Asymptotic Divergences and Strong Dichotomy

Xiang Huang [1] (joint work with Jack H. Lutz [1], Elvira Mayordomo [2], and Donald M. Stull [3])

March 12, 2019

[1]Computer Science Department, Iowa State University

[2]Departamento de Informática e Ingeniería de Sistemas, Instituto de Investigación en Ingeniería de Aragón, Universidad de Zaragoza

[3]Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

Overview

Schnorr-Stimm Dichotomy Theorem

Theorem (Schnorr and Stimm 72')

1. *If S is not normal, then there is a finite-state gambler that wins money at an infinitely-often exponential rate betting on S.*
2. *If S is normal, then any finite-state gambler betting on S loses money at an exponential rate betting on S.*

Schnorr-Stimm Dichotomy Theorem

Theorem (Schnorr and Stimm 72')

1. *If S is not normal, then there is a finite-state gambler that wins money at an infinitely-often exponential rate betting on S.*
2. *If S is normal, then any finite-state gambler betting on S loses money at an exponential rate betting on S.*

Question: What are these exponential rates? How to interpret them?

Strong Dichotomy Theorem

Modulo asymptotic caveats, We show that

1'. The infinitely-often exponential rate of winning in 1 is $2^{\mathrm{Div}(S||\alpha)|w|}$.

2'. The exponential rate of loss in 2 is $2^{-\mathrm{Risk}_G(w)}$.

Strong Dichotomy Theorem

Modulo asymptotic caveats, We show that

1'. The infinitely-often exponential rate of winning in 1 is $2^{\mathrm{Div}(S||\alpha)|w|}$.

2'. The exponential rate of loss in 2 is $2^{-\mathrm{Risk}_G(w)}$.

This is a **quantify** version of Schnorr-Stimm's dichotomy theorem.

Outline

1. Information theory: entropy and divergence
2. Asymptotic divergences
3. Finite-state gambler and strong dichotomy
4. Finite-state dimension

Information Theory: a Short Introduction

Self-information

A long time ago, in a galaxy far, far away...

Self-information

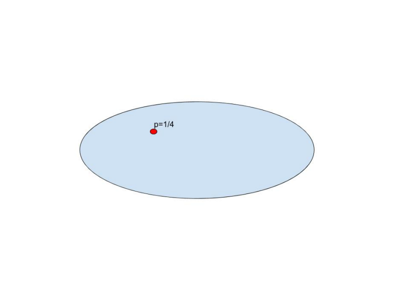
A long time ago, in a galaxy far, far away...

There is a point.

Self-information

A long time ago, in a galaxy far, far away...

There is a point.



Self-information

A long time ago, in a galaxy far, far away...

There is a point.

So he therefore believes we need to have $\log \frac{1}{1/4} = 2$ bits to encode the objects in his galaxy. And that is only **his** opinion.

We call $\log \frac{1}{p}$ this point's *self-information*.

Entropy

A *(discrete) probability measure* on a nonempty finite set Ω is a function $\pi : \Omega \to [0, 1]$ satisfying

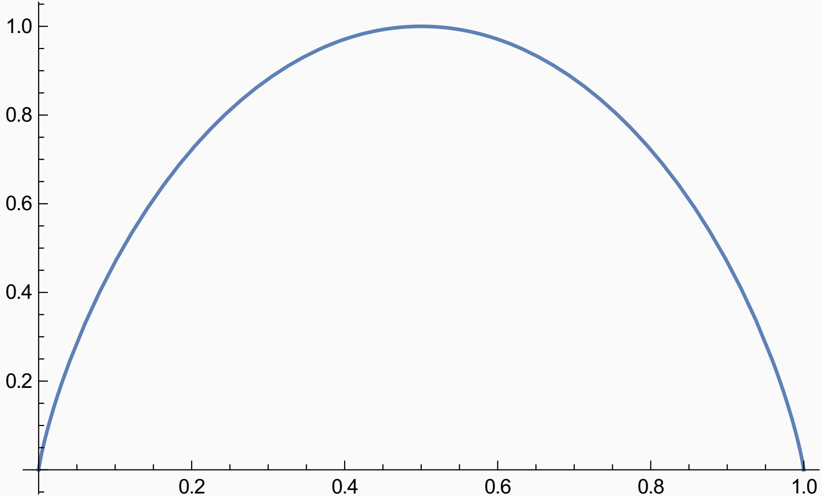
$$\sum_{\omega \in \Omega} \pi(\omega) = 1. \tag{1}$$

We write $\Delta(\Omega)$ for the set of all probability measures on Ω. Then the Shannon entropy of π is define as:

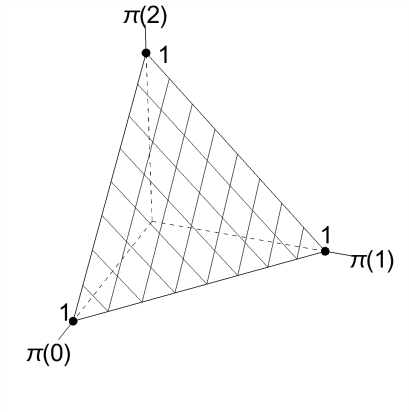
$$H(\pi) = \sum_{\omega \in \Omega} \pi(\omega) \log \frac{1}{\pi(\omega)} \tag{2}$$

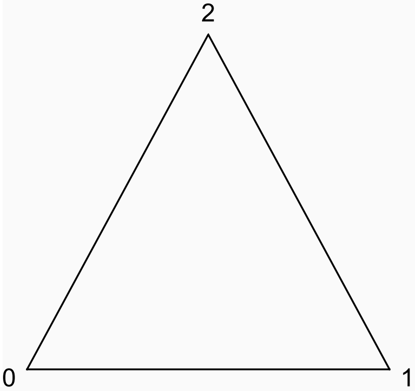
That is, a (weighted) average of the self-information of all points.

Entropy: 2D-plots



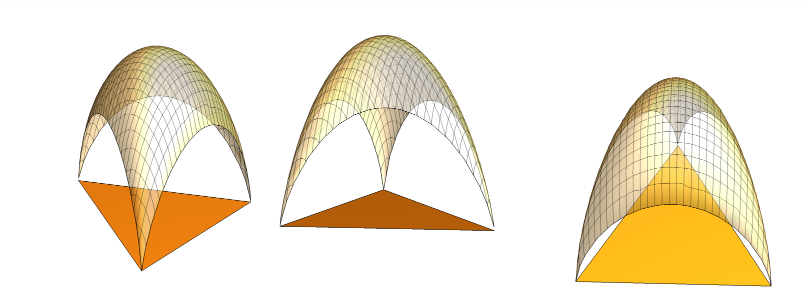
Entropy: 3D Probability Measures





Two views of the simplex $\Delta(\{0,1,2\})$. The simplex is part of the plane $x+y+z=1$.

Entropy: 3D-entropy



3D-entropy plotted on the previous equilateral triangle.

Kullback-Leibler Divergence

Definition
Let $\alpha, \beta \in \Delta(\Omega)$, where Ω is a nonempty finite set. The *Kullback-Leibler divergence* (or *KL-divergence*) *of* β *from* α is

$$D(\alpha||\beta) = \sum_{\omega\in\Omega} \alpha(\omega) \log \frac{\alpha(\omega)}{\beta(\omega)} = E_\alpha \log \frac{\alpha}{\beta}, \tag{3}$$

where the logarithm is base-2.

Kullback-Leibler Divergence

Definition
Let $\alpha, \beta \in \Delta(\Omega)$, where Ω is a nonempty finite set. The *Kullback-Leibler divergence* (or *KL-divergence*) *of* β *from* α is

$$D(\alpha||\beta) = \sum_{\omega \in \Omega} \alpha(\omega) \log \frac{\alpha(\omega)}{\beta(\omega)} = E_\alpha \log \frac{\alpha}{\beta}, \tag{3}$$

where the logarithm is base-2.

Note that we can view
$D(\alpha||\beta) = \sum_{\omega \in \Omega} \alpha(\omega) \log \frac{1}{\beta(\omega)} - \sum_{\omega \in \Omega} \alpha(\omega) \log \frac{1}{\alpha(\omega)}$

Kullback-Leibler Divergence

Definition
Let $\alpha, \beta \in \Delta(\Omega)$, where Ω is a nonempty finite set. The *Kullback-Leibler divergence* (or *KL-divergence*) *of* β *from* α is

$$D(\alpha||\beta) = \sum_{\omega \in \Omega} \alpha(\omega) \log \frac{\alpha(\omega)}{\beta(\omega)} = E_\alpha \log \frac{\alpha}{\beta}, \tag{3}$$

where the logarithm is base-2.

Note that we can view
$D(\alpha||\beta) = \sum_{\omega \in \Omega} \alpha(\omega) \log \frac{1}{\beta(\omega)} - \sum_{\omega \in \Omega} \alpha(\omega) \log \frac{1}{\alpha(\omega)}$

- $\beta(\omega)$ can be view as "subjective" probability, or a coding scheme.

Kullback-Leibler Divergence

Definition
Let $\alpha, \beta \in \Delta(\Omega)$, where Ω is a nonempty finite set. The *Kullback-Leibler divergence* (or *KL-divergence*) *of* β *from* α is

$$D(\alpha||\beta) = \sum_{\omega \in \Omega} \alpha(\omega) \log \frac{\alpha(\omega)}{\beta(\omega)} = E_\alpha \log \frac{\alpha}{\beta}, \tag{3}$$

where the logarithm is base-2.

Note that we can view
$D(\alpha||\beta) = \sum_{\omega \in \Omega} \alpha(\omega) \log \frac{1}{\beta(\omega)} - \sum_{\omega \in \Omega} \alpha(\omega) \log \frac{1}{\alpha(\omega)}$

- $\beta(\omega)$ can be view as "subjective" probability, or a coding scheme.
- $\alpha(\omega)$ can be view as the "real" probability.

Kullback-Leibler Divergence

Definition
Let $\alpha, \beta \in \Delta(\Omega)$, where Ω is a nonempty finite set. The *Kullback-Leibler divergence* (or *KL-divergence*) *of* β *from* α is

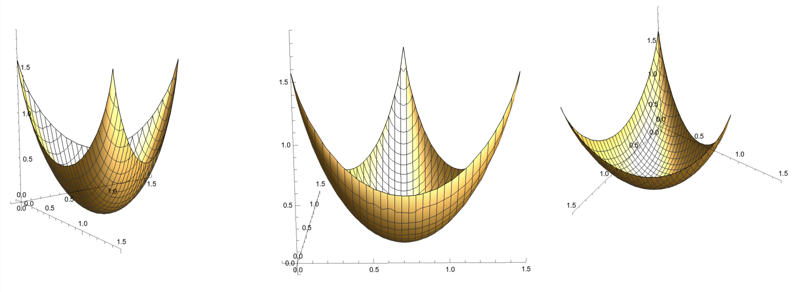
$$D(\alpha||\beta) = \sum_{\omega \in \Omega} \alpha(\omega) \log \frac{\alpha(\omega)}{\beta(\omega)} = E_\alpha \log \frac{\alpha}{\beta}, \tag{3}$$

where the logarithm is base-2.

Note that we can view
$D(\alpha||\beta) = \sum_{\omega\in\Omega} \alpha(\omega) \log \frac{1}{\beta(\omega)} - \sum_{\omega\in\Omega} \alpha(\omega) \log \frac{1}{\alpha(\omega)}$

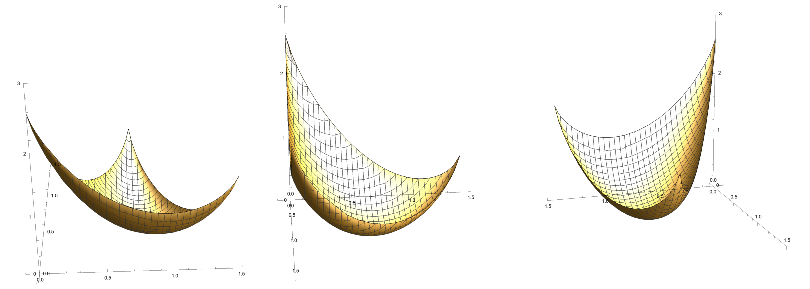
- $\beta(\omega)$ can be view as "subjective" probability, or a coding scheme.
- $\alpha(\omega)$ can be view as the "real" probability.
- $D(\alpha||\beta)$ measures the difference between the average code-length of a coding scheme and an optimal coding. So it is always ≥ 0.

Divergence: $D(\alpha||\beta)$, where $\beta = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$



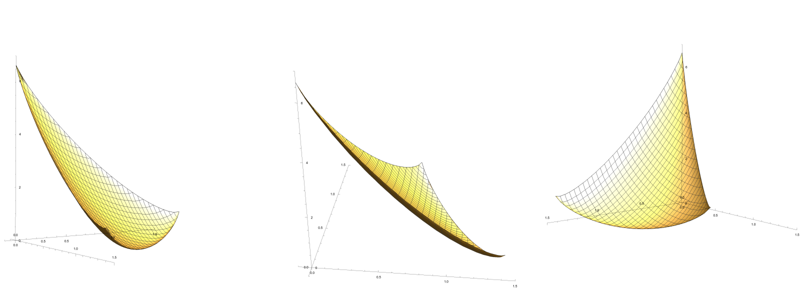
Three views of the function $D(\alpha||\beta)$, where $\beta = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. It is the upside-down copy of the 3D-entropy function.

Divergence: $D(\alpha||\beta)$, where $\beta = (\frac{1}{6}, \frac{1}{3}, \frac{1}{2})$



Three views of the function $D(\alpha||\beta)$, where $\beta = (\frac{1}{6}, \frac{1}{3}, \frac{1}{2})$.

Divergence: $D(\alpha||\beta)$, where $\beta = (\frac{1}{100}, \frac{1}{2}, \frac{49}{100})$



Three views of the function $D(\alpha||\beta)$, where $\beta = (\frac{1}{100}, \frac{1}{2}, \frac{49}{100})$.

We have three peaks in the previous plots. Which one is the highest one? And how talk it is?

Proposition (radius of β[1])
The highest peak of $D(\alpha||\beta)$ (with β fixed) equals to
$c = \log\left(1/\min_{\omega\in\Omega}\beta(\omega)\right)$.

[1]By several COM S 633 students, 17'Fall.

Divergence: the Highest Peak

We have three peaks in the previous plots. Which one is the highest one? And how talk it is?

Proposition (radius of β[1])
The highest peak of $D(\alpha||\beta)$ (with β fixed) equals to
$c = \log\left(1/\min_{\omega\in\Omega}\beta(\omega)\right)$.

Remark
The number c will show up later in this talk again.

[1] By several COM S 633 students, 17'Fall.

Asymptotic entropies and divergences

Block Counting

We work in a finite alphabet Σ with $2 \leq |\Sigma| < \infty$. For the purpose of this talk, one can treat $\Sigma = \{0, 1\}$.

Block Counting

We work in a finite alphabet Σ with $2 \leq |\Sigma| < \infty$. For the purpose of this talk, one can treat $\Sigma = \{0, 1\}$.

We write Σ^ℓ for the set of strings of length ℓ over Σ, $\Sigma^* = \bigcup_{\ell=0}^{\infty} \Sigma^\ell$ for the set of (finite) *strings* over Σ, Σ^ω for the set of (infinite) *sequences* over Σ, and $\Sigma^{\leq\omega} = \Sigma^* \cup \Sigma^\omega$.

Block Counting

We work in a finite alphabet Σ with $2 \leq |\Sigma| < \infty$. For the purpose of this talk, one can treat $\Sigma = \{0, 1\}$.

We write Σ^ℓ for the set of strings of length ℓ over Σ, $\Sigma^* = \bigcup_{\ell=0}^{\infty} \Sigma^\ell$ for the set of (finite) *strings* over Σ, Σ^ω for the set of (infinite) *sequences* over Σ, and $\Sigma^{\leq\omega} = \Sigma^* \cup \Sigma^\omega$.

For nonempty strings $w, x \in \Sigma^*$, we write

$$\#_\Box(w, x) = \left| \left\{ m \leq \frac{|x|}{|w|} - 1 \;\middle|\; x[m|w|..(m+1)|w| - 1] = w \right\} \right|$$

for the *number of block occurrences* of w in x. Note that $0 \leq \#_\Box(w, x) \leq \frac{|x|}{|w|}$.

Block Frequency

For each $S \in \Sigma^\omega$, $n \in \mathbb{Z}^+$, and $\lambda \neq w \in \Sigma^*$, the n-th *block frequency* of w in S is

$$\pi_{S,n}(w) = \frac{\#_\square(w, S[0..n|w|-1])}{n} \tag{4}$$

Block Frequency

For each $S \in \Sigma^\omega$, $n \in \mathbb{Z}^+$, and $\lambda \neq w \in \Sigma^*$, the n-th *block frequency* of w in S is

$$\pi_{S,n}(w) = \frac{\#_\Box(w, S[0..n|w| - 1])}{n} \tag{4}$$

Remark
Note that the above w can be of any length. We now fixed a length, and consider the set of string Σ^ℓ. We denote $\pi^{(\ell)}_{S,n} = \pi_{S,n} \upharpoonright \Sigma^\ell$ be the restriction of the function $\pi_{S,n}$ to the set Σ^ℓ of strings of length ℓ.

Block Frequency

For each $S \in \Sigma^\omega$, $n \in \mathbb{Z}^+$, and $\lambda \neq w \in \Sigma^*$, the n-th *block frequency* of w in S is

$$\pi_{S,n}(w) = \frac{\#_\square(w, S[0..n|w|-1])}{n} \tag{4}$$

Remark

Note that the above w can be of any length. We now fixed a length, and consider the set of string Σ^ℓ. We denote $\pi^{(\ell)}_{S,n} = \pi_{S,n} \restriction \Sigma^\ell$ be the restriction of the function $\pi_{S,n}$ to the set Σ^ℓ of strings of length ℓ.

Remark

$\pi^{(\ell)}_{S,n}$ is a rational-valued probability measure on Σ^ℓ.

Block Frequency

For each $S \in \Sigma^\omega$, $n \in \mathbb{Z}^+$, and $\lambda \neq w \in \Sigma^*$, the n-th *block frequency* of w in S is

$$\pi_{S,n}(w) = \frac{\#_\Box(w, S[0..n|w| - 1])}{n} \tag{4}$$

Remark

Note that the above w can be of any length. We now fixed a length, and consider the set of string Σ^ℓ. We denote $\pi^{(\ell)}_{S,n} = \pi_{S,n} \upharpoonright \Sigma^\ell$ be the restriction of the function $\pi_{S,n}$ to the set Σ^ℓ of strings of length ℓ.

Remark

$\pi^{(\ell)}_{S,n}$ is a rational-valued probability measure on Σ^ℓ.

Remark

Intuitively, this is like we are using Σ^ℓ as our alphabet instead. And $\pi^{(\ell)}_{S,n}$ becomes the frequency of each "symbol" of the new alphabet.

Asymptotic divergences

A probability measure $\alpha \in \Delta(\Sigma)$ naturally induces, for each $\ell \in \mathbb{Z}^+$, a probability measure $\alpha^{(\ell)} \in \Delta(\Sigma^\ell)$ defined by

$$\alpha^{(\ell)}(w) = \prod_{i=0}^{|w|-1} \alpha(w[i]). \tag{5}$$

Asymptotic divergences

A probability measure $\alpha \in \Delta(\Sigma)$ naturally induces, for each $\ell \in \mathbb{Z}^+$, a probability measure $\alpha^{(\ell)} \in \Delta(\Sigma^\ell)$ defined by

$$\alpha^{(\ell)}(w) = \prod_{i=0}^{|w|-1} \alpha(w[i]). \tag{5}$$

Definition
Let $\ell \in \mathbb{Z}^+$, $S \in \Sigma^\omega$, and $\alpha \in \Delta(\Sigma)$.

1. The *lower ℓ-divergence of α from S* is $\mathrm{div}_\ell(S||\alpha) = \liminf_{n\to\infty} D(\pi^{(\ell)}_{S,n}||\alpha^{(\ell)})$.

Asymptotic divergences

A probability measure $\alpha \in \Delta(\Sigma)$ naturally induces, for each $\ell \in \mathbb{Z}^+$, a probability measure $\alpha^{(\ell)} \in \Delta(\Sigma^\ell)$ defined by

$$\alpha^{(\ell)}(w) = \prod_{i=0}^{|w|-1} \alpha(w[i]). \tag{5}$$

Definition
Let $\ell \in \mathbb{Z}^+$, $S \in \Sigma^\omega$, and $\alpha \in \Delta(\Sigma)$.

1. The *lower ℓ-divergence of α from S* is $\mathrm{div}_\ell(S||\alpha) = \liminf_{n\to\infty} D(\pi_{S,n}^{(\ell)}||\alpha^{(\ell)})$.
2. The *upper ℓ-divergence of α from S* is $\mathrm{Div}_\ell(S||\alpha) = \limsup_{n\to\infty} D(\pi_{S,n}^{(\ell)}||\alpha^{(\ell)})$.

Asymptotic divergences

A probability measure $\alpha \in \Delta(\Sigma)$ naturally induces, for each $\ell \in \mathbb{Z}^+$, a probability measure $\alpha^{(\ell)} \in \Delta(\Sigma^\ell)$ defined by

$$\alpha^{(\ell)}(w) = \prod_{i=0}^{|w|-1} \alpha(w[i]). \tag{5}$$

Definition
Let $\ell \in \mathbb{Z}^+$, $S \in \Sigma^\omega$, and $\alpha \in \Delta(\Sigma)$.

1. The *lower ℓ-divergence of α from S* is $\mathrm{div}_\ell(S||\alpha) = \liminf_{n\to\infty} D(\pi^{(\ell)}_{S,n}||\alpha^{(\ell)})$.
2. The *upper ℓ-divergence of α from S* is $\mathrm{Div}_\ell(S||\alpha) = \limsup_{n\to\infty} D(\pi^{(\ell)}_{S,n}||\alpha^{(\ell)})$.
3. The *lower divergence of α from S* is $\mathrm{div}(S||\alpha) = \sup_{\ell\in\mathbb{Z}^+} \mathrm{div}_\ell(S||\alpha)/\ell$.

Asymptotic divergences

A probability measure $\alpha \in \Delta(\Sigma)$ naturally induces, for each $\ell \in \mathbb{Z}^+$, a probability measure $\alpha^{(\ell)} \in \Delta(\Sigma^\ell)$ defined by

$$\alpha^{(\ell)}(w) = \prod_{i=0}^{|w|-1} \alpha(w[i]). \tag{5}$$

Definition
Let $\ell \in \mathbb{Z}^+$, $S \in \Sigma^\omega$, and $\alpha \in \Delta(\Sigma)$.

1. The *lower ℓ-divergence of α from S* is $\mathrm{div}_\ell(S||\alpha) = \liminf_{n\to\infty} D(\pi^{(\ell)}_{S,n}||\alpha^{(\ell)})$.
2. The *upper ℓ-divergence of α from S* is $\mathrm{Div}_\ell(S||\alpha) = \limsup_{n\to\infty} D(\pi^{(\ell)}_{S,n}||\alpha^{(\ell)})$.
3. The *lower divergence of α from S* is $\mathrm{div}(S||\alpha) = \sup_{\ell\in\mathbb{Z}^+} \mathrm{div}_\ell(S||\alpha)/\ell$.
4. The *upper divergence of α from S* is $\mathrm{Div}(S||\alpha) = \sup_{\ell\in\mathbb{Z}^+} \mathrm{Div}_\ell(S||\alpha)/\ell$.

Definition
Let $\ell \in \mathbb{Z}^+$ and $S,\ T \in \Sigma^\omega$.

1. The *lower ℓ-divergence of T from S* is $\mathrm{div}_\ell(S||T) = \liminf_{n\to\infty} D(\pi^{(\ell)}_{S,n}||\,\pi^{(\ell)}_{T,n})$.
2. The *upper ℓ-divergence of T from S* is $\mathrm{Div}_\ell(S||T) = \limsup_{n\to\infty} D(\pi^{(\ell)}_{S,n}||\,\pi^{(\ell)}_{T,n})$.
3. The *lower divergence of T from S* is $\mathrm{div}(S||T) = \sup_{\ell\in\mathbb{Z}^+} \mathrm{div}_\ell(S||T)/\ell$.
4. The *upper divergence of T from S* is $\mathrm{Div}(S||T) = \sup_{\ell\in\mathbb{Z}^+} \mathrm{Div}_\ell(S||T)/\ell$.

Normality

The following notions are essentially due to Borel.

Definition
Let $\alpha \in \Delta(\Sigma)$, $S \in \Sigma^\omega$, and $\ell \in \mathbb{Z}^+$.

1. S is α-ℓ-*normal* if, for all $w \in \Sigma^\ell$,

$$\lim_{n\to\infty} \pi_{S,n}(w) = \alpha^{(\ell)}(w).$$

2. S is α-*normal* if, for all $\ell \in \mathbb{Z}^+$, S is α-ℓ-normal.

Characterization of Normality

Theorem (divergence characterization of normality)

For all $\alpha \in \Delta(\Sigma)$ and $S \in \Sigma^\omega$, the following conditions are equivalent.

(1) *S is α-normal.*

(2) $\mathrm{Div}(S||\alpha) = 0$.

(3) *For every α-normal sequence $T \in \Sigma^\omega$,* $\mathrm{Div}(S||T) = 0$.

(4) *There exists an α-normal sequence $T \in \Sigma^\omega$ such that* $\mathrm{Div}(S||T) = 0$.

Finite-state gambler and strong dichotomy

Finite-state Gambler

Definition
A *finite-state gambler (FSG)* is a 4-tuple

$$G = (Q, \delta, s, B),$$

where Q is a finite set of *states*, $\delta : Q \times \Sigma \to Q$ is the *transition function*, $s \in Q$ is the *start state*, and $B : Q \to \Delta_{\mathbb{Q}}(\Sigma)$ is the *betting function*.

Finite-state Gambler

Definition
A *finite-state gambler (FSG)* is a 4-tuple

$$G = (Q, \delta, s, B),$$

where Q is a finite set of *states*, $\delta : Q \times \Sigma \to Q$ is the *transition function*, $s \in Q$ is the *start state*, and $B : Q \to \Delta_{\mathbb{Q}}(\Sigma)$ is the *betting function*.

The payoffs in the betting are determined by a *payoff probability measure* $\alpha \in \Delta(\Sigma)$.

Finite-state Gambler

Definition
A *finite-state gambler (FSG)* is a 4-tuple

$$G = (Q, \delta, s, B),$$

where Q is a finite set of *states*, $\delta : Q \times \Sigma \to Q$ is the *transition function*, $s \in Q$ is the *start state*, and $B : Q \to \Delta_{\mathbb{Q}}(\Sigma)$ is the *betting function*.

The payoffs in the betting are determined by a *payoff probability measure* $\alpha \in \Delta(\Sigma)$.

We write $d_{G,\alpha}(w)$ for the gambler G's capital (amount of money) after betting on the successive bits of a prefix $w \sqsubseteq S$, and we assume that the initial capital is $d_{G,\alpha}(\lambda) = 1$.

Finite-state Gambler

Definition
A *finite-state gambler (FSG)* is a 4-tuple

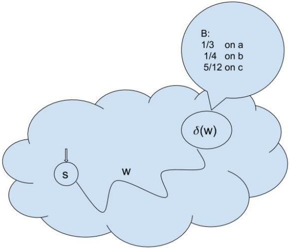
$$G = (Q, \delta, s, B),$$

where Q is a finite set of *states*, $\delta : Q \times \Sigma \to Q$ is the *transition function*, $s \in Q$ is the *start state*, and $B : Q \to \Delta_{\mathbb{Q}}(\Sigma)$ is the *betting function*.

The payoffs in the betting are determined by a *payoff probability measure* $\alpha \in \Delta(\Sigma)$.

We write $d_{G,\alpha}(w)$ for the gambler G's capital (amount of money) after betting on the successive bits of a prefix $w \sqsubseteq S$, and we assume that the initial capital is $d_{G,\alpha}(\lambda) = 1$.

Finite-state Gambler



Let our capital after seeing $w \sqsubseteq S$ be $d_{G,\alpha}(w)$. The gambler's bets $B(\delta(w))(a)$ of the money that the next symbol of S is an a. If it then turns out to be the case, the gambler's capital will be

$$d_{G,\alpha}(wa) = d_{G,\alpha}(w)\frac{B\big(\delta(w)\big)(a)}{\alpha(a)}. \tag{6}$$

α-martingale

The payoffs in (6) are fair with respect to α, which means that the conditional α-expectation

$$\sum_{a \in \Sigma} \alpha(a) d_{G,\alpha}(wa)$$

of $d_{G,\alpha}(wa)$, given that $w \sqsubseteq S$, is exactly $d_{G,\alpha}(w)$. This says that the function $d_{G,\alpha}$ is an α-*martingale*.

Strong Dichotomy: Winning

Theorem
Let $\alpha \in \Delta(\Sigma)$, $S \in \Sigma^\omega$, and $\gamma < 1$. If S is not α-normal, then there is a finite-state gambler G such that, for infinitely many prefixes $w \sqsubseteq S$,

$$d_{G,\alpha}(w) > 2^{\gamma\,\mathrm{Div}(S||\alpha)|w|}.$$

Proof.
(Sketch)

Step 1 : Let $r < 1$.By the definition of $\mathrm{Div}(S||\alpha)$ there must exist ℓ such that

$$\mathrm{Div}_\ell(S||\alpha) > r\,\mathrm{Div}(S||\alpha) > 0. \tag{7}$$

Remark
This step get us fix an ℓ.

Step 2 : Fix some (rational) π_0 such that

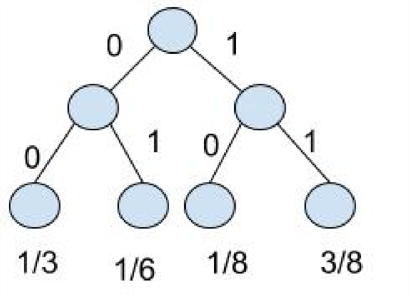
$$D(\pi_0||\alpha^{(\ell)}) \approx \ell\,\mathrm{Div}_\ell(S||\alpha) > 0. \tag{8}$$

Winning: cont.

Remark
The advantage of π_0 is that it is a rational probability. It can be used to implement a finite-state gambler. And $D(\pi_0||\alpha^{(\ell)})/\ell$ is closed to $\mathrm{Div}_\ell(S||\alpha)$ *and hence* $\mathrm{Div}(S||\alpha)$*. The closer, the better.*

Step 3: Use π_0 as a gambling strategy. Note that π_0 is probability measure on Σ^ℓ. But our finite-state gambler need to gambler on Σ.

Play the conditional probability trick to fill this gap.



Winning: cont.

Step 4: the resulting gambler can be viewed as gamble on Σ^ℓ. Since a string w can be view as (assuming $|w|$ is a multiple of ℓ)

$$w = u_0 u_1 \cdots u_{n-1}, \quad \text{where } |u_i| = \ell \text{ for } 0 \le i \le n-1.$$

Then

$$d_{G,\alpha}(w) = \prod_{0}^{n-1} \frac{\pi_0(u_i)}{\alpha(u_i)}. \tag{9}$$

Winning: cont.

Step 4: the resulting gambler can be viewed as gamble on Σ^ℓ. Since a string w can be view as (assuming $|w|$ is a multiple of ℓ)

$$w = u_0 u_1 \cdots u_{n-1}, \quad \text{where } |u_i| = \ell \text{ for } 0 \leq i \leq n-1.$$

Then

$$d_{G,\alpha}(w) = \prod_{0}^{n-1} \frac{\pi_0(u_i)}{\alpha(u_i)}. \tag{9}$$

$$\begin{aligned}
\log d_{G,\alpha}(w) &= \sum_{i=0}^{n-1} \log \frac{\pi_0(u_i)}{\alpha^{(\ell)}(u_i)} = \sum_{|u|=\ell} \#_{\Box}(u,w) \log \frac{\pi_0(u)}{\alpha^{(\ell)}(u)}, \\
&= n \sum_{|u|=\ell} \frac{\#_{\Box}(u,w)}{n} \log \frac{\pi_0(u)}{\alpha^{(\ell)}(u)} = n \sum_{|u|=\ell} \pi_{S,n}^{(\ell)}(u) \log \frac{\pi_0(u)}{\alpha^{(\ell)}(u)}, \\
&= n \sum_{|u|=\ell} \left[\pi_{S,n}^{(\ell)}(u) \log \frac{\pi_{S,n}^{(\ell)}(u)}{\alpha^{(\ell)}(u)} - \pi_{S,n}^{(\ell)}(u) \log \frac{\pi_{S,n}^{(\ell)}(u)}{\pi_0(u)} \right] \\
&= n \left(D(\pi_{S,n}^{(\ell)} || \alpha^{(\ell)}) - D(\pi_{S,n}^{(\ell)} || \pi_0) \right) \ldots > \gamma Divg(S||\alpha)|w|
\end{aligned}$$

Strong Dichotomy: Losing

We define the *risk* that G takes in a state $q \in Q$ to be

$$\mathrm{risk}_G(q) = D\big(\alpha || B(q)\big).$$

i.e., the divergence of $B(q)$ from not betting. We also define the *total risk* that G takes *along* a string $w \in \Sigma^*$ to be

$$\mathrm{Risk}_G(w) = \sum_{u \sqsubset w} \mathrm{risk}_G\left(\delta(u)\right).$$

Strong Dichotomy: Losing

We define the *risk* that G takes in a state $q \in Q$ to be

$$\mathrm{risk}_G(q) = D\big(\alpha || B(q)\big).$$

i.e., the divergence of $B(q)$ from not betting. We also define the *total risk* that G takes *along* a string $w \in \Sigma^*$ to be

$$\mathrm{Risk}_G(w) = \sum_{u \sqsubset w} \mathrm{risk}_G\left(\delta(u)\right).$$

Theorem (strong dichotomy theorem)
Let $\alpha \in \Delta(\Sigma)$, $S \in \Sigma^\omega$, and $\gamma < 1$.

If S is α-normal, then, for every finite-state gambler G, for all but finitely many prefixes $w \sqsubseteq S$,

$$d_{G,\alpha}(w) < 2^{-\gamma \,\mathrm{Risk}_G(w)}.$$

Remark

1. *By Agafonov theorem, If S is α-normal, then it behaves quite random. In the sense that no matter in what state q, the probability q sees an a is $\alpha(a)$.*

Remark

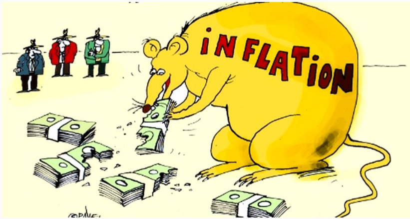
1. *By Agafonov theorem, If S is α-normal, then it behaves quite random. In the sense that no matter in what state q, the probability q sees an a is $\alpha(a)$.*
2. *A gambling strategy $B(q)$ that is not the same with α is consider taking a risk of* $\mathrm{risk}_G(q)$.

Remark

1. *By Agafonov theorem, If S is α-normal, then it behaves quite random. In the sense that no matter in what state q, the probability q sees an a is $\alpha(a)$.*
2. *A gambling strategy $B(q)$ that is not the same with α is consider taking a risk of $\mathrm{risk}_G(q)$.*
3. *The rate of losing the the risk a gambler accumulates along the way of gambling S.*

Finite-state dimension

Hard times: the Age of Inflation



Finite-state dimension

For each $\alpha \in \Delta(\Sigma)$ and each $S \in \Sigma^\omega$, define the sets

$$\mathfrak{G}^\alpha(S) = \left\{ s \in [0,\infty) \;\middle|\; (\exists \text{FSG } G) \limsup_{w \to S} \alpha^{|w|}(w)^{1-s} d_{G,\alpha}(w) = \infty \right\}$$

Finite-state dimension

For each $\alpha \in \Delta(\Sigma)$ and each $S \in \Sigma^\omega$, define the sets

$$\mathfrak{G}^\alpha(S) = \left\{ s \in [0,\infty) \,\middle|\, (\exists \text{FSG } G) \limsup_{w \to S} \alpha^{|w|}(w)^{1-s} d_{G,\alpha}(w) = \infty \right\}$$

and

$$\mathfrak{G}^\alpha_{\text{str}}(S) = \left\{ s \in [0,\infty) \,\middle|\, (\exists \text{FSG } G) \liminf_{w \to S} \alpha^{|w|}(w)^{1-s} d_{G,\alpha}(w) = \infty \right\}$$

Definition (Lutz 11')
Let $\alpha \in \Delta(\Sigma)$ and $S \in \Sigma^\omega$.

1. The *finite-state α-dimension* of S is $\dim^\alpha_{\text{FS}}(S) = \inf \mathfrak{G}^\alpha(S)$.
2. The *finite-state strong α-dimension* of S is $\text{Dim}^\alpha_{\text{FS}}(S) = \inf \mathfrak{G}^\alpha_{\text{str}}(S)$

Asymptotic Entropy

Definition
Let $S \in \Sigma^\omega$. We define the following asymptotic entropy.

1. $H_+^{(\ell)}(S) = \frac{1}{\ell} \limsup_{n\to\infty} H(\pi_{S,n}^{(\ell)})$.
2. $H^+(S) = \sup_\ell H_+^{(\ell)}$.
3. $H_-^{(\ell)}(S) = \frac{1}{\ell} \liminf_{n\to\infty} H(\pi_{S,n}^{(\ell)})$.
4. $H^-(S) = \sup_\ell H_-^{(\ell)}$.

A Bound for Finite-state Dimension

Theorem
For all $\alpha \in \Delta(\Sigma)$ *and* $S \in \Sigma^\omega$. *Then,*

$$\dim_{\mathrm{FS}}^{\alpha}(S) \leq \frac{H^{+}(S)}{H^{+}(S) + \mathrm{Div}(S||\alpha)}$$

and

$$\mathrm{Dim}_{\mathrm{FS}}^{\alpha}(S) \leq \frac{H^{-}(S)}{H^{-}(S) + \mathrm{div}(S||\alpha)}.$$

A Bound for Finite-state Dimension

Theorem
For all $\alpha \in \Delta(\Sigma)$ *and* $S \in \Sigma^\omega$. *Then,*

$$\dim_{\mathrm{FS}}^{\alpha}(S) \leq \frac{H^{+}(S)}{H^{+}(S) + \mathrm{Div}(S||\alpha)}$$

and

$$\mathrm{Dim}_{\mathrm{FS}}^{\alpha}(S) \leq \frac{H^{-}(S)}{H^{-}(S) + \mathrm{div}(S||\alpha)}.$$

Proof:

- Unpack the $\alpha^{|w|}(w)^{1-s}$, express it with asymptotic divergence and entropy.
- Apply (the winning side of) strong dichotomy theorem to the $d_{G,\alpha}(w)$ part.

Proof Sketch Cont.

$$
\begin{aligned}
\alpha^{|w|}(w)^{1-s} d_{G,\alpha}(w) &= d_{G,\alpha}(w) 2^{(1-s)\log \alpha^{|w|}(w)} \\
&= d_{G,\alpha}(w) 2^{(1-s)\frac{|w|}{\ell}\sum_{|u|=\ell} \frac{\#_{\square}(u,w)}{|w|/\ell}\log \alpha(u)} \\
&= d_{G,\alpha}(w) 2^{(1-s)\frac{|w|}{\ell}\sum_{|u|=\ell} \pi^{(\ell)}_{S,n}(u)\log \alpha(u)} \\
&= d_{G,\alpha}(w) 2^{-(1-s)\frac{|w|}{\ell}\sum_{|u|=\ell} \pi^{(\ell)}_{S,n}(u)\left(\log \frac{\pi^{(\ell)}_{S,n}(u)}{\alpha(u)} - \log \pi^{(\ell)}_{S,n}(u)\right)} \\
&= d_{G,\alpha}(w) 2^{-(1-s)\frac{|w|}{\ell}\left(D(\pi^{(\ell)}_{S,n}||\alpha^{(\ell)}) + H(\pi^{(\ell)}_{S,n})\right)} \\
&> 2^{\gamma\,\mathrm{Div}(S||\alpha)|w| - (1-s)\frac{|w|}{\ell}\left(D(\pi^{(\ell)}_{S,n}||\alpha^{(\ell)}) + H(\pi^{(\ell)}_{S,n})\right)}
\end{aligned}
$$

The last inequality is the competition of “winning rate” and “inflation rate”. We just want the total rate be positive infinitely often.

Divergence Theorem for Normality and Finite-state Dimensions

Theorem (Lutz[2] 08')
If α and β are positive probability measure on Σ, then, for every β-normal sequence $S \in \Sigma^\omega$,

$$\dim_{\mathrm{FS}}^{\alpha}(S) = \mathrm{Dim}_{\mathrm{FS}}^{\alpha}(S) = \frac{H(\beta)}{H(\beta) + D(\beta||\alpha)}$$

[2] A Divergence Formula for Randomness and Dimension

Divergence Theorem for Normality and Finite-state Dimensions

Theorem (Lutz[2] 08')
If α and β are positive probability measure on Σ, then, for every β-normal sequence $S \in \Sigma^\omega$,

$$\dim_{\mathrm{FS}}^{\alpha}(S) = \mathrm{Dim}_{\mathrm{FS}}^{\alpha}(S) = \frac{H(\beta)}{H(\beta) + D(\beta||\alpha)}$$

Note that it is obvious that if S is β-normal, then

$$\frac{H(\beta)}{H(\beta) + D(\beta||\alpha)} = \frac{H^{+}(S)}{H^{+}(S) + \mathrm{Div}(S||\alpha)}$$

[2] A Divergence Formula for Randomness and Dimension

Divergence Theorem for Normality and Finite-state Dimensions

Theorem (Lutz[2] 08')
If α and β are positive probability measure on Σ, then, for every β-normal sequence $S \in \Sigma^\omega$,

$$\dim_{\mathrm{FS}}^{\alpha}(S) = \mathrm{Dim}_{\mathrm{FS}}^{\alpha}(S) = \frac{H(\beta)}{H(\beta) + D(\beta||\alpha)}$$

Note that it is obvious that if S is β-normal, then

$$\frac{H(\beta)}{H(\beta) + D(\beta||\alpha)} = \frac{H^{+}(S)}{H^{+}(S) + \mathrm{Div}(S||\alpha)}$$

That is, equality in the previous theorem holds if S is β-dimension sequence, for some $\beta \in \Delta(\Sigma)$. Hence the bound is tight.

[2] A Divergence Formula for Randomness and Dimension

Another Bound: the return of c

As a special case of the previous bound, we have

Theorem
For all $\alpha \in \Delta(\Sigma)$ *and* $S \in \Sigma^\omega$ *let* $c = \log(1/\min_{a\in\Sigma} \alpha(a))$. *Then,*

$$\dim_{\mathrm{FS}}^{\alpha}(S) \leq 1 - \mathrm{Div}(S||\alpha)/c$$

and

$$\mathrm{Dim}_{\mathrm{FS}}^{\alpha}(S) \leq 1 - \mathrm{div}(S||\alpha)/c.$$

Thank you for your time!