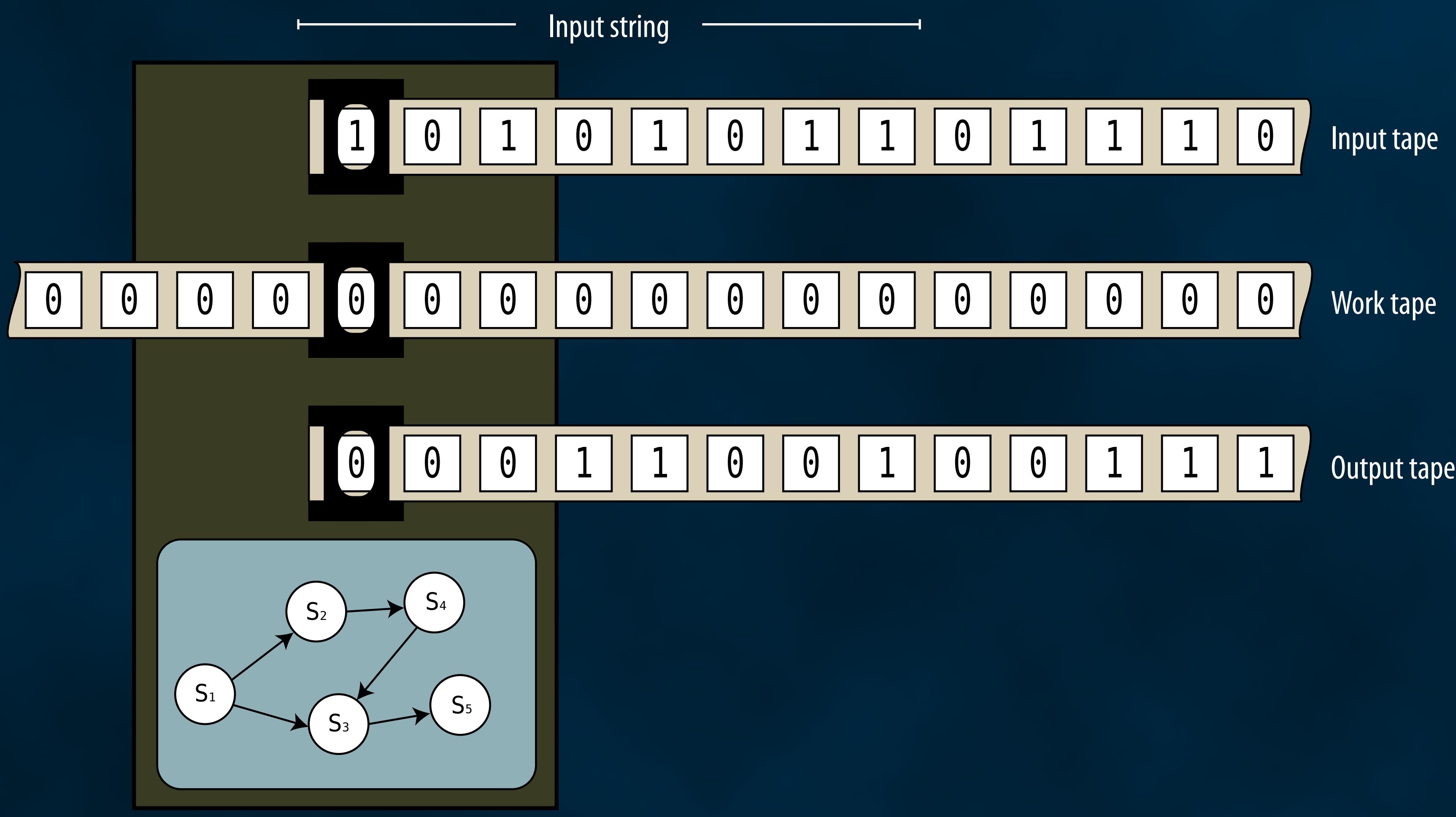
Minimum Message Length and Kolmogorov Complexity

C. S. Wallace and D. L. Dowe

Presentation by Alkis Gkotovos

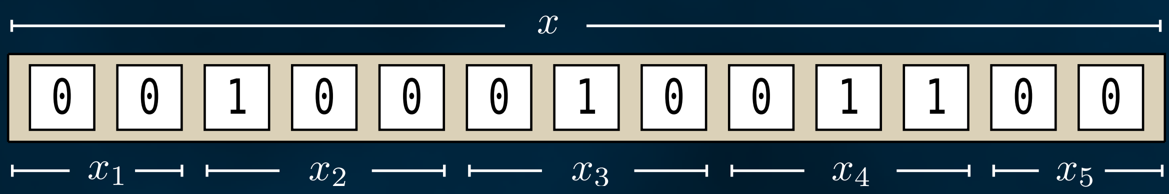
Overview

Turing Machines



Data & Hypotheses

Data string x is a representation of observational data from a real world phenomenon



$$L = \{00, 100, 010, 011\}$$

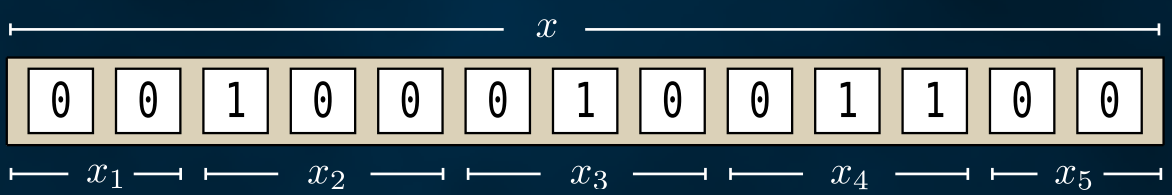
- "Sentences" $x_i \in L$, where L is a prefix-free set (data "language")
- Distinct sentences represent distinct real-world facts
- Sentences are conditionally independent given full knowledge of the phenomenon
- Strings are invariant to sentence permutation

Data & Hypotheses

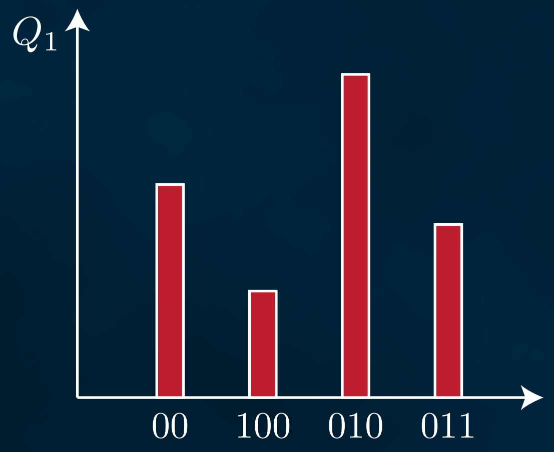
Hypothesis Q is a (computable) probability distribution over L

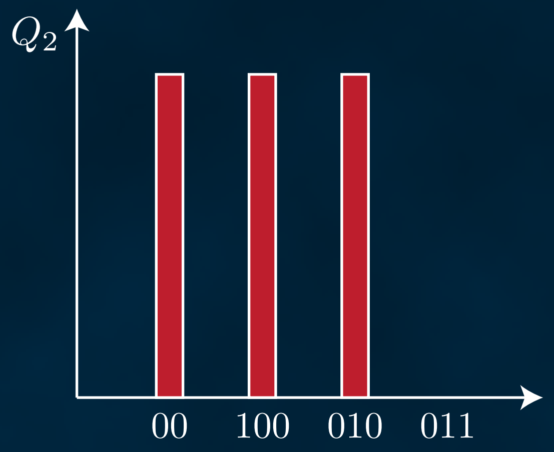
Conditional independence of sentences implies

$$x = x_1 \ldots x_n \Rightarrow Q(x) = Q(x_1) \times \ldots \times Q(x_n)$$



$$L = \{00, 100, 010, 011\}$$





Two-part encoding

How do we acquire a hypothesis-based encoding of data in the Algorithmic Complexity framework?

Idea

- Use conditional Kolmogorov complexity

$$K_T(x \mid y) = \min\{l(p) \mid T(\langle y, p \rangle) = x\}$$

  and interpret y as hypothesis and x as data

- Corresponding conditional algorithmic probability

$$P_T(x \mid y) = 2^{-K_T(x|y)}$$

Problem

Probability can never be 0, i.e. Popper-falsification not possible, because

$$K(x \mid y) < K(x) + O(1) \Rightarrow P_K(x \mid y) > P_K(x) + O(1)$$

Why? Hypothesis y acts as "extra info", instead of assertively

Proposal

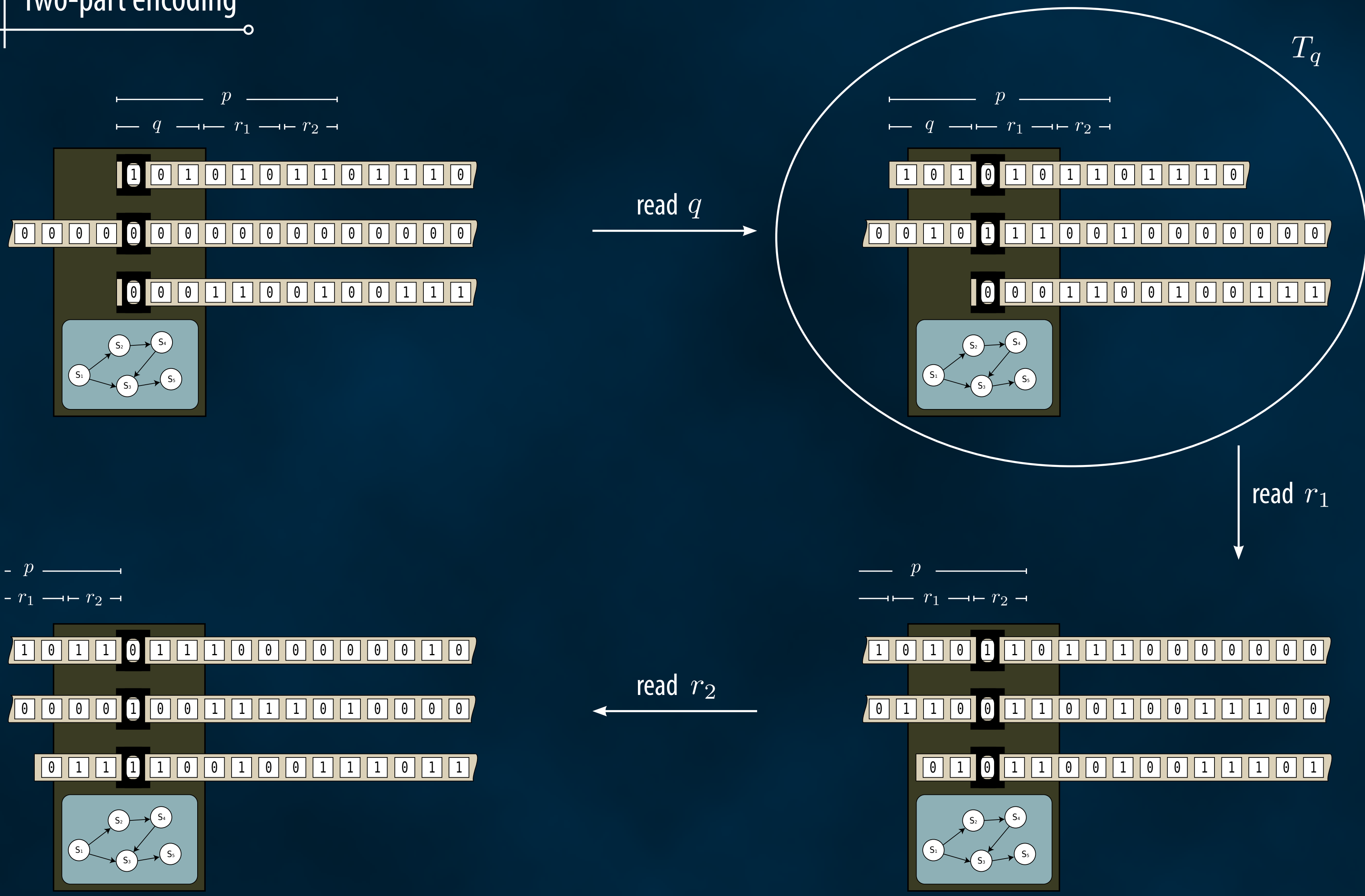
- Have hypothesis be a prefix of input string p
- Force intended two-part encoding by imposing conditions on p

Two-part encoding

Input p is an acceptable MML message encoding data string x, if

1)	$T(p) = x$	p encodes x
2)	$l(p) < l(x)$	some compression is achieved
3)	$p = qr$	two-part encoding
4)	$T(q) = \epsilon$	hypothesis q is does not determine data
5)	$T_q(rs) = xT_q(s)$	reading r does not alter the state of T
6)	$l(r) < K_T(x)$	hypothesis q is "significant"
7)	$x = x_1 \dots x_n \Rightarrow \begin{cases} r = r_1 \dots r_n \\ T_q(r_i) = x_i, \; i = 1 \dots n \end{cases}$	conditionally independent sentences
8)	$\begin{matrix} x' = x^{(1)}x^{(2)} \\ j' = j^{(1)}j^{(2)} \end{matrix} \Rightarrow \begin{matrix} T_q(j^{(1)}) = x^{(1)}, \; j^{(1)} < K_T(x^{(1)}) \\ T_q(j^{(2)}) = x^{(2)}, \; j^{(2)} < K_T(x^{(2)}) \end{matrix}$	hypothesis q is "general"
9)	No prefix of q satisfies all the above conditions	all of q is required

Two-part encoding



Two-part encoding

- The division of p into q and r is unique

- In what way exactly does hypothesis string q affect T?

  Remember $T \xrightarrow{q} T_q$

  T_q is a decoder of "second parts"

  $$T_q : S \to W$$

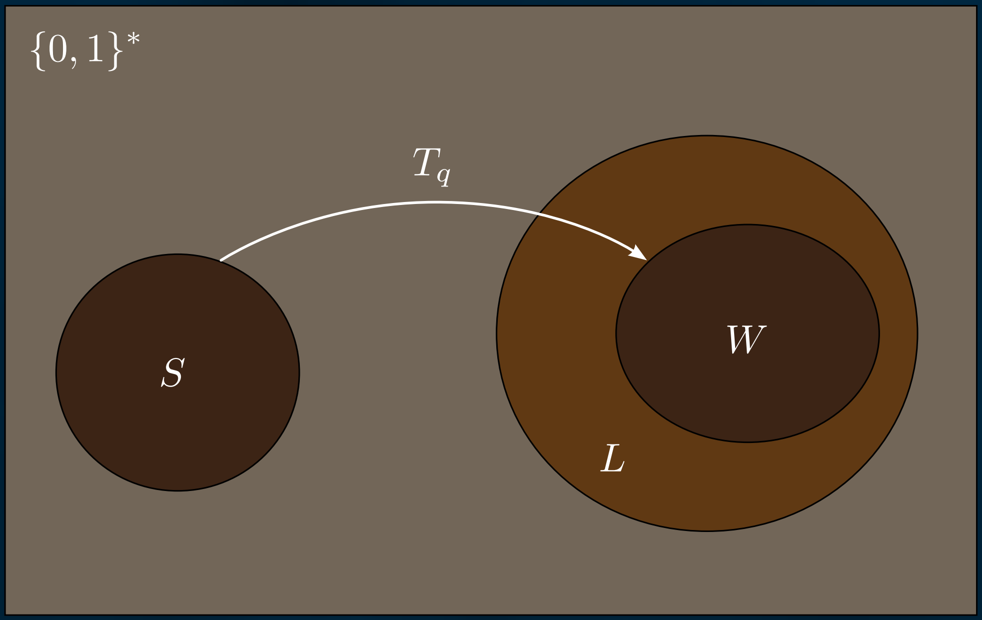
  Code words

  $$S = \{r_i \in \{0,1\}^* \mid T_q(r_i) \in L\}$$

  Subset of L that is coded

  $$W = \{x_i \in L \mid \exists r_i \in S : T_q(r_i) = x_i\}$$

  In fact, T_q decodes a prefix code (why?)



Two-part encoding

- What is the hypothesis (probability distribution) Q implied by hypothesis string q?

$$Q(x_i) = \begin{cases} 2^{-l(p)} & \text{, if } p \text{ is a shortest codeword for sentence } x_i \in L \\ 0 & \text{, if there is no codeword for sentence } x_i \in L \end{cases}$$

Because of prefix code

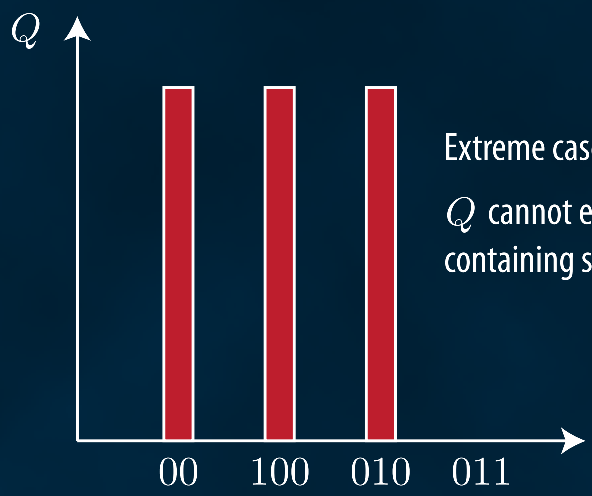
$$\sum_{x_i \in L} Q(x_i) = \sum_{x_i \in W} 2^{-l(p)} \overset{\text{Kraft}}{\leq} 1$$

- In this setting, hypotheses are falsifiable:

2) $l(p) < l(x) \Rightarrow l(r) < l(x)$

If Q assigns low probability (eq. high codeword length) to a sentence x_i, then adding enough such sentences to the data string will violate the above condition and falsify the hypothesis.

Can Q assign lower codeword length to every sentence?
(L is a complete prefix code for "data facts")



Extreme case:

Q cannot encode *any* data string containing sentence 011.