Arithmetic Coding and Adaptive Coding
Review

• Huffman Codes
• Entropy
Arithmetic coding

- Partitioning the unit segment.
- Identifying a part using a binary expansion.
How many bits?

- $p$ - the probability of the character
- $l$ = the length of the segment
- There segment must contain a dyadic number with $\log (1/p)$ bits
Coding more than one char

• An input stream: $x_1, x_2, x_3, ...$
• $x_1$ chooses a part $[a_1, b_1)$ in partition of $[0, 1)$
• $x_2$ chooses a part $[a_2, b_2)$ in partition of $[a_1, b_1)$
• $x_3$ chooses ....
When can we send the next bit?

- As soon as we know whether the segment is on the left or on the right of a dyadic partition.
- Unbounded delay ...

bit 1

Not yet...

bit1 = 1
Performance of arithmetic codes

The message: \( x_1, x_2, x_3, \ldots, x_n \)

Generated IID according to distribution \( p \)

\[
\ell = \left\lfloor \log_2 \left( \prod_{i=1}^{n} \frac{1}{p(x_i)} \right) \right\rfloor = \left\lfloor \sum_{i=1}^{n} \log_2 \frac{1}{p(x_i)} \right\rfloor < \sum_{i=1}^{n} \log_2 \frac{1}{p(x_i)} + 1
\]

\[
E(\ell) < n \sum_{x} p(x) \log_2 \frac{1}{p(x)} + 1 = nH(p) + 1
\]

At most one bit more than the Shannon lower bound for the whole message
Using the wrong distribution

- So far we assumed that we are coding using the correct distribution \( p \). Suppose that we are coding according to a dist \( q \neq p \)

\[
E(\ell) < n \sum_x p(x) \log_2 \frac{1}{q(x)} + 1 = \\
= n \left( \sum_x p(x) \log_2 \frac{1}{p(x)} + \sum_x p(x) \log_2 \frac{p(x)}{q(x)} \right) + 1 \\
= n \left( H(p) + D_{KL}(p||q) \right) + 1
\]
Two part codes

- Receiver does not know distribution
- Sender sends two pieces:
  1. Distribution parameters (Model)
  2. Message, coded using distribution (Data given model)
Non IID sources

\[ p(x_1, x_2, \ldots, x_t) \neq \prod_{s=1}^{t} p(x_s) \]

\[ p(x_t|x_{t-1}, \ldots, x_1) \neq p(x_t) \]

Arithmetic coding does not require characters to be IID
Adaptive Coding

Original Message

Arithmetic Coder

encoded message

Decoded Message

Arithmetic Decoder
performance of adaptive codes.

• Source is IID

• Predictor converges to correct distribution over time.

• Code length:  \[ \ell = \left\lceil \sum_{t=1}^{n} \log_2 \frac{1}{q_t(x_t)} \right\rceil \]

\[ E(\ell) < \sum_{t=1}^{n} \sum_{x} p(x) \log_2 \frac{1}{q_t(x)} + 1 \]
online prediction of probabilities

• A binary input stream: \( x_1, x_2, x_3, \ldots \)

• Generated IID according to a fixed but unknown distribution \((p, 1-p)\).

• Task: map \( x_1, x_2, x_3, \ldots, x_{t-1} \) to \( q_t \) so that \( q_t \rightarrow p \) quickly so as to minimize

\[
E_{x_1 \sim p, \ldots x_t \sim p} \left( \sum_{t=1}^{n} \log_2 \frac{1}{q_t(x_t)} \right)
\]
Laplace Law Of Succession

$$q_t = \frac{\#1 + 1}{t + 1}$$

\#1 = number of 1's in \(x_1, x_2, x_3, \ldots, x_{t-1}\)

$$E_{x_1 \sim p, \ldots, x_t \sim p} \left( \sum_{t=1}^{n} \log_2 \frac{1}{q_t(x_t)} \right) \leq tH(p) + \log_2 t$$
Kritchvski-Trofimov Prediction rule

\[ q_t = \frac{\#1 + 1/2}{t} \]

\#1 = number of 1’s in \( x_1, x_2, x_3, \ldots, x_{t-1} \)

\[ E_{x_1 \sim p, \ldots, x_t \sim p} \left( \sum_{t=1}^{n} \log_2 \frac{1}{q_t(x_t)} \right) \leq tH(p) + \frac{1}{2} \log_2 t \]

Best possible factor
Summary

- Arithmetic coding
- Adaptive coding
- Predictive coding
- Laplace Law of succession and the KT prediction rule.