Monotone Conditional Complexity Bounds on Future Prediction Errors

Alexey Chernov   Marcus Hutter

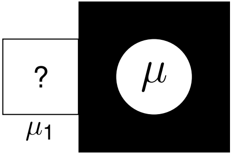
Istituto Dalle Molle di Studio sull'Intelligenza Artificiale
Lugano, Switzerland

Algorithmic Learning Theory 2005
Singapore

Supported by SNF grants 200020-107590/1 (to Jürgen Schmidhuber), 2100-67712 and 200020-107616

Outline

Sequence

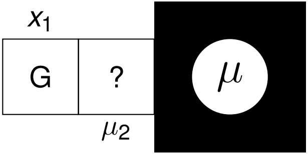


finite alphabet $\mathcal{X} \ni x_1, x_2, \ldots$

μ is a measure

$$\mu_{i+1}(\cdot) = \mu(\cdot | x_1 \ldots x_i) = \frac{\mu(x_1 \ldots x_i \cdot)}{\mu(x_1 \ldots x_i)}$$

Sequence

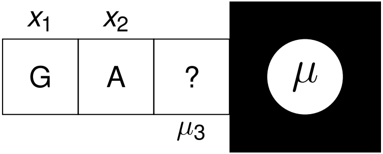


finite alphabet $\mathcal{X} \ni x_1, x_2, \ldots$

μ is a measure

$$\mu_{i+1}(\cdot) = \mu(\cdot | x_1 \ldots x_i) = \frac{\mu(x_1 \ldots x_i \cdot)}{\mu(x_1 \ldots x_i)}$$

Sequence

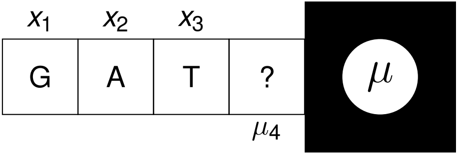


finite alphabet $\mathcal{X} \ni x_1, x_2, \ldots$

μ is a measure

$$\mu_{i+1}(\cdot) = \mu(\cdot | x_1 \ldots x_i) = \frac{\mu(x_1 \ldots x_i \cdot)}{\mu(x_1 \ldots x_i)}$$

Sequence

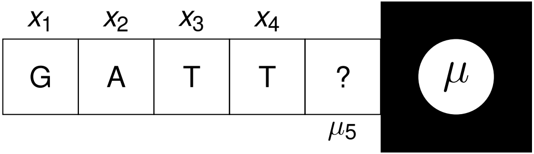


finite alphabet $\mathcal{X} \ni x_1, x_2, \ldots$

μ is a measure

$$\mu_{i+1}(\cdot) = \mu(\cdot | x_1 \ldots x_i) = \frac{\mu(x_1 \ldots x_i \cdot)}{\mu(x_1 \ldots x_i)}$$

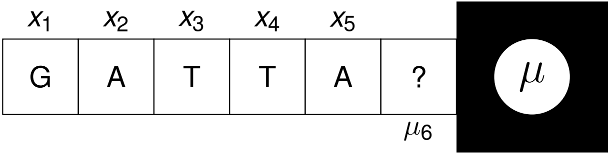
Sequence



finite alphabet $\mathcal{X} \ni x_1, x_2, \ldots$ μ is a measure

$$\mu_{i+1}(\cdot) = \mu(\cdot | x_1 \ldots x_i) = \frac{\mu(x_1 \ldots x_i \cdot)}{\mu(x_1 \ldots x_i)}$$

Sequence

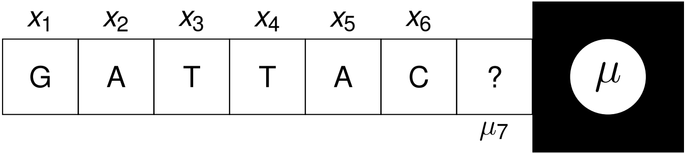


finite alphabet $\mathcal{X} \ni x_1, x_2, \ldots$

μ is a measure

$$\mu_{i+1}(\cdot) = \mu(\cdot|x_1 \ldots x_i) = \frac{\mu(x_1 \ldots x_i \cdot)}{\mu(x_1 \ldots x_i)}$$

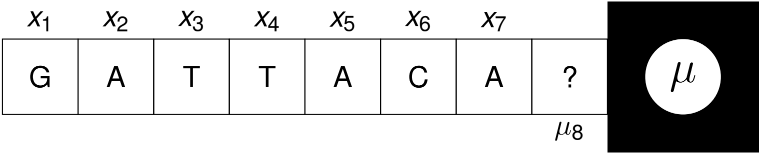
Sequence



finite alphabet $\mathcal{X} \ni x_1, x_2, \ldots$ μ is a measure

$$\mu_{i+1}(\cdot) = \mu(\cdot|x_1 \ldots x_i) = \frac{\mu(x_1 \ldots x_i \cdot)}{\mu(x_1 \ldots x_i)}$$

Sequence

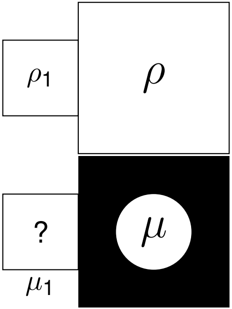


finite alphabet $\mathcal{X} \ni x_1, x_2, \ldots$　　μ is a measure

$$\mu_{i+1}(\cdot) = \mu(\cdot | x_1 \ldots x_i) = \frac{\mu(x_1 \ldots x_i \cdot)}{\mu(x_1 \ldots x_i)}$$

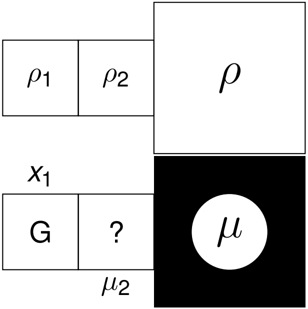
Predictor

Predictor ρ: $x_1, \ldots, x_i \quad \mapsto \quad \rho_{i+1}(\cdot) \approx \mu_{i+1}(\cdot)$



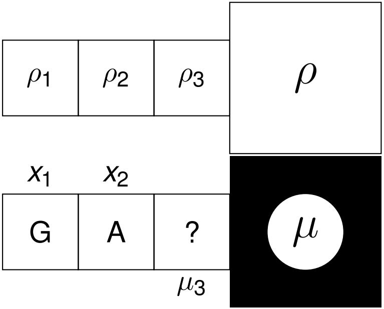
Predictor

Predictor ρ: $\quad x_1, \dots, x_i \quad \mapsto \quad \rho_{i+1}(\cdot) \approx \mu_{i+1}(\cdot)$



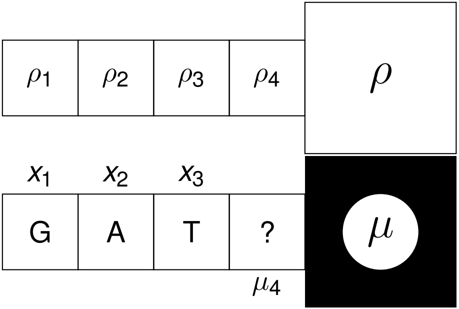
Predictor

Predictor ρ: $x_1, \ldots, x_i \quad \mapsto \quad \rho_{i+1}(\cdot) \approx \mu_{i+1}(\cdot)$



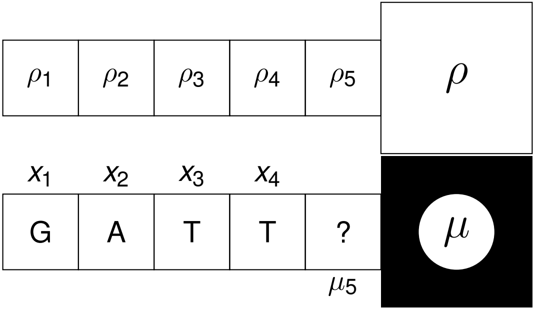
Predictor

Predictor ρ: $x_1, \ldots, x_i \quad \mapsto \quad \rho_{i+1}(\cdot) \approx \mu_{i+1}(\cdot)$



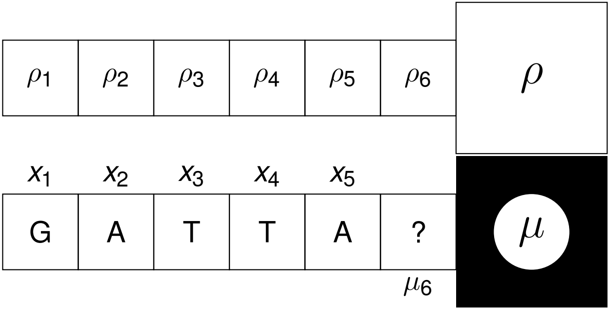
Predictor

Predictor ρ: $\quad x_1, \ldots, x_i \quad \mapsto \quad \rho_{i+1}(\cdot) \approx \mu_{i+1}(\cdot)$



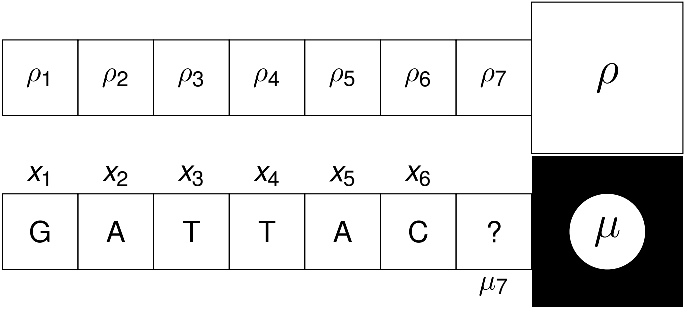
Predictor

Predictor ρ: $x_1, \ldots, x_i \quad \mapsto \quad \rho_{i+1}(\cdot) \approx \mu_{i+1}(\cdot)$



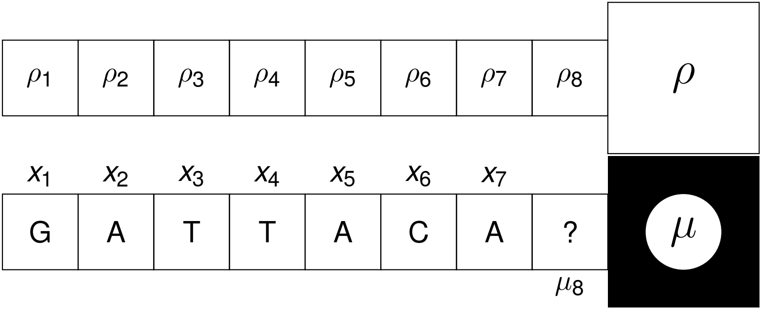
Predictor

Predictor ρ: $\quad x_1, \ldots, x_i \quad \mapsto \quad \rho_{i+1}(\cdot) \approx \mu_{i+1}(\cdot)$



Predictor

Predictor ρ: $\quad x_1, \ldots, x_i \quad \mapsto \quad \rho_{i+1}(\cdot) \approx \mu_{i+1}(\cdot)$



Quality of prediction

$\rho_i(\cdot) \approx \mu_i(\cdot)$ with high μ-probability:

$$\mathrm{Dist}(\rho, \mu) = \mathbf{E} \sum_{i=1}^{\infty} \mathrm{dist}_{x_1 \ldots x_i}(\rho_i, \mu_i) \quad = \sum_{x_1 x_2 \ldots} \mu(x_1 x_2 \ldots) \sum_{i=1}^{\infty} \mathrm{dist}_{x_1 \ldots x_i}(\rho_i, \mu_i)$$

$$\mathrm{dist}_{x_1 \ldots x_i}(\rho_i, \mu_i) = \frac{1}{\ln 2} \times$$

$$\sum_{a \in \mathcal{X}} (\rho_i(a) - \mu_i(a))^2 \quad \text{or} \quad \frac{1}{2} \left(\sum_{a \in \mathcal{X}} |\rho_i(a) - \mu_i(a)| \right)^2 \quad \text{or}$$

$$\sum_{a \in \mathcal{X}} \left(\sqrt{\rho_i(a)} - \sqrt{\mu_i(a)} \right)^2 \quad \text{or} \quad \sum_{a \in \mathcal{X}} \mu_i(a) \ln \frac{\mu_i(a)}{\rho_i(a)}$$

$$0 \leq \mathrm{Dist}(\rho, \mu) \leq D_\rho := \mathbf{E} \log_2 \frac{\mu(x_1 x_2 \ldots)}{\rho(x_1 x_2 \ldots)}$$

Intuitively: (for a deterministic μ)

$D_\rho \sim$ number of prediction errors

Solomonoff prior

$$\rho_i(\cdot) = \frac{M(x_1 \ldots x_i \cdot)}{M(x_1 \ldots x_i)} \qquad M(x) = \sum_{\mu} w_\mu \mu(x)$$

M is a Bayes mixture of all semi-computable semi-measures.

Theorem (Solomonoff 1964, 1978)

For any computable measure μ

$$\mathrm{Dist}(M, \mu) \leq D_M \overset{+}{\leq} K(\mu)$$

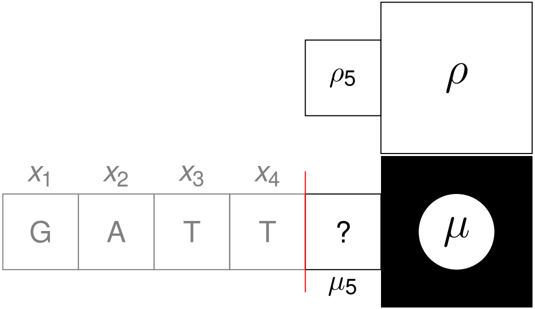
$K(\mu)$ is Kolmogorov complexity of μ

$\sim$ quantity of information in μ

$\sim$ the size of the shortest description of μ

Outline

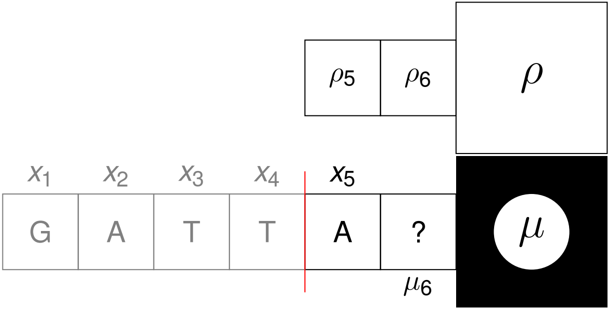
Prediction with a priori information



$x_1, \ldots, x_n$ are fixed

$$\mathrm{Dist}(\rho, \mu | x_1 \ldots x_n) = \underset{x_{n+1} x_{n+2} \ldots}{\mathbf{E}} \sum_{i=n+1}^{\infty} \mathrm{dist}_{x_1 \ldots x_i}(\rho_i, \mu_i)$$

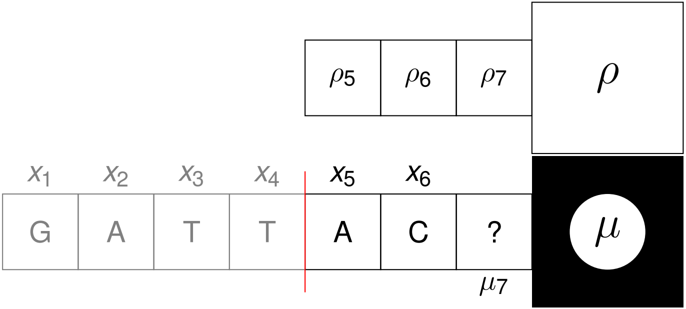
Prediction with a priori information



$x_1, \ldots, x_n$ are fixed

$$\mathrm{Dist}(\rho, \mu | x_1 \ldots x_n) = \mathop{\mathbf{E}}_{x_{n+1} x_{n+2} \ldots} \sum_{i=n+1}^{\infty} \mathrm{dist}_{x_1 \ldots x_i}(\rho_i, \mu_i)$$

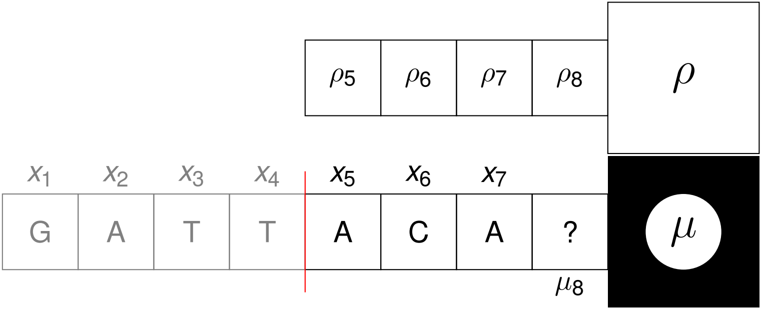
Prediction with a priori information



$x_1, \ldots, x_n$ are fixed

$$\mathrm{Dist}(\rho, \mu | x_1 \ldots x_n) = \mathop{\mathbf{E}}_{x_{n+1} x_{n+2} \ldots} \sum_{i=n+1}^{\infty} \mathrm{dist}_{x_1 \ldots x_i}(\rho_i, \mu_i)$$

Prediction with a priori information



$x_1, \ldots, x_n$ are fixed

$$\mathrm{Dist}(\rho, \mu | x_1 \ldots x_n) = \mathop{\mathbf{E}}_{x_{n+1} x_{n+2} \ldots} \sum_{i=n+1}^{\infty} \mathrm{dist}_{x_1 \ldots x_i}(\rho_i, \mu_i)$$

The problem

$$x = x_1 \dots x_n \qquad \mathrm{Dist}(M, \mu | x) \leq D_M(x) := \mathbf{E}_y \log_2 \frac{\mu(y_1 y_2 \dots | x)}{M(y_1 y_2 \dots | x)}$$

For any computable measure μ, for any word x

$$\frac{\mu(y|x)}{M(y|x)} \leq \quad ?$$

We know

$$\log_2 \frac{\mu(y)}{M(y)} \stackrel{+}{\leq} K(\mu)$$

We want

$$\log_2 \frac{\mu(y|x)}{M(y|x)} \stackrel{+}{\leq} K(\mu|x)$$

If x contains a lot of information about μ ($K(\mu|x)$ is small), prediction is easy.

The problem

$x = x_1 \ldots x_n \qquad \mathrm{Dist}(M, \mu|x) \leq D_M(x) := \mathbf{E}_y \log_2 \frac{\mu(y_1 y_2 \ldots|x)}{M(y_1 y_2 \ldots|x)}$

For any computable measure μ, for any word x

$$\frac{\mu(y|x)}{M(y|x)} \quad \leq \quad ?$$

We know

$$\log_2 \frac{\mu(y)}{M(y)} \overset{+}{\leq} K(\mu)$$

We want

$$\log_2 \frac{\mu(y|x)}{M(y|x)} \overset{+}{\leq} K(\mu|x)$$

If x contains a lot of information about μ ($K(\mu|x)$ is small), prediction is easy.

Prefix Kolmogorov complexity

Definition

$x, y \in \mathcal{X}^*$, U is a universal prefix machine, $p \in \{0,1\}^*$

$$K(y|x) = \min\{\ell(p) | U(p*, x) = y\}$$

Prefix machine

U gets finite x and infinite sequence α,
reads a finite part p of α, and halts with output y

Universal machine

for any other machine V: there is constant C

$$V(q,x) = y \quad \Rightarrow \quad \ell(q) \leq \ell(p) + C, \qquad \text{where } U(p,x) = y$$

Prefix Kolmogorov complexity

Definition

$x, y \in \mathcal{X}^*$, U is a universal prefix machine, $p \in \{0,1\}^*$

$$K(y|x) = \min\{\ell(p) | U(p*, x) = y\}$$

Prefix machine

U gets finite x and infinite sequence α,
reads a finite part p of α, and halts with output y

Universal machine

for any other machine V: there is constant C

$$V(q, x) = y \quad \Rightarrow \quad \ell(q) \leq \ell(p) + C, \qquad \text{where } U(p, x) = y$$

Prefix Kolmogorov complexity

Definition

$x, y \in \mathcal{X}^*$, U is a universal prefix machine, $p \in \{0,1\}^*$

$$K(y|x) = \min\{\ell(p) | U(p*, x) = y\}$$

Prefix machine

U gets finite x and infinite sequence α,
reads a finite part p of α, and halts with output y

Universal machine

for any other machine V: there is constant C

$$V(q,x) = y \quad \Rightarrow \quad \ell(q) \leq \ell(p) + C, \qquad \text{where } U(p,x) = y$$

$K(\mu|x)$ bound

Theorem

For any computable measure μ and any $x, y \in \mathcal{X}^*$

$$\log_2 \frac{\mu(y|x)}{M(y|x)} \overset{+}{\leq} K(\mu|x) + K(\ell(x))$$

Corollary

1. $\mathrm{Dist}(M, \mu|x_1 \dots x_n) \overset{+}{\leq} K(\mu|x_1 \dots x_n) + K(n)$
2. $\mathrm{Dist}(M, \mu) \overset{+}{\leq} \min_n \{\mathbf{E}_{\ell(x)=n} K(\mu|x) + K(n) + \frac{2}{\ln 2} n\}$

$K(\mu|x)$ bound

Theorem

For any computable measure μ and any $x, y \in \mathcal{X}^*$

$$\log_2 \frac{\mu(y|x)}{M(y|x)} \stackrel{+}{\leq} K(\mu|x) + K(\ell(x))$$

Corollary

1. $\mathrm{Dist}(M, \mu|x_1 \ldots x_n) \stackrel{+}{\leq} K(\mu|x_1 \ldots x_n) + K(n)$
2. $\mathrm{Dist}(M, \mu) \stackrel{+}{\leq} \min_n \{\mathbf{E}_{\ell(x)=n} K(\mu|x) + K(n) + \frac{2}{\ln 2} n\}$

Example

$$\mathrm{Dist}(M,\mu) \overset{+}{\leq} \min_n \{ \mathbf{E}_{\ell(x)=n} K(\mu|x) + K(n) + \frac{2}{\ln 2} n \}$$

x_1

?

"number of errors" $\sim \mathrm{Dist}(M,\mu)$

Solomonoff bound:

$$\mathrm{Dist}(M,\mu) \lesssim K(\mu) \sim \text{"size of the image"} \approx 10^5$$

New bound:

$$\mathrm{Dist}(M,\mu) \lesssim K(\mu|x_1) + K(1) + \frac{2}{\ln 2} \sim \text{"small constant"}$$

Example

$$\mathrm{Dist}(M,\mu) \overset{+}{\leq} \min_n \{\mathbf{E}_{\ell(x)=n} K(\mu|x) + K(n) + \tfrac{2}{\ln 2} n\}$$

x_1 x_2

?

“number of errors” $\sim \mathrm{Dist}(M,\mu)$

Solomonoff bound:

$$\mathrm{Dist}(M,\mu) \lesssim K(\mu) \sim \text{“size of the image”} \approx 10^5$$

New bound:

$$\mathrm{Dist}(M,\mu) \lesssim K(\mu|x_1) + K(1) + \tfrac{2}{\ln 2} \sim \text{“small constant”}$$

Example

$$\mathrm{Dist}(M,\mu) \overset{+}{\leq} \min_n \{ \mathbf{E}_{\ell(x)=n} K(\mu|x) + K(n) + \frac{2}{\ln 2} n \}$$

x_1

x_2

x_3

?

"number of errors" $\sim \mathrm{Dist}(M,\mu)$

Solomonoff bound:

$$\mathrm{Dist}(M,\mu) \lesssim K(\mu) \sim \text{"size of the image"} \approx 10^5$$

New bound:

$$\mathrm{Dist}(M,\mu) \lesssim K(\mu|x_1) + K(1) + \frac{2}{\ln 2} \sim \text{"small constant"}$$

Example

$$\mathrm{Dist}(M,\mu) \overset{+}{\leq} \min_n \{\mathbf{E}_{\ell(x)=n} K(\mu|x) + K(n) + \frac{2}{\ln 2} n\}$$

x_1 x_2 x_3 x_4




?

"number of errors" $\sim \mathrm{Dist}(M,\mu)$

Solomonoff bound:

$$\mathrm{Dist}(M,\mu) \lesssim K(\mu) \sim \text{"size of the image"} \approx 10^5$$

New bound:

$$\mathrm{Dist}(M,\mu) \lesssim K(\mu|x_1) + K(1) + \frac{2}{\ln 2} \sim \text{"small constant"}$$

Example

$$\mathrm{Dist}(M,\mu) \overset{+}{\leq} \min_n \{\mathbf{E}_{\ell(x)=n} K(\mu|x) + K(n) + \tfrac{2}{\ln 2} n\}$$

x_1

x_2

x_3

x_4

x_5

?

"number of errors" $\sim \mathrm{Dist}(M,\mu)$

Solomonoff bound:

$$\mathrm{Dist}(M,\mu) \lesssim K(\mu) \sim \text{“size of the image”} \approx 10^5$$

New bound:

$$\mathrm{Dist}(M,\mu) \lesssim K(\mu|x_1) + K(1) + \tfrac{2}{\ln 2} \sim \text{“small constant”}$$

Example

$$\mathrm{Dist}(M,\mu) \overset{+}{\leq} \min_n \{\mathbf{E}_{\ell(x)=n} K(\mu|x) + K(n) + \tfrac{2}{\ln 2} n\}$$

x_1 x_2 x_3 x_4 x_5

“number of errors” $\sim \mathrm{Dist}(M,\mu)$

Solomonoff bound:

$$\mathrm{Dist}(M,\mu) \lesssim K(\mu) \sim \text{“size of the image”} \approx 10^5$$

New bound:

$$\mathrm{Dist}(M,\mu) \lesssim K(\mu|x_1) + K(1) + \tfrac{2}{\ln 2} \sim \text{“small constant”}$$

Outline

Structure of the bounds

If x is not μ-typical ($\mu(x) \approx 0$), then x is disinformation

prediction errors	$\Leftarrow$	information	$+$	quantity of disinformation
$\mathrm{Dist}(M, \mu)$	$\stackrel{+}{\leq}$	$K(\mu)$		
$\mathrm{Dist}(M, \mu\|x)$	$\stackrel{+}{\leq}$	$K(\mu\|x)$	$+$	$K(\ell(x))$
$\mathrm{Dist}(M, \mu\|x)$	$\stackrel{+}{\leq}$	$K(\mu)$	$+$	$K(d_\mu(x))$
$\mathrm{Dist}(M, \mu\|x)$	$\stackrel{+}{\leq}$	$K_*(\mu\|x*)$	$+$	$K(d_\mu(x))$

Randomness deficiency: $d_\mu(x) = \log_2 \frac{M(x)}{\mu(x)}$

d_μ is a measure of non-typicalness, $d_\mu(x)$ is small for most x
$d_\mu(x) = \ell(x) - K(x)$ for uniform μ

Structure of the bounds

If x is not μ-typical ($\mu(x) \approx 0$), then x is disinformation

prediction errors	$\Leftarrow$	information	$+$	quantity of disinformation
$\mathrm{Dist}(M, \mu)$	$\stackrel{+}{\leq}$	$K(\mu)$		
$\mathrm{Dist}(M, \mu\|x)$	$\stackrel{+}{\leq}$	$K(\mu\|x)$	$+$	$K(\ell(x))$
$\mathrm{Dist}(M, \mu\|x)$	$\stackrel{+}{\leq}$	$K(\mu)$	$+$	$K(d_\mu(x))$
$\mathrm{Dist}(M, \mu\|x)$	$\stackrel{+}{\leq}$	$K_*(\mu\|x*)$	$+$	$K(d_\mu(x))$

Randomness deficiency: $d_\mu(x) = \log_2 \frac{M(x)}{\mu(x)}$

d_μ is a measure of non-typicalness, $d_\mu(x)$ is small for most x
$d_\mu(x) = \ell(x) - K(x)$ for uniform μ

Structure of the bounds

If x is not μ-typical ($\mu(x) \approx 0$), then x is disinformation

prediction errors	$\Leftarrow$	information	$+$	quantity of disinformation
$\mathrm{Dist}(M,\mu)$	$\overset{+}{\leq}$	$K(\mu)$		
$\mathrm{Dist}(M,\mu\vert x)$	$\overset{+}{\leq}$	$K(\mu\vert x)$	$+$	$K(\ell(x))$
$\mathrm{Dist}(M,\mu\vert x)$	$\overset{+}{\leq}$	$K(\mu)$	$+$	$K(d_\mu(x))$
$\mathrm{Dist}(M,\mu\vert x)$	$\overset{+}{\leq}$	$K_*(\mu\vert x*)$	$+$	$K(d_\mu(x))$

Randomness deficiency: $d_\mu(x) = \log_2 \frac{M(x)}{\mu(x)}$

d_μ is a measure of non-typicalness, $d_\mu(x)$ is small for most x
$d_\mu(x) = \ell(x) - K(x)$ for uniform μ

Structure of the bounds

If x is not μ-typical ($\mu(x) \approx 0$), then x is disinformation

prediction errors	$\Leftarrow$	information	$+$	quantity of disinformation
$\mathrm{Dist}(M, \mu)$	$\stackrel{+}{\leq}$	$K(\mu)$		
$\mathrm{Dist}(M, \mu\|x)$	$\stackrel{+}{\leq}$	$K(\mu\|x)$	$+$	$K(\ell(x))$
$\mathrm{Dist}(M, \mu\|x)$	$\stackrel{+}{\leq}$	$K(\mu)$	$+$	$K(d_\mu(x))$
$\mathrm{Dist}(M, \mu\|x)$	$\stackrel{+}{\leq}$	$K_*(\mu\|x*)$	$+$	$K(d_\mu(x))$

Randomness deficiency: $d_\mu(x) = \log_2 \frac{M(x)}{\mu(x)}$

d_μ is a measure of non-typicalness, $d_\mu(x)$ is small for most x
$d_\mu(x) = \ell(x) - K(x)$ for uniform μ

Structure of the bounds

If x is not μ-typical ($\mu(x) \approx 0$), then x is disinformation

prediction errors	$\Leftarrow$	information	$+$	quantity of disinformation
$\mathrm{Dist}(M, \mu)$	$\stackrel{+}{\leq}$	$K(\mu)$		
$\mathrm{Dist}(M, \mu \mid x)$	$\stackrel{+}{\leq}$	$K(\mu \mid x)$	$+$	$K(\ell(x))$
$\mathrm{Dist}(M, \mu \mid x)$	$\stackrel{+}{\leq}$	$K(\mu)$	$+$	$K(d_\mu(x))$
$\mathrm{Dist}(M, \mu \mid x)$	$\stackrel{+}{\leq}$	$K_*(\mu \mid x*)$	$+$	$K(d_\mu(x))$

Randomness deficiency: $d_\mu(x) = \log_2 \frac{M(x)}{\mu(x)}$

d_μ is a measure of non-typicalness, $d_\mu(x)$ is small for most x
$d_\mu(x) = \ell(x) - K(x)$ for uniform μ

Structure of the bounds

If x is not μ-typical ($\mu(x) \approx 0$), then x is disinformation

prediction errors	$\Leftarrow$	information	$+$	quantity of disinformation
$\mathrm{Dist}(M, \mu)$	$\overset{+}{\leq}$	$K(\mu)$		
$\mathrm{Dist}(M, \mu \vert x)$	$\overset{+}{\leq}$	$K(\mu \vert x)$	$+$	$K(\ell(x))$
$\mathrm{Dist}(M, \mu \vert x)$	$\overset{+}{\leq}$	$K(\mu)$	$+$	$K(d_\mu(x))$
$\mathrm{Dist}(M, \mu \vert x)$	$\overset{+}{\leq}$	$K_*(\mu \vert x*)$	$+$	$K(d_\mu(x))$

Randomness deficiency: $$d_\mu(x) = \log_2 \frac{M(x)}{\mu(x)}$$

d_μ is a measure of non-typicalness, $d_\mu(x)$ is small for most x
$d_\mu(x) = \ell(x) - K(x)$ for uniform μ

Structure of the bounds

If x is not μ-typical ($\mu(x) \approx 0$), then x is disinformation

prediction errors	$\Leftarrow$	information	$+$	quantity of disinformation
$\mathrm{Dist}(M, \mu)$	$\stackrel{+}{\leq}$	$K(\mu)$		
$\mathrm{Dist}(M, \mu\|x)$	$\stackrel{+}{\leq}$	$K(\mu\|x)$	$+$	$K(\ell(x))$
$\mathrm{Dist}(M, \mu\|x)$	$\stackrel{+}{\leq}$	$K(\mu)$	$+$	$K(d_\mu(x))$
$\mathrm{Dist}(M, \mu\|x)$	$\stackrel{+}{\leq}$	$K_*(\mu\|x*)$	$+$	$K(d_\mu(x))$

Randomness deficiency: $$d_\mu(x) = \log_2 \frac{M(x)}{\mu(x)}$$

d_μ is a measure of non-typicalness, $d_\mu(x)$ is small for most x
$d_\mu(x) = \ell(x) - K(x)$ for uniform μ

New conditional complexity

Definition (Prefix complexity monotone in conditions)

$x, y \in \mathcal{X}^*$, U is a universal twice-prefix machine, $p \in \{0,1\}^*$

$$K_*(y|x*) = \min\{\ell(p) | U(p*, x*) = y\}$$

Recall: conditional prefix complexity

$x, y \in \mathcal{X}^*$, U is a universal prefix machine, $p \in \{0,1\}^*$

$$K(y|x) = \min\{\ell(p) | U(p*, x) = y\}$$

New conditional complexity

Definition (Prefix complexity monotone in conditions)

$x, y \in \mathcal{X}^*$, U is a universal twice-prefix machine, $p \in \{0,1\}^*$

$$K_*(y|x*) = \min\{\ell(p) | U(p*, x*) = y\}$$

$$K_*(y|xz*) \leq K_*(y|x*)$$

$K_*(\mu|x*)$ bound

Theorem

For any computable measure μ and any $x, y \in \mathcal{X}^*$

$$\log_2 \frac{\mu(y|x)}{M(y|x)} \overset{+}{\leq} K_*(\mu|x*) + K(\lceil d_\mu(x) \rceil)$$

Corollary

$$\mathrm{Dist}(M, \mu|x_1 \dots x_n) \overset{+}{\leq} \min_{i \leq n}\{K(\mu|x_1 \dots x_i) + K(i) + K(d_\mu(x_1 \dots x_i))\}$$

For μ-typical x, $\mathrm{Dist}(M, \mu|x) \leq K(\mu|x') + O(\log \ell(x'))$

$K_*(\mu|x*)$ bound

Theorem

For any computable measure μ and any $x, y \in \mathcal{X}^*$

$$\log_2 \frac{\mu(y|x)}{M(y|x)} \stackrel{+}{\leq} K_*(\mu|x*) + K(\lceil d_\mu(x) \rceil)$$

Corollary

$$\mathrm{Dist}(M, \mu|x_1 \ldots x_n) \stackrel{+}{\leq} \min_{i \leq n}\{K(\mu|x_1 \ldots x_i) + K(i) + K(d_\mu(x_1 \ldots x_i))\}$$

For μ-typical x, $\mathrm{Dist}(M, \mu|x) \leq K(\mu|x') + O(\log \ell(x'))$

Conclusion and Open Problems

We extended the Solomonoff results on online sequence prediction to the case when some initial part of the sequence is given.

- Informative initial segment reduces the future loss; this gives us improved total loss bounds if the alphabet is large
- The future loss $\Leftarrow$ information $+$ quantity of disinformation

Directions for further research:

- Future loss bounds for general Bayes mixtures
- Online classification instead of sequence prediction
- Technical properties of the new complexity

Conclusion and Open Problems

We extended the Solomonoff results on online sequence prediction to the case when some initial part of the sequence is given.

- Informative initial segment reduces the future loss; this gives us improved total loss bounds if the alphabet is large
- The future loss $\Leftarrow$ information $+$ quantity of disinformation

Directions for further research:

- Future loss bounds for general Bayes mixtures
- Online classification instead of sequence prediction
- Technical properties of the new complexity