The Source Coding Theorem

Or: How low can we go?

Section 1

Block coding and generic Lossy Compression

Lossy Compression

We take a simple model for lossy compression.

$x \in A_X$	$P(x)$	$c(x)$
a	0.25	00
b	0.25	01
c	0.24	10
d	0.24	11
e	0.02	—

so small we won't see it much.

More formally:

$$P(x \in S_\delta) \geq 1-\delta$$

smallest δ-sufficient subset S_δ

$\delta = 0.02$

$S_\delta = \{a, b, c, d\}$

$$H_\delta(x) = \log |S_\delta|$$

Essential bit content

Blocks of Symbols

Losing an entire symbol is clearly problematic, but what if we considered <u>blocks</u> of symbols?

$$A_x = \{a, b, c\}$$

$$A_x^2 = \{aa, ab, ac, ba, bb, bc, ca, cb, cc\}$$

Now we can remove a subset of the c-pairs.

↓

A_x^N has lots of freedom to throw things away.

E.g.

A_x	P_x
a	0.25
b	0.25
c	0.25
d	0.25

Throwing this away is very problematic

A_x^3	P_x^3
aaa	0.25^3
⋮	⋮
ccb	0.25^3
ccc	0.25^3

} Can throw these away at cost 2×0.25^3

Compressing Blocks

$X \to X^N$ N RVs

$\overleftarrow{a, c, d \ldots x}^{N}$

$A_x = \{a, b\}$ $A_x^N : \{aa, ab, ba, bb\}$

$$H(X,Y) = -\sum_{xy} P(x,y) \log P(x,y)$$

$$= -\sum_{xy} P(x)P(y) \log P(x)P(y)$$

$$= -\sum_{x} P(x) \log P(x) - \sum_{y} P(y) \log P(y)$$

$$= H(X) + H(Y)$$

$$\therefore \boxed{H(X^N) = N H(X)}$$

$H(x) = -\sum P_x \log P_x$ $\quad P(x^N) = P(x)^N$

$H(x^N) = -\sum P(x,x^2,x^3) \log P(\log$ $\quad P(a,a) \log P(a,a)$

$= -\sum P(x)^N \log P(x)^N$ $\quad P(a)^2 \log P(a)^2$

$2P(a)^2 \log P(a)^2$.

Compressing Blocks

If we now assign codewords to blocks and compress as before

Before compression we need $\frac{1}{N} H(X^N)$ bits per symbol.

After compression $\frac{1}{N} H_\delta(X^N)$ bits per symbol.

Our interest is the limit s.t.

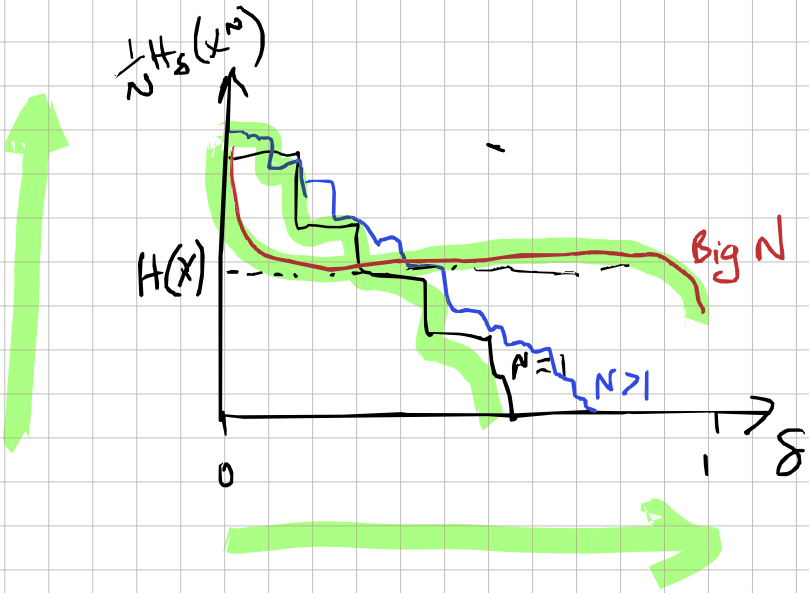
$$\frac{1}{N} H(X^N) \sim \frac{1}{N} H_\delta(X^N)$$

$$H(x) = -\sum P_x \log P_x \qquad P(x^N) = P(x)^N$$

$$H(X^N) = -\sum P(x,x^2,x^3) \log P(\log \qquad P(a,a) \log P(a,a)$$

$$= -\sum P(x)^N \log P(x)^N \qquad P(a)^2 \log P(a)^2$$

$$2P(a)^2 \log P(a)^2.$$


$\frac{1}{N} H_\delta(X^N)$
$H(X)$
Big N
N=1
N>1
0
1
δ

The Source Coding Theorem

Let X be an ensemble with entropy $H(X) = H$

Given $\epsilon > 0$ and $0 < \delta < 1$ there exists a positive integer N_0 s.t. for $N > N_0$

$$\left| \frac{1}{N} H_\delta(X^N) - H \right| < \epsilon$$

Now read Mackay pg. 74-78.

Section 2

The Typical Set.

Probabilities of groups of outcomes

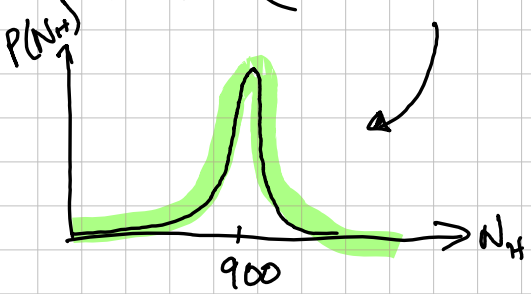
Take a biased coin with $P(H) = 0.9$ and flip it 1,000 x.

What is the most probable sequence?

↳ 1,000 heads, with probability 0.9^{1000}

But 0.9^{1000} is tiny - there is $1 - 0.9^{1000}$ chance it will be something else. i.e. it's almost certainly something else.

$$P(N_H) = \left(\text{No of sequences with } N_H\right) \times P(\text{sequence with } N_H)$$

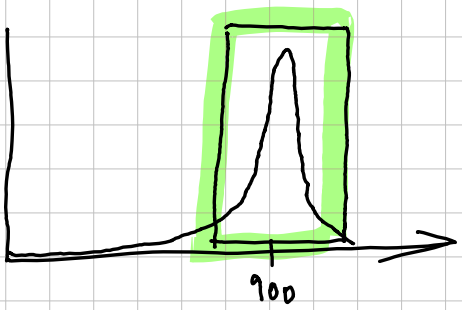


$$\binom{1000}{N_H} \times 0.9^{N_H} \times 0.1^{1,000 - N_H}$$

Dominates

Typicality

We find that the probability is concentrated around $N_H = 900$. As N increases, it gets tighter and tighter around $0.9N$



Idea: pick the sequences 'close' to $N_{1+} = 900$ and assign codewords only to them.

Draw x from A_x^N $\quad A_x = \{1, 2, 3, \ldots\}$

$P(x) = P_1^{n_1} P_2^{n_2} P_3^{n_3} \ldots P_k^{n_k}$ $\quad n_i$ = number of event i

$P(x)_{typical} = P_1^{NP_1} P_2^{NP_2} P_3^{NP_3} \ldots P_k^{NP_k}$

$\log P(x)_{typ} = N(P_1 \log P_1 + P_2 \log P_2 + \ldots + P_k \log P_k)$

$= -N H(x) = -NH(x)$

OR $\quad P(x)_{typ} = 2^{-NH}$ for the typical values.

Soooo.. define the typical set to be near this

$$2^{-N(H+\beta)} < P(x) < 2^{-N(H-\beta)}$$

constant leeway

"Typical set"

$$T_{N\beta} = \left\{ x \in A_x^N : \left| \frac{1}{N} \log_2 \frac{1}{P(x)} - H \right| < \beta \right\}$$

$$= \left\{ x \in A_x^N : \left(\frac{1}{N} \log_2 \frac{1}{P(x)} - H \right)^2 < \beta^2 \right\}$$

Now read Mackey 79, 80

(up to Asymptotic Equipartition)

Reminder : Chebyshev's Inequality

$$P\left((x-\bar{x})^2 \geq \alpha\right) \leq \frac{\sigma_x^2}{\alpha}$$

x is a random variable.

Reminder : Weak Law Large Numbers

x goes from a random variable to the average of N random variables, $\{h_1, h_2 \ldots h_N\}$.

$$\Rightarrow \quad P\left((x-\bar{h})^2 \geq \alpha\right) \leq \frac{\sigma_{\bar{h}}^2}{\alpha N}$$

$$T_{N\beta} = \left\{ x \in A_x^N : \left[\frac{1}{N} \log_2 \frac{1}{P(x)} - H \right]^2 < \beta^2 \right\}$$

what is $P(x \notin T_{N\beta})$?

$$= P\left(\left(\frac{1}{N} \log \frac{1}{P(x)} - H \right)^2 \geq \beta^2 \right)$$

$$= \frac{\sigma^2}{\beta^2 N} \quad \text{by WLLN}$$

$$\therefore \quad P(x \in T_{N\beta}) = 1 - \frac{\sigma^2}{\beta^2 N}$$

Const.

Block size.

Const

So as $N \to \infty$ the probability is almost entirely concentrated in $T_{N\beta}$

Now read Mackey 80-82
up to Part I

Section 3

We can compress down to $H(X)$

Goal How much can we compress?

$$\frac{H_\delta(x^N)}{N} \to H(x) \qquad N \to \infty$$

i) S_δ $\quad P(x \in S_\delta) \geq 1-\delta$

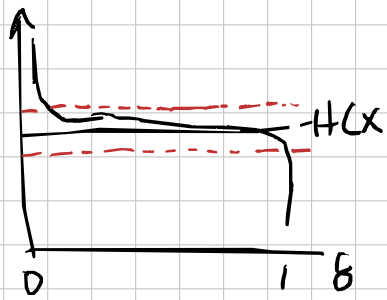
$H_\delta = \log_2 |S_\delta|$

} Not related to $H(x)$

ii) Typical set $\quad P(x \in T_{N\beta}) \geq 1 - \frac{\sigma^2}{\beta^2 N}$

For $x \in T_{N\beta}$ $\quad 2^{-N(H+\beta)} < P(x) < 2^{-N(H-\beta)}$

} Not related to H_δ



$T_{N\beta}$ might <u>not</u> be the best for compression but if we show it is good, the real optimal set can only be better.

We know $|T_{N\beta}| \times 2^{-N(H+\beta)} < 1$

Everything in $T_{N\beta}$ has this prob. or greater

Sum of prob. must not exceed 1

So $|T_{N\beta}| < 2^{N(H+\beta)}$

This needs $\log |T_{N\beta}| = N(H+\beta)$ bits

β can be arbitrarily small $\therefore$ we can compress to arbitrarily close to $H(X)$ per symbol.

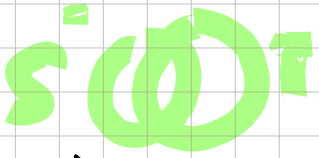
Now read Mackey §2

Section 4

But you can't go below $H(X)$

Imagine a set S' that is used to compress instead of $T_{N\beta}$, and $|S'| < N^H$.

E.g. $|S'| \leq 2^{N(H-2\beta)}$



S' contains prob. $P(S') = P(S' \cap T_{N\beta}) + P(S' \cap \overline{T_{N\beta}})$

$P(S' \cap T_{N\beta})$: Bits of S' in $T_{N\beta}$ (Typical)

$P(S' \cap \overline{T_{N\beta}})$: Rest of $T_{N\beta}$ (Atypical)

$|S'|_{max} = 2^{N(H-2\beta)}$

Max prob. of any $T_{N\beta}$ sequence is $2^{-N(H-\beta)}$

$\therefore P(S') < 2^{N(H-2\beta)} \cdot 2^{-N(H-\beta)}$

$< 2^{-N\beta}$

$\to 0$ as $N \to \infty$

Negligible since we showed Prob of things outside $T_{N\beta}$ is negligible

$$P(S') = (\text{tiny number}) + (\text{tiny number})$$

$\Rightarrow$ Any set significantly smaller than 2^{nH} will fail to capture the data

$\Rightarrow$ Entropy $H(x)$ is the hard limit of lossless compression

Now read Mackey 83

Source Coding Theorem Restated

N i.i.d. random variables, each with entropy $H(X)$ can be compressed into more than $NH(X)$ bits with negligible risk of information loss as $N \to \infty$! Conversely if they are compressed into fewer than $NH(X)$ bits it is virtually certain that information will be lost.

(Mackay pg 81)