Failure of symmetry of information
for a non-negligible fraction of pairs (x,y) of strings produced by a **samplable distribution** is all we need to construct key cryptographic primitives and protocols

What about rK^{poly} ?

Symmetry of Information and One-Way Functions

Theorem:

Quasipoly-time secure one-way functions do not exist	⟺	"Average-case" SoI for $\mathrm{rK}^{\mathrm{quasipoly}}$ holds

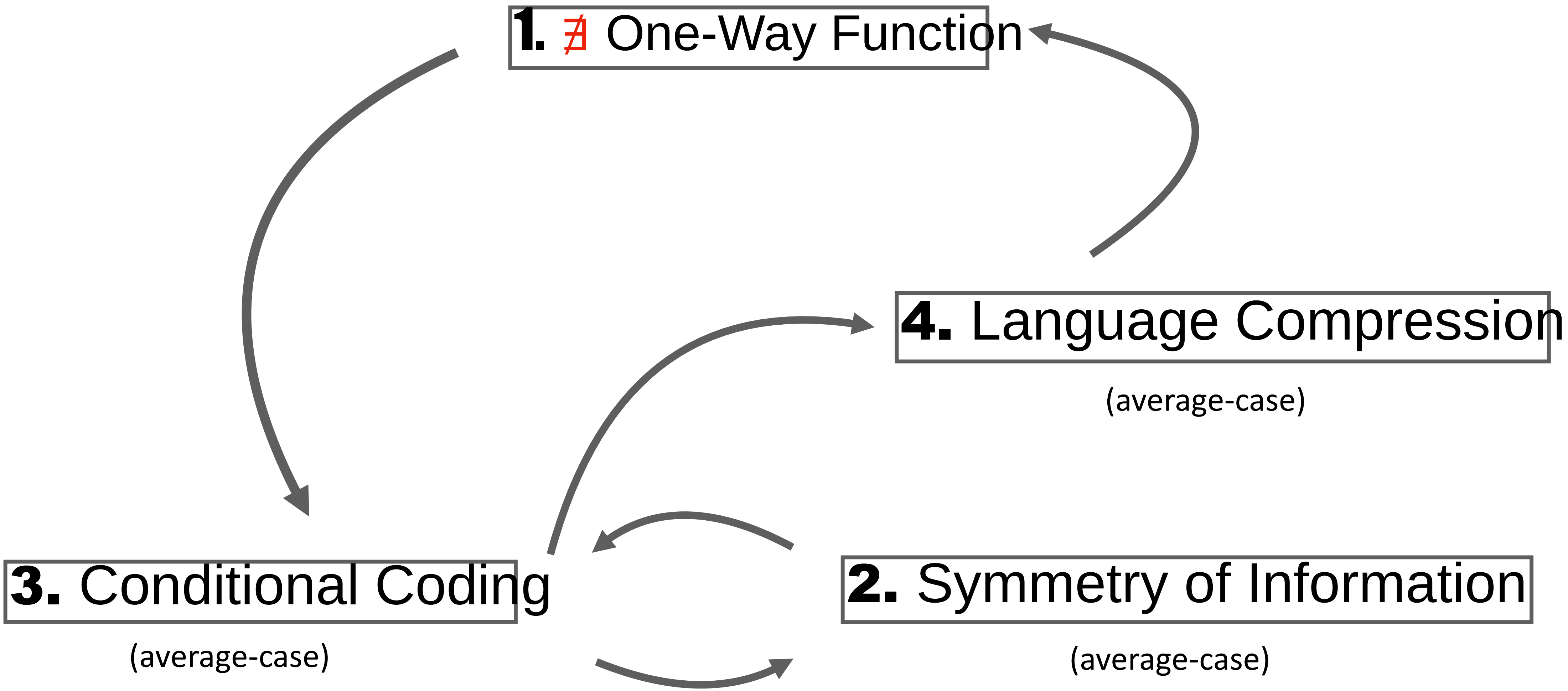
The following are equivalent:

- Quasipoly-time secure one-way functions do not exist.
- For every poly-time-samplable distribution $\{D_n\}$ over $\{0,1\}^n \times \{0,1\}^n$, constant $c \geq 0$, and sufficiently large $t \geq \mathrm{quasipoly}(n)$, there are infinitely many n such that

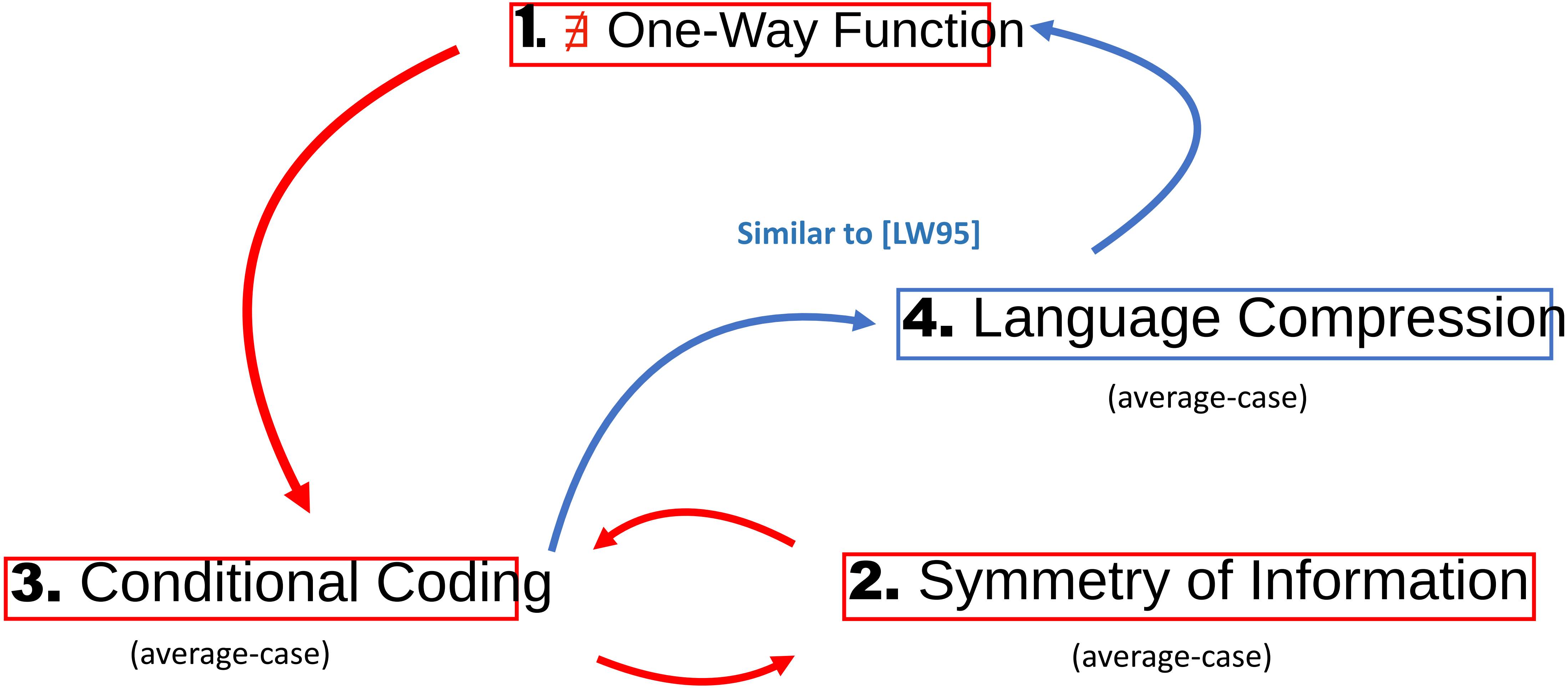
$$\Pr_{(x,y)\sim D_n}[\mathrm{rK}^t(x,y) \geq \mathrm{rK}^t(x) + \mathrm{rK}^t(y \mid x) - \log t(n)] \geq 1 - 1/\exp(\log^c n)$$

Techniques (pK^{poly})

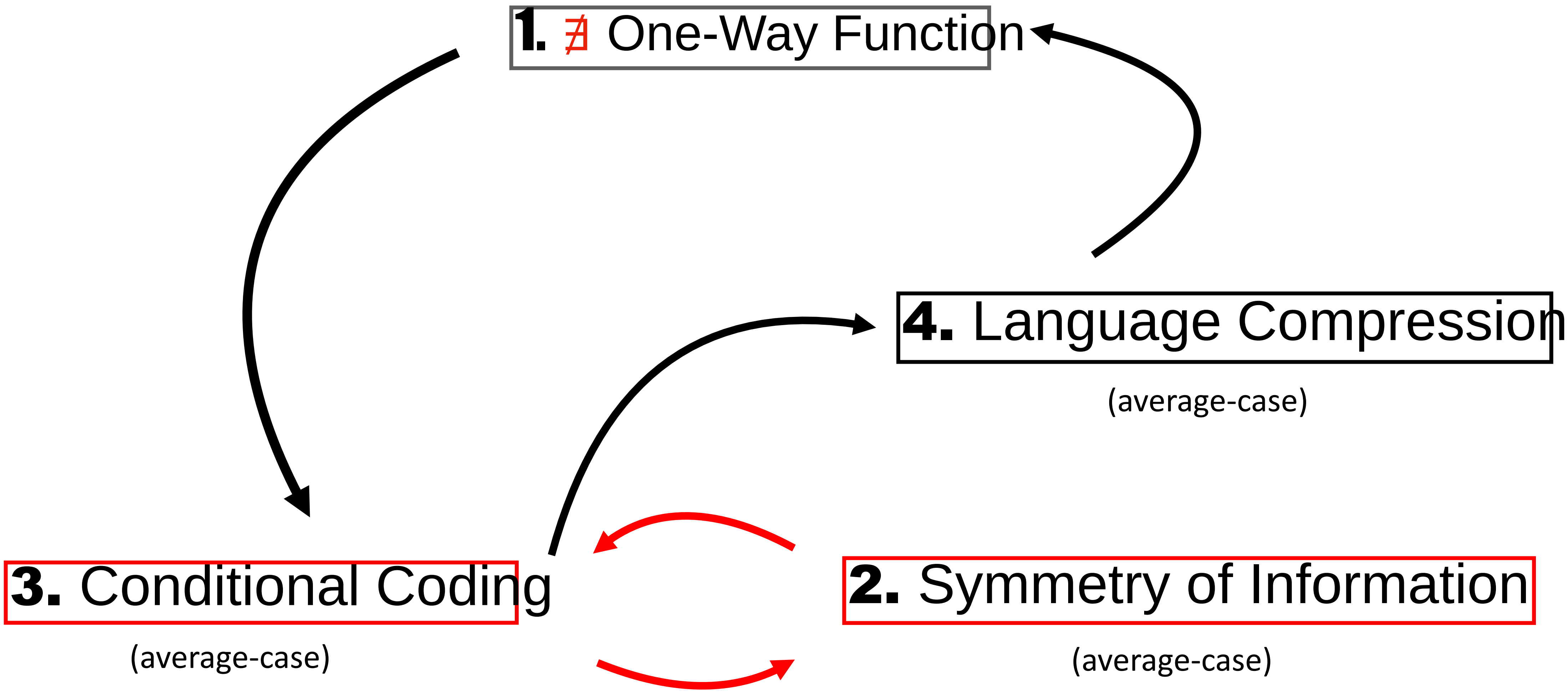
Map of Proofs for $\mathrm{pK}^{\mathrm{poly}}$



Map of Proofs for $\mathrm{pK}^{\mathrm{poly}}$



Map of Proofs for $\mathrm{pK}^{\mathrm{poly}}$



Coding Theorem

Coding theorem for **time-unbounded** Kolmogorov complexity:

A computable distribution D that samples x with probability $D(x)$	$\Longrightarrow$	$\mathrm{K}(x) \lesssim \log\left(\frac{1}{D(x)}\right)$

Coding Theorem

Coding theorem for **time-unbounded** Kolmogorov complexity:

A computable distribution D that samples x with probability $D(x)$	$\Rightarrow$	$\mathrm{K}(x) \lesssim \log\left(\frac{1}{D(x)}\right)$

We don't have a coding theorem for $\mathbf{K}^{\mathbf{poly}}$

Coding Theorem

<u>Coding theorem for **time-unbounded** Kolmogorov complexity:</u>

A computable distribution D that samples x with probability $D(x)$ $\Rightarrow$ $\mathrm{K}(x) \lesssim \log\left(\frac{1}{D(x)}\right)$

We don't have a coding theorem for $\mathbf{K}^{\mathbf{poly}}$

An efficiently samplable distribution D that samples x with probability $D(x)$ $\overset{?}{\Rightarrow}$ $\mathrm{K}^{\mathrm{poly}}(x) \lesssim \log\left(\frac{1}{D(x)}\right)$

Coding Theorem

Theorem [Lu-Oliveira-Zimand'22]:

For every poly-time-samplable distribution $\{D_n\}$ over $\{0,1\}^n$, and every $x \in \mathrm{support}(D_n)$

$$\mathrm{pK}^{\mathrm{poly}}(x) \leq \log\left(\frac{1}{D_n(x)}\right) + O(\log n)$$

Coding Theorem

Theorem [Lu-Oliveira-Zimand'22]:

For every poly-time-samplable distribution $\{D_n\}$ over $\{0,1\}^n$, and every $x \in \mathrm{support}(D_n)$

$$\mathrm{pK}^{\mathrm{poly}}(x) \leq \log\left(\frac{1}{D_n(x)}\right) + O(\log n)$$

Incompressibility (extension of counting argument):

$$\mathrm{K}(x) \gtrsim \log(1/D_n(x)) \text{ w.h.p over } x \sim D_n$$

Coding Theorem

Theorem [Lu-Oliveira-Zimand'22]:

For every poly-time-samplable distribution $\{D_n\}$ over $\{0,1\}^n$, and every $x \in \mathrm{support}(D_n)$

$$\mathrm{pK}^{\mathrm{poly}}(x) \leq \log\left(\frac{1}{D_n(x)}\right) + O(\log n)$$

Incompressibility (extension of counting argument):

$$\mathrm{K}(x) \gtrsim \log(1/D_n(x)) \text{ w.h.p over } x \sim D_n$$

Corollary:

$$\mathrm{pK}^{\mathrm{poly}}(x) \approx \log(1/D_n(x)) \text{ w.h.p over } x \sim D_n$$

Conditional Coding

Definition (Conditional Coding for pK):

For every poly-time-samplable distribution $\{D_n\}$ over $\{0,1\}^n \times \{0,1\}^n$, and every $(x, y) \in \mathrm{support}(D_n)$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D_n(x \mid y)}\right)$$

Conditional Coding

Definition (Conditional Coding for pK):

For every poly-time-samplable distribution $\{D_n\}$ over $\{0,1\}^n \times \{0,1\}^n$, and every $(x, y) \in \text{support}(D_n)$

$$\text{pK}^{\text{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D_n(x \mid y)}\right)$$

Definition (Average-Case Conditional Coding for pK):

For every poly-time-samplable distribution $\{D_n\}$ over $\{0,1\}^n \times \{0,1\}^n$,

$$\Pr_{(x,y)\sim D_n}\left[\text{pK}^{\text{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D_n(x \mid y)}\right)\right] \geq 1 - \frac{1}{\text{poly}(n)}$$

Conditional Coding and SoI

Lemma:

Average-case **conditional coding** holds ⟺ Average-case **SoI** holds

Conditional Coding and SoI

Lemma:

Average-case conditional coding holds ⟹ Average-case SoI holds

Proof:

$$\Pr_{(x,y)\sim D_n}\left[\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D(x \mid y)}\right)\right] \geq 1 - 1/n^{O(1)}$$

$$\Pr_{(x,y)\sim D_n}\left[\mathrm{pK}^{\mathrm{poly}}(x, y) \gtrsim \mathrm{pK}^{\mathrm{poly}}(y) + \mathrm{pK}^{\mathrm{poly}}(x \mid y)\right] \geq 1 - 1/n^{O(1)}$$

Conditional Coding and SoI

Lemma:

Average-case **conditional coding** holds $\Rightarrow$ Average-case **SoI** holds

Proof: $(x, y) \sim D$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D(x \mid y)}\right)$$

$$\mathrm{pK}^{\mathrm{poly}}(x, y) \gtrsim \mathrm{pK}^{\mathrm{poly}}(y) + \mathrm{pK}^{\mathrm{poly}}(x \mid y)$$

Conditional Coding and SoI

Lemma:

Average-case **conditional coding** holds $\Rightarrow$ Average-case **SoI** holds

Proof: $(x, y) \sim D$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D(x \mid y)}\right)$$

$$\mathrm{pK}^{\mathrm{poly}}(x, y) \gtrsim \mathrm{pK}^{\mathrm{poly}}(y) + \mathrm{pK}^{\mathrm{poly}}(x \mid y)$$

Conditional Coding and SoI

Lemma:

Average-case **conditional coding** holds $\Rightarrow$ Average-case **SoI** holds

Proof: $(x, y) \sim D$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D(x \mid y)}\right)$$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \mathrm{pK}^{\mathrm{poly}}(x, y) - \mathrm{pK}^{\mathrm{poly}}(y)$$

Conditional Coding and SoI

Lemma:

Average-case **conditional coding** holds ➡ Average-case **SoI** holds

Proof: $(x, y) \sim D$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D(x \mid y)}\right)$$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{D^{(2)}(y)}{D(x, y)}\right)$$

Because $D(x \mid y) = \frac{D(x,y)}{D^{(2)}(y)}$.

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \mathrm{pK}^{\mathrm{poly}}(x, y) - \mathrm{pK}^{\mathrm{poly}}(y)$$

Conditional Coding and SoI

Lemma:

Average-case **conditional coding** holds $\Rightarrow$ Average-case **SoI** holds

Proof: $(x, y) \sim D$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D(x \mid y)}\right)$$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{D^{(2)}(y)}{D(x, y)}\right)$$

Because $D(x \mid y) = \frac{D(x,y)}{D^{(2)}(y)}$.

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D(x, y)}\right) - \log\left(\frac{1}{D^{(2)}(y)}\right)$$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \mathrm{pK}^{\mathrm{poly}}(x, y) - \mathrm{pK}^{\mathrm{poly}}(y)$$

Conditional Coding and SoI

Lemma:

Average-case **conditional coding** holds	$\Rightarrow$	Average-case **SoI** holds

Proof: $(x, y) \sim D$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D(x \mid y)}\right)$$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{D^{(2)}(y)}{D(x, y)}\right)$$

Because $D(x \mid y) = \frac{D(x,y)}{D^{(2)}(y)}$.

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D(x, y)}\right) - \log\left(\frac{1}{D^{(2)}(y)}\right)$$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \mathrm{pK}^{\mathrm{poly}}(x, y) - \mathrm{pK}^{\mathrm{poly}}(y)$$

Because by coding theorem, $\mathrm{pK}^{\mathrm{poly}}(x, y) \approx \log\left(\frac{1}{D(x,y)}\right)$ and $\mathrm{pK}^{\mathrm{poly}}(y) \approx \log\left(\frac{1}{D^{(2)}(y)}\right)$ for most $(x, y) \sim D$

Conditional Coding and SoI

Lemma:

Average-case **conditional coding** holds $\Leftarrow$ Average-case **SoI** holds

Proof: $(x, y) \sim D$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D(x \mid y)}\right)$$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{D^{(2)}(y)}{D(x, y)}\right)$$

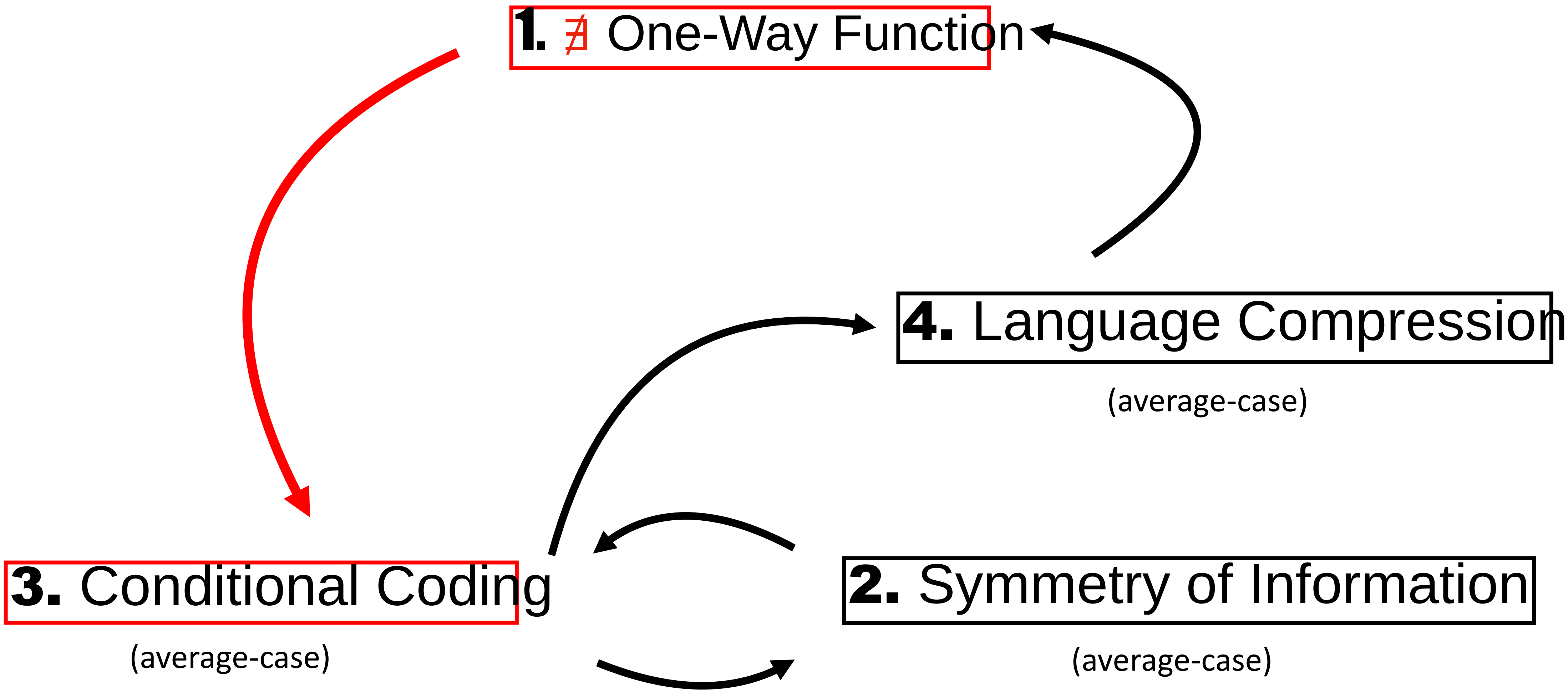
Because $D(x \mid y) = \frac{D(x,y)}{D^{(2)}(y)}$.

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D(x, y)}\right) - \log\left(\frac{1}{D^{(2)}(y)}\right)$$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \mathrm{pK}^{\mathrm{poly}}(x, y) - \mathrm{pK}^{\mathrm{poly}}(y)$$

Because by coding theorem, $\mathrm{pK}^{\mathrm{poly}}(x, y) \approx \log\left(\frac{1}{D(x,y)}\right)$ and $\mathrm{pK}^{\mathrm{poly}}(y) \approx \log\left(\frac{1}{D^{(2)}(y)}\right)$ for most $(x, y) \sim D$

Map of Proofs for $\mathrm{pK}^{\mathrm{poly}}$



OWFs and Average-Case Conditional Coding

Lemma:

Infinitely-often one-way functions do not exist	$\Rightarrow$	Average-case conditional coding holds

Proof Sketch: Assume we can invert OWFs, we want to show w.h.p over $(x, y) \sim D$

$$\boxed{\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D(x \mid y)}\right)}$$

OWFs and Average-Case Conditional Coding

Lemma:

Infinitely-often one-way functions do not exist $\Rightarrow$ Average-case conditional coding holds

Proof Sketch: Assume we can invert OWFs, we want to show w.h.p over $(x, y) \sim D$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D(x \mid y)}\right)$$

Theorem (Extrapolators) [Impagliazzo-Luby'89, Impagliazzo-Levin'90]:

One-way functions do not exist $\Rightarrow$ Efficiently sample $D(\cdot \mid y)$ (approximately) for most $y \sim D^{(2)}$

OWFs and Average-Case Conditional Coding

Lemma:

Infinitely-often one-way functions do not exist $\Rightarrow$ Average-case conditional coding holds

Proof Sketch: Assume we can invert OWFs, we want to show w.h.p over $(x, y) \sim D$

$$\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D(x \mid y)}\right)$$

Theorem (Extrapolators) [Impagliazzo-Luby'89, Impagliazzo-Levin'90]:

One-way functions do not exist $\Rightarrow$ Efficiently sample $D(\cdot \mid y)$ (approximately) for most $y \sim D^{(2)}$

Idea:

- Use the efficient extrapolator as a proxy for the conditional distribution
- Apply the **original** coding theorem!

Extrapolation

Theorem [Impagliazzo-Luby'89, Impagliazzo-Levin'90]:

One-way functions do not exist $\Rightarrow$ Efficient simulation of $D(\cdot \mid y)$ for most $y \sim D^{(2)}$

If infinitely-often OWFs do not exist, then for every poly-time-samplable $\{D_n\}$ over $\{0,1\}^n \times \{0,1\}^n$ and $c > 0$, there is a poly-time randomized algorithm EXT, such that for all n

$$\Pr_{y \sim D_n^{(2)}}\left[\mathrm{L}_1(\mathrm{EXT}(y), D_n(\cdot \mid y)) \leq \frac{1}{n^c}\right] \geq 1 - \frac{1}{n^c}$$

OWFs and Average-Case Conditional Coding

Lemma:

Infinitely-often one-way functions do not exist $\Rightarrow$ Average-case conditional coding holds

Proof Sketch:

- W.h.p over $y \sim D^{(2)}$, we have $\mathrm{L}_1(\mathrm{EXT}(y), D_n(\cdot \mid y))$ is small.
- $\mathrm{EXT}(y)$ runs polynomial-time, so it yields some poly-time-samplable distribution D'_y
- We can show $\mathrm{L}_1\left(D'_y, D_n(\cdot \mid y)\right)$ implies $D'_y(x) \approx D_n\ (x \mid y)$ for most $x \sim D(\cdot \mid y)$
- By the original coding theorem for $\mathrm{pK}^{\mathrm{poly}}$, for most $(x, y) \sim D$
    - $\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D'_y(x)}\right) \approx \log\left(\frac{1}{D_n(x \mid y)}\right)$

OWFs and Average-Case Conditional Coding

Lemma:

Infinitely-often one-way functions do not exist $\Rightarrow$ Average-case conditional coding holds

Proof Sketch:

- W.h.p over $y \sim D^{(2)}$, we have $\mathrm{L}_1(\mathrm{EXT}(y), D_n(\cdot \mid y))$ is small.
- $\mathrm{EXT}(y)$ runs polynomial-time, so it yields some poly-time-samplable distribution D'_y
- We can show $\mathrm{L}_1\left(D'_y, D_n(\cdot \mid y)\right)$ implies $D'_y(x) \approx D_n\,(x \mid y)$ for most $x \sim D(\cdot \mid y)$
- By the original coding theorem for $\mathrm{pK}^{\mathrm{poly}}$, for most $(x, y) \sim D$
  - $\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D'_y(x)}\right) \approx \log\left(\frac{1}{D_n(x \mid y)}\right)$

OWFs and Average-Case Conditional Coding

Lemma:

Infinitely-often one-way functions do not exist $\Rightarrow$ Average-case conditional coding holds

Proof Sketch:

- W.h.p over $y \sim D^{(2)}$, we have $\mathrm{L}_1(\mathrm{EXT}(y), D_n(\cdot \mid y))$ is small.
- $\mathrm{EXT}(y)$ runs polynomial-time, so it yields some poly-time-samplable distribution D'_y
- We can show $\mathrm{L}_1\left(D'_y, D_n(\cdot \mid y)\right)$ implies $D'_y(x) \approx D_n\ (x \mid y)$ for most $x \sim D(\cdot \mid y)$
- By the original coding theorem for $\mathrm{pK}^{\mathrm{poly}}$, for most $(x, y) \sim D$
  - $\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D'_y(x)}\right) \approx \log\left(\frac{1}{D_n(x \mid y)}\right)$

OWFs and Average-Case Conditional Coding

Lemma:

Infinitely-often one-way functions do not exist $\Rightarrow$ Average-case conditional coding holds

Proof Sketch:

- W.h.p over $y \sim D^{(2)}$, we have $\mathrm{L}_1(\mathrm{EXT}(y), D_n(\cdot \mid y))$ is small.
- $\mathrm{EXT}(y)$ runs polynomial-time, so it yields some poly-time-samplable distribution D'_y
- We can show $\mathrm{L}_1\left(D'_y, D_n(\cdot \mid y)\right)$ implies $D'_y(x) \approx D_n\ (x \mid y)$ for most $x \sim D(\cdot \mid y)$
- By the original coding theorem for $\mathrm{pK}^{\mathrm{poly}}$, for most $(x, y) \sim D$
  - $\mathrm{pK}^{\mathrm{poly}}(x \mid y) \lesssim \log\left(\frac{1}{D'_y(x)}\right) \approx \log\left(\frac{1}{D_n(x \mid y)}\right)$

Techniques ($rK^{quasipoly}$)

Key Difficulty: We don't have a coding theorem for rK^{poly}

Main Technique for $\mathrm{rK}^{\mathrm{quasipoly}}$ (Key Perspective: Meta-Complexity)

Quasipoly-time secure one-way functions do not exist

$\Rightarrow$

"Average-case" SoI for $\mathrm{rK}^{\mathrm{quasipoly}}$ holds

[Ilango-Ren-Santhanam'22]

Quasipoly-time algorithm for approximating $\mathrm{K}(x \mid y)$ on average (A distinguisher!)

Generator construction with $\mathrm{rK}^{\mathrm{quasipoly}}$ reconstruction

$\mathrm{rK}^{\mathrm{quasipoly}}(x \mid y) \leq \mathrm{K}(x \mid y) + \mathrm{polylog}(n)$ for almost all $(x, y) \sim D$

General Theory

	$\Longleftrightarrow$	$\Longrightarrow$
$\nexists$ i. o. OWFs	• Average-case conditional coding • Average-case conditional language compression • Average-case SoI	

	$\Longleftrightarrow$	$\Longrightarrow$
$\nexists$ i.o. OWFs	• Average-case conditional coding • Average-case conditional language compression • Average-case SoI	
$\mathrm{NP} \subseteq \mathrm{BPP}$	• Worst-case conditional coding • Worst-case conditional language compression	• Worst-case SoI ($\mathrm{NP} \subseteq \mathrm{AvgBPP}$ suffices)

	$\Longleftrightarrow$	$\Longrightarrow$
$\nexists$ i. o. OWFs	• Average-case conditional coding • Average-case conditional language compression • Average-case SoI	
$\mathrm{NP} \subseteq \mathrm{BPP}$	• Worst-case conditional coding • Worst-case conditional language compression	• Worst-case SoI ($\mathrm{NP} \subseteq \mathrm{AvgBPP}$ suffices)
$\mathrm{NP} \subseteq \mathrm{HeurBPP}$	• "Independent average-case" conditional coding • "Independent average-case" conditional language compression	• "Independent average-case" SoI

	$\Longleftrightarrow$	$\Longrightarrow$
$\nexists$ i.o. OWFs	• Average-case conditional coding • Average-case conditional language compression • Average-case SoI	
NP $\subseteq$ BPP	• Worst-case conditional coding • Worst-case conditional language compression	• Worst-case SoI (NP $\subseteq$ AvgBPP suffices)
NP $\subseteq$ HeurBPP	• "Independent average-case" conditional coding • "Independent average-case" conditional language compression	• "Independent average-case" SoI

Capturing average-case complexity

Theorem [This Work]:

DistNP ⊆ HeurBPP

⟺

"Independent" Average-case **Conditional Coding** for $\mathrm{pK}^{\mathrm{poly}}$ holds

Capturing average-case complexity

Theorem [This Work]:

DistNP $\subseteq$ HeurBPP $\Longleftrightarrow$ "Independent" Average-case **Conditional Coding** for $\mathrm{pK}^{\mathrm{poly}}$ holds

The following are equivalent:

- DistNP $\subseteq$ HeurBPP .
- For every poly-time-samplable distributions $\{D_n\}$ over $\{0,1\}^n \times \{0,1\}^n$ and $\{C_n\}$ over $\{0,1\}^n$, and for every polynomial *q*, there is a polynomial *p* such that for all large enough n

$$\Pr_{y \sim C_n,\, x \sim D_n(.|y)}\left[\mathrm{pK}^{p(n)}(x \mid y) \leq \log\left(\frac{1}{D_n(x \mid y)}\right) + \log p(n)\right] \geq 1 - 1/q(n)$$

Open Problems

	$\Longleftrightarrow$	$\Longrightarrow$
$\nexists$ i.o. OWFs	• Average-case conditional coding • Average-case conditional language compression • Average-case SoI	
$\mathrm{NP} \subseteq \mathrm{BPP}$	• Worst-case conditional coding • Worst-case conditional language compression	• Worst-case SoI ($\mathrm{NP} \subseteq \mathrm{AvgBPP}$ suffices)
$\mathrm{NP} \subseteq \mathrm{HeurBPP}$	• "Independent average-case" conditional coding • "Independent average-case" conditional language compression	• "Independent average-case" SoI

1. Understand the role of **SoI** in complexity theory:
   Is there a natural computational assumption equivalent to worst-case **SoI**?

2. Applications of these characterizations?

Main Reference for Lecture 2:

Paper: **“A duality between OWFs and average-case symmetry of information”** (2023)

(Joint work with S. Hirahara, R. Ilango, Z. Lu, and M. Nanashima)

Thank you

Conditional Coding

1. **(Worst-Case Conditional Coding)** There exists a polynomial p such that for all n, and $(x, y) \in \mathsf{Support}(\mathcal{D}_n)$
$$\mathsf{pK}^{p(n)}(x \mid y) \le \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n).$$

2. **(Independent Average-Case Conditional Coding)** Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ be samplable distribution families, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and each $\mathcal{C}_n$ is over the support of the second half of $\mathcal{D}_n$. For every polynomial q, there exists a polynomial p such that for all n,
$$\Pr_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot\mid y)}\left[\mathsf{pK}^{p(n)}(x \mid y) \le \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n)\right] \ge 1 - \frac{1}{q(n)}.$$

3. **(Average-Case Conditional Coding)** Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be samplable distribution family, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$. For every polynomial q, there exists a polynomial p such that for all n,
$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{pK}^{p(n)}(y \mid x) \le \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n)\right] \ge 1 - \frac{1}{q(n)}.$$

Language Compression

1. **(Worst-Case Language Compression)** Let $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_{n\in\mathbb{N}}$ be a polynomial-time computable set. There exists a polynomial p such that for all n, and $y \in \{0,1\}^n$,

$$x \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \leq \log |L_y| + \log p(n).$$

2. **(Independent Average-Case Language Compression)** Let $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_{n\in\mathbb{N}}$ be a recursively enumerable set. Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ be samplable distribution families, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and each $\mathcal{C}_n$ is over the support of the second half of $\mathcal{D}_n$. For every polynomial q, there exists a polynomial p such that for all n,

$$\Pr_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[x \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \leq \log |L_y| + \log p(n)\right] \geq 1 - \frac{1}{q(n)}.$$

3. **(Average-Case Language Compression)** Let $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_{n\in\mathbb{N}}$ be a recursively enumerable set. Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be samplable distribution family, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$. For every polynomial q, there exists a polynomial p such that for all n,

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[x \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \leq \log |L_y| + \log p(n)\right] \geq 1 - \frac{1}{q(n)}.$$

Symmetry of Information

1. **(Worst-Case Symmetry of Information)** There exists a polynomial p such that for all $t \geq 2n$ and for all n and all $x, y \in \{0,1\}^n$,

$$\mathsf{pK}^t(x,y) \geq \mathsf{pK}^{p(t)}(x \mid y) + \mathsf{pK}^{p(t)}(y) - \log p(t).$$

2. **(Independent Average-Case Symmetry of Information)** Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ be samplable distribution families, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and each $\mathcal{C}_n$ is over the support of the second half of $\mathcal{D}_n$. For every polynomial q, there exists a polynomial p such that for every computable time bound $t\colon \mathbb{N} \to \mathbb{N}$ with $t(n) \geq p(n)$ and for all n,

$$\Pr_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[\mathsf{pK}^{t(n)}(x,y) \geq \mathsf{pK}^{t(n)}(x \mid y) + \mathsf{pK}^{t(n)}(y) - \log t(n)\right] \geq 1 - \frac{1}{q(n)}.$$

3. **(Average-Case Symmetry of Information)** Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be samplable distribution family, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$. For every polynomial q, there exists a polynomial p such that for every computable time bound $t\colon \mathbb{N} \to \mathbb{N}$ with $t(n) \geq p(n)$ and for all n,

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{pK}^{t(n)}(x,y) \geq \mathsf{pK}^{t(n)}(x \mid y) + \mathsf{pK}^{t(n)}(y) - \log t(n)\right] \geq 1 - \frac{1}{q(n)}.$$