The Bernoulli Cell Problem in Algorithmic Information Theory

Aman Sharma
amansh@iitk.ac.in

November 28, 2012

Abstract

Kolomogrov complexity has been an extensively researched topic in Algorithmic information theory.We study the relation between Kolomogrov Complexity and Shannon Entropy for bernoulli sequences by considering their bernoulli parameters n(length of string) and k(number of 1s in string).We consider a fixed point estimator(E) given by Vladmir Vovk and study a few results that provide us more insight into the problem.

1 Introduction

1.1 Algorithmic Information Theory

"Algorithmic information theory is the result of putting Shannon's information theory and Turings computability theory into a cocktail shaker and shaking vigorously."
—G. J. Chaitin

- AIT is a subfield of both information theory and computer science
- (Almost) simultaneously and independently developed by
  - ◇ 1960: introduced by R. J. Solomonoff as part of work on inductive inference
  - ◇ 1965: A. N. Kolmogorov
  - ◇ 1966: G. J. Chaitin (while an 18-year old undergraduate!)
- It (generally) deals with the methods/coding schemes used in describing strings

2 Kolomogrov Complexity

2.1 Definition

- The Kolmogorov complexity of a string x is the length of the smallest program that outputs x, relative to some model of computation. That is,

$$C_f(x) = \min_p \{|p| : f(p) = x\} \tag{1}$$

for some computer f

- Informally, C (x) measures the information content & degree of randomness of x.

2.2 Invariance Theorem

- There exists a universal description method ψ_0, such that: $C_{\psi 0} \leq C_{\psi} + c$ for some constant c that depends on ψ and ψ_0 (but not on x).
- (Proof Idea) Every Turing machine can be simulated on a Universal Turing Machine with a simulation code that is independent of x.

2.3 Conditional Complexity

- The conditional Kolmogorov complexity of a string x, relative to a string y and a model of computation f , is:
  - $C_f(x|y) = min\{|p| : C_f(p, y) = x\}$
  - $C_f(x) = C_f(x|\epsilon)$
- C$(x|y)$ is the size of the minimal program for x when started with input y
- C$(x : y\ )$ = C (x) C $(x|y)$ describes the information y contains about x
- When C $(x : y)$ = C (x), x and y are algorithmically independent

3 Some Results

3.1 Incompressibility

- A string x is incompressible if
  - $K(x) \geq |x|$
- Short programs encode patterns in non-random strings
- These strings are Algorithmially Random
- Algorithmic randomness is not identical to the intuitive concept of randomness

3.2 Incompressibilty Theorem

- For all n, there exists an incompressible string of length n
- There are 2^n strings of length n and fewer than 2n descriptions that are shorter than n
  $\sum_{i=0}^{n} 2^i = 2^{n-1} < 2^n$

3.3 Non-Computability Theorem

- *Kolmogrov Complexity K(x) of a string is uncomputable.*
- Proof
  - ◇ Consider the string, ' the smallest string x with $C(x) > n$'(Berry's Paradox)
  - ◇ If K(x) is computable,we can keep querying strings in ascending order until we find our string.
  - ◇ Now,above string has a description of the order $K(n) \leq log(n) < n$
  - ◇ Hence K(x) is uncomputable

4 Shannon Entropy

4.1 Definition

- Let X be a discrete random variable taking a finite number of possible values $x_1, x_2, ..., x_n$ with probabilities $p_1, p_2, ..., p_n$ respectively such that $p_i \geq 0, i = 1, 2, ..., n, \sum_{i=1}^{n} p(i)=1$

  $H_n(p_i, p_2....p_n)=\sum_{i=1}^{n} -p_i log(p_i)$
- The entropy of a random variable is a good measure of randomness or uncertainty.Greater randomness implies greater entropy and vice versa.

4.2 Relation with Kolomogrov Complexity

- Shannon Entropy is related to Kolmogrov complexity in the following way

  $$K(x) = H(A) \pm c$$

  (for most elements $x \in A$)
- More Formally
  - ◇ *If the entropy of a recursive probability distribution P is* $H(P) = -\sum P(x) log P(x)$

    $$0 \leq \sum_{x}^{n} (P(x)K(x) - H(P)) \leq c_p$$

    where c_P *is a constant that depends only on the length of the program necessary to compute the probability distribution P*

5 Bernoulli Cell Problem

5.1 Introduction

- Suppose we were to restrict our domain to a particular Bernoulli cell ie binary strings of length n with k 1's.Would the relation between Kolmogrov Complexity and Shannon Entopy still hold?
- In more formal terms,does the following equation hold?
  - $\diamond\ KP_\theta x = -log(P_\theta^n(x)) + K(\frac{\theta}{n})$
- No solution to the above question has been found so far

5.2 Previous Research

- This problem has been around for about 50 years,since Andrei Kolmogrov thought of it
- His student,Vladmir Vovk who currently in the University of London has been studying the problem extensively
- We survey a computable point estimator given by Vovk and a few lemmas associated to it.

5.3 Estimation

- Estimation theory is a branch of statistics and signal processing that deals with estimating the values of parameters based on measured/empirical data that has a random component.
- An estimator attempts to approximate the unknown parameters using the measurements
- For the problem,Vovk considers 'nets' of the form
  $\theta_n(a) = sin^2(\frac{a}{\sqrt{n}})a \in 1, ..., \lfloor \frac{\pi\sqrt{n}}{2} - 1 \rfloor - 1$
- Let x be a sequence of the length n and k be the total number of 1s in it.Then the estimator $E_n(k) = E(k)$ is defined as the element of the net closest to $\frac{k}{n}$

5.4 Lemma 1

- When $n \in N, \alpha \in [1/2, \pi n^{1/2}/2 - 1/2]$ and $a, b \in [0, \pi n^{1/2}/2]$ so that $a \leq \alpha \leq b$ and $1/2 \leq b - a \leq 2$,we have,

$$sin^2(\frac{b}{\sqrt{n}}) - sin^2(\frac{a}{\sqrt{n}}) = \frac{1}{\sqrt{n}} sin(\frac{\alpha}{\sqrt{n}}) cos(\frac{\alpha}{\sqrt{n}}) \tag{4}$$

Proof

- Equivalent transformations of the equation yield

$$(sin(\frac{b}{\sqrt{n}}) - sin(\frac{a}{\sqrt{n}}))(sin(\frac{b}{\sqrt{n}}) + sin(\frac{a}{\sqrt{n}})) = \frac{1}{\sqrt{n}} sin(\frac{\alpha}{\sqrt{n}}) cos(\frac{\alpha}{\sqrt{n}}) \quad (5)$$

$$(cos(\frac{b+a}{2\sqrt{n}}) sin(\frac{b+a}{2\sqrt{n}}) sin(\frac{b-a}{2\sqrt{n}}) cos(\frac{b-a}{2\sqrt{n}}) = \frac{1}{\sqrt{n}} sin(\frac{\alpha}{\sqrt{n}}) cos(\frac{\alpha}{\sqrt{n}}) \quad (6)$$

- The problem is reduced to proving that

$$cos(\frac{b+a}{2\sqrt{n}}) =^{\cdot} cos(\frac{\alpha}{\sqrt{n}}) \quad (7)$$

$$sin(\frac{b-a}{2\sqrt{n}}) =^{\cdot} \frac{1}{\sqrt{n}} \quad (8)$$

$$cos(\frac{b-a}{2\sqrt{n}}) =^{\cdot} 1 \quad (9)$$

$$sin(\frac{b+a}{2\sqrt{n}}) =^{\cdot} sin(\frac{\alpha}{\sqrt{n}}) \quad (10)$$

- Equalities 2 & 4 follow from the fact that $\frac{1}{2} \leq b - a \leq 2$

- Equalities 1 & 3 reduce from the view that $a \leq \alpha \leq b$ and $\alpha \in [1/2, \pi\sqrt{n}/2 - 1/2]$ to

$$cos(\frac{\pi}{2} - \frac{1}{2\sqrt{n}}) = cos(\frac{\pi}{2} - \frac{1}{4\sqrt{n}}) \quad (11)$$

$$sin(\frac{1}{2\sqrt{n}}) = sin(\frac{1}{4\sqrt{n}}) \quad (12)$$

- These two relations are equivalent and the second one is obviously true.

5.5 Log-Likelihood Function

- The likelihood of a set of parameter values given some observed outcomes is equal to the probability of those observed outcomes given those parameter values.
- For many applications involving likelihood functions, it is more convenient to work in terms of the natural logarithm of the likelihood function
- Because the logarithm is a monotonically increasing function, the logarithm of a function achieves its maximum value at the same points as the function itself.
- We use a log likelihood function $G_a(n, k)$ where

$$G_a(n, k) = ln(sin^{2k}(\frac{a}{\sqrt{(n)}}) cos^{2(n-k)}(\frac{a}{\sqrt{(n)}})) \quad (13)$$

- We use the notation $\hat{a}$ for the maximum likelihood estimate of the parameter a

$$\hat{a}(n, k) = \max_a \{G_a(n, k)\} \quad (14)$$

5.6 Lemma 2

When $n \geq 1, \alpha \in [1, \pi n^{1/2}/2 - 1]$, and $k \in 1, ..., n-1$ range so that $|a - \hat{a}(n,k)| < 1$

$$G_{n,k}(a) =^{+} G_{n,k}(\hat{a}(n,k)) \tag{15}$$

Proof
Denote $\hat{a} = \hat{a}(n,k)$. It suffices to prove that the values

$$\sup_a |\frac{d^2 G_{n,k}(a)}{da^2}| \tag{16}$$

where a ranges over $[1, \frac{\pi\sqrt{n}}{2} - 1] \cap [\hat{a}-1, \hat{a}+1]$ do not exceed some bound.Calculating the second derivative,we rewrite the above equation as

$$2 \sup_a \left(\frac{\frac{k}{n}}{sin^2 \frac{a}{\sqrt{n}}} + \frac{1 - \frac{k}{n}}{cos^2 \frac{a}{\sqrt{n}}} \right) \tag{17}$$

Note that

$$\frac{k}{n} = sin^2 \left(\frac{\hat{a}}{\sqrt{n}} \right) \tag{18}$$

$$1 - \frac{k}{n} = cos^2 \left(\frac{\hat{a}}{\sqrt{n}} \right) \tag{19}$$

, so it suffices to prove that,

$$\sup_a \left(\frac{sin^2(\frac{a+1}{\sqrt{n}})}{sin^2(\frac{a}{\sqrt{n}})} \right) and \tag{20}$$

$$\sup_a \left(\frac{cos^2(\frac{a-1}{\sqrt{n}})}{cos^2(\frac{a}{\sqrt{n}})} \right) \tag{21}$$

are bounded above by some constant.It is easy to see that both suprema are equal to,

$$\frac{sin^2(2n^{-1/2})}{sin^2(n^{-1/2})} \to 4 (n \to \infty) \tag{22}$$

5.7 Lemma 3

Let $n \geq 1$,$a \in [0, \pi n^{1/2}/2]$, and $k \in 1, ..., n-1$.For some constant $\epsilon > 0$,

$$G_{n,k}(\hat{a}(n,k)) - G_{n,k}(a) \geq^{+} \epsilon |a - \hat{a}(n,k)| \tag{23}$$

Proof
Denote $\hat{a} = \hat{a}(n,k)$ By the symmetry of the problem, we can suppose $a \geq \hat{a}$. Furthermore, we can consider only the case $a \geq \hat{a} + 1/2$. Since $G_k''(a)$ is negative everywhere,it is sufficient to prove that $G_k'(\hat{a} + 1/2)$ is greater than some constant $\epsilon > 0$. We find

$$-G_k'(a) = \frac{2}{\sqrt{n}} \left((n-k) \frac{sin(an^{-1/2})}{cos(an^{-1/2})} - k \frac{cos(an^{-1/2})}{sin(an^{-1/2})} \right) \tag{24}$$

So,we are required to prove

$$(n-k)sin^2\left((\hat{a}+1/2)n^{-1/2}\right)-kcos^2\left((\hat{a}+1/2)n^{-1/2}\right)> \\ \frac{\epsilon}{2}n^{1/2}sin^2\left((\hat{a}+1/2)n^{-1/2}\right)cos^2\left((\hat{a}+1/2)n^{-1/2}\right) \quad (25)$$

The last inequality immediately follows from Lemma 4.
The following inequalities

$$G_{n,k}(\hat{a})-G_{n,k}(a)\geq G_{n,k}(\hat{a}+1/2)-G_{n,k}(a)= \\ -G'(a')(a-\hat{a}-1/2)\geq\epsilon(a-\hat{a})-\frac{\epsilon}{2} \quad (26)$$

where $\hat{a}+1/2<a'<a$,complete the proof.

5.8 Lemma 4

Let $n\geq 1$,$a\in[0,\pi n^{1/2}/2]$, and $k\in 1,...,n-1$. Then,

$$KP(\lfloor\hat{a}(n,k)\rfloor|n,a)\leq^+ (ln^{-1}2)(G_{n,k}(\hat{a}(n,k))-G_{n,k}(a)) \quad (27)$$

Proof
By the previous lemma,

$$G_{n,k}(\hat{a}(n,k))-G_{n,k}(a)\geq^+\epsilon|a-\hat{a}(n,k)| \quad (28)$$

Assertion of the lemma follows from:

$$KP(\lfloor\hat{a}(n,k)\rfloor|n,a)\leq^+ KP(\lfloor\hat{a}(n,k)\rfloor|n,\lfloor a\rfloor) \quad (29)$$
$$KP(\lfloor\hat{a}(n,k)\rfloor|n,\lfloor a\rfloor)\leq^+ KP(\lfloor\hat{a}(n,k)\rfloor-\lfloor a\rfloor|n) \quad (30)$$
$$KP(\lfloor\hat{a}(n,k)\rfloor-\lfloor a\rfloor|n)\leq^+ 2log|\lfloor\hat{a}(n,k)\rfloor-\lfloor a\rfloor| \quad (31)$$
$$2log|\lfloor\hat{a}(n,k)\rfloor-\lfloor a\rfloor|\leq^+\epsilon(ln^{-1}2)|\lfloor\hat{a}(n,k)\rfloor-\lfloor a\rfloor| \quad (32)$$
$$\epsilon(ln^{-1}2)|\lfloor\hat{a}(n,k)\rfloor-\lfloor a\rfloor|=^+\epsilon(ln^{-1}2)|\hat{a}(n,k)-a| \quad (33)$$

6 References

- V.V. Vyugin: Algorithmic complexity and stochastic properties of finite binary sequences
- V.G. Vovk: On the concept of the Bernoulli property
- Lance Fortnow: Kolmogorov Complexity
- Cover & Thomas: Elements of Information Theory
- CS687 Course Notes: Course at IITK taught by Dr. Satyadev Nandakumar
  Link : http://www.cse.iitk.ac.in/users/satyadev/a10/a10.html

7 Acknowledgements

- I would like to thank my mentor Dr. Satyadev Nandakumar for mentoring me throughout this course and being extremely helpful throughout
- I would also like to thank the DUGC of the CSE Department,Dr. Ajai Jain for allowing me to opt for this course