Mackay's bent coin lottery: solution

If $p = 0.1$, $N = 1000$ then this is ≈ 0.1

Suggested soln – buy all tix with $\leq \underbrace{N\Big(\overbrace{p}^{0.1} + 10\sqrt{\tfrac{p(1-p)}{N}}\Big)}$ ones

$\Rightarrow$ Guaranteed we win with prob ≥ 0.99

How many tix did we buy? Let $n = N\left(p + 10\sqrt{\frac{p(1-p)}{N}}\right)$

$$= \sum_{i=0}^{n} 2^{N H_2\left(\frac{i}{N}\right)} \approx 2^{N H_2\left(\frac{n}{N}\right)}$$

$H_2(0.2)$

$$= 2^{N H_2\left(p + \frac{10\sqrt{p(1-p)}}{\sqrt{N}}\right)}$$

$$\approx 2^{N H_2(p)} \quad (\text{as } N \to \infty)$$

(last term in summation $\gg$ sum of all other terms)

(lossy) source coding theorem for binary sources

given $X^N = (X_1 X_2 \dots X_N)$, where each $X_i \sim$Bernoulli(p)

δ-lossy compression

$L = \phi(X^N)$ defined only for $X^N \in \mathcal{S}_\delta$ s.t. $\mathbb{P}[\mathcal{S}_\delta] \geq 1 - \delta$

Good tix (typical seq)

00...00

00...01 '1'

$\left(\leq \eta \text{ '1's} \right)$

01...0 η '1s'

Bad tix

$\left(> \eta \text{ 1s} \right) \leftarrow 2^{N(1 - H_2(p))}$

but $\mathbb{P}[\text{bad tix}] \leq \delta$

$\leftarrow \approx 2^{N H_2\left(p + \frac{\sigma}{\sqrt{\delta N}}\right)} \Rightarrow N H_2\left(p + \frac{\sigma}{\sqrt{\delta N}}\right)$ bits

$\mathbb{P}[\text{good tix}] \geq 1 - \delta$

(lossy) source coding theorem for binary sources

given $X^N = (X_1 X_2 \dots X_N)$, where each $X_i \sim$Bernoulli(p)

δ-lossy compression

$L = \phi(X^N)$ defined only for $X^N \in \mathcal{S}_\delta$ s.t. $\mathbb{P}[\mathcal{S}_\delta] \geq 1 - \delta$

- δ-sufficient subset S_δ: smallest subset of $\{0,1\}^N$ s.t. $\mathbb{P}[\mathcal{S}_\delta] \geq 1 - \delta$
- essential information content in X^N: $H_\delta(X^N) \triangleq \log_2 |S_\delta|$

(lossy) source coding theorem for binary sources

given $X^N = (X_1 X_2 \dots X_N)$, where each $X_i \sim$Bernoulli(p)

δ-lossy compression

$L = \phi(X^N)$ defined only for $X^N \in \mathcal{S}_\delta$ s.t. $\mathbb{P}[\mathcal{S}_\delta] \geq 1 - \delta$

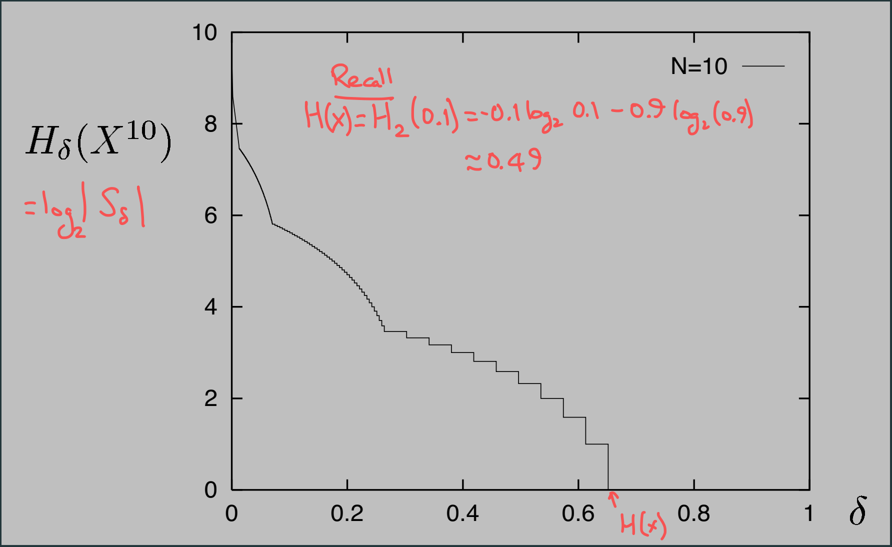
- δ-sufficient subset S_δ: smallest subset of $\{0,1\}^N$ s.t. $\mathbb{P}[\mathcal{S}_\delta] \geq 1 - \delta$
- essential information content in X^N: $H_\delta(X^N) \triangleq \log_2 |S_\delta|$ $\approx N H_2\left(p + \frac{k\sigma}{\sqrt{N}}\right)$

Shannon's source coding theorem (lossy version)

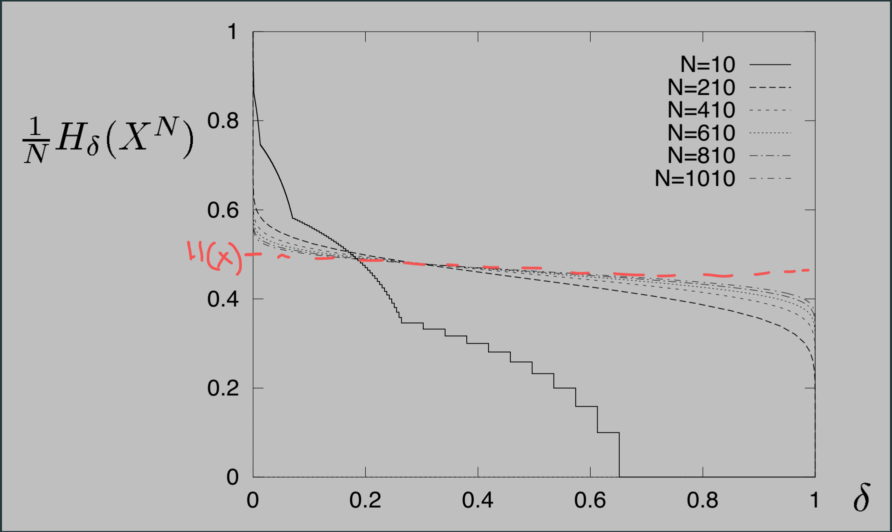
if X has entropy $H(X)$, then for any $\epsilon > 0$ and $0 < \delta < 1$, there exists N_0 s.t. for all $N > N_0$, we have

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

(lossy) source coding for binary sources: intuition



(lossy) source coding for binary sources: intuition



Next

- lossy $\rightarrow$ lossless
- 'practical' codes
  - symbol codes (Huffman code)
  - stream codes 'not practical'
- lower bounds

lossless source coding

from lossy to lossless compression

given $X^N = \underbrace{(X_1 X_2 \dots X_N)}_{D}$, where each $X_i \sim$Bernoulli(p)

$S_\delta = \{D_1, D_2, \dots, D_{|S_\delta|}\}, \quad |S_\delta| \approx 2^{NH(x)}$

Lossy scheme : - use $\log_2 |S_\delta| = N H(x)$ bits to encode each $D_i \in S_\delta$

- If $D_i \notin S_\delta$, send arbitrary symbol

wp $\geq 1-\delta$, can decode D using $NH(x)$ bits

Lossless scheme - use 'prefix' 0 to indicate $D \in S_\delta$, 1 to indicate $D \notin S_\delta$

- code : $\hat{L}(D) = \begin{cases} 0\,\overbrace{L}^{NH(x)\text{ bits}} & \text{if } D \in S_\delta \\ 1\,\underbrace{D}_{N\text{ bits}} & \text{if } D \notin S_\delta \end{cases}$

from lossy to lossless compression

given $X^N = (X_1 X_2 \dots X_N)$, where each $X_i \sim$Bernoulli(p)

Shannon's source coding theorem

if X has entropy $H(X)$, then for any $\epsilon > 0$ and $0 < \delta < 1$, there exists N_0 s.t. for all $N > N_0$, we have a lossless code $L = \phi(X^N)$ s.t.

$$\left| \frac{\mathbb{E}[L]}{N} - H(X) \right| \leq \epsilon$$

From last slide

$$\mathbb{E}[\hat{L}(D)] = (N H_\delta(X)) \cdot \overbrace{(1-\delta)}^{\mathbb{P}[D \in S_\delta]} + N \cdot \overbrace{(\delta)}^{\mathbb{P}[D \notin S_\delta]}$$

$$= N H_\delta(X) + \underbrace{N\delta(1 - H_\delta(X))}_{\searrow 0 \text{ as } \delta \searrow 0}$$

$$\approx N H(X)$$

(from lossy coding thm)

lossless compression via typical set encoding

typical set

$\mathcal{X} = \{a, b, \dots z\}$

iid source produces $X^N = (X_1 X_2 \dots X_n)$; each $X_i \in \mathcal{X}$ has entropy $H(X)$

then X^N is very likely to be one of $\approx 2^{NH(X)}$ typical strings, (typicality)

all of which have probability $\approx 2^{NH(X)}$ (asymptotic equipartition)

$X \longrightarrow X^N$

a • p_a

b • p_b

• p_c

⋮

z o p_z

Prob = 99%

Typical sequence

All equally likely

Size = $2^{NH(x)}$

Atypical sequence

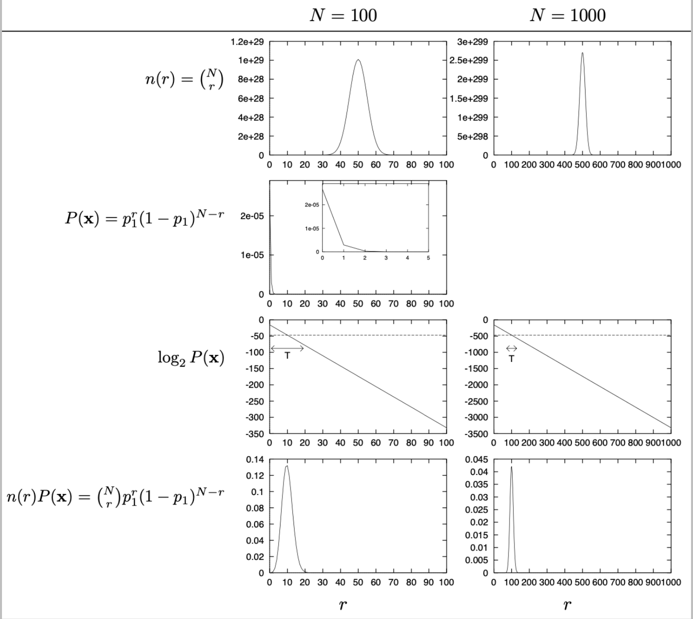
Prob = 1%

Size $2^{N(\log_2|x| - H(x))}$

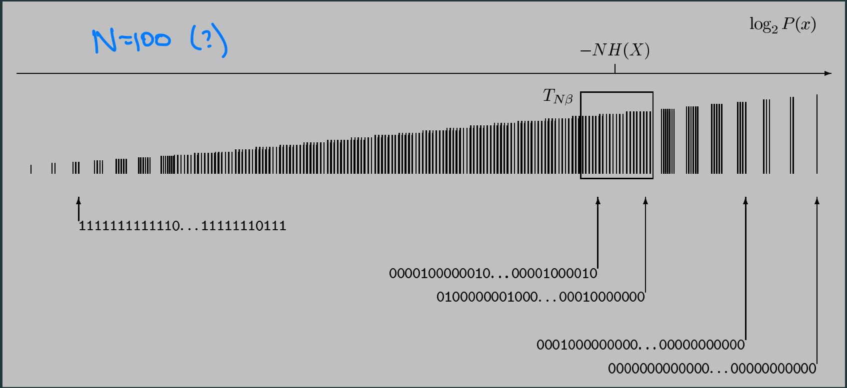
(since equally likely)

$\Rightarrow H(X^N) = NH(X)$

visualizing the typical set



visualizing 'asymptotic equipartition'



practical source coding solutions

symbol codes

$$X_1X_2\dots X_n \quad \rightarrow \quad \phi(X_1)\phi(X_2)\dots\phi(X_n)$$

Each letter gets its own codeword

stream codes

$$X_1X_2\dots X_n \quad \rightarrow \quad \phi(X_1)\phi(X_2|X_1)\phi(X_3|X_1X_2)\dots\phi(X_n|X_1X_2\dots X_{n-1})$$

Entire dataset gets its own codeword

Eg - LZW, arithmetic codes

Lempel-Ziv-Welch

Reading Assignment

symbol codes (Mackay chapter 5)

symbol codes

expected length of symbol code

let $X \sim \{p(x)\}_{x \in \mathcal{X}}$, and consider code $C(\cdot)$, and let $\ell(x) = |C(x)|$

the expected length of C is $\mathbb{E}[L(C, X)] = \sum_x p(x)\ell(x)$

what we want from symbol code C:

- unique decodability: $\forall\, x_1 x_2 \ldots x_n \neq y_1 y_2 \ldots y_n$, we have $C(x_1)C(x_2)\ldots C(x_n) \neq C(y_1)C(y_2)\ldots C(y_n)$ $\Leftrightarrow$ lossless
- easy to decode
- small $\mathbb{E}[L(C, X)]$

types of symbol codes

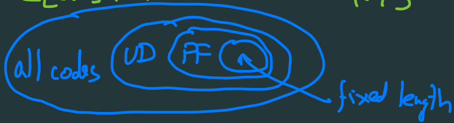
$H(x) = 1.75$

consider source producing $X \sim \{a, b, c, d\}$ with prob $\{\overset{1}{\frac{1}{2}}, \overset{2}{\frac{1}{4}}, \overset{3}{\frac{1}{8}}, \overset{3}{\frac{1}{8}}\}$

Symbols	info content ($h(x)$)	Code1 (one-hot encoding)	$l(x)$	Code2 (binary encoding)	$l(x)$
a	1	1000	4	00	2
b	2	0100	4	01	2
c	3	0010	4	10	2
d	3	0001	4	11	2
	$H(x) = 1.75$		$E[l(x)] = 4$		2

prefix codes (variable length, greedy encoding/decoding)

Symbol	info content $h(x)$	Code 3 (prefix-free)	$l(x)$	Code 4 (uniquely decodable)	$l(x)$
a	1	0	1	0	1
b	2	10	2	01	2
c	3	110	3	011	3
d	3	111	3	111	3
	$H(x)=1.75$		$E[l(x)]=1.75$		1.75



the limits of unique decodability

Kraft-McMillan inequality (conservation laws)

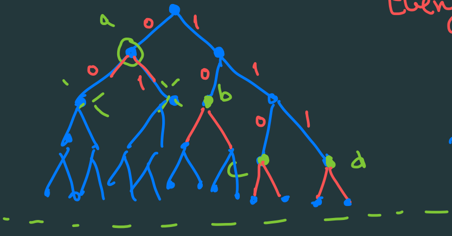
for any $C \equiv$ uniquely decodable binary code over $\mathcal{X}$, with $\ell(x) = |C(x)|$

$$\sum_{x \in \mathcal{X}} \underbrace{2^{-\ell(x)}}_{\text{fraction of leaf nodes in partition } x} \leq 1$$

moreover, for any $\{\ell(x)\}$ satisfying this, we can find a prefix code

- (Kraft's Ineq) Special case for prefix-free

Every prefix free code $\equiv$ subset of nodes in a binary tree



L_{max}

$2^{L_{max}}$ 'leaf node' partitioned among $|\mathcal{X}|$ symbols