# On Exotic Sequences

Samuel Epstein
samepst@jptheorygroup.org

November 5, 2023

# Contents

# Chapter 1

# Introduction

This manuscript contains a series of results regarding the amount of information shared between finite sequences and the halting information,$\mathcal{H}$. Sequences $\alpha$ with high $\mathbf{I}(\alpha;\mathcal{H})$ are considered exotic and non=realizable, and thus by proving sequences $\alpha$ with certain properties have high $\mathbf{I}(\alpha\mathcal{H})$ then this implies such properties cannot be realized in the physical universe. The results of this manuscript are as follows:

## Classification

Classification is the task of learning a binary function $c$ from $\mathbb{N}$ to bits $\{0,1\}$. The learner is given a sample consisting of pairs $(x,b)$ for string $x$ and bit $b$ and outputs a binary classifier $h:\mathbb{N}\to\{0,1\}$ that should match $c$ as much as possible. Occam's razor says that "the simplest explanation is usually the best one." Simple hypothesis are resilient against overfitting to the sample data. With certain probabilistic assumptions, learning algorithms that produce hypotheses of low Kolmogorov complexity are likely to correctly predict the target function [BEHW89]. The following theorem shows that the samples can be compressed to their count.

**Theorem.** *Given a set of samples* $\{(x_i,b_i)\}_{i=1}^n$*, there is a function* $f:\mathbb{N}\to\{0,1\}$ *such that* $f(x_i)=b_i$*, for* $i=1,\dots,n$*, and* $\mathbf{K}(f)<^{\log} n+\mathbf{I}(\{(x_i,b_i)\};\mathcal{H})$*.*

## Regression

A fundamental area of machine learning is regression, in which one is given a set of pairs $\{(x_i,y_i)\}$, $i=1\dots n$, and the goal is to find a function $f$, such that $f(x_i)=y_i$. Usually each $x_i$ and $y_i$ represents a point in Euclidean space, but for our purposes they are natural numbers. The goal is to use Occam's razor to find the simpliest function, to prevent overfitting to the random noise inherent in the sample data. This chapter presents the following bounds on the simplest total computable function completely consistent with the data.

**Theorem.** *For* $\{(x_i,y_i)\}_{i=1}^n$*, there exists* $f:\mathbb{N}\to\mathbb{N}$ *with* $f(x_i)=y_i$ *for* $i\in\{1,\dots,n\}$ *and* $\mathbf{K}(f)<^{\log}\sum_{i=1}^n\mathbf{K}(y_i|x_i)+\mathbf{I}(\{(x_i,y_i)\};\mathcal{H})$*.*

## Monotone EL Theorem

The EL Theorem [Lev16, Eps19] states that the algorithmic probability of a non-exotic set $\mathbf{m}(D) = \sum_{x\in D}\mathbf{m}(D)$ is concentrated on its simpliest member. A monotone variant of this theorem can be proved. A continuous semi-measure is a function such that $Q(\emptyset) = 0$ and $Q(x) \geq Q(x0) + Q(x1)$. For prefix free set $D$, $Q(D) = \sum_{x\in D} Q(x)$. There exists a universal lower computable continuous semi-measure $\mathbf{M}$. The monotone complexity of a string $x$ is $\mathbf{Km}(x) = \min\{\|p\| : x \sqsubseteq U(p)\}$.

**Theorem.** *For prefix free set $D$, $\min_{x\in D}\mathbf{Km}(x) <^{\log} -\log\mathbf{M}(D) + \mathbf{I}(\langle D\rangle : \mathcal{H})$.*

## Probabilities are Balanced

In this chapter, we look at probabilities over strings of length $n$, and prove that they must give measure to simple strings. This result also appears in the black holes section of the Algorithmic Physics manuscript at http://www.jptheorygroup.org.

**Theorem.** *There is a $c \in \mathbb{N}$ where for probability $p$ over $\{0,1\}^n$, for $m > \mathbf{K}(p) + c$, $p\{x : \mathbf{K}(x) < m\} > 2^{m-n-2\mathbf{I}(p;\mathcal{H})-3\mathbf{K}(n,m)-c}$.*

## The Kolmogorov Birthday Paradox

Let us say we select a random subset $D$ of size $2^{n/2}$ consisting of (possibly repeated) strings of length $n$, where each string is selected independently with a uniform probability. For the simple Kolmogorov birthday paradox, with overwhelming probability, there are two (possibly the same) strings $x, y \in D$, such that $\mathbf{K}(x|y) = O(1)$, for a large enough constant. This is due to reasoning from the classical birthday paradox. We now prove the general Kolmogorov birthday paradox. Let $P$ be any probability over sets $D$ consisting of $2^{n/2}$ (non repeated) strings of length $n$. Since $D \subset \{0,1\}^n$, for all $D$, $\max_{x,y\in D}\mathbf{K}(x/y) <^+ n$. The chapter gives the following result.

**Theorem.** $\Pr_{D\sim P}\left[\min_{x,y\in D, x\neq y}\mathbf{K}(x|y) <^{\log} \mathbf{I}(P;\mathcal{H}) + 2\mathbf{K}(n) + c\right] > 1 - 2^{-c}$.

## On the Conditional Complexity of Sets of Strings

We define a $(k,l)$ bunch $X$ to be a finite set of strings, where $k = \lceil \log|X|\rceil$, $l > k$, and for all $x, x' \in X$, $\mathbf{K}(x|x') \leq l$. If $l \gg k$, such as the bunch consisting of two large independent random strings, then it is difficult to proof properties about it. If $l \approx k$, then interesting properties emerge, such as the bunch theorem of this chapter.

**Theorem.** *For $(k,l)$ bunch $X$, $\min_{x\in X}\max_{x'\in X}\mathbf{K}(x|x') <^{\log} 2(l-k) + \mathbf{I}(X : \mathcal{H})$.*

We also prove a similar result using expectation instead maximum. We define a $(k,l)$ batch $X$ to be a finite set of strings, where $k = \lceil\log|X|\rceil$, $l > k$, and for all $x \in X$, $\mathbf{E}_{x'\in X}[\mathbf{K}(x|x')] \leq l$.

**Theorem.** *For $(k,l)$ batch $X$, $\min_{x\in X}\mathbf{E}_{x'\in X}[\mathbf{K}(x|x')] <^{\log} l - k + \mathbf{I}(X : \mathcal{H})$.*

# Extending Chaitin's Incompleteness Theorem

Gödel's famous incompleteness theorem states that any theory $\mathcal{F}$ that is consistent, recursively axiomatizable, and "sufficiently rich" (contains Robinson-arithmetic $\mathcal{Q}$, or $\mathcal{Q}$ can be interpreted in it) is incomplete, in that there exists true statements that cannot be proven in it. Chaitin's incompleteness theorem proves there exist no logical means to prove lower bounds on $\mathbf{K}$. Let $\mathcal{F}$ be as above, and significantly strong to make assertions about the Kolmogorov complexity of strings. Furthermore, let $\mathcal{F}$ be sound. Then we get the celebrated theorem.

**Theorem. (Chaitin's Incompleteness Theorem)** *For theory $\mathcal{F}$, there is a constant $c$ such that $\mathcal{F}$ does not prove $c < \mathbf{K}(x)$ for any $x$.*

However this theorem doesn't prohibit the existence of formal systems that prove $c < \mathbf{K}(x)$ for a finite but very large number of strings. Or for our purposes, the above theorem doesn't prohibit theories which prove $\mathbf{K}(x) = c$ for a large (but finite) number of strings. Such theories are not to be expected to be accessible by logicians. In this chapter, we prove such systems are exotic, and cannot exist in the physical world. To do so we prove the following theorem, which states $\mathbf{K}$ is uniformly uncomputable.

**Theorem.** *A relation $X \subset \mathbb{N} \times \mathbb{N}$ of $2^n$ unique pairs $(b, \mathbf{K}(b))$ has $n <^{\log} \mathbf{I}(X; \mathcal{H})$.*

The Extended Chaitin's Incompleteness Theorem follows from the fact that any formal system that can compute $2^n$ unique pairs $(b, \mathbf{K}(b))$ has high mutual information with the halting sequence and thus is exotic and non-realizable.

# A Small Theorem for Small m

In this chapter, we show that semi measures that majorize the algorithmic probability have infinite mutual information with the halting sequence.

**Theorem.** *If $\mathbf{w}$ is a semimeasure on $\{0,1\}^*$ and $\mathbf{m} < O(1)\mathbf{w}$ then $\mathbf{I}(\mathbf{w} : \mathcal{H}) = \infty$.*

# Chapter 2

# Conventions

We use $\{0,1\}$, $\{0,1\}^*$, $\{0,1\}^\infty$ $\mathbb{W}$, $\mathbb{N}$, $\mathbb{Q}$, and $\mathbb{R}$ to denote bits, finite strings, infinite sequences, whole numbers, natural numbers, rationals, and reals, respectively. Let $X_{\geq 0}$ and $X_{>0}$ be the sets of nonnegative and positive elements of $X$. $\{0,1\}^{*\infty} = \{0,1\}^* \cup \{0,1\}^\infty$. The positive part of a real is $\lceil a\rceil^+ = \max\{a, 0\}$. For string $x \in \{0,1\}^*$, $x0^- = x1^- = x$. For $x \in \{0,1\}^*$ and $y \in \{0,1\}^{*\infty}$, we use $x \sqsubseteq y$ if there is some string $z \in \{0,1\}^{*\infty}$ where $xz = y$. We say $x \sqsubset y$ if $x \sqsubseteq y$ and $x \neq y$. The indicator function of a mathematical statement $A$ is denoted by $[A]$, where if $A$ is true, then $[A] = 1$; otherwise, $[A] = 0$. The self-delimiting code of a string $x \in \{0,1\}^*$ is $\langle x\rangle = 1^{\|x\|}0x$. The encoding of (a possibly ordered) set $\{x_1, \ldots, x_m\} \subset \{0,1\}^*$ is $\langle m\rangle\langle x_1\rangle \ldots \langle x_m\rangle$.

Probability measures $Q$ over numbers are elementary if $|\mathrm{Support}(Q)| < \infty$ and $\mathrm{Range}(Q) \subset Q_{\geq 0}$. Elementary probability measures $Q$ with $\{x_1, \ldots, x_m\} = \mathrm{Support}(Q)$ are encoded by finite strings, with $\langle Q\rangle = \langle\{x_1, Q(x_1), \ldots, x_m, Q(x_m)\}\rangle$. For the nonnegative real function $f$, we use $<^+ f$, $>^+ f$, and $=^+ f$ to denote $< f + O(1)$, $> f - O(1)$, and $= f \pm O(1)$. We also use $<^{\log} f$ and $>^{\log} f$ to denote $< f + O(\log(f+1))$ and $> f - O(\log(f+1))$, respectively.

We use a universal prefix-free algorithm $U$, where we say $U_\alpha(x) = y$ if $U$, on main input $x$ and auxiliary input $\alpha$, outputs $y$. We define Kolmogorov complexity with respect to $U$, where if $x \in \{0,1\}^*$, $y \in \{0,1\}^{*\infty}$, then $\mathbf{K}(x/y) = \min\{\|p\| : U_y(p) = x\}$. The universal probability $\mathbf{m}$ is defined as $\mathbf{m}(x/y) = \sum_p [U_y(p) = x]2^{-\|p\|}$. By the coding theorem, $\mathbf{K}(x/y) =^+ -\log \mathbf{m}(x/y)$. By the chain rule, $\mathbf{K}(x, y) =^+ \mathbf{K}(x) + \mathbf{K}(y/x, \mathbf{K}(x))$. The halting sequence $\mathcal{H} \in \{0,1\}^\infty$ is the unique infinite sequence where $\mathcal{H}[i] = [U(i) \text{ halts}]$. The information that $x \in \{0,1\}^*$ has about $\mathcal{H}$, conditional on $y \in \{0,1\}^{*\infty}$, is $\mathbf{I}(x; \mathcal{H}/y) = \mathbf{K}(x/y) - \mathbf{K}(x/\langle y, \mathcal{H}\rangle)$. $\mathbf{I}(x; \mathcal{H}) = \mathbf{I}(x; \mathcal{H}/\emptyset)$.

A continuous semi-measure is a function such that $Q(\emptyset) = 0$ and $Q(x) \geq Q(x0) + Q(x1)$. For prefix free set $D$, $Q(D) = \sum_{x \in D} Q(x)$. There exists a universal lower computable continuous semi-measure $\mathbf{M}$. The monotone complexity of a string $x$ is $\mathbf{Km}(x) = \min\{\|p\| : x \sqsubseteq U(p)\}$. This differs from the standard definition in that the universal Turing machine $U$ is used and it must halt.

This paper uses notions of stochasticity in the field of algorithmic statistics. A string $x$ is stochastic, i.e., has a low $\mathbf{Ks}(x)$ score if it is typical of a simple probability distribution. The extended deficiency of the randomness function of a string $x$ with respect to an elementary probability measure $P$ conditional on $y \in \{0,1\}^*$ is $\mathbf{d}(x|P, y) = \lfloor -\log P(x)\rfloor - \mathbf{K}(x/\langle P\rangle, y)$. $\mathbf{d}(x|P) = \lfloor -\log P(x)\rfloor - \mathbf{K}(x/\langle P\rangle)$

**Definition 1 (Stochasticity)** *For $x, y \in \{0,1\}^*$, $\mathbf{Ks}(x/y) = \min\{\mathbf{K}(P/y) + 3\log\max\{\mathbf{d}(x|P, y), 1\} :$ $P$ is an elementary probability measure$\}$. $\mathbf{Ks}(x) = \mathbf{Ks}(x/\emptyset)$.*

**Theorem 1** *For program $q$ that computes probability $p$ over $\mathbb{N}$, $\mathbf{E}_{a\sim p}\left[2^{\mathbf{I}(\langle q,a\rangle;\mathcal{H})}\right] \stackrel{*}{<} 2^{\mathbf{I}(q;\mathcal{H})}$.*

**Proof.** The goal is to prove $\sum_a p(a)\mathbf{m}(a,q/\mathcal{H})/\mathbf{m}(a,q) \overset{*}{<} \mathbf{m}(q/\mathcal{H})/\mathbf{m}(q)$. Rewriting this inequality, it suffices to prove $\sum_a \big(\mathbf{m}(q)p(a)/\mathbf{m}(a,q)\big)\big(\mathbf{m}(a,q/\mathcal{H})/\mathbf{m}(q/\mathcal{H})\big) \overset{*}{<} 1$. The term $\mathbf{m}(q)p(a)/\mathbf{m}(a,q) \overset{*}{<} 1$ because $\mathbf{K}(q)-\log p(a) >^+ \mathbf{K}(a,q)$. Furthermore, it follows directly that $\sum_a \mathbf{m}(a,q/\mathcal{H})/\mathbf{m}(q/\mathcal{H}) \overset{*}{<} 1$. □

**Theorem 2** *For partial computable $f:\mathbb{N}\to\mathbb{N}$, for all $a\in\mathbb{N}$, $\mathbf{I}(f(a);\mathcal{H}) <^+ \mathbf{I}(a;\mathcal{H})+\mathbf{K}(f)$.*

**Proof.** Observe that,

$$\begin{aligned}\mathbf{I}(a;\mathcal{H}) &= \mathbf{K}(a)-\mathbf{K}(a|\mathcal{H})\\ &>^+ \mathbf{K}(a,f(a))-\mathbf{K}(a,f(a)|\mathcal{H})-\mathbf{K}(f)\end{aligned}$$

The chain rule ($\mathbf{K}(x,y) =^+ \mathbf{K}(x)+\mathbf{K}(y|\mathbf{K}(x),x)$) applied twice results in

$$\begin{aligned}\mathbf{I}(a;\mathcal{H})+\mathbf{K}(f) &>^+ \mathbf{K}(f(a))+\mathbf{K}(a|f(a),\mathbf{K}(f(a)))-(\mathbf{K}(f(a)|\mathcal{H})+\mathbf{K}(a|f(a),\mathbf{K}(f(a)|\mathcal{H}),\mathcal{H}))\\ &=^+ \mathbf{I}(f(a);\mathcal{H})+\mathbf{K}(a|f(a),\mathbf{K}(f(a)))-\mathbf{K}(a|f(a),\mathbf{K}(f(a)|\mathcal{H}),\mathcal{H})\\ &=^+ \mathbf{I}(f(a);\mathcal{H})+\mathbf{K}(a|f(a),\mathbf{K}(f(a)))-\mathbf{K}(a|f(a),\mathbf{K}(f(a)),\mathbf{K}(f(a)|\mathcal{H}),\mathcal{H})\\ &>^+ \mathbf{I}(f(a);\mathcal{H}).\end{aligned}$$

□

**Theorem 3** *For probability $p$ over $\mathbb{N}$, computed by program $q$, $\mathbf{E}_{a\sim p}[2^{\mathbf{I}(a;\mathcal{H})}] \overset{*}{<} 2^{\mathbf{I}(q;\mathcal{H})}$.*

**Proof.** This corollary follows from Theorems 1 and 2. □

**Corollary 1** *For probability $p$ over $\mathbb{N}$, computed by program $q$, $\Pr_{a\sim p}[\mathbf{I}(a;\mathcal{H}) > \mathbf{I}(q;\mathcal{H})+m] \overset{*}{<} 2^{-m}$.*

**Proof.** This corollary follows from Theorem 3. □

It is well known in the literature that non-stochastic objects have high mutual information with the halting sequence. In the following lemma, we reprove this fact, without using left-total machines, which was used in the original proof.

**Lemma 1** $\mathbf{Ks}(x) <^{\log} \mathbf{I}(x;\mathcal{H})$.

**Proof.** We dovetail all programs to the universal Turing machine $U$. For $p\in\mathrm{Domain}(U)$, $n(p)\in\mathbb{N}$ is the position in which the program $p\in\{0,1\}^*$ terminates. Let $\Omega^n = \sum_{p:n(p)<n} 2^{-\|p\|}$ and $\Omega=\Omega^\infty$ be Chaitin's Omega. Let $\Omega^n_t$ be $\Omega^n$ restricted to the first $t$ digits. Let $x^*\in\{0,1\}^{\mathbf{K}(x)}$, with $U(x^*)=x$ with minimum $n(x^*)$. Let $k(p)=\max\{\ell : \Omega^{n(p)}_\ell = \Omega_\ell\}$ and $k=k(x^*)$. We define the elementary probability measure $Q(x)=\max\{2^{-\|p\|+k} : k(p)=k, U(p)=x\}$, $Q(\emptyset)=$

$1 - Q(\{0,1\}^* \setminus \{\emptyset\})$.

$$\begin{aligned}
\mathbf{d}(x|Q) &= -\log Q(x) - \mathbf{K}(x|Q) <^+ (\mathbf{K}(x) - k) - \mathbf{K}(x|\Omega_k) \\
&<^+ (\mathbf{K}(x|\Omega_k) + \mathbf{K}(\Omega_k) - k) - \mathbf{K}(x|\Omega_k) <^+ (k + \mathbf{K}(k)) - k \\
&<^+ \mathbf{K}(k).
\end{aligned}$$

$$\begin{aligned}
\mathbf{K}(x|\mathcal{H}) &<^+ \ \mathbf{K}(x|Q) + \mathbf{K}(Q|\mathcal{H}) <^+ \mathbf{K}(x|Q) + \mathbf{K}(\Omega_k|\mathcal{H}) \\
&<^+ -\log Q(x) + \mathbf{K}(k) <^+ (\mathbf{K}(x) - k) + \mathbf{K}(k) \\
k &<^{\log} \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H})
\end{aligned}$$

$$\mathbf{Ks}(x) <^+ \mathbf{K}(Q) + O(\log\max\{\mathbf{d}(x|P), 1\}) <^+ k + O(\mathbf{K}(k)) <^{\log} \mathbf{I}(x;\mathcal{H}).$$

□

# Chapter 3

# Classification

A binary predicate is defined to be a function of the form $f : D \to \{0,1\}$, where $D \subseteq \mathbb{N}$. We say that binary predicate (or finite string) $\lambda$ is an extension of $\gamma$, if for all $i \in \mathrm{Dom}(\gamma)$, $\gamma(i) = \lambda(i)$. If a binary predicate has a domain of $\mathbb{N}$ and is an extension of binary predicate $\gamma$, then we say it is a complete extension of $\gamma$. The self-delimiting code for a binary predicate $\gamma$ with a finite domain is $\langle\{x_1, \lambda(x_1), \ldots, x_n, \lambda(x_n)\}\rangle$. The Kolmogorov complexity of a binary predicate $\lambda$ with an infinite sized domain is $\mathbf{K}(\lambda) = \mathbf{K}(f)$, where $f : \mathbb{N} \to \mathbb{N}$ is a partial computable function where $f(i) = \lambda(i)$ if $i \in \mathrm{Dom}(\lambda)$ and $f(i)$ is undefined otherwise. If there is no such partial computable function, then $\mathbf{K}(\lambda) = \infty$. In this chapter, we assume the universal Turing machine $U$, is "left-total", which is introduced in Chapter 7.

**Theorem 4 (EL Theorem [Eps19, Lev16])** *For set $D \subset \mathbb{N}$, $\min_{x \in D} \mathbf{K}(x) <^{\log} -\log \mathbf{m}(D) + \mathbf{I}(D;\mathcal{H})$.*

**Theorem 5** *For binary predicate $\gamma$ and the set $\Gamma$ of complete extensions of $\gamma$,*
*$\min_{g \in \Gamma} \mathbf{K}(g) <^{\log} |\mathrm{Dom}(\gamma)| + \mathbf{I}(\langle\gamma\rangle{:}\mathcal{H})$.*

**Proof.** We recall that $\mathbf{bb}(b) = \max\{U(p) : p \lhd b, \text{ or } p \sqsupseteq b\}$ is the largest number produced by a program that extends or is to the left of $b$. The theorem is meaningless if $|\mathrm{Dom}(\gamma)| = \infty$, so we can assume $q = |\mathrm{Dom}(\gamma)| < \infty$. Let $n = \max\{i : i \in \mathrm{Dom}(\gamma)\}$. Let $b$ be the shortest total string where $\mathbf{bb}(b) \geq n$. Let $N = \mathbf{bb}(b)$.

Let $D$ be the set of all strings of length $N$, that extends $\gamma$. Theorem 15, relative to $b$, gives $a \in D$ with

$$\mathbf{K}(a|b) <^{\log} -\log \mathbf{m}(D|b) + \mathbf{I}(D;\mathcal{H}|b)$$
$$\mathbf{K}(a) <^{\log} q + \mathbf{K}(b) + \mathbf{I}(D;\mathcal{H}|b).$$

It must be that $\mathbf{K}(b|D, \|b\|) = O(1)$ as there is a program that can enumerate, from the left, total strings of length $\|b\|$. This program returns the first total string $b'$ such that $D \subset \{0,1\}^{\mathbf{bb}(b')}$. This $b'$ is equal to $b$, otherwise $b' \lhd b$ and thus $\mathbf{bb}(b'^{-}) \geq \mathbf{bb}(b') \geq n$, contradicting the definition of $b$. Applying Lemma 3,

$$\mathbf{K}(a) <^{\log} q + \mathbf{I}(D;\mathcal{H}) + \mathbf{K}(b|D, \|b\|)$$
$$\mathbf{K}(a) <^{\log} q + \mathbf{I}(D;\mathcal{H}).$$

Furthermore, it must be $\mathbf{K}(D|\langle\gamma\rangle) <^+ \mathbf{K}(\|b\|)$ using the same reasoning as above. So, using lemma 2, and the fact that the left hand side $>^+ \mathbf{K}(b)$, where $b$ is a random string,

$$\mathbf{K}(a) <^{\log} q + \mathbf{I}(\langle\gamma\rangle;\mathcal{H}) + \mathbf{K}(\|b\|)$$
$$\mathbf{K}(a) <^{\log} q + \mathbf{I}(\langle\gamma\rangle;\mathcal{H})$$

Thus there exists a complete extension $g' \in \Gamma$, of $\gamma$, that is equal to $a[i]$ for all $i \leq \|a\|$, and 0 otherwise. This $g'$ can be computed with a program of size $<^+ \mathbf{K}(a)$, and thus,

$$\min_{g\in\Gamma} \mathbf{K}(g) \leq \mathbf{K}(g') <^+ \mathbf{K}(a) <^{\log} |\mathrm{Dom}(\gamma)| + \mathbf{I}(\gamma : \mathcal{H}).$$

□

# Chapter 4

# Regression

## 4.1 Introduction

One central area of machine learning is regression, in which one is given a set of pairs $\{(x_i, y_i)\}$, $i = 1 \dots n$, and the goal is to find a function $f$, such that $f(x_i) = y_i$. Usually each $x_i$ and $y_i$ represents a point in Euclidean space, but for our purposes they are natural numbers. The goal is to use Occam's razor to find the simpliest function, to prevent overfitting to the random noise inherent in the sample data. This chapter presents the following bounds on the simplest total computable function completely consistent with the data.

**Theorem.** *For* $\{(x_i, y_i)\}_{i=1}^n$, *there exists* $f : \mathbb{N} \to \mathbb{N}$ *with* $f(x_i) = y_i$ *for* $i \in \{1, \dots, n\}$ *and* $\mathbf{K}(f) <^{\log} \sum_{i=1}^n \mathbf{K}(y_i|x_i) + \mathbf{I}(\{(x_i, y_i)\}; \mathcal{H})$.

## 4.2 Conventions

For positive real functions $f$, by $<^+ f$, $>^+ f$, $=^+ f$, and $<^{\log} f$, $>^{\log} f$, $\sim f$ we denote $\leq f + O(1)$, $\geq f - O(1)$, $= f \pm O(1)$ and $\leq f + O(\log(f+1))$, $\geq f - O(\log(f+1))$, $= f \pm O(\log(f+1))$. $\mathbf{K}(x|y)$ is the conditional prefix Kolmogorov complexity. The chain rule states $\mathbf{K}(x, y) =^+ \mathbf{K}(x) + \mathbf{K}(y|\mathbf{K}(x), x)$. Let $[A] = 1$ if the mathematical statement $A$ is true, otherwise $[A] = 0$. Let $\mathbf{K}_t(x|y) = \inf\{\|p\| : U_y(p) = x \text{ in } t \text{ steps}\}$. The information the halting sequence $\mathcal{H}$ has about $x$ is $\mathbf{I}(x; \mathcal{H}|y) = \mathbf{K}(x|y) - \mathbf{K}(x|y, \mathcal{H})$. $\mathbf{I}(x; \mathcal{H}) = \mathbf{I}(x; \mathcal{H}|\emptyset)$. A probability measure is elementary if its support is finite and it has rational values. The deficiency of randomness of $x \in \{0,1\}^*$ with respect to elementary probability measure $Q$ is $\mathbf{d}(X|Q) = \lceil -\log Q(X) - \mathbf{K}(x|\langle Q \rangle) \rceil$. The stochasticity of $x$ is $\mathbf{Ks}(x) = \min_Q \mathbf{K}(Q) + 3 \log \max\{\mathbf{d}(X|Q), 1\}$.

## 4.3 Results

Let $\Omega = \sum\{2^{-\|p\|} : U(p) \text{ halts}\}$ be Chaitin's Omega, $\Omega_n \in \mathbb{Q}_{\geq 0}$ be be the rational formed from the first $n$ bits of $\Omega$, and $\Omega^t = \sum\{2^{-\|p\|} : U(p) \text{ halts in time } t\}$. For $n \in \mathbb{N}$, let $\mathbf{bb}(n) = \min\{t : \Omega_n < \Omega^t\}$. $\mathbf{bb}^{-1}(m) = \arg\min_n\{\mathbf{bb}(n-1) < m \leq \mathbf{bb}(n)\}$. Let $\Omega[n] \in \{0,1\}^*$ be the first $n$ bits of $\Omega$.

**Lemma 2** *For* $n = \mathbf{bb}^{-1}(m)$, $\mathbf{K}(\Omega[n]|m, n) = O(1)$.

**Proof.** For a string $x$, let $BB(x) = \inf\{t : \Omega^t > 0.x\}$. Enumerate strings of length $n$, starting with $0^n$, and return the first string $x$ such that $BB(x) \geq m$. This string $x$ is equal to $\Omega[n]$, otherwise let $y$ be the largest common prefix of $x$ and $\Omega[n]$. Thus $BB(y) = \mathbf{bb}(\|y\|) \geq BB(x) \geq m$, which means $\mathbf{bb}^{-1}(m) \leq \|y\| < n$, causing a contradiction. $\square$

**Theorem 6** *For* $\{(x_i, y_i)\}_{i=1}^n$, *there exists* $f : \mathbb{N} \to \mathbb{N}$ *with* $f(x_i) = y_i$ *for* $i \in \{1, \ldots, n\}$ *and* $\mathbf{K}(f) <^{\log} \sum_{i=1}^n \mathbf{K}(y_i|x_i) + \mathbf{I}(\{(x_i, y_i)\}; \mathcal{H})$.

**Proof.** Let $S = \{(x_i, y_i)\}$. Let $K = \sum_{i=1}^n \mathbf{K}(y_i|x_i)$. We have $T = \arg\min_t \sum_{i=1}^n \mathbf{K}_t(y_i|x_i) = K$. Let $N = \mathbf{bb}^{-1}(T)$ and $M = \mathbf{bb}(N)$ and we define $m(x|y) = 2^{-\mathbf{K}_M(x|y)}$, setting $m(\emptyset|y) = 1 - m(\mathbb{N}|y)$.

We condition all terms on $M$ and $K$, and later in the proof, we'll make this condition explicit. Let $Q$ be an elementary probability that realizes the stochasticity of $S$, where $d = \max\{\mathbf{d}(S|Q), 1\}$. Without loss of generality, we can assume the support of $Q$ consists entirely of samples $R = \{(x_j, y_j)\}_{j=1}^{n_R}$ (of potentially different sizes) such that $\prod_{j=1}^{n_R} m(y_j|x_j) = 2^{-M}$. Let

$$z = \max\{x : (x, y) \in R \in \text{Support}(Q)\}.$$

We define a probability measure $\kappa$ over $d2^K$ lists $\mathcal{L}$ of size $z$ over $\mathbb{N}$, where each $\ell \in \mathcal{L}$ is chosen independently, and for each $\ell \in \mathcal{L}$, $\ell(i)$ is chosen independently according to $m(\cdot|i)$. We say a sample $R = \{(x_j, y_j)\}$ is inconsistent with a list $\ell$, $R \ltimes \ell$, if there exists $j$, where $\ell(x_j) \neq y_j$. $\eta(R, \mathcal{L}) = [\forall \ell \in \mathcal{L}, E \ltimes \ell]$.

$$\mathbf{E}_{\mathcal{L}\sim\kappa}\mathbf{E}_{R\sim Q}[\eta(R, \mathcal{L})] = \mathbf{E}_{\{(x_j,y_j)\}\sim Q}\left(1 - \prod_j m(y_j|x_j)\right)^{d2^K} < \mathbf{E}_{R\sim Q}e^{-d} = e^{-d}.$$

Thus there exists a set of $d2^K$ lists $\mathcal{L}$, where $\mathbf{E}_{R\sim Q}[\eta(R, \mathcal{L})] < e^{-d}$. Thus let $t(R) = \eta(R, \mathcal{L})e^d$ be a $Q$-test, where $\mathbf{E}_{R\sim Q}[t(R)] \leq 1$. It must be $t(S) = 0$, otherwise we have

$$1.44d \leq \log t(S) <^+ \mathbf{d}(S|Q) <^+ d,$$

which is contradiction for large $d$, which we can assume without loss of generality. So there exists a list $\ell$ such that $\ell(x_i) = y_i$, for all $(x_i, y_i) \in S$. Thus one can construct a total computable function $f : \mathbb{N} \to \mathbb{N}$ from $\ell$ that is consistent with $S$, for example $f(x) = \ell(x)$ if $x \leq z$ and $f(x) = 1$ otherwise. Making the condition term $M$ explicit and keeping the condition term $K$ implicit we have,

$$\begin{aligned}
\mathbf{K}(f|M) &<^+ \mathbf{K}(\ell|M) \\
&<^+ \log|\mathcal{L}| + \mathbf{K}(\mathcal{L}|M) \\
&<^+ K + \log d + \mathbf{K}(Q, d|M) \\
&<^+ K + \mathbf{Ks}(S|M).
\end{aligned}$$

Using Lemma 1, we get, noting $M = \mathbf{bb}(N)$, and $\mathbf{bb}$ is computable relative to $\mathcal{H}$,

$$\begin{aligned}
\mathbf{K}(f|M) &<^{\log} K + \mathbf{I}(S; \mathcal{H} \mid M). \\
\mathbf{K}(f) &<^{\log} K + \mathbf{K}(S|M) + \mathbf{K}(M) - \mathbf{K}(S|\mathcal{H}) + \mathbf{K}(N).
\end{aligned}$$

So we have,

$$
\begin{aligned}
&\mathbf{K}(S|M)+\mathbf{K}(M)\\
&<^{+}\mathbf{K}(S|M,\mathbf{K}(M))+\mathbf{K}(\mathbf{K}(M)|M)+\mathbf{K}(M)\\
&<^{+}\mathbf{K}(S,M)+\mathbf{K}(\mathbf{K}(M)|M) && (4.1)\\
&<^{+}\mathbf{K}(S,N,M)+O(\log N) && (4.2)\\
&<^{+}\mathbf{K}(S,N)+O(\log N). && (4.3)\\
&<^{+}\mathbf{K}(S)+O(\log N).\\
\mathbf{K}(f) &<^{\log} K+\mathbf{K}(S)-\mathbf{K}(S|\mathcal{H})+O(\log N). && (4.4)
\end{aligned}
$$

Equation 9.2 is from the chain rule. Equation 9.3 is from the fact that $M=\mathbf{bb}(N)$. Equation 9.4 comes $\mathbf{K}(T|S,K)=O(1)$ and Lemma 7, which implies $\mathbf{K}(M|N,T)<^{+}\mathbf{K}(\Omega[N]|N,T)<^{+}O(1)$.

From $K$, and $S$, one can compute $T$, where $\mathbf{bb}^{-1}(T)=N$. Therefore by Lemma 7, $\mathbf{K}(\Omega[N]|S)<^{+}\mathbf{K}(N)$, so by Lemma 2,

$$N<^{\log}\mathbf{I}(\Omega[N];\mathcal{H})<^{\log}\mathbf{I}(S;\mathcal{H})+\mathbf{K}(N)<^{\log}\mathbf{I}(S;\mathcal{H}). \qquad (4.5)$$

The above equation used the common fact that the first $n$ bits of $\Omega$ had $n-O(\log n)$ bits of mutual information with $\mathcal{H}$. So combining Equations 9.5 and 9.6, we get

$$\mathbf{K}(f)<^{\log}K+\mathbf{I}(S;\mathcal{H}).$$

The proof is completed by noting the log precision, and the $K$ term in the equation removes the implicit conditioning of $K$. □

# Chapter 5

# Monotone EL Theorem

The EL Theorem states that the algorithmic probability of a non-exotic set is concentrated on its simpliest element, with

$$\min_{x\in D}\mathbf{K}(x) <^{\log} -\log \mathbf{m}(D) + \mathbf{I}(\langle D\rangle : \mathcal{H}).$$

There exists a montone variant to this theorem, using $\mathbf{M}$ instead of $\mathbf{m}$ and $\mathbf{Km}$ instead of $\mathbf{K}$. The two results are related, but neither one is readily entailed by the other. Chapter 3 is a direct consequence to the main theorem of this chapter.

## 5.1 Open Sets

The Kolmogorov complexity of an infinite sequence $\alpha \in \{0,1\}^\infty$ is $\mathbf{K}(\alpha)$, the size of the smallest program to a universal Turing machine that will output, without halting, $\alpha$ on the output tape. The uniform measure is $\mu(x) = 2^{-\|x\|}$.

**Theorem 7** *For clopen set $C \subseteq \{0,1\}^\infty$, $\min_{\alpha\in C}\mathbf{K}(\alpha) <^{\log} -\log\mu(C) + \mathbf{I}(\langle C\rangle;\mathcal{H})$.*

**Proof.** Let $s = \lceil -\log\mu(C)\rceil$ and we relativize the universal Turing machine $U$ to $s$. Let $P$ be an elementary probabilty measure that realizes $\mathbf{Ks}(\langle C\rangle)$. Let $n = \max\{\|x\| : x \in W \subset \{0,1\}^*, \langle W\rangle \in \mathrm{Supp}(P)\}$. Let $d = \max\{\mathbf{d}(\langle C\rangle|P), 1\}$ and $c \in \mathbb{N}$ be a constant to be chosen later. Let $\kappa$ be the uniform probability measure over lists $L$ of $cd2^{s+1}$ strings of length $n$, where $\kappa(L) = 2^{-ncd2^{s+1}}$. Let $t_L(\langle W\rangle)$ be a function, parameterized by a list $L \subseteq \{0,1\}^n$, over encoded clopen sets $W \subseteq \{0,1\}^\infty$, with $t_L(\langle W\rangle) = [\mu(W) \geq 2^{-s}, W \trianglelefteq L = \emptyset]e^{cd}$.

$$\begin{aligned}
\mathbf{E}_{L\sim\kappa}\mathbf{E}_{\langle W\rangle\sim P}[t_L(\langle W\rangle)] &\leq \sum_{\text{clopen } W\subseteq\{0,1\}^\infty} P(\langle W\rangle)\left(1-2^{-s}\right)^{cd2^{s+1}} e^{cd}\\
&\leq e^{-2^{-s}cd2^{s+1}}e^{cd} = e^{-cd}\\
&< 1.
\end{aligned}$$

Thus there exists a list $L$ of $cd2^{s+1}$ strings such that $\mathbf{E}_{\langle W\rangle\sim P}[t_L(\langle W\rangle)] < 1$. This $L$ can be found with brute force search, with $\mathbf{K}(L|c,d,P) = O(1)$. It must be that $C \trianglelefteq L \neq \emptyset$. Otherwise $t_L(\langle C\rangle) = e^{cd}$ and since $t_L(\cdot)P(\cdot)$ is a semi-measure, for large enough $c$ solely dependent on the universal Turing machine $U$, a contradiction occurs, with

$$\begin{aligned}
\mathbf{K}(C|c,d,\langle P\rangle) &< -\log t_L(\langle C\rangle)P(\langle C\rangle)+O(1)\\
\mathbf{K}(C|c,d,\langle P\rangle) &< -\log P(\langle C\rangle)-(\lg e)cd+O(1)\\
(\lg e)cd &< -\log P(\langle C\rangle)-\mathbf{K}(C|s,\langle P\rangle)+\mathbf{K}(d,c)+O(1)\\
(\lg e)cd &< d+\mathbf{K}(d,c)+O(1).
\end{aligned}$$

So there exists $x\in C\trianglelefteq L$, with

$$\begin{aligned}
\mathbf{K}(x) &<^{+} \log\left|L\right|+\mathbf{K}(L)\\
&<^{+} \log\left|L\right|+\mathbf{K}(d,P)\\
&<^{+} \log d+s+\mathbf{K}(d)+\mathbf{K}(P)\\
&< s+\mathbf{Ks}(\langle C\rangle).
\end{aligned}$$

Since $x\in C\trianglelefteq L$, $\Gamma_x\subseteq C$. Thus there is a program $g$ that outputs $x$ and then an infinite sequence of 0's. Since $x0^\infty\in C$ and $\|g\|<^{+}\mathbf{K}(x)$, and using Lemma 1,

$$\begin{aligned}
\min_{\alpha\in C}\mathbf{K}(\alpha)\leq\|g\|<^{+}\mathbf{K}(x) &< s+\mathbf{Ks}(\langle C\rangle)\\
&<^{\log} s+\mathbf{I}(\langle C\rangle;\mathcal{H}).
\end{aligned}$$

□

Theorem 7 can be generalized to arbitrary open sets of the Cantor space. Such sets $S$ can have encodings $\langle S\rangle$ that are infinite sequences. We recall that the information term between infinite sequences is $\mathbf{I}(\alpha:\beta)=\log\sum_{x,y\in\{0,1\}^*}\mathbf{m}(x|\alpha)\mathbf{m}(y|\beta)2^{\mathbf{I}(x:y)}$.

**Theorem 8** *For open set $S\subseteq\{0,1\}^\infty$, $\min_{\alpha\in S}\mathbf{K}(\alpha)<^{\log}-\log\mu(S)+\mathbf{I}(\langle S\rangle:\mathcal{H})$.*

**Proof.** Let $s=\lceil-\log\mu(S)\rceil$. Let $\{x_i\}_{i=1}^n=\{x:\Gamma_x\text{ is maximal in }S\}$, with $n\in\mathbb{N}\cup\infty$. Let $N\in\mathbb{N}$ be the smallest number such that $\sum_{i=1}^N 2^{-\|x_i\|}>2^{-s-1}$. Let $C=\bigcup_{i=1}^N\Gamma_{x_i}$ be a clopen set with $C\subseteq S$. By Theorem 7,

$$\min_{\alpha\in C}\mathbf{K}(\alpha)<^{\log} s+\mathbf{I}(\langle C\rangle;\mathcal{H}). \tag{5.1}$$

Based on the definition of $\mathbf{I}$:

$$\begin{aligned}
\mathbf{I}(\langle C\rangle;\mathcal{H}) &<^{+} \mathbf{I}(\langle S\rangle:\mathcal{H})+\mathbf{K}(\langle C\rangle|\langle S\rangle)\\
&<^{+} \mathbf{I}(\langle S\rangle:\mathcal{H})+\mathbf{K}(s).
\end{aligned}$$

So

$$\min_{\alpha\in S}\mathbf{K}(\alpha)<^{\log} s+\mathbf{I}(\langle S\rangle:\mathcal{H}).$$

□

## 5.2 Algorithmic Monotone Probability of Sets

A total computable function $\nu:\{0,1\}^*\to\{0,1\}^*$ is prefix-monotonic iff for all strings $x$ and $y$, $\nu(x)\sqsubseteq\nu(xy)$. Let $\overline{\nu}:\{0,1\}^*\cup\{0,1\}^\infty\to\{0,1\}^*\cup\{0,1\}^\infty$ be used to represent the unique extension of $\nu$ to infinite sequences. Its definition for all $\alpha\in\{0,1\}^*\cup\{0,1\}^\infty$ is $\overline{\nu}(\alpha)=\sup\{\nu(\alpha_{\leq n}):n\leq\|\alpha\|\}$, where the supremum is respect to the partial order derived with the $\sqsubseteq$ relation. The following theorem relates prefix monotone machines and continuous semi-measures. It is equivalent to Theorem 4.5.2 in [LV08], with the simple modification that the machine be total computable.

**Theorem 9** *For each lower-computable continuous semi-measure $\sigma$ over $\{0,1\}^\infty$, there is a prefix-monotonic function $\nu_\sigma$, where for prefix free $G\subset\{0,1\}^*$, $\lceil-\log\sigma(G)\rceil=^+\lceil-\log\mu\{\alpha{:}\overline{\nu_\sigma}(\alpha)\sqsupseteq x\in G\}\rceil$.*

Since there is a universal lower-semicomputable continuous semi-measure $\mathbf{M}$, there exists a prefix-monotonic function $\nu_{\mathbf{M}}$, with the following property.

**Corollary 2** *For finite prefix free set $G$, $-\log\mathbf{M}(G)=^+-\log\mu\{\alpha:x\sqsubseteq\overline{\nu_{\mathbf{M}}}(\alpha),\,\alpha\in\{0,1\}^\infty,x\in G\}$.*

The following corollary is equivalent to Theorem 7 in terms of finite strings instead of clopen sets. For finite prefix free set $G\subset\{0,1\}^*$, $\mu(G)=\sum_{x\in G}2^{-\|x\|}$.

**Corollary 3** *For finite prefix free $G\subset\{0,1\}^*$, $s=\lceil-\log\mu(G)\rceil$, and $h=\mathbf{I}(G;\mathcal{H})$, we have $\min_{y\sqsupseteq x\in G}\mathbf{K}(y)<^{\log}s+h$.*

**Theorem 10** *For finite prefix-free set $G\subset\{0,1\}^*$, $\min_{x\in G}\mathbf{Km}(x)<^{\log}-\log\mathbf{M}(G)+\mathbf{I}(G;\mathcal{H})$.*

**Proof.** Let $i=\lceil-\log\mathbf{M}(G)\rceil$ and $h=\mathbf{I}(G;\mathcal{H})$. Due to Theorem 9, there exists a finite prefix-free set $F\subset\{0,1\}^*$ such that

1. $-\log\mu(F)\leq i+1$,
2. for all $x\in F$, $\nu_{\mathbf{M}}(x)\sqsupseteq z\in G$,
3. $\mathbf{K}(F|G)<^+\mathbf{K}(i)$.

By Corollary 3, there exists $y\sqsupseteq x\in F$, with $\mathbf{K}(y)<^{\log}i+h'$, where $h'=\mathbf{I}(F;\mathcal{H})$. Using Lemma 2, we have that $\mathbf{K}(y)<^{\log}i+h$, noting that $\mathbf{K}(F|G)<^+\mathbf{K}(i)$. Thus there is a program $p$ of length $<^+\mathbf{K}(y)$ that computes $y$ and then outputs $\nu_{\mathbf{M}}(y)\sqsupseteq\nu_{\mathbf{M}}(x)\sqsupseteq z\in G$. So $\mathbf{Km}(G)<^+\|p\|<^+\mathbf{K}(y)<^{\log}i+h$. □

**Corollary 4** *For (potentially infinite) prefix-free set $G\subset\{0,1\}^*$, $\min_{x\in G}\mathbf{Km}(x)<^{\log}-\log\mathbf{M}(G)+\mathbf{I}(\langle G\rangle:\mathcal{H})$.*

The proof of this corollary follows analogously to the proof of Theorem 8, except $\mathbf{M}$ is used instead of $\mu$. □

# Chapter 6

# Probabilities are Balanced

## 6.1 Introduction

It has been proven that large sets of strings are exotic if they all have similar complexities. By exotic, we mean their encoding has high mutual information with the halting sequence. Similarly if one probability over infinite strings gives large measure to sequences with low deficiency of randomness with respect to a second probability, then it is exotic. In this chapter, we look at probabilities over strings of length $n$, and prove that they must give measure to simple strings. We first prove a simple bound. The main result is the tighter bound. This result also appears in the black holes section of the Algorithmic Physics manuscript at http://www.jptheorygroup.org.

**Proposition 1 (Simple Bound)** *There is a $c$ where for probability $p$ over $\{0,1\}^n$, for all $m > \mathbf{K}(p) + c$, $p\{x : \mathbf{K}(x) < m\} > 2^{m-2\mathbf{K}(m,p)-n-c}$.*

**Proof.** Order strings $x$ of size $n$ by $p(x)$ value, with largest values first, and breaking ties through any simple ordering on $\{0,1\}^n$. It must be the first $2^\ell$ strings $X$ has $p(X) \geq 2^{\ell-n-1}$ Otherwise the average value of $p(x)$, $x \in X$, is less than $2^{-n-1}$. Thus for the remaining $2^n - 2^\ell$ strings $Y$, $p(y) < 2^{-n-1}$, So

$$\begin{aligned} p(\{0,1\}^n) &= p(X) + P(Y) \\ &< 2^{\ell-n-1} + (2^n - 2^\ell)(2^{-n-1}) \\ &= 2^{\ell-n-1} + 2^{-1} - 2^{\ell-n-1} \\ &= 1/2, \end{aligned}$$

which is a contradiction. Furthermore, the first $2^\ell$ elements $x$ have complexity $\mathbf{K}(x|p) <^+ \ell + \mathbf{K}(\ell)$ or $\mathbf{K}(x) <^+ \mathbf{K}(p,\ell) + \ell$. Let $m = \ell + \mathbf{K}(\ell,p) + O(1)$. By Proposition 3, $m - 2\mathbf{K}(m,p) <^+ \ell$. $\square$

**Proposition 2** *For every $c, n \in \mathbb{N}$, there exists $c' \in \mathbb{N}$ where for all $a, b \in \mathbb{N}$, if $a < b + n\log a + c$ then $a < b + 2n\log b + c'$.*

**Proof.**

$$\begin{aligned} \log a &< \log b + \log\log a + \log cn \\ 2\log a - 2\log\log a &< 2\log b + 2\log cn \\ \log a &< 2\log b + 2\log dn. \end{aligned}$$

Combining with the original inequality

$$
\begin{aligned}
a &< b + n\log a + c\\
a &< b + n(2\log b + 2\log dn) + c\\
&= y + 2n\log y + c',
\end{aligned}
$$

where $c' = 2n\log cn + c$. □

**Proposition 3** *For all $d \in \mathbb{N}$ there is a $d' \in \mathbb{N}$ where if $x+\mathbf{K}(x,z)+d > y$ then $x+d' > y-2\mathbf{K}(y,z)$.*

**Proof.** If $x + d > y$, then the lemma is satisfied, so $x + f \leq d$. Thus $y - x < \mathbf{K}(x,z) + d$ implies $\mathbf{K}(y-x) <^+ 2\log\mathbf{K}(x,z)+2\log d$. Thus $\mathbf{K}(x,z) <^+ \mathbf{K}(y,z)+\mathbf{K}(y-x) <^+ \mathbf{K}(y,z)+2\log\mathbf{K}(x,z)+2\log d$. Applying Proposition 2, where $a = (x,z)$, $b = (y,z)$ and $c = 2\log d + O(1)$ and $n = 2$, we get a $c'$ dependent on $c$ and $n$ where $\mathbf{K}(x,z) < \mathbf{K}(y,z) + 4\log\mathbf{K}(y,z) + c' < 2\mathbf{K}(y,z) + c' + O(1)$. So

$$
\begin{aligned}
x + \mathbf{K}(x,z) + d &> y\\
x + (2\mathbf{K}(y,z) + d' + O(1)) + d &> y\\
x + d'' &> y - 2\mathbf{K}(y,z),
\end{aligned}
$$

where $d'' = d' + O(1) + d$. □

## 6.2 Tighter Bound

**Theorem 11 ([Eps23])** *For probability $p$ over $\{0,1\}^*$, $D \subset \mathbb{N}$, $|D| = 2^s$, $s < \max_{a\in D}\mathbf{d}(a|p) + \mathbf{I}(D;\mathcal{H}) + O(\log\mathbf{I}(D;\mathcal{H})) + \mathbf{K}(s) + O(\log\mathbf{K}(s,p))$.*

**Theorem 12** *There is a $c \in \mathbb{N}$ where for probability $p$ over $\{0,1\}^n$, for $m > \mathbf{K}(p) + c$, $p\{x : \mathbf{K}(x) < m\} > 2^{m-n-2\mathbf{I}(p;\mathcal{H})-O(\mathbf{K}(n,m))-c}$.*

**Proof.** Without loss of generality, $p$ can be assumed to have a range in powers of 2. Assume not, then there exist $\ell \in (\mathbf{K}(p) + c, n)$ such that $p\{x : \mathbf{K}(x) \leq \ell\} < 2^{-k}$, where $k = n - \ell - c - 2\mathbf{I}(p;\mathcal{H}) - O(\mathbf{K}(n,\ell))$ and $c$ solely depends on the universal Turing machine. $\mathbf{K}(k) <^+ \mathbf{K}(n,\ell,c,\mathbf{I}(p;\mathcal{H}),\mathbf{K}(n,\ell))$. Suppose $\max\{p(x) : \mathbf{K}(x) > \ell\} \geq 2^{-k}$. Then

$$
\mathbf{K}(p) + O(1) > \mathbf{K}\left(\arg\max_x p(x)\right) > \ell > \mathbf{K}(p) + c,
$$

causing a contradiction, for choice of $c$ dependent on $U$. Sample $2^{k-2}$ elements $D$ without replacement according to $p$. $p^*$ is the probability of $D$, where $\mathbf{K}(p^*) <^+ \mathbf{K}(p,n,\ell,c,\mathbf{I}(p;\mathcal{H}),\mathbf{K}(n,\ell))$. Even if every element $x$ chosen has $p(x) = 2^{-k-1}$, the total $p$ mass sampled is not greater than

$$
2^{k-1}2^{k-2} \leq 2^{-3}.
$$

The probability $q$ that all $x \in D$ has $\mathbf{K}(x) > \ell$ is

$$
q > \left(1 - 2^{-k}/(1 - 2^{-3} + 2^{-k})\right)^{2^{k-2}} > \left(1 - 2^{k+1}\right)^{2^{k-2}} = 1/2.
$$

Thus, by Theorems 2 and 3,

$$\Pr_{S\sim p^*}[\mathbf{I}(S;\mathcal{H}) > \mathbf{I}(p^*;\mathcal{H}) + m] \stackrel{*}{<} 2^{-m},$$

$$\Pr_{S\sim p^*}[\mathbf{I}(S;\mathcal{H}) > \mathbf{I}((p,n,\ell,c,\mathbf{I}(p;\mathcal{H}),\mathbf{K}(n,\ell));\mathcal{H}) + m] \stackrel{*}{<} 2^{-m}.$$

So by probabilistic arguments, there exists $D \subset \{0,1\}^n$, where for all $x \in D$, $\mathbf{K}(x) > \ell$ and

$$\mathbf{I}(D;\mathcal{H}) <^+ \mathbf{I}(p^*;\mathcal{H}) <^+ \mathbf{I}((p,\ell,c,\mathbf{I}(p;\mathcal{H}),\mathbf{K}(n,\ell));\mathcal{H}) <^+ \mathbf{I}(p;\mathcal{H}) + \mathbf{K}(\ell,n,\mathbf{I}(p;\mathcal{H}),\mathbf{K}(n,\ell),c).$$

So by Theorem 11, applied to $D$ and the uniform measure $U_n$ over strings of length $n$,

$$\begin{aligned} k &< \max_{a\in D}\mathbf{d}(a|U_n) + \mathbf{I}(D;\mathcal{H}) + O(\log\mathbf{I}(D;\mathcal{H})) + \mathbf{K}(k) + O(\log\mathbf{K}(U_n,k)) \\ n-\ell+\mathbf{K}(n)+c+O(\mathbf{K}(\ell,n))+2\mathbf{I}(p;\mathcal{H}) &< n-\ell+\mathbf{I}(p;\mathcal{H})+O(\mathbf{K}(\ell,n,\mathbf{K}(n,\ell),c))+O(\log(\mathbf{I}(p;\mathcal{H})) \\ c &< O(\mathbf{K}(c)). \end{aligned}$$

which is a contradiction for large enough $c$ dependent solely on the universal Turing machine $U$.□

# Chapter 7

# Kolmogorov Birthday Paradox

## 7.1 Introduction

We prove a Kolmogorov complexity version of the birthday paradox. If you randomly select $2^{n/2}$ strings of length $n$, then, with overwhelming probability, you will have selected at least two strings $x$ and $y$ with low $\mathbf{K}(x|y)$. This is true for all probabilities with low mutual information with the halting sequence. The function $\mathbf{K}$ is the prefix-free Kolmogorov complexity.

To prove this fact, we first prove an interesting property about bunches of finite strings. A $(k,l)$-bunch is a finite set of strings $X$ where $l > \max_{x,y\in X}\mathbf{K}(y|x)$ and $2^k < |X|$. Bunches were introduced in [Rom03], but we use a slightly different definition. Although bunches have only two parameters, they exhibit many interesting properties. Both [Rom03] and [Rom22] proved the existence of strings that are simple to each member of the bunches. That is, there exists a string $z$ such that $\mathbf{K}(z|x) < O(l-k) + \mathbf{K}(l)$ and $\mathbf{K}(x|z) < l + O(l-k) + \mathbf{K}(l)$, for all $x \in X$. In [Eps21c], it was proven that each bunch has a member that is simple relative to all members of the bunch, similar to the above definition. If not, then the bunch has high mutual information with the halting sequence. The mutual information between a string and the halting sequence is $\mathbf{I}(x;\mathcal{H}) = \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H})$. We prove that if a nonexotic bunch $X$ has many members and low $\max_{x,y\in X,x\neq y}\mathbf{K}(y|x)$, then it will have two elements $x, y$ with very low $\mathbf{K}(y|x)$. A string (or any object that it is represented by) is exotic if it has high mutual information with the halting sequence.

**Theorem.** *For $(k,l)$-bunch $X$, $\min_{x,y\in X,x\neq y}\mathbf{K}(y|x) <^{\log} \lceil l-2k\rceil^{+} + \mathbf{I}(X;\mathcal{H}) + 2\mathbf{K}(k,l)$.*

**The Kolmogorov Birthday Paradox**. Let us say we select a random subset $D$ of size $2^{n/2}$ consisting of (possibly repeated) strings of length $n$, where each string is selected independently with a uniform probability. For the simple Kolmogorov birthday paradox, with overwhelming probability, there are two (possibly the same) strings $x, y \in D$, such that $\mathbf{K}(x|y) = O(1)$, for a large enough constant. This is due to reasoning from the classical birthday paradox. We now prove the general Kolmogorov birthday paradox. Let $P$ be any probability over sets $D$ consisting of $2^{n/2}$ (non repeated) strings of length $n$. Since $D \subset \{0,1\}^n$, for all $D$, $\max_{x,y\in D}\mathbf{K}(x|y) <^{+} n$. By Corollary 1, $\Pr_{D\sim P}\left[\mathbf{I}(D;\mathcal{H}) > \mathbf{I}(P;\mathcal{H}) + m\right] \overset{*}{<} 2^{-m}$. Combining these facts with the above theorem, with $l = n + O(1)$ and $k = .5n - 1$, we obtain the following result.

**Corollary.** $\Pr_{D\sim P}\left[\min_{x,y\in D,x\neq y}\mathbf{K}(x|y) <^{\log} \mathbf{I}(P;\mathcal{H}) + 2\mathbf{K}(n) + c\right] > 1 - 2^{-c}$.

Obviously, the bound loosens if $P$ samples sets of smaller size, mirroring the classical birthday


Figure 7.1: The domain of a Turing machine $T$ can be interpreted as the $[0, 1]$ interval, and the strings for which $T$ halts can be seen as a collection of dyadic subintervals. A left-total machine $L$ has the property that if $L$ halts on a string $x$, then it will halt on a string $y$ whose binary interval is smaller (i.e., to the left of $x$). The infinite sequence $B$ is called the border sequence and is the binary expansion of Chaitin's Omega. This paper uses a left-total universal Turing machine.

paradox.

## 7.2 Related Work

The study of Kolmogorov complexity originated from the work of [Kol65]. The canonical self-delimiting form of Kolmogorov complexity was introduced in [ZL70] and treated later in [Cha75]. The universal probability $\mathbf{m}$ was introduced in [Sol64]. More information about the history of the concepts used in this paper can be found in textbook [LV08].

The main result of this paper is an inequality including the mutual information of the encoding of a finite set with the halting sequence. A history of the origin of the mutual information of a string with the halting sequence can be found in [VV04].

A string is stochastic if it is typical of a simple elementary probability distribution. A string is typical of a probability measure if it has a low deficiency of randomness. The deficiency of randomness of a number $a \in \mathbb{N}$ with respect to a probability $P$ is $\mathbf{d}(a|P) = -\log P(a) - \mathbf{K}(a|\langle P\rangle)$. It is a measure of the extent of the refutation against the hypothesis $P$ given the result $a$ [G21]. Thus, the stochasticity, $\mathbf{Ks}(a)$, of a string $a$ is roughly $\min_{\text{probability } P} \mathbf{K}(P) + O(\log \mathbf{d}(a|P))$.

In the proof of Theorem 13, the stochasticity measure of encodings of finite sets is used. The notion of the deficiency of randomness with respect to a measure follows from the work of [She83] and is also studied in [KU87, V'Y87, She99]. Aspects involving stochastic objects were studied in [She83, She99, V'Y87, V'Y99].

This work uses the notion of left-total machine (see Figure 7.1) and the notion of the infinite "border" sequence, which is equal to the binary expansion of Chaitin's Omega (see Section 8.3). The works of [VV04, GTV01] introduced the notion of using the prefix of the border sequence to define strings into a two-part code. This paper uses the lemmas found in [Eps21a].

This paper can be seen as a conditional variant to the main result in [Lev16]. [Lev16] proved that for nonexotic sets $D$, the a priori probability, $\mathbf{m}$, of a set is concentrated on a single element.

**Theorem.** ([Lev16]) $-\log \max_{x\in D} \mathbf{m}(x) <^{\log} -\log \sum_{x\in D} \mathbf{m}(x) + \mathbf{I}(D; \mathcal{H})$.

There is a simple proof for this theorem in [She12]. The proof of Theorem 13 is similar to that of the main result in [Lev16], in that they both first prove stochasticity, $\mathbf{Ks}(O)$, of an object $O$ with certain properties and then show that this object has high $\mathbf{I}(O; \mathcal{H})$. In [Lev16], $O$ is equal to a set, and in this paper, $O$ is equal to a (sub)graph. Theorem 14 is not directly implied by the theorem in [Lev16] because this paper addresses conditional complexities between elements of a set. In addition, Theorem 14 is not a generalization of the main theorem in [Lev16] because it relies on the parameters of bunches and not the a priori probability $\mathbf{m}$.

## 7.3 Labeled Graph, Warm Up

In Section 7.4, a property of a complete subgraph of a labeled graph is proven. A labeled graph is a directed graph such that each vertex has a unique string attached to it. Given certain properties of the graph $G = (G_E, G_V)$, where $G_E$ are the directed edges, $G_V$ are the vertices, and subgraph $J = (J_E, V_V)$, Theorem 13 in Section 7.4 proves that $J$ is guaranteed to have an edge $(x, y) \in J_E$ with low $\mathbf{K}(x|y)$. In this section, we describe the overall arguments in the proof of this theorem.

We specify a vertex interchangeably with the string assigned to it. The general argument for the proof of Theorem 13 is as follows. Given a labeled graph $G$, if there is a random subgraph $F = (F_E, F_V)$ that is large enough, then it will probably share an edge with most large complete subgraphs $J$ of $G$. Thus, large complete subgraphs of $G$ with an empty intersection with $F$ will be considered atypical. If $F$ shares an edge with complete subgraph $J \subseteq G$, then

$$\min_{(x,y)\in J_E} \mathbf{K}(y|x) \lessapprox \log \max_{x\in F_V} \mathrm{OutDegree}(x) + \mathbf{K}(F).$$

This inequality follows from the fact that given a description of $F$ describing $\{(x, y) : (x, y) \in F_E\}$ and an $x \in F$, each $y \in \{y : (x, y) \in F_E\}$ can be described relative to $x$ with $\lceil \log\ \mathrm{OutDegree}(x)\rceil$ bits. In this section, instead of using random subgraphs, we use random lists of vertices $L_\bullet$, indexed by $x \in G_V$. Thus, for each $x \in G_V$, $L_x$ is a list of vertices, possibly with repetition. This allows for easier manipulation.

The warm-up arguments are as follows. Let $G = (G_E, G_V)$ be a graph of max degree $2^l$ and $\mathcal{J}$ be the set of complete subgraphs of $G$ of size $2^k$. We assume $l > 2k$. Each vertex $x \in G_V$ has a random list $L_x$ of $2^{l-2k}$ vertices, where for $i \in [1, 2^{l-2k}]$, $\Pr(y = L_x[i]) = [(x, y) \in G_E]2^{-l}$ and $\Pr(\emptyset = L_x[i]) = 1 - \mathrm{OutDegree}(x)2^{-l}$. For $J \in \mathcal{J}$, indexed list $L_\bullet$,

$$\mathrm{Miss}(J, L_\bullet) \text{ is true iff } \forall x, \forall y \in J_V, y \notin L_x.$$

For each $J \in \mathcal{J}$,

$$\begin{aligned}
\Pr(\mathrm{Miss}(J, L_\bullet)) &= \prod_{x\in J_V} \Pr(\forall y \in J_V, y \notin L_x) \\
&\leq \prod_{x\in J_V} \prod_{i\in[1,2^{l-2k}]} \Pr[\forall y \in J_V, y \neq L_x[i]] \\
&\leq \prod_{x\in J_V} \left(1 - 2^{k-l}\right)^{|L_x|} \\
&\leq \prod_{x\in J_V} \left(1 - 2^{k-l}\right)^{2^{l-2k}} \\
&\leq \left((1 - 2^{k-l})^{2^{l-2k}}\right)^{|J|} \\
&\leq \left(e^{-2^{-k}}\right)^{|J|} < e^{-1} < 1.
\end{aligned}$$

Now assume that $|L_x| = b2^{l-2k}$ for all $x \in G_V$, i.e., $b$ times more than before. It is not hard to see that $\Pr(\text{Miss}(J, L_\bullet)) < e^{-b}$ for each $J \in \mathcal{J}$. We assume a uniform distribution $\mathcal{U}$ over $\mathcal{J}$ (i.e. complete subgraphs of size $2^k$). Under this assumption,

$$\mathbf{E}\Big[[\text{Miss}(J, L_\bullet)]\Big] < \sum_{J\in\mathcal{J}} |\mathcal{J}|^{-1} e^{-b} = e^{-b}.$$

Thus, given all the parameters, $G$, $k$, $l$, and $b$, using brute force search, one can find a set of lists $L'_\bullet$ of size $b2^{l-2k}$ indexed by $x \in G_V$, such that less than $e^{-b}$ of members $J$ of $\mathcal{J}$ have $\text{Miss}(J, L'_\bullet)$. If $\text{Miss}(J, L'_\bullet)$ is true for $J \in \mathcal{J}$, then it must be atypical of $\mathcal{U}$ because $\mathbf{E}_{J\sim\mathcal{U}}[\text{Miss}(J, L'_\bullet)] < e^{-b}$. One can construct a $\mathcal{U}$-test using $L'_\bullet$. A $\mathcal{U}$-test is any function $t : \{0,1\}^* \to \mathbb{R}_{\geq 0}$ such that $\sum_{J\in\mathcal{J}} t(J)\mathcal{U}(J) \leq 1$. Thus, $t \cdot \mathcal{U}$ is a semimeasure, and therefore,

$$\mathbf{K}(J|t,\mathcal{U}) <^+ -\log t(J)\mathcal{U}(J). \tag{7.1}$$

Thus, the function $t(J) = [\text{Miss}(J, L'_\bullet)]e^b$ is a $\mathcal{U}$-test, with $\sum_{J\in\mathcal{J}} t(J)\mathcal{U}(J) < 1$. We set aside the parameters $(G, k, l, b, \mathcal{U})$ because they complicate the discussion. That is, we roll the parameters into the additive constants of the inequalities. By the definition of randomness deficiency,

$$\begin{aligned}
\mathbf{d}(J|\mathcal{U}) &= -\log \mathcal{U}(J) - \mathbf{K}(J|\mathcal{U}) \\
&>^+ \log|\mathcal{J}| - \mathbf{K}(J|L'_\bullet) && (7.2)\\
&>^+ \log|\mathcal{J}| - \mathbf{K}(J|t) && (7.3)\\
&>^+ \log|\mathcal{J}| + \log t(J)\mathcal{U}(J) && (7.4)\\
&>^+ \log|\mathcal{J}| + \log t(J)|\mathcal{J}|^{-1} \\
&>^+ b\log e.
\end{aligned}$$

Equation 7.2 has two components. The first term $\log|\mathcal{J}|$ is equal to $-\log \mathcal{U}(J)$ because $\mathcal{U}$ is the uniform distribution over all $\mathcal{J} \ni J$, the set of all complete subgraphs of $G$ of size $2^k$. The second term is due to the additive equalities

$$\mathbf{K}(J|L'_\bullet) = \mathbf{K}(J|L'_\bullet, G, k, l, b, \mathcal{U}) =^+ \mathbf{K}(J|G, k, l, b, \mathcal{U}) =^+ \mathbf{K}(J|\mathcal{U}),$$

in that given all the hidden parameters $(G, k, l, b, \mathcal{U})$, one can compute $L'_\bullet$ using brute force search, as described above. Equation 7.3 derives from the test $t$ being constructed from $L'_\bullet$ (and the hidden parameters). Equation 7.4 is due to the properties of the tests, as shown in Equation 7.1.

Thus, all complete subgraphs $J \in \mathcal{J}$ of $G$ for which $\text{Miss}(J, L'_\bullet)$ is true will be atypical of $\mathcal{U}$, with randomness deficiency $\mathbf{d}(J|\mathcal{U})$ greater than $b$. Thus, if a subgraph $J \in \mathcal{J}$ is $b$-typical, then there exists $(x, y) \in J_E$, with $y \in L_x$. Therefore, $b$-typical subgraphs $J \in \mathcal{J}$ will have

$$\min_{(x,y)\in J_E} \mathbf{K}(y|x) <^+ \log\big|L_x\big| <^+ l - 2k + \log b. \tag{7.5}$$

For Theorem 13, the uniform probability measure $\mathcal{U}$ is replaced by a special computable measure $P$ that realizes the stochasticity $\mathbf{Ks}$ of the subgraph $J$. In addition, $b$ is chosen to equal $b \approx \mathbf{d}(J|P)$ so that the subgraph $J$ is guaranteed to be typical of $P$, so $\text{Miss}(J)$ is false. This means that Equation 7.5 holds for $J$. In addition, in the next section, the parameters $(G, k, l, b)$ must be taken into account.

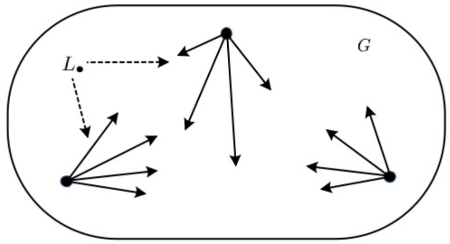


Figure 7.2: The above diagram is a graphical representation of $\kappa$ and $L_\bullet$, assuming that $cd2^{\ell-2k} = 4$. Each vertex has four edges chosen at random, where each particular edge is chosen with probability $2^{-\ell}$.

## 7.4 Labeled Graphs

In this section, we study exotic subgraphs of simple labeled graphs. A subgraph $J$ is exotic if consists of labeled edges $(x, y) \in J_E$, such that the conditional complexity $\mathbf{K}(y|x)$ is high. The proof of the following theorem uses stochasticity $\mathbf{Ks}$. An example proof that uses $\mathbf{Ks}$ and mirrors the proof of Theorem 13 can be found in Appendix 7.8. Note that the lemma in Appendix 7.8 is just an exercise to demonstrate reasoning with $\mathbf{Ks}$. The lemma is not used in the paper.

**Theorem 13** *For graph $G = (G_E, G_V)$, complete subgraph $J = (J_E, J_V)$; if $2^l > \max \text{Outdegree}(G)$, $2^k < |J|$, then we have $\min_{(x,y)\in J_E} \mathbf{K}(y|x) <^{\log} \lceil l - 2k\rceil^+ + \mathbf{I}(J; \mathcal{H}|G, k) + \mathbf{K}(G, k)$.*

**Proof.** We put $(G, k)$ on an auxiliary tape to the universal Turing machine $U$. Thus, all algorithms have access to $(G, k)$, and all complexities implicitly have $(G, k)$ as conditional terms.

Let $\ell = \max\{l, 2k\}$. Let $P$ be the probability that realizes $\mathbf{Ks}(J)$ and the deficiency of randomness $d = \max\{\mathbf{d}(J|P), 1\}$. Let $V : G \times G \to \mathbb{R}_{\geq 0}$ be a conditional probability measure where $V(y|x) = [(x, y) \in G_E]2^{-\ell}$ and $V(\emptyset|x) = 1 - \text{OutDegree}(x)2^{-\ell}$. We define a conditional probability measure over lists $L$ of $cd2^{\ell-2k}$ vertices of $G$, with $\kappa : G \times G^{cd2^{\ell-2k}} \to \mathbb{R}_{\geq 0}$, where $\kappa(L|x) = \prod_{y\in L} V(y|x)$. The constant $c \in \mathbb{N}$ will be determined later. Let $L_\bullet$ be an indexed list of $cd2^{\ell-2k}$ elements, indexed by $x \in G$, where each list is denoted by $L_x$ for $x \in G_V$. Let $\kappa(L_\bullet) = \prod_{x\in G} \kappa(L_x|x)$. A graphical representation of $\kappa$ and $L_\bullet$ can be found in Figure 7.2. For indexed list $L_\bullet$ and graph $H = (H_E, H_V)$, we use the indicator $\mathbf{i}(L_\bullet, H) = [\text{Complete } H \subseteq G, 2^k < |H_V|, \forall(x, y) \in H_E, y \notin L_x]$.

$$\begin{aligned}
\mathbf{E}_{L_\bullet\sim\kappa}\mathbf{E}_{H\sim P}[\mathbf{i}(L_\bullet, H)] &\leq \sum_H P(H) \Pr_{L_\bullet\sim\kappa}(\forall(x, y) \in H_E, y \notin L_x, |H_V| > 2^k,\ \text{Complete } H \subseteq G)\\
&\leq \sum_H P(H)[|H_V| > 2^k] \prod_{x\in H_V} (1 - 2^{k-\ell})^{|L_x|}\\
&\leq \sum_H P(H)[|H_V| > 2^k] \prod_{x\in H_V} (1 - 2^{k-\ell})^{cd2^{\ell-2k}}\\
&\leq \sum_H P(H)[|H_V| > 2^k] \prod_{x\in H_V} e^{-cd2^{-k}}\\
&< \sum_H P(H)e^{-cd}\\
&= e^{-cd}.
\end{aligned}$$

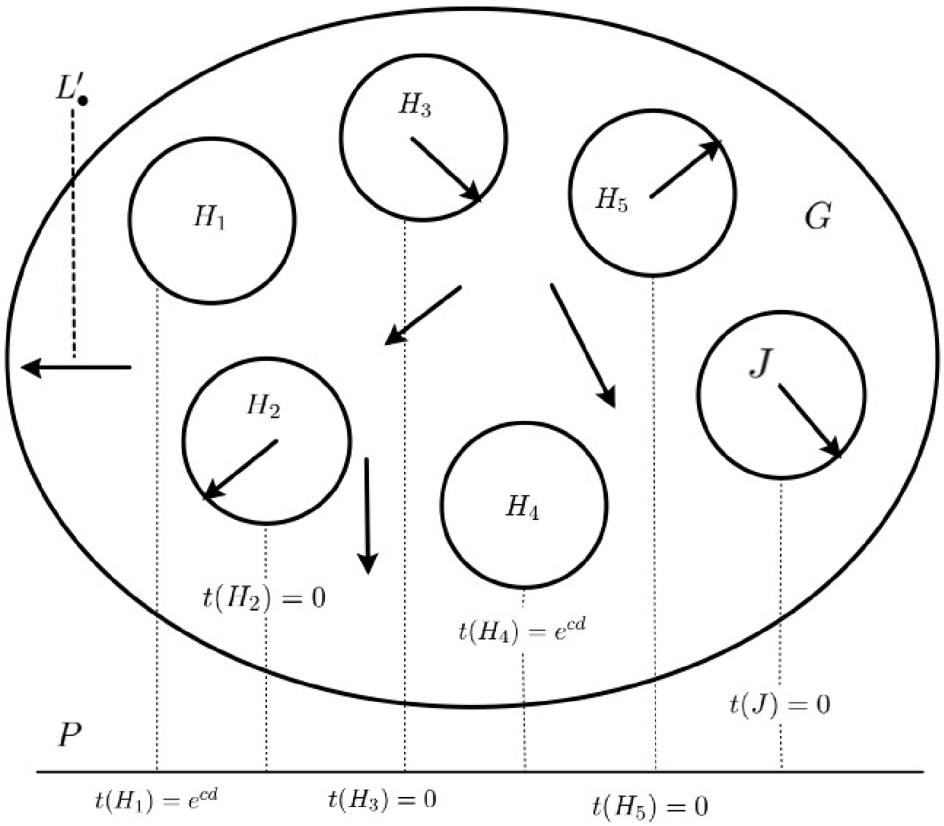


Figure 7.3: The above diagram is a graphical representation of the concepts used in the proof of Theorem 13. The main ellipse models the graph $G$, and the circles in the graph represent complete subgraphs (labeled $H_1$ to $H_5$ and $J$) with $> 2^k$ vertices. Each subgraph is in the support of probability $P$, represented by the dotted lines. The set $L'_\bullet$ represents a collection of selected edges. If a subgraph $H_i$ does not contain an edge in $L'_\bullet$, then $H_i$ is *atypical* and has a high score $t(H_i)$. By design, $J$ is typical and thus shares an edge with $L'_\bullet$.

Thus, there exists an $L'_\bullet$ such that $\mathbf{E}_{H\sim P}[\mathbf{i}(L'_\bullet, H)] < e^{-cd}$. This $L'_\bullet$ can be found with brute force search with all the parameters, with

$$\mathbf{K}(L'_\bullet|P, c, d) = O(1). \tag{7.6}$$

Thus, $t(H) = \mathbf{i}(L'_\bullet, H)e^{cd}$ is a $P$ test, where $\mathbf{E}_{H\sim P}[t(H)] \leq 1$. This test $t$ gives a high score to complete subgraphs of $G$ of size $> 2^k$ that have no intersecting edges with $L'_\bullet$. A diagram of the components used in this proof can be found in Figure 7.3. Furthermore,

$$\mathbf{K}(t|P, c, d) =^+ \mathbf{K}(t|L'_\bullet, P, c, d) = O(1).$$

It must be that there is an $(x, y) \in J_E$ where $y \in L_x$. Otherwise, $t_{L_\bullet}(J) = e^{cd}$ and

$$\begin{aligned}
\mathbf{K}(J|P,c,d) &<^+ \mathbf{K}(J|t,P,c,d)\\
\mathbf{K}(J|P,c,d) &<^+ -\log t(J)P(J) \qquad (7.7)\\
&<^+ -(\log e)cd - \log P(J)\\
(\log e)cd &<^+ -\log P(J) - \mathbf{K}(J|P,c,d)\\
(\log e)cd &<^+ -\log P(J) - \mathbf{K}(J|P) + \mathbf{K}(c,d)\\
(\log e)cd &<^+ d + \mathbf{K}(c,d),
\end{aligned}$$

which is a contradiction for large enough $c$ solely dependent on the universal Turing machine $U$. Equation 7.7 is due to Equation 7.1. The constant $c$ is folded into the additive constants of the inequalities of the rest of the proof. Thus, since there exists $(x,y) \in J_E$ where $y \in L_x$,

$$\begin{aligned}
\mathbf{K}(y|x) &<^+ \log|L'_x| + \mathbf{K}(L'_\bullet)\\
&<^+ \lceil l-2k\rceil^+ + \log d + \mathbf{K}(L'_\bullet|P,d) + \mathbf{K}(P,d)\\
&<^+ \lceil l-2k\rceil^+ + \log d + \mathbf{K}(P,d) \qquad (7.8)\\
&<^+ \lceil l-2k\rceil^+ + 3\log d + \mathbf{K}(P)\\
&<^+ \lceil l-2k\rceil^+ + \mathbf{Ks}(D) \qquad (7.9)
\end{aligned}$$

Equation 7.8 is due to Equation 7.6. Equation 7.9 is due to the definition of stochasticity. We now make the relativization of $(G,k)$ explicit, with

$$\begin{aligned}
\mathbf{K}(y|x,G,k) &<^+ \lceil l-2k\rceil^+ + \mathbf{Ks}(J|G,k)\\
&<^{\log} \lceil l-2k\rceil^+ + \mathbf{I}(J;\mathcal{H}|G,k) \qquad (7.10)\\
\mathbf{K}(y|x) &<^{\log} \lceil l-2k\rceil^+ + \mathbf{I}(J;\mathcal{H}|G,k) + \mathbf{K}(G,k).
\end{aligned}$$

Equation 7.10 is due to Lemma 10 in [Eps21a], which states $\mathbf{Ks}(x) < \mathbf{I}(x:\mathcal{H}) + O(\mathbf{K}(\mathbf{I}(x:\mathcal{H})))$. □

## 7.5 Warm Up for the Main Theorem of the Paper

Theorem 13 can be used to prove results about the minimum conditional complexity between two elements of a bunch. This section gives a broad overview of the arguments used in the proof of Theorem 14. Let $X \subset \{0,1\}^*$ be a $(k,l)$-bunch, where $|X| > 2^k$, and $\max_{x,y\in X}\mathbf{K}(y|x) < l$.

Let $\mathbf{K}^r(x|y) = \min\{\|p\| : U_y(p) = x \text{ in time } r\}$ be the conditional complexity of $x$ given $y$ at time $r$. Therefore, given a number $r$, $\mathbf{K}^r$ is computable. We also assume $\mathbf{K}^r(x|y) = \infty$ if $\|y\| > r$ to ensure that $\mathbf{K}^r$ has finite $\{(x,y) : \mathbf{K}^r(x|y) < \infty, x,y \in \mathbb{N}\}$ for each $r$. Let $G^r = (G^r_E, G^r_V)$ be a graph defined by $(x,y) \in G^r_E$ iff $\mathbf{K}^r(x|y) < l$.

Let $s$ be the smallest number where $\mathbf{K}^s(x|y) < l$, for all $x,y \in X$. Let $G = (G_E, G_V) = G^s$. Since $X$ is a $(k,l)$-bunch, $X$ can be viewed as a complete subgraph of $G$ of size $>2^k$. Invoking Theorem 13, we obtain

$$\min_{(x,y)\in X, x\neq y}\mathbf{K}(y|x) <^{\log} \lceil l-2k\rceil^+ + \mathbf{I}(X;\mathcal{H}|G,k) + \mathbf{K}(G,k). \qquad (7.11)$$

We have $\mathbf{K}(s|G) <^+ \mathbf{K}(l)$ because $s = \min\{r : G = G^r\}$. Therefore,

$$\mathbf{K}(X|G) <^+ \mathbf{K}(X|s) + \mathbf{K}(s|G) <^+ \mathbf{K}(X|s) + \mathbf{K}(l). \qquad (7.12)$$

Due to the definition of $G = G^s$,

$$\mathbf{K}(G|s) <^+ \mathbf{K}(l). \tag{7.13}$$

By the definition of $\mathbf{I}$,

$$\begin{aligned}
\mathbf{I}(X;\mathcal{H}|G,k) &= \mathbf{K}(X|G,k) - \mathbf{K}(X|G,k,\mathcal{H}) \\
&= \mathbf{K}(X|G) - \mathbf{K}(X|G,\mathcal{H}) + O(\mathbf{K}(k)) \\
&<^+ \mathbf{K}(X|s) - \mathbf{K}(X|G,\mathcal{H}) + O(\mathbf{K}(k,l)) && (7.14)\\
&< \mathbf{K}(X|s) - \mathbf{K}(X|s,\mathcal{H}) + \mathbf{K}(G|s) + O(\mathbf{K}(k,l)) \\
&< \mathbf{I}(X;\mathcal{H}|s) + O(\mathbf{K}(k,l)). && (7.15)
\end{aligned}$$

Equation 7.14 is due to Equation 7.12. Equation 7.15 is due to Equation 7.13. Using $\mathbf{K}(G) <^+ \mathbf{K}(s) + \mathbf{K}(l)$ and Equation 7.15, we obtain

$$\mathbf{I}(X;\mathcal{H}|G,k) + \mathbf{K}(G,k) < \mathbf{I}(X;\mathcal{H}|s) + \mathbf{K}(s) + O(\mathbf{K}(k,l)). \tag{7.16}$$

Combining Equations 7.11 and 7.16, we obtain

$$\min_{(x,y)\in J_E} \mathbf{K}(y|x) <^{\log} \lceil l-2k\rceil^+ + \mathbf{I}(X;\mathcal{H}|s) + \mathbf{K}(s) + O(\mathbf{K}(k,l)). \tag{7.17}$$

This inequality is close to the form of Theorem 14. The main difference is that the number $s$ appears in Equation 7.17. This can be rectified if we use a different notion of a computational resource. In the next section, we introduce left-total universal machines, and the resource used is not a number $s$ but a so-called total string $b$. Then, Lemma 3, defined in Section 8.3, can be used to remove the $b$ factor from the final inequality.

## 7.6 Left-Total Machines

We recall that for $x \in \{0,1\}^*$, $\Gamma_x = \{x\beta : \beta \in \{0,1\}^\infty\}$ is the interval of $x$. The notions of total strings and the "left-total" universal algorithm are needed in this paper. We say $x \in \{0,1\}^*$ is total with respect to a machine if the machine halts on all sufficiently long extensions of $x$. More formally, $x$ is total with respect to $T_y$ for some $y \in \{0,1\}^{*\infty}$ iff there exists a finite prefix-free set of strings $Z \subset \{0,1\}^*$ where $\sum_{z\in Z} 2^{-\|z\|} = 1$ and $T_y(xz) \neq \perp$ for all $z \in Z$. We say (finite or infinite) string $\alpha \in \{0,1\}^{*\infty}$ is to the "left" of $\beta \in \{0,1\}^{*\infty}$ and use the notation $\alpha \lhd \beta$ if there exists an $x \in \{0,1\}^*$ such that $x0 \sqsubseteq \alpha$ and $x1 \sqsubseteq \beta$. A machine $T$ is left-total if for all auxiliary strings $\alpha \in \{0,1\}^{*\infty}$ and for all $x, y \in \{0,1\}^*$ with $x \lhd y$, one has that $T_\alpha(y) \neq \perp$ implies that $x$ is total with respect to $T_\alpha$. Left-total machines were introduced in [Lev16]. An example can be seen in Figure 7.4.

For the remainder of this paper, we can and will change the universal self-delimiting machine $U$ into a universal left-total machine $U'$ by the following definition. The algorithm $U'$ orders all strings $p \in \{0,1\}^*$ by the running time of $U$ when given $p$ as an input. Then, $U'$ assigns each $p$ an interval $i_p \subseteq [0,1]$ of width $2^{-\|p\|}$. The intervals are assigned "left to right", where if $p \in \{0,1\}^*$ and $q \in \{0,1\}^*$ are the first and second strings in the ordering, then they will be assigned the intervals $[0, 2^{-\|p\|}]$ and $[2^{-\|p\|}, 2^{-\|p\|} + 2^{-\|q\|}]$, respectively.

Let the target value of $p \in \{0,1\}^*$ be $(p) \in \mathbb{W}$, which is the value of the string in binary. For example, the target value of both strings 011 and 0011 is 3. The target value of 0100 is 4. The target interval of $p \in \{0,1\}^*$ is $\Gamma(p) = ((p)2^{-\|p\|}, ((p){+}1)2^{-\|p\|})$.

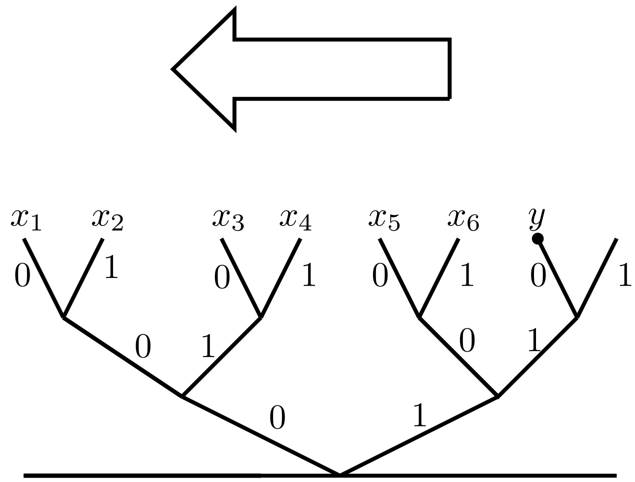


Figure 7.4: The above diagram represents the domain of a left-total machine $T$ with the 0 bits branching to the left and the 1 bits branching to the right. For $i \in \{1,\dots,5\}$, $x_i \lhd x_{i+1}$ and $x_i \lhd y$. Assuming $T(y)$ halts, each $x_i$ is total. This also implies that each $x_i^-$ is total.

The universal machine $U'$ outputs $U(p)$ on input $p'$ if the intervals $\Gamma(p')$ are strictly contained in $i_p$, with $\Gamma(p') \subset i_p$, and $\Gamma(p'^-)$ are not strictly contained in $i_p$, with $\Gamma(p'^-) \not\subset i_p$. The same definition applies to machines $U'_\alpha$ and $U_\alpha$ over all $\alpha \in \{0,1\}^{*\infty}$.

Recall that a function $f : \mathbb{N} \to \mathbb{N}$ is partially computable with respect to $U$ if there is a string $t \in \{0,1\}^*$ such that $f(x) = U(t\langle x\rangle)$ when $f(x)$ is defined and $U(t\langle x\rangle)$ does not halt otherwise. Similarly, a function $f : \mathbb{N} \to \mathbb{N}$ is partially computable with respect to $U'$ if there is $t \in \{0,1\}^*$, such that whenever $f(x)$ is defined, there is an interval $i_{t\langle x\rangle}$ and for any string $p$ where $\Gamma(p)$ and not that of $\Gamma(p^-)$ is contained in $i_{t\langle x\rangle}$, then $U'(p) = f(x)$. Otherwise, if $f(x)$ is not defined, the interval $i_{t\langle x\rangle}$does not exist. The following proposition was used without being proven in [Lev16].

**Proposition 4** $\mathbf{K}_U(x|y) =^+ \mathbf{K}_{U'}(x|y)$.

**Proof.** It must be that $\mathbf{K}_U(x|y) <^+ \mathbf{K}_{U'}(x|y)$ because there is a Turing machine that computes $U'$. Therefore, due to the universality of $U$, there is a $t \in \{0,1\}^*$, such that $U_y(tx) = U'_y(x)$, thus proving the minimality of $\mathbf{K}_U$. It must be that $\mathbf{K}_{U'}(x|y) <^+ \mathbf{K}_U(x|y)$. This is because if $U(x) = z$, then there is interval $i_x$ such that for all strings $p$ where $\Gamma(p)$ and not that of $\Gamma(p^-)$ that are strictly contained in $i_x$ has $U'_y(p) = U_y(x)$. Thus, we have that $\|p\| \leq \|x\| + 2$. This implies that $\mathbf{K}_{U'}(x|y) \leq \mathbf{K}_U(x|y) + 2$. □

For the rest of the paper, we now set $U$ to be equal to $U'$, so the universal Turing machine can be considered to be left-total. Without loss of generality, as shown in Proposition 4, the complexity terms of this paper are defined with respect to the universal left-total machine $U$.

**Proposition 5** *There exists a unique infinite sequence $\mathcal{B}$ with the following properties.*

1. *All the finite prefixes of $\mathcal{B}$ have total and nontotal extensions.*

2. *If a finite string has total and nontotal extensions, then it is a prefix of $\mathcal{B}$.*

3. *If a string $b$ is total and $b^-$ is not, then $b^- \sqsubset \mathcal{B}$.*

**Proof.**

1. Let $\Omega \in \mathbb{R}$ be Chaitin's Omega, the probability that a random sequence of bits halts when given to $U$, with $\Omega = \sum_{p\in\{0,1\}^*}[U(p) \text{ halts}]2^{-\|p\|}$. Thus, $\Omega$ characterizes the domain of $U$, with $\bigcup_{p\in\{0,1\}^*} i_p = [0,\Omega)$. Let $\mathcal{B} \in \{0,1\}^\infty$ be the binary expansion of $\Omega$, which is an ML random string. For each $n \in \mathbb{N}$, let $b_n \sqsubset \mathcal{B}$, $\|b_n\| = n$. Let $m \in \mathbb{W}$ be the smallest whole number such that $b_n1^{(m)}0 \sqsubset \mathcal{B}$. Then, $b_n1^{(m+1)}$ is a nontotal string because $[0,\Omega] \cap \Gamma(b_n1^{(m+1)}) = \emptyset$. Furthermore, let $m \in \mathbb{W}$ be the smallest whole number such that $b_n0^{(m)}1 \sqsubset \mathcal{B}$. Then, $b_n0^{(m+1)}$ is a total string because $\Gamma(b_n0^{(m+1)}) \subset [0,\Omega)$.

2. Assume there are two strings $x$ and $y$ of length $n$ that have total and nontotal extensions, with $x \lhd y$. Since $y$ has total extensions, there exist $z$ such that $U'(yz)$ halts. Since $x \lhd yz$, by the definition of left-total machines, $x$ is total, causing a contradiction.

3. This is because $b^-$ has total and nontotal extensions.

$\square$

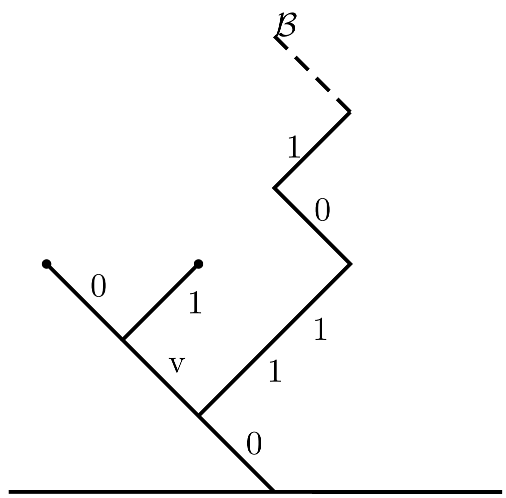


Figure 7.5: The above diagram represents the domain of the universal left-total algorithm $U'$, with 0 bits branching to the left and 1 bits branching to the right. The strings in the above diagram, $0v0$ and $0v1$, are halting inputs to $U'$ with $U(0v0) \neq\perp$ and $U(0v1) \neq\perp$. Therefore, $0v$ is a total string. The infinite border sequence $\mathcal{B} \in \{0,1\}^\infty$ represents the unique infinite sequence such that all its finite prefixes have total and nontotal extensions. All finite strings branching to the right of $\mathcal{B}$ will cause $U'$ to diverge.

We call this infinite sequence $\mathcal{B}$, "border" because for any string $x \in \{0,1\}^*$, $x \lhd \mathcal{B}$ implies that $x$ is total with respect to $U$ and $\mathcal{B} \lhd x$ implies that $U$ will never halt when given $x$ as an initial input. Figure 7.5 shows the domain of $U'$ with respect to $\mathcal{B}$. We now set $U$ to be equal $U'$. Without loss of generality, as shown in Proposition 4, the complexity terms of this paper are defined with respect to the universal left-total machine $U$.

For total string $b$, we define the busy beaver function, $\mathbf{bb}(b) = \max\{\|x\| : U(p) = x, p \lhd b \text{ or } p \sqsupseteq b\}$. For total string $b$, the $b$-computable complexity of string $x$ with respect to string $y \in \{0,1\}^{*\infty}$ is $\mathbf{K}_b(x|y) = \min\{\|p\| : U_y(p) = x \text{ in } \mathbf{bb}(b) \text{ time and} \|y\| \leq \mathbf{bb}(b)\}$. If $b$ and $c$ are total, and $b \lhd c$, then $\mathbf{K}_b \geq \mathbf{K}_c$. In addition, if $b$ and $b^-$ are total, then $\mathbf{K}_b \geq \mathbf{K}_{b^-}$.

The following lemma shows that if a prefix of the border sequence is simple relative to a string $x$, then it will be the common information between $x$ and the halting sequence $\mathcal{H}$.

**Lemma 3** *If $b \in \{0,1\}^*$ is total and $b^-$ is not, and $x \in \{0,1\}^*$, then $\mathbf{K}(b) + \mathbf{I}(x;\mathcal{H}|b) <^{\log} \mathbf{I}(x\,;\mathcal{H}) + \mathbf{K}(b|x)$.*

**Proof.** By Proposition 5, $b^- \sqsubset \mathcal{B}$ is a prefix of the border sequence and thus $\|b\| <^+ \mathbf{K}(b)$. Since $\mathcal{B}$ is computable from the halting sequence $\mathcal{H}$, we have that $b$ is computable from $\|b\|$ and $\mathcal{H}$, with $\mathbf{K}(b|\mathcal{H}) <^+ \mathbf{K}(\|b\|)$. The chain rule gives the equality $\mathbf{K}(b)+\mathbf{K}(x|b,\mathbf{K}(b)) =^+ \mathbf{K}(x)+\mathbf{K}(b|x,\mathbf{K}(x))$. Combined with the inequalities $\mathbf{K}(x|b) <^+ \mathbf{K}(x|b,\mathbf{K}(b)) + \mathbf{K}(\mathbf{K}(b))$ and $\mathbf{K}(b|x,\mathbf{K}(x)) <^+ \mathbf{K}(b|x)$, we get

$$\mathbf{K}(b) + \mathbf{K}(x|b) <^+ \mathbf{K}(x) + \mathbf{K}(b|x) + \mathbf{K}(\mathbf{K}(b)).$$

Subtracting $\mathbf{K}(x|b,\mathcal{H})$ from both sides results in

$$\begin{aligned}
\mathbf{K}(b) + \mathbf{K}(x|b) - \mathbf{K}(x|b,\mathcal{H}) &<^+ \mathbf{K}(x) + \mathbf{K}(b|x) + \mathbf{K}(\mathbf{K}(b)) - \mathbf{K}(x|b,\mathcal{H})\\
&<^+ \mathbf{K}(x) + \mathbf{K}(b|x) + \mathbf{K}(\mathbf{K}(b)) - \mathbf{K}(x|\mathcal{H}) + \mathbf{K}(b|\mathcal{H}).\\
&<^+ \mathbf{I}(x;\mathcal{H}) + \mathbf{K}(b|x) + \mathbf{K}(\mathbf{K}(b)) + \mathbf{K}(b|\mathcal{H})\\
&< \mathbf{I}(x;\mathcal{H}) + \mathbf{K}(b|x) + O(\log\|b\|)\\
&<^{\log} \mathbf{I}(x;\mathcal{H}) + \mathbf{K}(b|x).
\end{aligned}$$

□

## 7.7 Minimum Conditional Complexity

We recall that a $(k,l)$-bunch $X$ is a finite set of strings where $2^k < |X|$ and for all $x, x' \in X$, $\mathbf{K}(x|x') < l$. If $l \gg k$, such as the $(k,l)$-bunch consisting of two large independent random strings, then it is difficult to prove properties about it. If $l \approx 2k$, then interesting properties emerge.

**Theorem 14** *For $(k,l)$-bunch $X$, $\min_{x,y\in X, x\neq y} \mathbf{K}(y|x) <^{\log} \lceil l-2k\rceil^+ + \mathbf{I}(X;\mathcal{H}) + 2\mathbf{K}(k,l)$.*

**Proof.** We assume that the universal Turing machine $U$ is left-total. Let $b$ be a shortest total string such that $\mathbf{K}_b(y|x) < l$ for all $x, y \in X$. We have

$$\mathbf{K}(b|X) <^+ \mathbf{K}(\|b\|, l), \tag{7.18}$$

as there is a program that, when enumerating total strings of length $\|b\|$ from left to right, returns the first string with the desired properties. The first total string found is $b$, as shown in Figure 7.6. Thus, $b^-$ is not total, and by Proposition 5, $b^- \sqsubset B$ is a prefix of the border. For open parameter total string $c$, let $G^c$ be the graph defined by $(x,y) \in G_E$ iff $\mathbf{K}_c(y|x) < l$. Let $G = (G_E, G_V) = G^b$. Thus if $x, y \in X$, then $(x,y) \in G_E$. We have

$$\mathbf{K}(G|b) <^+ \mathbf{K}(l) \tag{7.19}$$
$$\mathbf{K}(b|G) <^+ \mathbf{K}(\|b\|, l). \tag{7.20}$$

Equation 7.19 is because $G = G^b$. Equation 7.20 is due to the existence of a program that enumerates total strings of length $\|b\|$ (from left to right) and returns the first total string $c$ such that

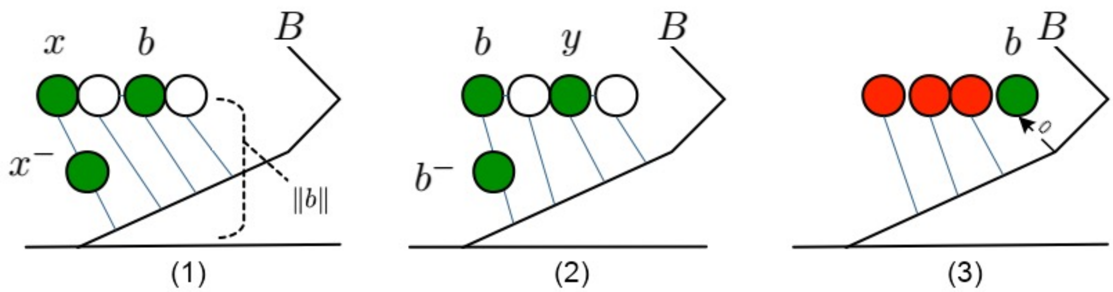


Figure 7.6: The above diagram represents the domain of the universal left-total Turing machine $U$ and uses the same conventions as Figure 7.5, with 0s branching to the left and 1s branching to the right. It shows all the total strings of length $\|b\|$, including $b$. The large diagonal line is the border sequence, $B$. A string $c$ is marked green if $\mathbf{K}_c(y|x) < l$ for all $x, y \in X$. By definition, $b$ is a shortest green string. If $x$ is green and total, and $x \lhd y$, and $y$ is total, then $y$ is green, since $\mathbf{K}_x \geq \mathbf{K}_y$. Furthermore, if $x$ is green and total and $x^-$ is total, then $x^-$ is green, as $\mathbf{K}_x \geq \mathbf{K}_{x^-}$. It cannot be that there is a green $x \lhd b$ with $\|x\| = \|b\|$. Otherwise, $x^-$ is total, and thus, it is green, causing a contradiction because it is shorter than $b$. This is shown in part (1). Furthermore, there cannot be a green $y$, with $b \lhd y$ and $\|y\| = \|b\|$. Otherwise, $b^-$ is total and thus green, contradicting the definition of $b$. This is shown in part (2). Thus, $b$ is unique, and since $b^-$ is not total, by Proposition 5, $b^-$ is a prefix of the border, as shown in part (3). Thus, an algorithm returning a green string of length $\|b\|$ will return $b$.

$G)E \subseteq G^c_E$. It cannot be that there is a total string $c$ shorter than $b$ with $G \subseteq G^c$. Otherwise, $G^c_E \supseteq G_E \supseteq \binom{X}{2}$, contradicting the definition of $b$ being a shortest total string with $G^b \supseteq X$. Thus, using this impossibility and the reasoning detailed in Figure 7.6, where $y$ is green if $G \subseteq G^y$, the program returns $b$. Theorem 13 gives $x, y \in X$, where

$$\mathbf{K}(y|x) <^{\log} \lceil l - 2k \rceil^+ + \mathbf{I}(X; H|G, k) + \mathbf{K}(G, k) \tag{7.21}$$

The rest of the proof is a straightforward sequence of application of inequalities. We have

$$\begin{aligned} \mathbf{K}(X|G) &<^+ \mathbf{K}(X|b) + \mathbf{K}(b|G) \\ &<^+ \mathbf{K}(X|b) + \mathbf{K}(\|b\|, l), \end{aligned} \tag{7.22}$$

where Equation 7.22 is due to Equation 7.20. We also have

$$\begin{aligned} \mathbf{K}(X|b, \mathcal{H}) &< \mathbf{K}(X|G, \mathcal{H}) + \mathbf{K}(G|b, \mathcal{H}), \\ &< \mathbf{K}(X|G, \mathcal{H}) + \mathbf{K}(l), \end{aligned} \tag{7.23}$$

where Equation 7.23 is due to Equation 7.19. Therefore,

$$\begin{aligned} \mathbf{I}(X; \mathcal{H}|G) &= \mathbf{K}(X|G) - \mathbf{K}(X|G, \mathcal{H}) \\ &<^+ \mathbf{I}(X; \mathcal{H}|b) + \mathbf{K}(l) + \mathbf{K}(\|b\|, l). \end{aligned} \tag{7.24}$$

Combining Equations 7.21 and 7.24,

$$\begin{aligned} \mathbf{K}(y|x) &<^{\log} \lceil l - 2k \rceil^+ + \mathbf{I}(X; \mathcal{H}|b) + \mathbf{K}(G) + \mathbf{K}(\|b\|) + O(\mathbf{K}(k, l)) \\ &<^{\log} \lceil l - 2k \rceil^+ + \mathbf{I}(X; \mathcal{H}|b) + \mathbf{K}(b) + \mathbf{K}(\|b\|) + O(\mathbf{K}(k, l)) && (7.25) \\ &<^{\log} \lceil l - 2k \rceil^+ + \mathbf{I}(X; \mathcal{H}|b) + \mathbf{K}(b) + O(\mathbf{K}(k, l)). && (7.26) \end{aligned}$$

Equation 7.25 is due to Equation 7.19. Equation 7.26 is because the precision is $(<^{\log})$. Furthermore, since $b$ is total and $b^-$ is not, by Proposition 5, $b^- \sqsubset B$. The border $B$ is the binary expansion of Chaitin's Omega (see Proposition 5), so $b$ is random, with $\mathbf{K}(\|b\|) = O(\log \mathbf{K}(b))$. Using Lemma 3 on Equation 7.26, we obtain

$$\begin{aligned}\mathbf{K}(y|x) &<^{\log} \lceil l-2k\rceil^+ + \mathbf{I}(X;\mathcal{H}) + \mathbf{K}(b|X,\|b\|) + O(\mathbf{K}(k,l)) \\ &<^{\log} \lceil l-2k\rceil^+ + \mathbf{I}(X;\mathcal{H}) + O(\mathbf{K}(k,l)) \qquad (7.27)\end{aligned}$$

where Equation 7.27 is due to Equation 7.18. Adding $(k,l)$ to the conditional on all terms results in

$$\begin{aligned}\mathbf{K}(y|x,k,l) &<^{\log} \lceil l-2k\rceil^+ + \mathbf{I}(X;\mathcal{H}|k,l) + O(\mathbf{K}(k,l|k,l)) \\ &<^{\log} \lceil l-2k\rceil^+ + \mathbf{I}(X;\mathcal{H}|k,l) \\ \mathbf{K}(y|x) &<^{\log} \lceil l-2k\rceil^+ + \mathbf{I}(X;\mathcal{H}) + 2\mathbf{K}(k,l).\end{aligned}$$

□

## 7.8 Warm-up Exercise in Stochasticity

The following proof demonstrates how the stochasticity term **Ks** can be used in mathematical arguments. The general structure of the proof parallels the proof in Theorem 13. This lemma first appeared (in a slightly different form) as Lemma 5 in [Eps21b]. The lemma itself is just an exercise and is not used in the paper.

**Lemma 4** *For $D\subseteq\{0,1\}^n$, $|D|=2^s$, $\min_{x\in D}\mathbf{K}(x) <^{\log} n-s+\mathbf{Ks}(D)+O(\mathbf{K}(s,n))$.*

**Proof.** We put $(n,s)$ on an auxiliary tape to the universal Turing machine $U$. Thus, all algorithms have access to $(n,s)$, and all complexities implicitly have $(n,s)$ as conditional terms. This can be done because the precision of the lemma is $O(\mathbf{K}(s,n))$. Let $Q$ realize $\mathbf{Ks}(D)$, with $d = \max\{\mathbf{d}(D|Q),1\}$. Thus, $Q$ is an elementary probability measure over $\{0,1\}^*$ and $D\in\text{Support}(Q)$, with randomness deficiency $d$.

Let $F\subseteq\{0,1\}^n$ be a random set where each element $a\in\{0,1\}^n$ is selected independently with probability $cd2^{-s}$, where $c\in\mathbb{N}$ is chosen later. Let $\mathcal{U}_n$ be the uniform measure over $\{0,1\}^n$. $\mathbf{E}[\mathcal{U}_n(F)]\le cd2^{-s}$. Furthermore,

$$\mathbf{E}[Q(\{G:|G|=2^s,G\subseteq\{0,1\}^n,G\cap F=\emptyset\})]\le\sum_G Q(G)(1-cd2^{-s})^{2^s}<e^{-cd}.$$

Thus, by the Markov inequality, $W\subseteq\{0,1\}^n$ can be chosen such that $\mathcal{U}_n(W)\le 2cd2^{-s}$ and $Q(\{G:|G|=2^s,G\subseteq\{0,1\}^n,G\cap W=\emptyset\})\le e^{1-cd}$.

$$\mathbf{K}(W|Q,d,c)=O(1). \qquad (7.28)$$

It must be that $D\cap W\neq\emptyset$. Otherwise, we obtain a contradiction with the following reasoning. Let $t:\{0,1\}^*\to\mathbb{R}_{\ge 0}$ be a $Q$-test, with $t(G)=[|G|=2^s,G\subseteq\{0,1\}^n,G\cap W=\emptyset]e^{cd-1}$, and $\sum_G Q(G)t(G)\le 1$. Thus, $t$ gives a high score to sets $G$ that do not intersect $W$. Therefore,

$t(D) = e^{cd-1}$. We have

$$
\begin{aligned}
\mathbf{K}(D|Q,d,c) &<^+ \mathbf{K}(D|W,Q,d,c) && (7.29)\\
&<^+ \mathbf{K}(D|t,W,Q,d,c) && (7.30)\\
&<^+ -\log Q(D)t(D) && (7.31)\\
&<^+ -\log Q(D) - (\log e)cd\\
(\log e)cd &<^+ -\log Q(D) - \mathbf{K}(D|Q) + \mathbf{K}(d,c)\\
&<^+ d + \mathbf{K}(d,c),
\end{aligned}
$$

which is a contradiction for a large enough $c$ dependent solely on the universal Turing machine. Equation 7.29 is due to Equation 7.28. Equation 7.30 is because the test $t$ can be computed from $(W,c,d)$. Equation 7.31 is due to Equation 7.1. Thus, there is an $x \in D \cap W$. Thus, since $\mathcal{U}_n(W) \leq 2cd2^{-s}$, the function $q(a) = [a \in W](2^s/cd)\mathcal{U}_n(a)$ is a semimeasure. Therefore, we have

$$\mathbf{K}(x) <^+ -\log q(x) + \mathbf{K}(q) <^+ n + \log d - s + \mathbf{K}(d) + \mathbf{K}(Q) <^+ n - s + \mathbf{Ks}(D).$$

□

# Chapter 8

# On the Conditional Complexity of Sets of Strings

## 8.1 Introduction

In [Rom03], a criteria for the amount of algorithmic information that can be extracted from a triplet of strings was established. In that paper, the notion of bunches was introduced. A $(k, l, n)$ bunch is a finite set of strings $X$ such that

1. $|X| = 2^k$,
2. $\mathbf{K}(x_1|x_2) < l$ for all $x_1, x_2 \in X$,
3. $\mathbf{K}(x) < n$ for all $x \in X$.

The term $\mathbf{K}$ used above represents the conditional Kolmogorov complexity. In [Rom03], Theorem 5, it was shown that common information could be extracted from bunches.

**Theorem 5.** [Rom03] *For any $(k, l, n)$ bunch $X$, there exists a string $z$ such that $\mathbf{K}(x|z) \leq l + O(|l-k| + \log n)$ and $\mathbf{K}(z|x) = O(|l-k| + \log n)$ for any $x \in X$.*

In this chapter, we revisit bunches and show that every bunch that is not exotic has an element that is simple conditional to all other members. We show this over the class of non-exotic bunches, that is bunches whose encoding has low mutual information with the halting sequence. We also prove a similar result for a structure we call batches, which are defined in terms of expectation instead of max. In this chapter, we use a slightly different definition of bunches (and batches), where there are no assumptions about the Kolmogorov complexity of its elements. We define a $(k, l)$ bunch $X$ to be a finite set of strings, where $k = \lceil \log |X| \rceil$, $l > k$, and for all $x, x' \in X$, $\mathbf{K}(x|x') \leq l$. If $l \gg k$, such as the bunch consisting of two large independent random strings, then it is difficult to proof properties about it. If $l \approx k$, then interesting properties emerge, such as the bunch theorem of this chapter. This theorem states when $l \approx k$, then for non-exotic bunches, there exists common information in the form of a member of this bunch which is simple relative to all other strings of the bunch. Otherwise the bunch is exotic, in that it has high mutual information with the halting sequence. The bunch theorem of this chapter is as follows.

**Theorem.** *For $(k, l)$ bunch $X$, $\min_{x \in X} \max_{x' \in X} \mathbf{K}(x|x') <^{\log} 2(l-k) + \mathbf{I}(X : \mathcal{H})$.*

We also prove a similar result using expectation instead maximum. We define a $(k, l)$ batch $X$ to be a finite set of strings, where $k = \lceil \log |X| \rceil$, $l > k$, and for all $x \in X$, $\mathbf{E}_{x' \in X}[\mathbf{K}(x|x')] \leq l$.

**Theorem.** *For $(k, l)$ batch $X$, $\min_{x \in X} \mathbf{E}_{x' \in X}[\mathbf{K}(x|x')] <^{\log} l - k + \mathbf{I}(X : \mathcal{H})$.*

The halting sequence is $\mathcal{H}$ and the information that a sring $x$ has with $\mathcal{H}$ is $\mathbf{I}(X : \mathcal{H})$. An example of an exotic bunch is $R_n$, the set of all random strings of size $n$, where $x \in R_n$ iff $\|x\| = n$ and $\mathbf{K}(x) >^+ n$. It is not hard to see that for all $x, x' \in R_n$, $\mathbf{K}(x|x') <^{\log} n$. So $R_n$ is a $(n - O(1), n + O(\log n))$ bunch. In addition, because $R_n$ contains all random strings of size $n$, $\min_{x \in X} \max_{x' \in X} \mathbf{K}(x|x') >^{\log} n$. Thus $R_n$ does not have such a conditionally simple element, and this implies it is exotic, because, due to the bunch theorem introduced above, $n <^{\log} \mathbf{I}(R_n : \mathcal{H})$. This bound is easily verifiable using the definition of $R_n$, since $\mathbf{K}(R_n) >^+ n$ and $\mathbf{K}(R_n|\mathcal{H}) <^+ \mathbf{K}(n)$, because given the halting sequence and $n$, there exists a simple program that can produce all random strings of size $n$.

Another example of a bunch is the set $S_{x,m}$, where $x$ is a string of arbitrary length, and $S_{x,m} = \{xy : y \text{ is a string of length } m\}$. This bunch is usually not exotic. It must be that for $\max_{y,x' \in S_{x,m}} \mathbf{K}(y|x') <^+ m + \mathbf{K}(m)$ as all strings in $S_{x,m}$ differ by a substring of size $m$. Furthermore $\#S_{x,m} = m$. Therefore $S_{x,m}$ is a $(m, m + \mathbf{K}(m) + O(1))$ bunch. Since $x$ and $m$ can be recovered from an encoding of the the set $S_{x,m}$, and of course $S_{x,m}$ can be created from $x$ and $m$, we have that $\mathbf{I}(S_{x,m} : \mathcal{H}) =^+ \mathbf{I}(x, m : \mathcal{H}) < \mathbf{I}(x : \mathcal{H}) + O(\mathbf{K}(m))$. So by the above bunch theorem, $\min_{y \in S_{x.m}} \max_{x' \in S_{x,m}} \mathbf{K}(y|x') <^{\log} 2\mathbf{K}(m) + \mathbf{I}(S_{x,m} : \mathcal{H}) <^{\log} \mathbf{I}(x : \mathcal{H}) + O(\mathbf{K}(m))$. Most $x$ has negligible information with the halting sequence, relative to its length. Furthermore it can be seen independently that $\min_{y \in S_{x,m}} \max_{x' \in S_{x,m}} \mathbf{K}(y|x') <^+ \mathbf{K}(m)$, because for $y = x0^m \in S_{x,m}$, there is a program that given any member of $S_{x,m}$ and a program for $m$, can output $y$.

## 8.2 Related Work

The study of Kolmogorov complexity originated from the work of [Kol65]. The canonical self-delimiting form of Kolmogorov complexity was introduced in [ZL70] and treated later in [Cha75]. The universal probability $\mathbf{m}$ was introduced in [Sol64]. More information about the history of the concepts used in this chapter can be found in the textbook [LV08].

The two main results of this chapter, involving bunches and batches, are inequalities including the mutual information of the encoding of a finite set with the halting sequence. A history of the origin of the mutual information of a string with the halting sequence can be found in [VV04].

A string is stochastic if it is typical of a simple elementary probability distribution. A string is typical of a probability measure if it has a low deficiency of randomness. In the proofs of Theorems 16 and 17, the stochasticity measure of encodings of finite sets is used. The notion of the deficiency of randomness with respect to a measure follows from the work of [She83], and also studied in [KU87, V'Y87, She99]. Aspects involving stochastic objects were studied in [She83, She99, V'Y87, V'Y99].

This work uses the notion of left total machine and the notion of the infinite "border" sequence, which is equal to the binary expansion of Chaitin's Omega, (see Section 8.3). The works of [VV04, GTV01] introduced the notion of using the prefix of the border sequence to define strings into a two part code.

This chapter can be seen as an update to main result in [Eps19, Lev16], focusing on conditional complexity instead of algorithmic probability. An accessible game-theoretic proof to [EL11, Lev16] can be found in [She12]. Bunches were first introduced by [Rom03], who used them to prove properties of common information of strings.

## 8.3 Left-Total Machines

The notions of total strings and the "left-total" universal algorithm are needed in the remaining sections of the paper. We say $x \in \{0,1\}^*$ is total with respect to a machine if the machine halts on all sufficiently long extensions of $x$. More formally, $x$ is total with respect to $T_y$ for some $y \in \{0,1\}^{*\infty}$ iff there exists a finite prefix free set of strings $Z \subset \{0,1\}^*$ where $\sum_{z\in Z} 2^{-\|z\|} = 1$ and $T_y(xz) \neq \perp$ for all $z \in Z$. We say (finite or infinite) string $\alpha \in \{0,1\}^{*\infty}$ is to the "left" of $\beta \in \{0,1\}^{*\infty}$, and use the notation $\alpha \lhd \beta$, if there exists a $x \in \{0,1\}^*$ such that $x0 \sqsubseteq \alpha$ and $x1 \sqsubseteq \beta$. A machine $T$ is left-total if for all auxiliary strings $\alpha \in \{0,1\}^{*\infty}$ and for all $x, y \in \{0,1\}^*$ with $x \lhd y$, one has that $T_\alpha(y) \neq \perp$ implies that $x$ is total with respect to $T_\alpha$. A detailed discussion of Left-total machines can be found in Chapter 7.

## 8.4 Stochasticity

In algorithmic statistics, a string is stochastic if it is typical of a simple probability measure. Properties of stochastic (and non-stochastic) strings can be found in the survey [VS17]. The deficiency of randomness of $x$ with respect to *elementary* probability measure $Q$ and $v \in \{0,1\}^*$ is $\mathbf{d}(x|Q,v) = \lceil -\log Q(x)\rceil - \mathbf{K}(x|v)$. The function $\mathbf{d}(\cdot|Q,v)$ is a $Q$-test (up to an additive constant). It is also universal, in that for any lower semicomputable test $d$, and $v \in \{0,1\}^*$, for all $x \in \{0,1\}^*$, $d(x|Q,v) <^+ \mathbf{d}(x|Q,v) + \mathbf{K}(d|v) + \mathbf{K}(Q|v)$, as shown in [G21].

For some $j,k \in \mathbb{N}$, we say that $x \in \mathbb{N}$ is $(j,k)$-stochastic if there exists $v \in \{0,1\}^j$, with $U(v) = Q$, $Q$ being an elementary probability measure, and $\mathbf{d}(x|Q,v) \le k$. The stochasticity of $x \in \mathbb{N}$, is measured by $\mathbf{Ks}(x) = \min\{j + 3k : x \text{ is } (j,k) \text{ stochastic}\}$. The conditional stochasticity form[1] is represented by $\mathbf{Ks}(x|\alpha)$, for $\alpha \in \{0,1\}^{*\infty}$. The definition of stochasticity can be changed to minimizing $j + ck$ for any constant $c \ge 3$, and the proofs will still hold.

Stochasticity follows non-growth laws; a total computable function cannot increase the stochasticity of a string by more than a constant factor dependent on its complexity. Lemma 5 illustrates this point. Another variant of the same idea can be found in Proposition 5 in [VS17].

**Lemma 5** *Given total computable function $f : \{0,1\}^* \to \{0,1\}^*$, $\mathbf{Ks}(f(x)) < \mathbf{Ks}(x) + O(\mathbf{K}(f))$.*

**Proof.** Let $v \in \{0,1\}^*$ realize $\mathbf{Ks}(x)$, with $U(v) = Q$, $\|v\| + 3\max\{\mathbf{d}(x|Q,v), 1\} = \mathbf{Ks}(x)$. Let $f(Q)$ be the image distribution of $Q$ with respect to $f$. Thus $f(Q)(a) = \sum_{b:f(b)=a} Q(b)$. The function $\mathbf{d}(f(\cdot)|f(Q),v)$ is a $Q$ test (relative to $v$ and up to an additive constant), because

$$\sum_{a\in\{0,1\}^*} 2^{\mathbf{d}(f(a)|f(Q),v)} Q(a) = \sum_{b\in\{0,1\}^*} 2^{\mathbf{d}(b|f(Q),v)} f(Q(b)) < O(1).$$

Also $\mathbf{d}(f(\cdot)|f(Q),v)$ is lower semi-computable given $v$, with $\mathbf{K}(\mathbf{d}(f(\cdot)|f(Q),v)|v) <^+ \mathbf{K}(f|v)$. So due to the universality of $\mathbf{d}$, $\mathbf{d}(f(x)|f(Q),v) <^+ \mathbf{d}(x|Q,v) + \mathbf{K}(f|v) <^+ \mathbf{d}(x|Q,v) + \mathbf{K}(f)$. Let $v' = v_0 v v_f \in \{0,1\}^*$ compute $f(Q)$, where $v_0$ is helper code of size $O(1)$ and $v_f$ is a shortest program that computes $f$, with $\|v_f\| = \mathbf{K}(f)$ . So $\|v'\| <^+ \|v\| + \mathbf{K}(f)$. Since $\mathbf{K}(x|v) <^+$

[1] This is formally represented as $\mathbf{Ks}(x|\alpha) = \min\{j + 3k : \exists v \in \{0,1\}^j, U_\alpha(v) = \langle Q\rangle, \mathbf{d}(x|Q,\langle v,\alpha\rangle) \le k \in \mathbb{N}\}$.

$\mathbf{K}(x|v') + \mathbf{K}(v'|v) <^+ \mathbf{K}(x|v') + \mathbf{K}(f)$, we have that $\mathbf{d}(f(x)|f(Q), v') <^+ \mathbf{d}(x|Q, v) + O(\mathbf{K}(f))$. So

$$\begin{aligned}\mathbf{Ks}(f(x)) &\leq \|v'\| + 3\max\{\mathbf{d}(f(x)|f(Q), v'), 1\}\\ &<^+ \|v\| + 3\max\{\mathbf{d}(f(x)|f(Q), v'), 1\} + \mathbf{K}(f)\\ &< \|v\| + 3\max\{\mathbf{d}(x|Q, v), 1\} + O(\mathbf{K}(f))\\ &\leq \mathbf{Ks}(x) + O(\mathbf{K}(f)).\end{aligned}$$

□

The following theorem is from [Eps19, Lev16]. Another proof of this theorem can be found in [She12]. It states that sets that are not exotic, i.e. sets with low mutual information with the halting sequence, have simple members that contain a large portion of the algorithmic weight of the sets. It is compatible with this paper's stochasticity definition because the term **Ks** used in this paper is larger than the stochasticity measure used in [Eps19, Lev16].

**Theorem 15** *For finite set $D \subset \{0,1\}^*$, $\min_{x\in D} \mathbf{K}(x) <^+ \lceil -\log \mathbf{m}(D)\rceil + 2\mathbf{K}(\lceil -\log \mathbf{m}(D)\rceil) + \mathbf{Ks}(D)$.*

## 8.5 Batches

Recall that a $(k,l)$ batch $X$ is a finite set of strings, where $k = \#X$, $l > k$, and for all $x \in X$, $\mathbf{E}_{x'\in X}[\mathbf{K}(x|x')] \leq l$. The following theorem states that for non-exotic batches, there is an element of $X$ that is simple, on average, conditional to all other members of $X$.

**Theorem 16** *For $(k,l)$ batch $X$, $\min_{x\in X} \mathbf{E}_{x'\in X}[\mathbf{K}(x|x')] <^{\log} l - k + \mathbf{I}(X : \mathcal{H})$.*

**Proof.**

We can assume that $k > 2$, otherwise the theorem is trivially proven. Let $b$ be the shortest total string where $\max_{y\in X} \mathbf{E}_{x'\in X}[\mathbf{K}[b](y|x')] < l + 2$, dubbed property $A$. Thus $\mathbf{K}(b|X) <^+ \mathbf{K}(\|b\|, (l-k))$. This is because, firstly, $l$ can be constructed from $(l-k)$ and $X$. This is because from $X$, one can compute $\#X$, and thus $k$. Then from a program that computes $(l-k)$ and $k$, one can compute $l$. Secondly there exists a program that can enumerate all total strings of length $\|b\|$ from "left" to "right". For each enumerated total string $h$ of length $\|b\|$, one can compute $\mathbf{K}[h]$ for all strings, and thus $\max_{y\in X} \mathbf{E}_{x'\in X}[\mathbf{K}[h](y|x')]$. This program can select the first one with property $A$. The first one selected will be $b$, otherwise there exists a $b' \lhd b$, $\|b'\| = \|b\|$, with property $A$. This implies there exists a total $b'^-$ with $\mathbf{K}[b'^-] \leq \mathbf{K}[b']$. Thus property $A$ holds for $b'^-$, contradicting the minimal length of $b$. This also implies $b^-$ is not total.

Let $S = \mathrm{Supp}(\mathbf{m}[b])$ be the support of $\mathbf{m}[b]$, which is finite. Let $\mathcal{G}$ be the infinite set of all functions $g : S \to \mathbb{N}$. Since $S$ is finite, each $g \in \mathcal{G}$ can be encoded in an explicit finite string. Let $\kappa : \mathcal{G} \to \mathbb{R}_{\geq 0}$ be a probabilility measure where $\kappa(g) = \prod_{a\in S} 2^{-g(a)}$. So for all $a \in S$, it must be that $\kappa(\{g : g(a) = n\}) = 2^{-n}$ and $\kappa(\{g : g(a) \geq n\}) = 2^{-n+1}$.

For any finite set $H \subset \{0,1\}^*$, $\#H > 2$, let $\mathcal{G}_1^H$ be the set of functions $g \in \mathcal{G}$, where there exists $x_g \in H$ with $g(x_g) = \#H - 2$. Using the fact that $(1-m)e^m \leq 1$ for $m \in [0,1]$, we have that

$$\kappa(\mathcal{G} \setminus \mathcal{G}_1^H) \leq \prod_{a\in H}\left(1 - 2^{-\#H+2}\right) \leq \left(1 - 2^{-\#H+2}\right)^{2^{\#H-1}} \leq e^{-2^{-\#H+2}2^{\#H-1}} = e^{-2} < 0.25.$$

So $\kappa(\mathcal{G}_1^H) > 0.75$.

We use measures $P'_g(y|x') : \{0,1\}^* \to \mathbb{R}_{\geq 0}$, indexed by $g \in \mathcal{G}$ and $x' \in S$. The measure $P'$ is defined as $P'_g(y|x') = [\delta_g(y,x') > 0]2^{-\delta_g(y,x')}\delta_g(y,x')^{-2} + [\delta_g(y,x') \leq 0]$, where $\delta_g(y,x') = \mathbf{K}[b](y|x') - g(y)$. By the definition of measurement, for a set $B \subseteq \{0,1\}^*$, we have that $P'_g(B|x') = \sum_{a\in B} P'_g(a|x')$.

We define a second set of functions $\mathcal{G}_2^H = \{g : \mathbf{E}_{x'\in H}[P'_g(S|x')] \leq 8, g \in \mathcal{G}\}$. The bound of 8 is chosen to satisfy a Markov inequality later in the proof. So

$$
\begin{aligned}
&\mathbf{E}_{g\sim\kappa}\mathbf{E}_{x'\in H}[P'_g(S|x')] \\
&= |H|^{-1} \sum_{x'\in H, y\in S} \mathbf{E}_{g\sim\kappa}[P'_g(y|x')] \\
&= |H|^{-1} \sum_{x'\in H} \sum_{y\in S} \left( \sum_{c=1}^{\mathbf{K}[b](y|x')-1} 2^{c-\mathbf{K}[b](y|x')}(\mathbf{K}[b](y|x') - c)^{-2}\kappa(\{g : g \in \mathcal{G}, g(y) = c\}) \right. \\
&\qquad\qquad \left. + \kappa(\{g : g \in \mathcal{G}, g(y) \geq \mathbf{K}[b](y|x')\}) \right) \\
&\leq |H|^{-1} \sum_{x'\in H} \sum_{y\in S} \left( \mathbf{m}[b](y|x') \sum_{c=1}^{\mathbf{K}[b](y|x')-1} (\mathbf{K}[b](y|x') - c)^{-2} \right) + 2^{-\mathbf{K}[b](y|x')+1} \\
&\leq |H|^{-1} \sum_{x'\in H} 2\mathbf{m}[b](S|x') + 2\mathbf{m}[b](S|x') < 4.
\end{aligned}
$$

So by the Markov inequality, $\kappa(\mathcal{G}_2^H) \geq 0.5$. So for all finite $H \subset \{0,1\}^*$, $\#H > 2$, $\kappa(\mathcal{G}_1^H \cap \mathcal{G}_2^H) > 0.25$.

We use the following probability measure $P_g(y|x')$, indexed by $g \in \mathcal{G}$ and $x' \in S$, defined as $P_g(y|x') = [y \in S]P'_g(y|x')/P'_g(S|x')$. Thus $P_g(\{0,1\}^*|x') = 1$ for all $x' \in S$, $g \in \mathcal{G}$. So for any $g \in \mathcal{G}_1^H \cap \mathcal{G}_2^H$, there exists $x_g \in H$ where $g(x_g) = \#H - 2$ and also

$$
\begin{aligned}
&\mathbf{E}_{x'\in H}[-\log P_g(x_g|x')] \\
&= \mathbf{E}_{x'\in H}[-\log P'_g(x_g|x') + \log P'_g(S|x')] && (8.1)\\
&= \mathbf{E}_{x'\in H}[-\log P'_g(x_g|x')] + \mathbf{E}_{x'\in H}[\log P'_g(S|x')] \\
&\leq \mathbf{E}_{x'\in H}[-\log P'_g(x_g|x')] + \log \mathbf{E}_{x'\in H}[P'_g(S|x')] \\
&<^+ \mathbf{E}_{x'\in H}[-\log P'_g(x_g|x')] && (8.2)\\
&=^+ \mathbf{E}_{x'\in H}\left[[\delta_g(x_g,x') > 0](-\log 2^{-\delta_g(x_g,x')}\delta_g(x_g,x')^{-2}) + [\delta_g(x_g,x') \leq 0]\right] && (8.3)\\
&<^+ \mathbf{E}_{x'\in H}[\max\{\delta_g(x_g,x') + 2\log\delta_g(x_g,x'), O(1)\}] \\
&<^+ \max\{\mathbf{E}_{x'\in H}[\delta_g(x_g,x')] + 2\log\mathbf{E}_{x'\in H}[\delta_g(x_g,x')], O(1)\} \\
&<^+ \max\{\mathbf{E}_{x'\in H}[\mathbf{K}[b](x_g|x') - g(x_g)] + 2\log\mathbf{E}_{x'\in H}[\mathbf{K}[b](x_g|x') - g(x_g)], O(1)\} && (8.4)\\
&<^{\log} \max\{\mathbf{E}_{x'\in H}[\mathbf{K}[b](x_g|x')] - \#H, O(1)\}. && (8.5)
\end{aligned}
$$

Equation 8.1, follows from definition of $P_g$. Equation 8.2 follows from the fact that $g \in \mathcal{G}_2^H$, and thus $\mathbf{E}_{x'\in H}[P'_g(S|x')] \leq 8$. Equation 8.3 follows from the definition of $P'_g$. Equation 8.4 follows from the definition of $\delta_g$. Equation 8.5 follows from $g \in \mathcal{G}_1^H$ and thus $g(x_g) = \#H - 2$.

Let $\{G_i\}$ be a computable enumeration of all finite subsets of $\mathcal{G}$. Let $f$ be a function that when given a set $H \subset \{0,1\}^*$, $\#H > 2$, outputs an encoding of the first finite subset $W \subset \mathcal{G}$ in the list

$\{G_i\}$ such that $W \subset \mathcal{G}_1^H \cap \mathcal{G}_2^H$ and $\kappa(W) > 0.25$. On all other inputs which are not an encoding of a finite set $H \subset \{0,1\}^*$ with $\#H > 2$, $f$ outputs the empty string. The function $f$ is total computable relative to $b$, with $\mathbf{K}(f|b) = O(1)$, because given $H$ and $b$, it is computable to determine whether a given function $g \in \mathcal{G}$ is in $\mathcal{G}_1^H \cap \mathcal{G}_2^H$.

Let $D = f(X)$. Invoking Theorem 15, conditional to $b$, gives $g \in D$, where $\mathbf{K}(g|b) <^+ \lceil -\log \mathbf{m}(D|b)\rceil + 2\mathbf{K}(\lceil -\log \mathbf{m}(D|b)\rceil) + \mathbf{Ks}(D|b)$. Since $\lceil -\log \mathbf{m}(D|b)\rceil <^+ -\log \kappa(D) + \mathbf{K}(\kappa|b) < O(1)$, we have that $\mathbf{K}(g|b) <^+ \mathbf{Ks}(D|b)$. Lemma 5, relativized to $b$, using total computable function $f$, gives $\mathbf{K}(g|b) <^+ \mathbf{Ks}(X|b)$. Lemma 1, gives

$$\mathbf{K}(g|b) <^{\log} \mathbf{I}(X : \mathcal{H}|b). \tag{8.6}$$

Since $g \in D \subset \mathcal{G}_1^X \cap \mathcal{G}_2^X$, there exists $x_g \in X$ where, due to Equation 8.5,

$$\mathbf{E}_{x' \in X}[-\log P_g(x_g|x')] <^{\log} \max\{\mathbf{E}_{x'\in X}[\mathbf{K}[b](x_g|x')] - \#X, O(1)\} <^{\log} l - k. \tag{8.7}$$

So we have that

$$\begin{aligned}
\mathbf{E}_{x'\in X}[\mathbf{K}(x_g|b, x')] &<^+ \mathbf{E}_{x'\in X}[\mathbf{K}(x_g|b, g, x') + \mathbf{K}(g|b)] \\
&=^+ \mathbf{E}_{x'\in X}[\mathbf{K}(x_g|b, g, x')] + \mathbf{K}(g|b) \\
&< \mathbf{E}_{x'\in X}[-\log P_g(x_g|x')] + \mathbf{I}(X : \mathcal{H}|b) + O(\log \mathbf{I}(X : \mathcal{H}|b)) && (8.8) \\
&< l - k + \mathbf{I}(X : \mathcal{H}|b) + O(\log \mathbf{I}(X : \mathcal{H}|b) + \log(l-k)) && (8.9) \\
\mathbf{E}_{x'\in X}[\mathbf{K}(x_g|x') - \mathbf{K}(b)] &< l - k + \mathbf{I}(X : \mathcal{H}|b) + O(\log \mathbf{I}(X : \mathcal{H}|b) + \log(l-k)) && (8.10) \\
\mathbf{E}_{x'\in X}[\mathbf{K}(x_g|x')] &< l - k + \mathbf{K}(b) + \mathbf{I}(X : \mathcal{H}|b) + O(\log(\mathbf{I}(X : \mathcal{H}|b) + \mathbf{K}(b)) + \log(l-k))
\end{aligned}$$

Equation 8.8 is due to Equation 8.6. Equation 8.9 is due to Equation 8.7. Equation 8.10 follows that for all $x' \in X$, $\mathbf{K}(x_g|x') <^+ \mathbf{K}(x_g|b, x') + \mathbf{K}(b)$.

$$\mathbf{E}_{x'\in X}[\mathbf{K}(x_g|x')] <^{\log} l - k + \mathbf{I}(X : \mathcal{H}) + \mathbf{K}(b|X) \tag{8.11}$$

$$\mathbf{E}_{x'\in X}[\mathbf{K}(x_g|x')] <^{\log} l - k + \mathbf{I}(X : \mathcal{H}) + \mathbf{K}(\langle \|b\|, (l-k)\rangle). \tag{8.12}$$

$$\mathbf{E}_{x'\in X}[\mathbf{K}(x_g|x')] <^{\log} l - k + \mathbf{I}(X : \mathcal{H}). \tag{8.13}$$

Equation 8.11 is due to the invocation of Lemma 3. Equation 8.12 is due to the fact that $\mathbf{K}(b|X) <^+ \mathbf{K}(\langle \|b\|, (l-k)\rangle)$. Equation 8.13 is because $a <^{\log} b + O(\log a)$ implies $a <^{\log} b$, where $a = \|b\| <^+ \mathbf{K}(b)$ and $b = \mathbf{I}(X : \mathcal{H}) + O(\log \|b\|)$. □

## 8.6 Bunches

Recall that a $(k, l)$ bunch $X$ is a finite set of strings, where $k = \#X$, $l > k$, and for all $x, x' \in X$, $\mathbf{K}(x|x') \leq l$. The following theorem states that for non-exotic bunches, there is an element of $X$ that is simple conditional to all other members of $X$.

**Theorem 17** *For $(k, l)$ bunch $X$,* $\min_{x\in X} \max_{x'\in X} \mathbf{K}(x|x') <^{\log} 2(l-k) + \mathbf{I}(X : \mathcal{H})$.

**Informal Proof.**

*This proof starts with the definition of elementary probability measure $Q$ that realizes the stochasticity of $X$. Using probabilistic arguments, we define a $Q$-test $t_g$ that gives a high score to a set*

*$Y$ if there does not exist $a \in Y$ such that $g(a) \gtrapprox \#Y$. A measure is defined by $P_g(x|y) \approx [g(x) \geq \mathbf{K}[b](x|y) - z]2^{-z}$, where $z = l - k$. A second test $t'_g$ gives a set $Y$ a zero score if more than half of $x' \in Y$ makes $P_g(\cdot|x')$ a semi-measure. By probabilistic arguments, there exists a function $g$ such that $t_g$ and $t'_g$ are $Q$-tests. Furthermore, since $X$ is typical of $Q$, $t_g(X) = t'_g(X) = 0$. Thus there exist $x_g \in X$ where $g(x_g) \geq \#X \geq \mathbf{K}[b](x|y) - z$, for all $x \in X$.. This means that $P_g(x_g|y) \approx 2^{-z}$ for all $y \in X$. By the fact that $t'(X) = 0$, for more than half $x' \in X' \subseteq X$, $P_g(\cdot|x')$ is a semimeasure, and thus $\mathbf{K}(x_g|x') \lessapprox -\log P_g(x_g|x')$. For $x' \in X'$, the bound of theorem is achieved. For $y' \in X \setminus X'$, there exists $\approx 2^k$ programs from $y'$ to $y \in X'$, and then there is a short program from $y$ to $x_g$ using $P_g$. Thus the algorithmic probability of $\mathbf{m}(x_g|y')$ is large and the bounds for $y' \in X \setminus X'$ is achieved. The remainder of the proof uses Lemma 1 to replace stochasticity with mutual information with the halting sequence and Lemma 3 to remove the total string $b$.*

**Proof.**

**(1.)** *The first step of the proof is to find a total string $b$ such that $X$ is a bunch with computable complexity $\mathbf{K}[b]$, with $\max_{x,x' \in X} \mathbf{K}[b](x|x') <^+ l$. This enables the proof to move forward with computable complexity and probability. The total string $b$ is factored out at the end of the proof. In this section, the probability measure $Q$ that realizes the stochasticity of $X$ is defined.*

Let $z = l - k$ and let $b$ be the shortest total string where $\max_{x,x' \in X} \mathbf{K}[b](x|x') < l + 2$, which we call satisfying property A. Thus $\mathbf{K}(b|X) <^+ \mathbf{K}(\langle z, \|b\| \rangle)$ and $b^-$ is not total, using arguments in the first paragraph of the proof of Theorem 16. Let $s = \langle b, z \rangle$. Let $v \in \{0,1\}^*$ and elementary probability measure $Q$ minimize $\mathbf{Ks}(X|s)$, where $U_s(v) = Q$. Recall that elementary measures are introduced in Chapter 2. Let $d = \max\{\mathbf{d}(X|Q, \langle v, s \rangle), 1\}$. Let $S = \bigcup\{Y : \langle Y \rangle \in \mathrm{Supp}(Q)\}$ be the union of all sets encoded in the support of $Q$. Since $Q$ is elementary, $|S| < \infty$. Let $\mathcal{G}$ be the set of all functions $g : S \to \mathbb{N}$. Since $S$ is finite, each $g \in \mathcal{G}$ can be encoded with an explicit finite string.

**(2.)**. *We define a probability measure $\kappa$ over functions $g \in \mathcal{G}$ from the union of the support of $Q$ to natural numbers, where functions with low values will have a higher probability.*

Let $\kappa : \mathcal{G} \to \mathbb{R}_{\geq 0}$ be a probability measure over $\mathcal{G}$, where $\kappa(g) = \prod_{a \in S} 2^{-g(a)}$. So for all $a \in S$, $\kappa(\{g : g(a) \geq n\}) = 2^{-n+1}$. Let $c \in \mathbb{N}$ be a constant solely dependent on $U$ to be determined later.

**(3.)** *The proof only works with $X$ having a minimum number of elements. Otherwise the theorem is trivially solved. This is a boundary case that can be skipped on first reading.*

We assume that $|X| > 16(c + d)$. Otherwise, $k <^+ \log d$, and then $\min_{x \in X} \max_{x' \in X} \mathbf{K}(x|x', s) \leq l <^+ z + \log d <^+ 2z + \mathbf{Ks}(X|s)$. From this point, the reasoning starting at Equation 8.14 can be used to prove the theorem.

**(4.)** *We define the first of two tests, $t_g$, parameterized by a function $g \in \mathcal{G}$. We will show later in the proof there is a $g$ such that $t_g$ is a $Q$ test. $t_g$ gives a high score to sets $Y$ such that all their elements $a \in Y$ have low $g$ score.*

We define the following function $t$ over $\mathrm{Supp}(Q)$, parameterized by $g \in \mathcal{G}$. Let $\mathcal{B}$ be the set of sets $G$ such that for all $x, x' \in G$, $\mathbf{K}[b](x|x') < \#G + z + 2$. Let $t_g(Y) = e^{2(d+c)-1}$ if $Y \cap \{a : g(a) \geq \lfloor \log(|Y|/(c + d)) \rfloor\} = \emptyset$ and $Y \in \mathcal{B}$, otherwise $t_g(Y) = 0$.

**(5.)** *Using probabilistic arguments, it is shown that the expectation of $t_g$ over $Q$ and $\kappa$ is small. This is required for probabilistic arguments to show the existence of a $g \in \mathcal{G}$ with $t_g$ being a $Q$-test.*

So, using the fact that $(1-m)e^m \leq 1$,

$$\begin{aligned}
&\mathbf{E}_{g\sim\kappa}\mathbf{E}_{Y\sim Q}[t_g(Y)]\\
&= \sum_{Y\in\mathcal{B}} Q(Y)\kappa(\{g : \forall_{a\in Y}, g(a) < \lfloor \log(|Y|/(c+d))\rfloor\})e^{2(d+c)-1}\\
&= \sum_{Y\in\mathcal{B}} Q(Y)\prod_{a\in Y}\kappa(\{g : g(a) < \lfloor \log(|Y|/(c+d))\rfloor\})e^{2(d+c)-1}\\
&= \sum_{Y\in\mathcal{B}} Q(Y)\prod_{a\in Y}\left(1-2^{-\lfloor \log(|Y|/(c+d))\rfloor+1}\right)e^{2(d+c)-1}\\
&\leq \sum_{Y\in\mathcal{B}} Q(Y)(1-2(c+d)/|Y|)^{|Y|}e^{2(d+c)-1}\\
&\leq \sum_{Y\in\mathcal{B}} Q(Y)e^{-2(c+d)}e^{2(c+d)-1} < 0.5.
\end{aligned}$$

**(6.)** *The measure $P_g$ is defined, parameterized by $g \in \mathcal{G}$ gives $P_g(x|y)$ a score of $\approx 2^{-z}$ if $g(x) \gtrapprox \mathbf{K}[b](x|y) - z$ and 0 otherwise. The constants and max function ensure proper boundary conditions and can be discounted on a first reading. By definition, the expecion of $P_g(S|y)$, over $g$ distributed by $\kappa$ is small.*

For each $x, y \in S$, $g \in \mathcal{G}$, we define the following function

$$P_g(x|y) = [g(x) \geq \max\{\mathbf{K}[b](x|y) - z - \lceil \log(c+d)\rceil - 3, 1\}]2^{-z-2(d+c)}.$$

Thus $P_g(x|y)$ is only one of two values, either 0 or $2^{-z-2(d+c)}$. $P_g(S|y) = \sum_{x\in S} P_g(x|y)$. So for all $y \in S$, we have

$$\begin{aligned}
&\mathbf{E}_{g\sim\kappa}[P_g(S|y)]\\
&= 2^{-z-2(d+c)}\sum_{x\in S}\kappa(\{g : g(x) \geq \max\{\mathbf{K}[b](x|y) - z - \lceil \log(d+c)\rceil - 3, 1\})\\
&= 2^{-z-2(d+c)}\sum_{x\in S} 2^{-\max\{\mathbf{K}[b](x|y)-z-\lceil \log(d+c)\rceil-3,1\}+1}\\
&\leq 2^{-z-2(d+c)}\sum_{x\in S}\mathbf{m}[b](x|y)2^{z+\lceil \log(d+c)\rceil+4}\\
&\leq 2^{-(d+c)+4}.
\end{aligned}$$

**(7.)** *We define an indicator function $\mathbf{I}_g(y)$ which is 0 iff $P_g(\cdot|y)$ is a semi-measure, and $\mathbf{I}_g(Y)$ counts the number of non semi-measures using $y \in Y$. Using bounds of the previous section, an upper bound on the expectaction of $\mathbf{I}_g$ is given.*

For all functions $g \in \mathcal{G}$, we define the following indicator function, with $\mathbf{I}_g(y) = [P_g(S|y) > 1]$. Furthermore, we extend the domain $\mathbf{I}$ to be over sets $Y \in \mathrm{Supp}(Q)$, with $\mathbf{I}_g(Y) = \sum_{y\in Y}\mathbf{I}_g(y)$. Thus $\mathbf{I}_g(y) = 0$ iff $P_g(\cdot|y)$ is a semimeasure where each $x \in \mathrm{Supp}(P_g(\cdot|y))$ can be identified by code of size $=^+ -\log P_g(x|y)$. For each such $y \in S$, the expectation of $\mathbf{I}$ with respect to $\kappa$ is small, and for $Y \in \mathrm{Supp}(Q)$, we have

$$\mathbf{E}_{g\sim\kappa}[\mathbf{I}_g(y)] \leq \mathbf{E}_{g\sim\kappa}[P_g(S|y)] \leq 2^{-(c+d)+4}$$
$$\mathbf{E}_{g\sim\kappa}[\mathbf{I}_g(Y)] \leq |Y|2^{-(c+d)+4}.$$

**(8.)** *We define the second test function $t'$, parameterized by $g \in \mathcal{G}$. It gives a set a zero score if $P_g$ is a semi-measure for at least half its elements. Otherwise it gives the set a high score. Through probabilistic arguments $t'_g$ has low $Q$ expectation when $g$ is distributed by $\kappa$. Note that since $\mathbf{E}_{g\sim\kappa}[\mathbf{I}_g(Y)] \leq |Y|2^{-(c+d)+4}$, by the Markov inequality $\kappa(g : \mathbf{I}_g(Y) \geq 0.5|Y|) \leq 2^{-(c+d)+5}$.*

We define the function $t' : \mathrm{Supp}(Q) \to \mathbb{R}_{\geq 0}$, parameterized by $g \in \mathcal{G}$, which will give a set $Y$ a zero score iff $P_g(\cdot|y)$ is a semi-measure for at least half of the elements $y \in Y$. Otherwise $t'_g$ gives $Y$ a high score. More formally, let $t'_g(Y) = 0$ if $\mathbf{I}_g(Y) < .5|Y|$ and $t'_g(Y) = 2^{(d+c)-7}$, otherwise. Thus we have that, due to the Markov inequality,

$$\begin{aligned}
\mathbf{E}_{g\sim\kappa}\mathbf{E}_{Y\sim Q}[t'_g(Y)] &= \sum_Y Q(Y)\mathbf{E}_{g\sim\kappa}[[g : \mathbf{I}_g(Y) \geq 0.5|Y|]]2^{c+d-7} \\
&= \sum_Y Q(Y)\kappa(\{g : \mathbf{I}_g(Y) \geq 0.5|Y|\})2^{c+d-7} \\
&\leq \sum_Y Q(Y)2^{-(c+d)+5}2^{c+d-7} \\
&= 0.25.
\end{aligned}$$

**(9.)** *Since the $\kappa$-expectation of $t_g$ and $t'_g$ are small, by probabilistic arguments, there is a $g \in \mathcal{G}$ where $t_g$ and $t'_g$ are both $Q$-tests. Using similar arguments to that in the proof of Theorem 16, it is proven that $t_g(X) = t'_g(X) = 0$.*

By probabilistic arguments, there exists $g \in \mathcal{G}$, such that $\mathbf{E}_{Y\sim Q}[t_g(Y)] \leq 1$ and $\mathbf{E}_{Y\sim Q}[t'_g(Y)] \leq 1$. So both $t_g(\cdot)Q(\cdot)$ and $t'_g(\cdot)Q(\cdot)$ are semi-measures. Furthermore, $\mathbf{K}(g|c,d,v,s) = O(1)$. It must be that $t_g(X) = 0$. Otherwise, for proper choice of $c$ solely dependent on $U$,

$$\begin{aligned}
d &= \mathbf{d}(X|Q,v,s) \\
&= \lceil -\log Q(X) \rceil - \mathbf{K}(X|v,s) \\
&> -\log Q(X) - (-\log t_g(X)Q(X) + \mathbf{K}(t_g(\cdot)Q(\cdot)|v,s)) - O(1) \\
&> -\log Q(X) - (-\log t_g(X)Q(X) + \mathbf{K}(g,Q|v,s)) - O(1) \\
&> 2(c+d)(\log e) - \mathbf{K}(c,d) - O(1) \\
&> d,
\end{aligned}$$

causing a contradiction. Thus $c$ is chosen to be large enough to have the property $c > \mathbf{K}(c) + O(1)$, where the additive constant is depedent solely on the universal Turing machine. The same reasoning can be used to show that $t'_g(X) = 0$. We roll $c$ into the additive constants of the theorem and remove it from consideration for the rest of the proof.

**(10.)** *Since $t_g(X) = 0$, there exists $a \in X$ where $g(a)$ has a high score, with $g(a) \gtrapprox \mathbf{K}[b](a|y) - z$, for all $y \in X$. The inequality follows from $k \geq k + \mathbf{K}[b](a|y) - l = \mathbf{K}[b](a|y) - z$. This ensures that $P_g(a|y) \approx 2^{-z}$ for all $y \in X$.*

Therefore, since $t_g(X)=0$, there exists $a \in X$ where for all $y \in X$, using the fact that $|Y| > 16(c+d)$,

$$\begin{aligned} g(a) &\geq \lfloor \log(|Y|/(d+c) \rfloor \\ &\geq \lfloor \log|Y| \rfloor - \lceil \log(d+c) \rceil \\ &\geq k-1-\lceil \log(d+c) \rceil \\ &\geq \max\{\mathbf{K}[b](a|y) - z - \lceil \log(d+c) \rceil - 3, 1\}. \end{aligned}$$

This ensures that $P_g(a|y) > 0$ for all $y \in X$, due to the definition of $P_g$.

**(11.)** *Since $t'_g(X)=0$, $P_g(\cdot|y)$ is a semimeasure for more than half $X'$ of $y \in X$. Thus $P_g(a|y)$ can be used to identify a given $y$ in this subset $X'$ and the desired bound on $\mathbf{K}(a|y)$ is achieved. Otherwise for $y' \notin X'$, a program can be created that computes some $y \in X'$ from $y'$ (bounded by $l$) and then use the bound proved of $\mathbf{K}(a|y)$. Since there is a lot of $y \in X'$, there is a lot of such programs, meaning the algorithmic probability of $\mathbf{m}(a|y')$ is large, and thus the bound is achieved.*

Furthermore, since $t'_g(X)=0$, there is a subset $X' \subseteq X$, $|X'| > 2^{k-2}$, where for all $y \in X'$, $P_g(\cdot|y)$ is a semimeasure. For such $y$, $\mathbf{K}(a|y,s) <^+ -\log P_g(a|y) + \mathbf{K}(g|d,v,s) + \mathbf{K}(d,v|s) <^+ z + 3d + \|v\| <^+ z + \mathbf{Ks}(X|s)$. Therefore for all $y' \in X \setminus X'$,

$$\begin{aligned} \mathbf{K}(a|y',s) &<^+ -\log \sum_{y \in X'} 2^{-\mathbf{K}(a|y,s) - \mathbf{K}(y|y',s)} \\ &<^+ -\log \sum_{y \in X'} 2^{-l-z-\mathbf{Ks}(X|s)} \\ &<^+ 2z + \mathbf{Ks}(X|s). \end{aligned}$$

**(12.)** *The following theorem removes the stochasticity term and the total string $b$, similarly to the proof of Theorem 16.*

So for all $x \in X$,

$$\begin{aligned} \mathbf{K}(a|x,s) &<^+ 2z + \mathbf{Ks}(X|s) && (8.14) \\ \mathbf{K}(a|x) &<^+ 2z + \mathbf{K}(s) + \mathbf{Ks}(X|s) \\ &< 2z + \mathbf{K}(b) + \mathbf{Ks}(X|s) + O(\log z) \\ &< 2z + \mathbf{K}(b) + \mathbf{I}(X:\mathcal{H}|s) + O(\log z + \log \mathbf{I}(X:\mathcal{H}|s)) && (8.15) \\ &< 2z + \mathbf{K}(b) + \mathbf{I}(X:\mathcal{H}|b) + O(\log z + \log(\mathbf{I}(X:\mathcal{H}|b) + \mathbf{K}(b))) \\ &<^{\log} 2z + \mathbf{I}(X:\mathcal{H}) + \mathbf{K}(b|X) && (8.16) \\ &<^{\log} 2z + \mathbf{I}(X:\mathcal{H}) + \mathbf{K}(\langle \|b\|, z \rangle) \\ &<^{\log} 2z + \mathbf{I}(X:\mathcal{H}). && (8.17) \end{aligned}$$

Equation 8.15 is due to the application of Lemma 1. Equation 8.16 is due to the application of Lemma 3. Equation 8.17 uses the same logic as Equation 8.13 in the proof of Theorem 16. □

# Chapter 9

# Extending Chaitin's Incompleteness Theorem

## 9.1 Introduction

Gödel's famous incompleteness theorem states that any theory $\mathcal{F}$ that is consistent, recursively axiomatizable, and "sufficiently rich" (contains Robinson-arithmetic $\mathcal{Q}$, or $\mathcal{Q}$ can be interpreted in it) is incomplete, in that there exists true statements that cannot be proven in it.

It is well known that there is no recursive method to determine a non constant lower bound on Kolmogorov complexity, $\mathbf{K}$. Chaitin's incompleteness theorem proves there exist no logical means to prove lower bounds on $\mathbf{K}$. Let $\mathcal{F}$ be as above, and significantly strong to make assertions about the Kolmogorov complexity of strings. Furthermore, let $\mathcal{F}$ be sound. Then we get the celebrated theorem.

**Theorem. (Chaitin's Incompleteness Theorem)** *For theory $\mathcal{F}$, there is a constant $c$ such that $\mathcal{F}$ does not prove $c < \mathbf{K}(x)$ for any $x$.*

The proof is straightforward. Assume otherwise. Take any $c$ and enumerate proofs of $\mathcal{F}$ until it proves the statement $c < \mathbf{K}(x)$ for some $x$. Then return $x$. This implies that $\mathbf{K}(x) < O(\log c)$, causing a contradiction for large enough $c$.

However this theorem doesn't prohibit the existence of formal systems that prove $c < \mathbf{K}(x)$ for a finite but very large number of strings. Or for our purposes, the above theorem doesn't prohibit theories which prove $\mathbf{K}(x) = c$ for a large (but finite) number of strings. Such theories are not to be expected to be accessible by logicians. In this chapter, we prove such systems are exotic, and cannot exist in the physical world. To do so we use two steps. The first step proves the following theorem, which states $\mathbf{K}$ is uniformly uncomputable.

**Theorem.** *A relation $X \subset \mathbb{N} \times \mathbb{N}$ of $2^n$ unique pairs $(b, \mathbf{K}(b))$ has $n <^{\log} \mathbf{I}(X; \mathcal{H})$.*

The term $\mathcal{H}$ is the halting sequence. The information term is $\mathbf{I}(x; \mathcal{H}) = \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H})$. The second part involves invoking the Indepedence Postulate (**IP**), introduced in [Lev84, Lev13]. **IP** is an unprovable statement that physical sequences are independent from mathematical ones. Among other applications, **IP** can be interpreted as a finitary Church-Turing thesis. The statement is as follows.

**IP:** *Let $\alpha$ be a sequence defined with an $n$-bit mathematical statement (e.g., in Peano*

*Arithmetic), and a sequence $\beta$ can be located in the physical world with a $k$-bit instruction set (e.g., ip-address). Then $\mathbf{I}(\alpha:\beta) < k+n+c$, for some small absolute constant $c$.*

We rework **IP** so that $x=\alpha\in\{0,1\}^*$, $\beta$ is equal to the halting sequence $\mathcal{H}$, and the information term is equal to $\mathbf{I}(x;\mathcal{H}) = \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H})$. Since $\mathcal{H}$ can be described by an $O(1)$ bit mathematical sequence, we get

$$\mathbf{I}(x;\mathcal{H}) <^+ \mathbf{Address}(x).$$

Let $\mathcal{F}$ be a formal system defined in Chaitin's Incompleteness Theorem. Assume that $\mathcal{F}$ can be used to prove $\mathbf{K}(x_i)=c_i$ for $2^n$ unique strings $x_i$. Then by Theorem 19, Lemma 2, and **IP**,

$$n <^{\log} \mathbf{I}(\{(x_i,c_i)\};\mathcal{H}) <^{\log} \mathbf{I}(\mathcal{F};\mathcal{H}) <^{\log} \mathbf{Address}(\mathcal{F}).$$

Thus as the number strings with proved Kolmogorov complexities grows, the formal system $\mathcal{F}$ becomes exotic and by **IP**, inaccessible in the physical world. For related work, in [Lev13], it was shown that consistent completions of PA have infinite mutual information with $\mathcal{H}$ and thus have infinite addresses. This paper extends this result by proving the existence of theories with finite mutual information with the halting sequence. Note that Theorem 19 can be generalized to binary relations that approximate Kolmogorov complexity.

## 9.2 Results

Let $\Omega = \sum\{2^{-\|p\|} : U(p) \text{ halts}\}$ be Chaitin's Omega, $\Omega_n \in \mathbb{Q}_{\geq 0}$ be be the rational formed from the first $n$ bits of $\Omega$, and $\Omega^t = \sum\{2^{-\|p\|} : U(p) \text{ halts in time } t\}$. For $n\in\mathbb{N}$, let $\mathbf{bb}(n) = \min\{t : \Omega_n < \Omega^t\}$. $\mathbf{bb}^{-1}(m) = \arg\min_n\{\mathbf{bb}(n-1) < m \leq \mathbf{bb}(n)\}$. Let $\Omega[n]\in\{0,1\}^*$ be the first $n$ bits of $\Omega$.

**Lemma 6** *For $n=\mathbf{bb}^{-1}(m)$, $\mathbf{K}(\Omega[n]|m,n) = O(1)$.*

**Proof.** For a string $x$, let $BB(x) = \inf\{t : \Omega^t > 0.x\}$. Enumerate strings of length $n$, starting with $0^n$, and return the first string $x$ such that $BB(x)\geq m$. This string $x$ is equal to $\Omega[n]$, otherwise let $y$ be the largest common prefix of $x$ and $\Omega[n]$. Thus $BB(y) = \mathbf{bb}(\|y\|) \geq BB(x) \geq m$, which means $\mathbf{bb}^{-1}(m) \leq \|y\| < n$, causing a contradiction. □

**Theorem 18** *A relation $X\subset\mathbb{N}\times\mathbb{N}$ of $2^n$ unique pairs $(b,\mathbf{K}(b))$ has $n <^{\log} \mathbf{I}(X;\mathcal{H})$.*

**Proof.** We relativize the universal Turing machine to $n$. Let $X = \{x_i,c_i\}_{i=1}^{2^n}$, and $T = \min\{t : \mathbf{K}_t(x_i) = c_i = \mathbf{K}(x_i), \text{ for } i=1,\dots,n\}$. Let $N=\mathbf{bb}^{-1}(T)$ and $B=\mathbf{bb}(N)$. We relativize the universal Turing machine to $B$. Later on, we will make this relativization explicit. We also assume that $c_i > n$. If this is not the case, then one can construct $X'\subset X$ of size $2^{n-1}$ with $c_i > n-1$ and use $X'$ instead.

Let $m(x) = 2^{-\mathbf{K}_B(x)}$. Let $Q$ be an elementary probability measure that realizes $\mathbf{Ks}(X)$ and $d=\max\{\mathbf{d}(X|Q),1\}$. Without loss of generality, the support of $Q$ is restricted to finite binary relations $B\subset\mathbb{N}\times\mathbb{N}$ of size $2^n$. Let $B_1 = \bigcup\{y : (y,c)\in B\}$. Let $S=\bigcup\{B_1 : B\in\text{Support}(Q)\}$. We randomly select each string in $S$ to be in a set $R$ independently with probability $d2^{-n}$. Thus $\mathbf{E}[m(R)] \leq d2^{-n}$. For $B\in\text{Support}(Q)$,

$$\begin{aligned}&\mathbf{E}_R\mathbf{E}_{B\sim Q}[[R\cap B_1=\emptyset]]\\ =&\mathbf{E}_{B\sim Q}\Pr(R\cap B_1=\emptyset)\\ =&(1-d2^{-n})^{2^n} < e^{-d}.\end{aligned}$$

Thus there exists a set $R \subseteq S$ such that $\mathbf{m}(R) \leq 2{\cdot}2^{-n}$ and $\mathbf{E}_{B\sim Q}[[R \cap B_1 = \emptyset]] < 2e^{-d}$. Let $t(B) = .5[R \cap B_1 = \emptyset]2^d$. $t$ is a $Q$-test, with $\mathbf{E}_{B\sim Q}[t(B)] \leq 1$. It must be that $t(X) \neq 0$, otherwise,

$$1.44d - 1 < \log t(X) <^+ \mathbf{d}(X|Q) + \mathbf{K}(t|Q) <^+ d + \mathbf{K}(d),$$

which is a contradiction for large enough $d$, which one can assume without loss of generality. Thus $t(X) \neq 0$ and $R \cap X_1 \neq \emptyset$. Furthermore, if $y \in R$, $\mathbf{K}(y) <^+ -\log m(x) - n + \log d + \mathbf{K}(m, R)$. So for $x \in R \cap X_1$, making the relativization of $B$ explicit. By Lemma 1,

$$\begin{aligned}
\mathbf{K}(x|B) &<^+ -\log m(x) - n + \log d + \mathbf{K}(m, R|B) \\
\mathbf{K}(x) - \mathbf{K}(B) &<^+ \mathbf{K}(x) - n + \log d + \mathbf{K}(S|B) \\
n &<^+ \mathbf{K}(B) + \log d + \mathbf{K}(d, Q|B) \\
n &<^+ \mathbf{K}(B) + \mathbf{Ks}(X|B) \\
n &<^{\log} \mathbf{K}(B) + \mathbf{I}(X; \mathcal{H}|B) \\
n &<^{\log} \mathbf{K}(B) + \mathbf{K}(X|B) - \mathbf{K}(X|\mathcal{H}) + O(\log N) \qquad (9.1)
\end{aligned}$$

Equation 9.1 is due to the fact that $B$ is computable from $\Omega[N]$, thus it is computable from $\mathcal{H}$ and $N$. So we have,

$$\begin{aligned}
&\mathbf{K}(X|B) + \mathbf{K}(B) \\
&<^+ \mathbf{K}(X|B, \mathbf{K}(B)) + \mathbf{K}(\mathbf{K}(B)|B) + \mathbf{K}(B) \\
&<^+ \mathbf{K}(X, B) + \mathbf{K}(\mathbf{K}(B)|B) \qquad (9.2) \\
&<^+ \mathbf{K}(X, N, B) + O(\log N) \qquad (9.3) \\
&<^+ \mathbf{K}(X, N) + O(\log N). \qquad (9.4) \\
&<^+ \mathbf{K}(X) + O(\log N). \\
n &<^{\log} \mathbf{K}(X) - \mathbf{K}(X|\mathcal{H}) + O(\log N). \qquad (9.5)
\end{aligned}$$

Equation 9.2 is from the chain rule. Equation 9.3 is from the fact that $M = \mathbf{bb}(N)$. Equation 9.4 comes from $\mathbf{K}(T|X) = O(1)$ and Lemma 6, which implies $\mathbf{K}(B|N, T) <^+ \mathbf{K}(\Omega[N]|N, T) <^+ O(1)$.

From $X$, one can compute $T$, where $\mathbf{bb}^{-1}(T) = N$. Therefore by Lemma 7, $\mathbf{K}(\Omega[N]|X) <^+ \mathbf{K}(N)$, so by Lemma 2,

$$N <^{\log} \mathbf{I}(\Omega[N]; \mathcal{H}) <^{\log} \mathbf{I}(X; \mathcal{H}) + \mathbf{K}(N) <^{\log} \mathbf{I}(X; \mathcal{H}). \qquad (9.6)$$

The above equation used the common fact that the first $n$ bits of $\Omega$ had $n - O(\log n)$ bits of mutual information with $\mathcal{H}$. So combining Equations 9.5 and 9.6, we get

$$n <^{\log} \mathbf{I}(X; \mathcal{H}).$$

$\square$

# Chapter 10

# A Small Theorem for Small m

## 10.1 Introduction

In this chapter, we show that semi measures that majorize the algorithmic probability have infinite mutual information with the halting sequence.

For a probability $p$ over $\{0,1\}^*$, $[p] \subset \{0,1\}^\infty$ is the set of infinite sequences $\beta \in [p]$ such that $U_x(\beta)$ outputs the bit representation of $p(x)$. The algorithm $U$ is a standard universal Turing machine. $\mathbf{I}(x:y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x,y)$ is the mutual information between two strings. For infinite sequences $\alpha, \beta \in \{0,1\}^\infty$, $\mathbf{I}(\alpha:\beta) = \log \sum_{x,y\in\{0,1\}^*} \mathbf{m}(x|\alpha)\mathbf{m}(y|\beta)2^{\mathbf{I}(x:y)}$ [Lev74].

**Theorem.** *If* $\mathbf{w}$ *is a semimeasure on* $\{0,1\}^*$ *and* $\mathbf{m} < O(1)\mathbf{w}$ *then* $\mathbf{I}(\mathbf{w}:\mathcal{H}) = \infty$.

## 10.2 Kolmogorov Complexity is Exotic

We cover material on busy beaver functions. Let $\Omega = \sum\{2^{-\|p\|} : U(p) \text{ halts}\}$ be Chaitin's Omega, $\Omega_n \in \mathbb{Q}_{\geq 0}$ be be the rational formed from the first $n$ bits of $\Omega$, and $\Omega^t = \sum\{2^{-\|p\|} : U(p) \text{ halts in time } t\}$. For $n \in \mathbb{N}$, let $\mathbf{bb}(n) = \min\{t : \Omega_n < \Omega^t\}$. $\mathbf{bb}^{-1}(m) = \arg\min_n\{\mathbf{bb}(n-1) < m \leq \mathbf{bb}(n)\}$. Let $\Omega[n] \in \{0,1\}^*$ be the first $n$ bits of $\Omega$. For $t \in \mathbb{N}$ define the function $\mathbf{m}^t(x) = \sum\{2^{-\|p\|} : U(p) = x \text{ in } t \text{ steps}\}$ and for $n \in \mathbb{N}$, we have $\mathbf{m}_n(x) = \sum\{2^{-\|p\|} : U(p) = x \text{ in } \mathbf{bb}(n) \text{ steps}\}$.

**Lemma 7** *For* $n = \mathbf{bb}^{-1}(m)$, $\mathbf{K}(\Omega[n]|m,n) = O(1)$.

**Proof.** For a string $x$, let $BB(x) = \inf\{t : \Omega^t > 0.x\}$. Enumerate strings of length $n$, starting with $0^n$, and return the first string $x$ such that $BB(x) \geq m$. This string $x$ is equal to $\Omega[n]$, otherwise let $y$ be the largest common prefix of $x$ and $\Omega[n]$. Thus $BB(y) = \mathbf{bb}(\|y\|) \geq BB(x) \geq m$, which means $\mathbf{bb}^{-1}(m) \leq \|y\| < n$, causing a contradiction. □

**Lemma 8** *A relation* $X = \{(x_i, c_i)\}_{i=1}^{2^n} \subset \{0,1\}^* \times \mathbb{N}$, $|\mathbf{K}(x_i) - c_i| \leq s$, *has* $n <^{\log} 2s + 2\mathbf{I}(X;\mathcal{H})$.

**Proof.** We relativize the universal Turing machine to $(n,s)$, which can be done due to the precision of the theorem. Let $T = \min\{t : \lceil -\log \mathbf{m}_t(x_i)\rceil - c_i < s+1\}$. Let $N = \mathbf{bb}^{-1}(T)$ and $M = \mathbf{bb}(N)$. So for all $x_i$, $-\log \mathbf{m}_M(x_i) - \mathbf{K}(x_i) <^+ 2s$. Let $Q$ be an elementary probability measure that realizes $\mathbf{Ks}(X)$ and $d = \max\{\mathbf{d}(X|Q), 1\}$. Without loss of generality, the support of $Q$ is restricted to binary relations $B \subset \{0,1\}^* \times \mathbb{N}$ of size $2^n$. Let $B_1 = \bigcup\{y : (y,c) \in B\}$. Let $S = \bigcup\{B_1 : B \in \text{Support}(Q)\}$.

We randomly select each string in $S$ to be in a set $R$ independently with probability $d2^{-n}$. Thus $\mathbf{E}[\mathbf{m}_M(R)] \leq d2^{-n}$. For $B \in \text{Support}(Q)$,

$$\mathbf{E}_R\mathbf{E}_{B\sim Q}[[R \cap B_1 = \emptyset]] = \mathbf{E}_{B\sim Q}\Pr(R \cap B_1 = \emptyset) = (1 - d2^{-n})^{2^n} < e^{-d}.$$

Thus there exists a set $R \subseteq S$ such that $\mathbf{m}_M(R) \leq 2{\cdot}2^{-n}$ and $\mathbf{E}_{B\sim Q}[[R \cap B_1 = \emptyset]] < 2e^{-d}$. Let $t(B) = .5[R \cap B_1 = \emptyset]2^d$. $t$ is a $Q$-test, with $\mathbf{E}_{B\sim Q}[t(B)] \leq 1$. It must be that $t(X) \neq 0$, otherwise,

$$1.44d - 1 < \log t(X) <^+ \mathbf{d}(X|Q) + \mathbf{K}(t|Q) <^+ d + \mathbf{K}(d),$$

which is a contradiction for large enough $d$, which one can assume without loss of generality. Thus $t(X) \neq 0$ and $R \cap X_1 \neq \emptyset$. Furthermore, if $y \in R$, $\mathbf{K}(y) <^+ -\log \mathbf{m}_M(x) - n + \log d + \mathbf{K}(d, M, R)$. So for $x \in R \cap X_1$, .

$$\begin{aligned}
\mathbf{K}(x) &<^+ -\log \mathbf{m}_M(x) - n + \log d + \mathbf{K}(d, M, R)\\
\mathbf{K}(x) &<^+ \mathbf{K}(x) + 2s - n + \log d + \mathbf{K}(M) + \mathbf{K}(R, d)\\
n &<^+ 2s + \mathbf{K}(M) + \log d + \mathbf{K}(Q, d)\\
n &<^+ 2s + \mathbf{K}(\Omega[N]) + \mathbf{Ks}(X)\\
n &<^+ 2s + \mathbf{K}(\Omega[N]) + \mathbf{I}(X; \mathcal{H}) \qquad (10.1)
\end{aligned}$$

From Lemma 7, $\mathbf{K}(\Omega[N]|T, N) =^+ \mathbf{K}(\Omega[N]|X, N) = O(1)$. Furthermore it is well known for the bits of Chaitin's Omega, $N <^+ \mathbf{K}(\Omega[N])$ and $\mathbf{K}(\Omega[N]|\mathcal{H}) <^+ \mathbf{K}(N)$. So, using Lemma 2,

$$N <^+ \mathbf{K}(\Omega[N]) <^{\log} \mathbf{I}(\Omega[N]; \mathcal{H}) <^{\log} \mathbf{I}(X; \mathcal{H}) + \mathbf{K}(N) <^{\log} \mathbf{K}(X; \mathcal{H}). \qquad (10.2)$$

So combining Equations 10.1 and 10.2, one gets

$$n <^{\log} 2s + 2\mathbf{I}(X; \mathcal{H}).$$

□

## 10.3 Results

**Theorem 19** *If $\mathbf{w}$ is a semimeasure on $\{0,1\}^*$ and $\mathbf{m} \overset{*}{<} \mathbf{w}$ then $\mathbf{I}(\mathbf{w} : \mathcal{H}) = \infty$.*

**Proof.** Note that $\mathbf{w}$ has full support since $\mathbf{m}$ does. One can also assume that for each $x \in \{0,1\}^*$, $-\log \mathbf{w}(x) \in \mathbb{N}$. Let $N \subseteq \mathbb{N}$ be a set of numbers $n$ such that $\mathbf{w}(\{0,1\}^n) < 1/n$. Obviously $|N| = \infty$. Fix $n \in N$. We have $X_n = \{x : \mathbf{w}(x) < 2^{-n-\log n + O(1)}\}$. Some simple math shows that $|X_n| \overset{*}{>} 2^n$. So for each $x \in X_n$, $\mathbf{K}(x) >^+ -\log \mathbf{w}(x) >^+ n + \log n$. We also have that for each $x \in \{0,1\}^n$, $\mathbf{K}(x) <^+ n + \mathbf{K}(n)$. Let $Y_n = \{(x, n + \log n) : x \in X_n\}$. So for each $(x, c) \in Y_n$, $|\mathbf{K}(x) - c| <^+ \log n$. So applying Lemma 8 to $Y_n$, we get $n <^{\log} \mathbf{I}(Y_n; \mathcal{H}) <^{\log} \mathbf{I}(\mathbf{w} : \mathcal{H}) + \mathbf{K}(n) <^{\log} \mathbf{I}(\mathbf{w} : \mathcal{H})$. Since $|N| = \infty$, $\mathbf{I}(\mathbf{w} : \mathcal{H}) = \infty$. □

# Bibliography


[BEHW89] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, 36(4):929–965, 1989.

[Cha75] G. J. Chaitin. A Theory of Program Size Formally Identical to Information Theory. *Journal of the ACM*, 22(3):329–340, 1975.

[EL11] Samuel Epstein and Leonid Levin. On sets of high complexity strings. *CoRR*, abs/1107.1458, 2011.

[Eps19] S. Epstein. On the algorithmic probability of sets. *CoRR*, abs/1907.04776, 2019.

[Eps21a] Samuel Epstein. All sampling methods produce outliers. *IEEE Transactions on Information Theory*, 67(11):7568–7578, 2021.

[Eps21b] Samuel Epstein. A note on the outliers theorem. *CoRR*, 2203.08733, 2021. v2.

[Eps21c] Samuel Epstein. On the conditional complexity of sets of strings. *CoRR*, 1907.01018, 2021.

[Eps23] Samuel Epstein. On the Existence of Anomalies. *CoRR*, abs/2302.05972, 2023.

[G2́1] Peter Gács. Lecture notes on descriptional complexity and randomness. *CoRR*, abs/2105.04704, 2021.

[Gei12] Philipp Geiger. *Mutual information and Gödel incompleteness.* PhD thesis, Heidelberg University, 10 2012.

[GTV01] P. Gács, J. Tromp, and P. Vitányi. Algorithmic Statistics. *IEEE Transactions on Information Theory*, 47(6):2443–2463, 2001.

[Kol65] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems in Information Transmission*, 1:1–7, 1965.

[KU87] A. N. Kolmogorov and V. A. Uspensky. Algorithms and Randomness. *SIAM Theory of Probability and Its Applications*, 32(3):389–412, 1987.

[Lev74] L. A. Levin. Laws of Information Conservation (Non-growth) and Aspects of the Foundations of Probability Theory. *Problemy Peredachi Informatsii*, 10(3):206–210, 1974.

[Lev84] L. A. Levin. Randomness conservation inequalities; information and independence in mathematical theories. *Information and Control*, 61(1):15–37, 1984.

[Lev13] L. A. Levin. Forbidden information. *J. ACM*, 60(2), 2013.

[Lev16] L. A. Levin. Occam bound on lowest complexity of elements. *Annals of Pure and Applied Logic*, 167(10):897–900, 2016. And also: S. Epstein and L.A. Levin, Sets have simple members, arXiv preprint arXiv:1107.1458, 2011.

[LV08] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer Publishing Company, Incorporated, 3 edition, 2008.

[Rom03] Andrei E. Romashchenko. Extracting the mutual information for a triple of binary strings. In *IEEE Conference on Computational Complexity*, pages 221–229. IEEE Computer Society, 2003.

[Rom22] A. Romashchenko. Clustering with respect to the information distance. *Theoretical Computer Science*, 2022.

[She83] A. Shen. The concept of (alpha,beta)-stochasticity in the Kolmogorov sense, and its properties. *Soviet Mathematics Doklady*, 28(1):295–299, 1983.

[She99] A. Shen. Discussion on Kolmogorov Complexity and Statistical Analysis. *The Computer Journal*, 42(4):340–342, 1999.

[She12] A. Shen. Game Arguments in Computability Theory and Algorithmic Information Theory. In *Proceedings of 8th Conference on Computability in Europe* , volume 7318 of *LNCS*, pages 655–666, 2012.

[Sol64] R. J. Solomonoff. A Formal Theory of Inductive Inference, Part l. *Information and Control*, 7:1–22, 1964.

[VS17] Nikolay K. Vereshchagin and Alexander Shen. Algorithmic statistics: Forty years later. In *Computability and Complexity*, pages 669–737, 2017.

[VV04] N. Vereshchagin and P. Vitányi. Kolmogorov's Structure Functions and Model Selection. *IEEE Transactions on Information Theory*, 50(12):3265 – 3290, 2004.

[V'Y87] V.V. V'Yugin. On Randomness Defect of a Finite Object Relative to Measures with Given Complexity Bounds. *SIAM Theory of Probability and Its Applications*, 32:558–563, 1987.

[V'Y99] V.V. V'Yugin. Algorithmic complexity and stochastic properties of finite binary sequences. *The Computer Journal*, 42:294–317, 1999.

[ZL70] A. K. Zvonkin and L. A. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Math. Surveys*, page 11, 1970.