CO 432 Spring 2025:
Lecture Notes

Lecture notes taken, unless otherwise specified, by myself during the Spring 2025 offering of CO 432, taught by Vijay Bhattiprolu.

Lectures

Chapter 1

Introduction

1.1 Entropy

TODO

Lecture 1
May 6

Definition 1.1.1 (entropy)
For a random variable X which is equal to i with probability p_i, the entropy $H(\mathsf{X}) := \sum_i p_i \log \frac{1}{p_i}$.

1.1.1 Axiomatic view of entropy

Lecture 2
May 8

We want $S : [0,1] \to [0, \infty)$ to capture how "surprised" we are $S(p)$ that an event with probability p happens. We want to show that under some natural assumptions, this is the only function we could have defined as entropy. In particular:

1. $S(1) = 0$, a certainty should not be surprising
2. $S(q) > S(p)$ if $p > q$, less probable should be more surprising
3. $S(p)$ is continuous in p
4. $S(pq) = S(p) + S(q)$, surprise should add for independent events. That is, if I see something twice, I should be twice as surprised.

Proposition 1.1.2
If $S(p)$ satisfies these 4 axioms, then $S(p) = c \cdot \log_2(1/p)$ for some $c > 0$.

Proof. Suppose a function $S : [0,1] \to [0, \infty)$ exists satisfying the axioms. Let $c := S(\frac{1}{2}) > 0$.

By axiom 4 (addition), $S(\frac{1}{2^k}) = kS(\frac{1}{2})$. Likewise, $S(\frac{1}{2^{1/k}} \cdots \frac{1}{2^{1/k}}) = S(\frac{1}{2^{1/k}}) + \cdots + S(\frac{1}{2^{1/k}}) = kS(\frac{1}{2^{1/k}})$.

Then, $S(\frac{1}{2^{m/n}}) = \frac{m}{n} S(\frac{1}{2}) = \frac{m}{n} \cdot c$ for any rational m/n.

By axiom 3 (continuity), $S(\frac{1}{2^z}) = c \cdot z$ for all $z \in [0, \infty)$ because the rationals are dense in the reals. In particular, for any $p \in [0, 1]$, we can write $p = \frac{1}{2^z}$ for $z = \log_2(1/p)$ and we get

$$S(p) = S\left(\frac{1}{2^z}\right) = c \cdot z = c \cdot \log_2(1/p)$$

as desired. □

We can now view entropy as expected surprise. In particular,

$$\sum_i p_i \log_2 \frac{1}{p_i} = \mathop{\mathbb{E}}_{x \sim \mathsf{X}} [S(p_x)]$$

for a random variable $\mathsf{X} = i$ with probability p_i.

1.1.2 Entropy as optimal lossless data compression

Suppose we are trying to compress a string consisting of n symbols drawn from some distribution.

Problem 1.1.3

What is the expected number of bits you need to store the results of n independent samples of a random variable X?

We will show this is $nH(\mathsf{X})$.

Notice that we assume that the symbols we are drawn independently, which is violated by almost all data we actually care about.

Definition 1.1.4

Let $C : \Sigma \to (\Sigma')^*$ be a code. We say C is uniquely decodable if there does not exist a collision $x, y \in \Sigma^*$, with identical encoding $C(x_1)C(x_2)\cdots C(x_k) = C(y_1)C(y_2)\cdots C(y_{k'})$.

Also, C is prefix-free (sometimes called instantaneous) if for any distinct $x, y \in \Sigma$, $C(x)$ is not a prefix of $C(y)$.

Proposition 1.1.5

Prefix-freeness is sufficient for unique decodability.

Example 1.1.6. Let $C : \{A, B, C, D\} \to \{0, 1\}^*$ where $C(A) = 11$, $C(B) = 101$, $C(C) = 100$, and $C(D) = 00$. Then, C is prefix-free and uniquely decodable.

We can easily parse 1011100001100 unambiguously as 101.11.00.00.11.00 ($BADDAD$).

Recall from CS 240 that a prefix-free code is equivalent to a trie, and we can decode it by traversing the trie in linear time.

Theorem 1.1.7 (Kraft's inequality)
A prefix-free binary code $C : \{1, \dots, n\} \to \{0,1\}^*$ with codeword lengths $\ell_i = |C(i)|$ exists if and only if
$$\sum_{i=1}^{n} \frac{1}{2^{\ell_i}} \leq 1.$$

Proof. Suppose $C : \{1, \dots, n\} \to \{0,1\}^*$ is prefix-free with codeword lengths ℓ_i. Let T be its associated binary tree and let W be a random walk on T where 0 and 1 have equal weight (stopping at either a leaf or undefined branch).

Define E_i as the event where W reaches i and $E_\varnothing$ where W falls off. Then,

$$\begin{aligned}
1 &= \Pr(E_\varnothing) + \sum_i \Pr(E_i) \\
&= \Pr(E_\varnothing) + \sum_i \frac{1}{2^{\ell_i}} && \text{(by independence)} \\
&\geq \sum_i \frac{1}{2^{\ell_i}} && \text{(probabilities are non-negative)}
\end{aligned}$$

Conversely, suppose the inequality holds for some ℓ_i. WLOG, suppose $\ell_1 < \ell_2 < \cdots < \ell_n$.

Start with a complete binary tree T of depth ℓ_n. For each $i = 1, \dots, n$, find any unassigned node in T of depth ℓ_i, delete its children, and assign it a symbol.

Now, it remains to show that this process will not fail. That is, for any loop step i, there is still some unassigned node at depth ℓ_i.

Let $P \leftarrow 2^{\ell_n}$ be the number of leaves of the complete binary tree of depth ℓ_n. After the i^{th} step, we decrease P by $2^{\ell_n - \ell_i}$. That is, after n steps,

$$\begin{aligned}
P &= 2^{\ell_n} - \sum_{i=1}^{n} \frac{2^{\ell_n}}{2^{\ell_i}} \\
&= 2^{\ell_n} - 2^{\ell_n} \sum_{i=1}^{n} \frac{1}{2^{\ell_i}} \\
&\geq 0
\end{aligned}$$

by the inequality. □