Lesson 5
Coding
5-1: Optimal Codes

Jan Reimann

Math 574, Topics in Logic
Penn State, Spring 2014

Codes

We have seen that the Kolmogorov complexity of a string σ can be seen as the length of an optimal code for σ.

Now we want to see if an analogue interpretation holds for (probabilistic) entropy.

Setting: X is an A-valued random variable. We want to assign code words to each outcome of X, so that the expected code length is minimal.

DEF: A (source) code for X is a mapping $c : A \to D^{<\mathbb{N}}$, where D is a finite alphabet.

$D = \{1, \dots, D\}$

We identify D with its cardinality.

The expected length $L(c)$ of a code for X with distribution P is given as

$$L(c) = \sum_{a \in A} P(a)|c(a)|.$$

Examples

$$L(C) = \sum_i P(a_i)\,|C(a_i)|$$

$$A = \{0, 1, 2, 3\}$$

$$P(0) = \frac{1}{2} \qquad P(2) = \frac{1}{8}$$

$$P(1) = \frac{1}{4} \qquad P(3) = \frac{1}{8}$$

$$D = \{0, 1\}$$

C_1:

$$0 \longmapsto 00$$
$$1 \longmapsto 01$$
$$2 \longmapsto 10$$
$$3 \longmapsto 11$$

C_2:

$$0 \longmapsto 0$$
$$1 \longmapsto 10$$
$$2 \longmapsto 110$$
$$3 \longmapsto 111$$

$$L(C_1) = 2$$

$$L(C_2) = \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{8}\cdot 3 + \frac{1}{8}\cdot 3$$

$$= 1.75$$

$$H(X) = 1.75$$

Prefix Codes

A code c is non-singular if it is one-one.

suffices for an unambiguous description of a single value of X

We want to encode a data stream produced by X, i.e. A-strings (sequences).

Extension of c: $c^*(a_1 \ldots a_n) = c(a_1)^\frown c(a_2)^\frown \ldots {}^\frown c(a_n)$.

A code is uniquely decodable if its extension is non-singular.

One way to ensure this: code words for c are prefix-free.

prefix code instantaneous code

Kraft Inequality

Which code lengths are possible for prefix codes?

THM: Let $W \subseteq D^{<\mathbb{N}}$ be prefix-free (possibly infinite). Then

$$\sum_{\sigma \in W} D^{-|\sigma|} \leqslant 1$$

Conversely, given any sequence $l_0, l_1, l_2, \ldots$ of non-negative integers satisfying

$$\sum_{i} D^{-l_i} \leqslant 1,$$

Then there exists a prefix-free set $\{\sigma_0, \sigma_1, \ldots\} \subseteq D^{<\mathbb{N}}$ of code words such that $|\sigma_i| = l_i$.

Proving the Kraft Inequality

PENNSTATE
Eberly College of Science

$$\sum D^{-|\sigma_i|} \leq 1$$

Proof: $\Rightarrow$

- Let $\boldsymbol{\lambda}$ be the measure on $D^{\mathbb{N}}$ induced by $\boldsymbol{\lambda}[\boldsymbol{\sigma}] = D^{-|\sigma|}$.
- Identifying strings with cylinders, a prefix-free set corresponds to a disjoint collection of open sets.
- By countable additivity and monotonicity of measures,

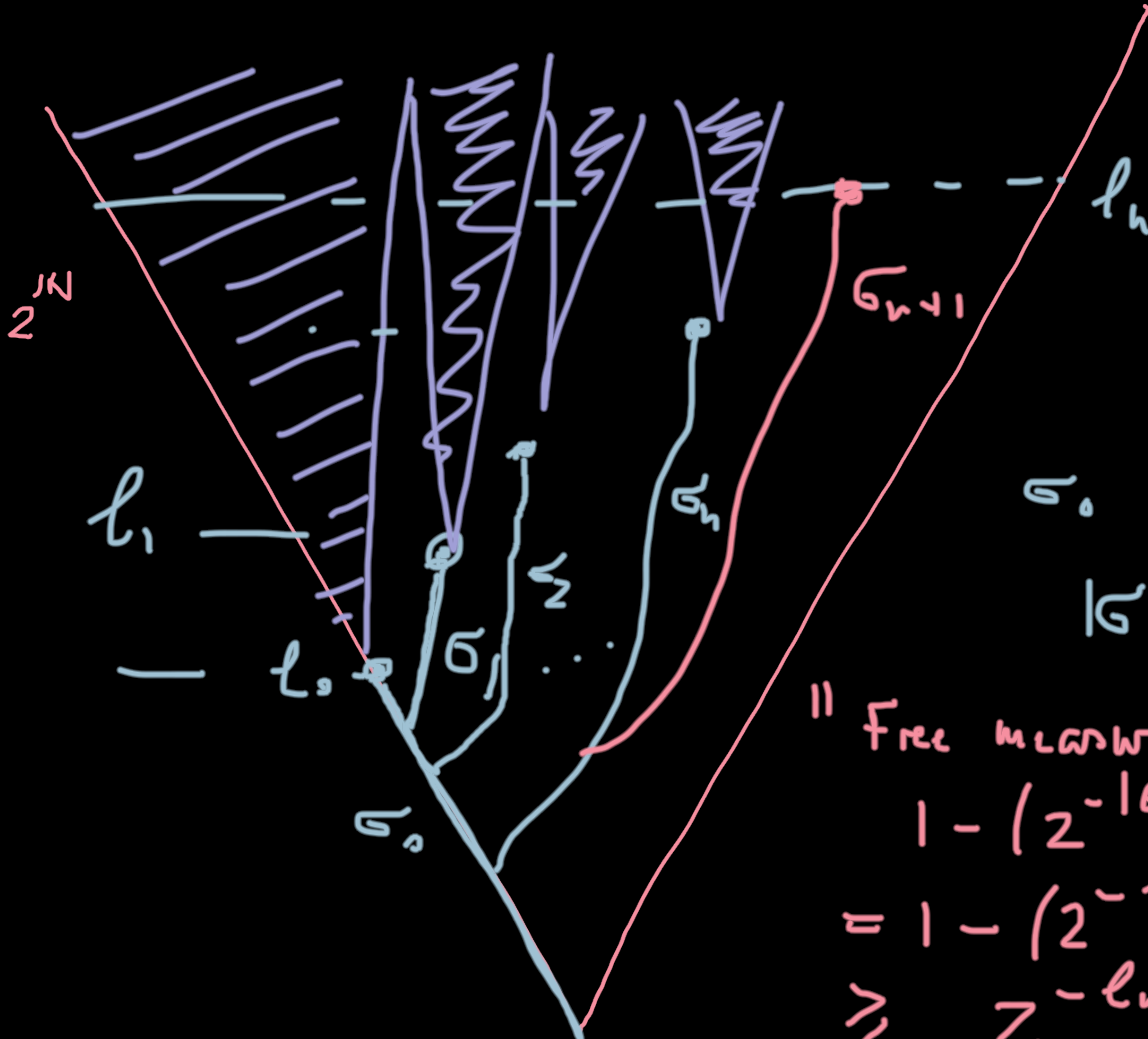
$$\sum_{\sigma \in W} \overset{D}{\cancel{2}}{}^{-|\sigma|} = \lambda\left(\bigcup_{\sigma} [\sigma]\right) \leqslant \lambda(D^{\mathbb{N}}) = 1.$$

$D = \{0, 1\}$

$\lambda[\sigma] = 2^{-|\sigma|}$

Proving the Kraft Inequality

Proof: $\Leftarrow$ Assume $\ell_0 \le \ell_1 \le \ell_2 \le \dots$



$$\sum_i 2^{-\ell_i} \le 1$$

$$\sigma_0, \sigma_1, \sigma_2, \dots, \sigma_n$$

$$|\sigma_i| = \ell_i$$

"Free measure":

$$1 - \left(2^{-|\sigma_0|} + \dots + 2^{-|\sigma_n|}\right)$$
$$= 1 - \left(2^{-\ell_0} + \dots + 2^{-\ell_n}\right)$$
$$\geq 2^{-\ell_{n+1}}$$

Optimal Codes

The Kraft Inequality puts some restraints on the nature of a prefix code.

Subject to this restraint, what is the minimum expected length of a code for X (A-valued)?

Suppose $A = \{1, \ldots, m\}$. Let $p_i = P(X = i)$. We want to find code lengths $l_1, \ldots l_m$.

We have to minimize

$$L = \sum p_i l_i$$

over all integers $l_1, \ldots l_m$ satisfying

$$\sum D^{-l_i} \leqslant 1.$$

Optimal Codes Via Calculus

Let us allow the l_i to be arbitrary non-negative reals, and assume the constraint holds with equality:

$$\sum D^{-l_i} = 1.$$

We can then use Lagrange multipliers: Find critical points of partial derivatives of

$$\Lambda = \sum p_i l_i + \lambda \left(\sum D^{-l_i} - 1 \right).$$

We have

$$\frac{\partial \Lambda}{\partial l_i} = p_i - \lambda D^{-l_i} \ln D.$$

Put $\frac{\partial \Lambda}{\partial l_i} = 0$, and we obtain

$$D^{-l_i} = \frac{p_i}{\lambda \ln D}.$$

Optimal Codes Via Calculus

Plug $D^{-l_i} = \frac{p_i}{\lambda \ln D}$ back in the constraint $\sum D^{-l_i} = 1$:

$$\sum p_i = \lambda \ln D, \quad \text{hence} \quad \lambda = 1/\ln D.$$

This in turn implies

$$D^{-l_i} = \frac{p_i}{\lambda \ln D} = p_i,$$

and thus

$$l_i^* = -\log_D p_i.$$

The expected code length is then

$$L^* = \sum p_i l_i^* = -\sum p_i \log_D(p_i) = H_D(X).$$

Optimality of Entropy Codes

We verify directly that the previous bound is indeed a global minimum: no integer-length prefix code has expected length less than entropy.

THM: Let L be the expected length of a D-ary prefix code of a random variable X. Then

$$L \geqslant H_D(X),$$

where equality holds iff $D^{-l_i} = p_i$.

Optimality of Entropy Codes

Proof:

≥ 0

$$L - H_D(X) = \sum p_i l_i + \sum p_i \log_D(p_i)$$
$$= -\sum p_i \log_D D^{-l_i} + \sum p_i \log_D(p_i).$$

Put $c = \sum D^{-l_i}$ and $r_i = D^{-l_i}/c$. Then

$= r_i$

$$L - H_D(X) = \sum p_i \log_D(p_i) - \sum p_i \log_D \frac{D^{-l_i}}{c} c$$
$$= \sum p_i \log_D \frac{p_i}{r_i} - \log_D c$$

KL-divergence

$c \leq 1$

$$\underbrace{= \frac{1}{\log D} \underbrace{D(\vec{p} \parallel \vec{r})}_{\geq 0} + \log_D(1/c)}_{\geq 0} \geq 0$$