Lecture 8: Optimal Code Length, Roof Code

- Gambling interpretation
- Roof code
- Coding over sequence
- Price of wrong code
- Universal source coding

Optimal code length

- Encode source X with pdf p_i
- Find code length l_i to minimize expected code length L
- Uniquely decodable code should satisfy Kraft-McMillan inequality

$$\begin{aligned} \text{minimize}_{l_i} \quad & \sum_{i=1}^{m} p_i l_i \\ \text{subject to} \quad & \sum_{i=1}^{m} D^{-l_i} \leq 1. \end{aligned}$$

Horse racing

- m horses run in a race
- ith horse wins with probability p_i
- if i wins, receive r dollar for 1 dollar bet
- if i loses, lose the 1 dollar

Gambler's problem

- initially have w dollars
- $b_i \geq 0$: fraction of money invested on horse i, fixed in each round

$$\sum_{i=1}^{m} b_i \leq 1$$

- in first round round, your return is

$$S_1 = rwb_{I_1}$$

I_1: index of winning horse in round 1

- wealth in nth round

$$S_n = rS_{n-1}b_{I_n}$$

I_1: index of winning horse in round 1

- after N rounds, total wealth

$$S_N = r^N w b_{I_N} \cdots b_{I_1}$$

- using law of large number for product, when $N \to \infty$

$$\sqrt[N]{S_N} \to r\,\sqrt[N]{w} D^{E \log_D (b_{I_1})} = r\,\sqrt[N]{w} D^{\sum_i p_i \log_D b_i}$$

Optimal gambling strategy

- maximize return given a budget constraint

$$
\begin{aligned}
\text{maximize}_{x_i} \quad & \sum_{i=1}^{m} p_i \log_D b_i \\
\text{subject to} \quad & \sum_{i=1}^{m} b_i \leq 1.
\end{aligned}
$$

- let $b_i = D^{-l_i}$, $\Rightarrow \log_D b_i = -l_i$

- equivalent to optimal code length problem

Roof code

- solution to optimal code length problem $l_i^* = -\log_D p_i$
- consider code length $l_i = -\lceil \log_D p_i \rceil$
- this satisfies Kraft inequality

$$\sum D^{-\lceil \log_D (1/p_i) \rceil} \le \sum D^{\log_D p_i} = \sum_i p_i = 1$$

- we can construct a instantaneous code from Kraft

- expected code length of roof code

$$\sum p_i \lceil \log_D(1/p_i) \rceil < \sum p_i (\log_D(1/p_i) + 1) = H_D(X) + 1$$

- we have shown that $L \geq H_D(X)$ ($D(p||q) \geq 0$ and Kraft inequality)
- the expected code length of roof code at most one bit more than optimal code
- optimal code must be better than roof code.
- expected length of optimal code L^*

$$H_D(X) \leq L^* < H_D(X) + 1$$

Sometimes “roof” can be quite bad

- code a biased coin flip with $p = 1/4$ using binary code
- $l_1 = \lceil \log_2 1/p \rceil = 2$, $l_2 = \lceil \log_2 1/(1-p) \rceil = \lceil 0.415 \rceil = 1$
- $C(1) = 01$, $C(2) = 1$, $L = 1.25$
- but obviously $C(1) = 0$ and $C(2) = 1$ is better, with $L = 1$
- if send 100 symbols using roof code, we will use $0.25 \times 100 = 25$ extra bits, that is 25% more than needed!

Coding over sequence

- $-\log_D p_i$ is not integer $\Rightarrow$ there is an overhead of at most 1 bit
- Can reduce overhead per symbol by coding over a sequence
- Design a system to send n symbols from source X
- Example: coding for single symbol of horse racing:

$$C(3) = 110, \quad C(1) = 0, \quad C(7) = 111110$$

coding for a sequence of outcomes:

$$C(\begin{bmatrix} 3 & 1 & 7 \end{bmatrix}) = 10010010$$

Coding over i.i.d. sequence

- $X_1, X_2, \ldots, X_n$ are i.i.d. with pdf $p(x)$
- Coding using joint pdf: $p(x_1, \ldots, x_n)$
- Expected code length per symbol

$$L_n = \frac{1}{n} \sum p(x_1, \ldots, x_n) l(x_1, \ldots, x_n)$$

- $H(X_1, \ldots, X_n) \le nL_n < H(X_1, \ldots, X_n) + 1$

$$H(X) \le L_n < H(X) + \frac{1}{n}$$

- Increase n, lower overhead

Coding over dependent process

- $X_1, X_2, \ldots, X_n$, not necessarily dependent
- Coding using joint pdf $p(x_1, \ldots, x_n)$
- $H(X_1, \ldots, X_n) \leq nL_n < H(X_1, \ldots, X_n) + 1$
- $\frac{1}{n}H(X_1, \ldots, X_n) \leq L_n < \frac{1}{n}H(X_1, \ldots, X_n) + \frac{1}{n}$
- $n \to \infty$, both sides becomes entropy rate $H(\mathcal{X})$
- entropy rate $H(\mathcal{X})$ is the compression limit of stochastic process

Price of wrong code

- we do not really know $p(x)$
- if we code using an estimate $q(x)$

$$l(x) = \lceil \log_D (1/q(x)) \rceil$$

- how much do we pay in expected code length?
- $H(p) + D(p||q) \leq L(X) < H(p) + D(p||q) + 1$
- Incur a penalty $D(p||q)$

Further topics

Universal source coding

- Our source may consist of several possible object: text, image, voice
- How do we build a code robust to mismatch errors?
- If there are two possible pdfs $p_1(x)$ and $p_2(x)$, equally likely
- Find encoding pdf $q(x)$ such that $D(q||p_1) = D(q||p_1)$
- Ingeneral, find

$$q(x) = \min_q \max_n D(q||p_n)$$

Coding over large alphabet: Many common sources: text or images, have essentially infinite alphabets

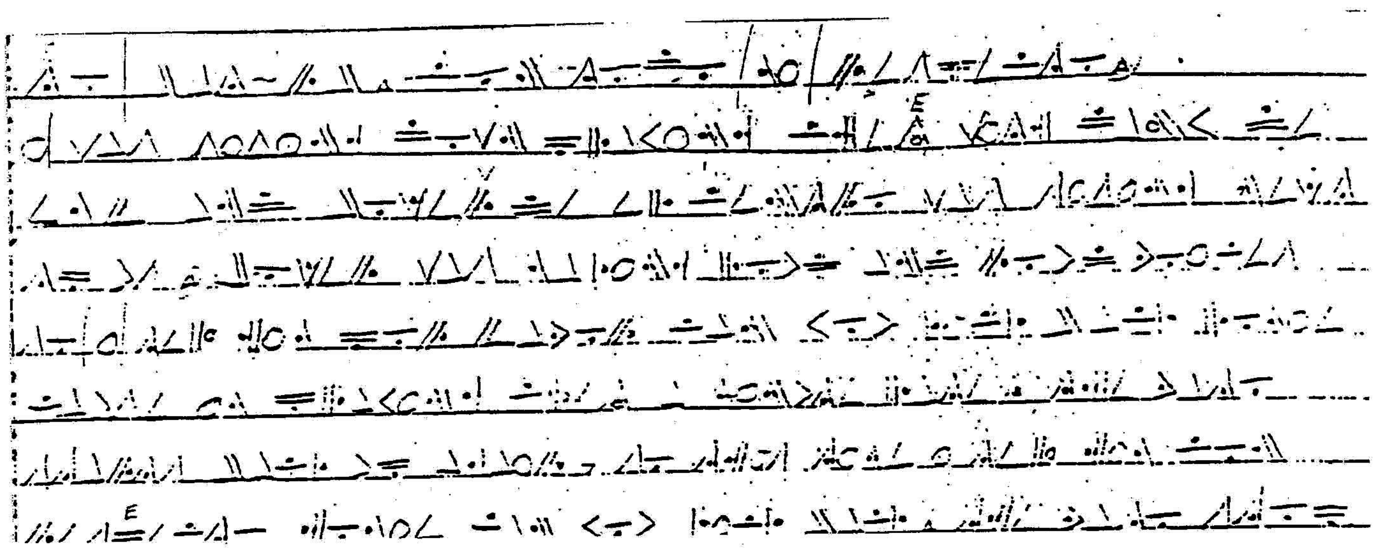
Entropy rate of English

- Shannon guessing game:
- Human subject is given a sample of English text, and asked to guess the next letter
- An optimal subject will estimate the probabilities of the next letter and guess the most likely, and then second most likely...
- Record the number of guesses
- Entropy of the guess sequence $\geq$ entropy rate of English
- About 1.3 bits per symbol

Cryptography

- Decoding without knowing encoding function

$$f : \{\text{code space}\} \rightarrow \{\text{Usual alphabet}\}$$



P. Diaconis, The Markov chain Monte Carlo revolution.

Summary

- Optimal gambling is the dual of finding shortest code length
- Roof code: a simply construction, incurs at most 1 bit overhead per symbol
- Coding over sequence to reduce the overhead
- Cost of using estimated pdf $D(p||q)$