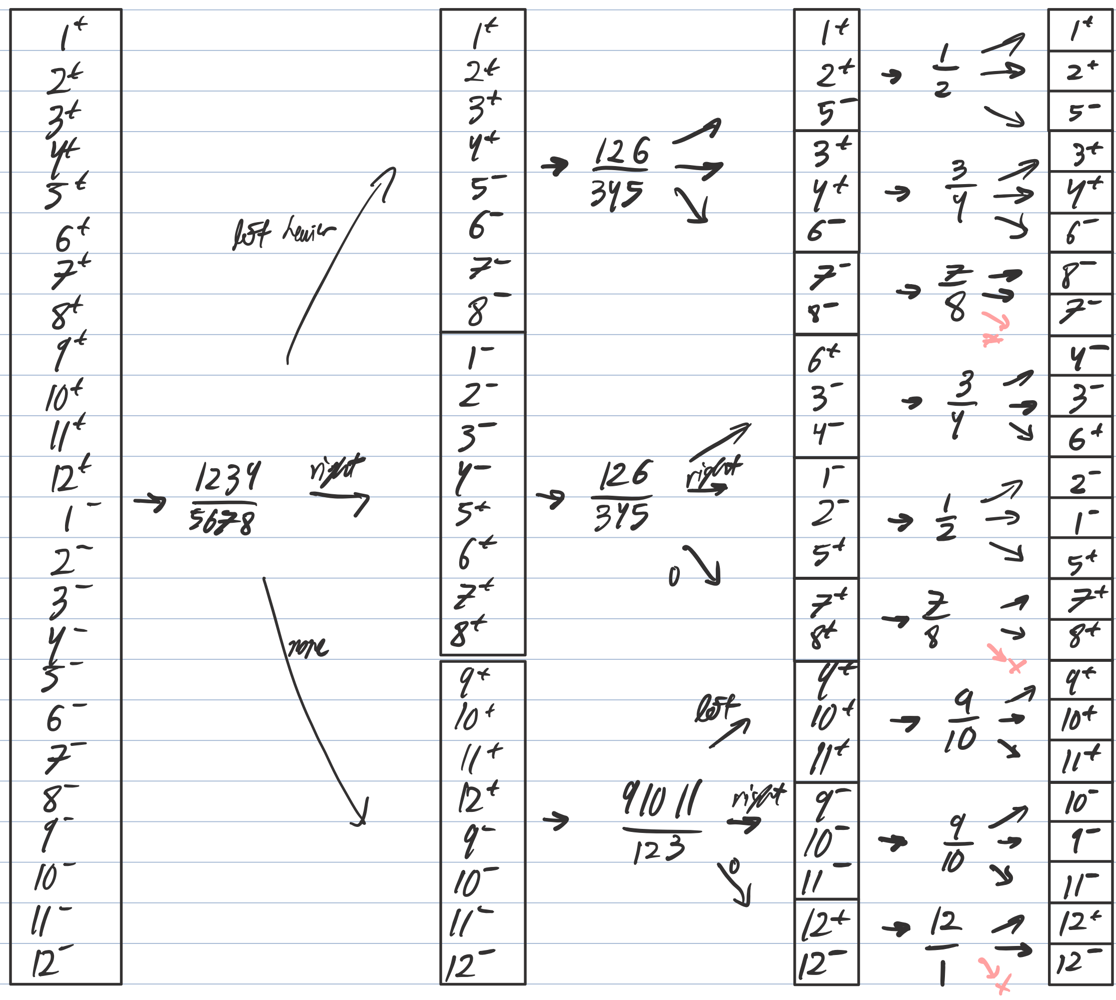
Chapter 4

4.1 24 hypotheses $\log_2 3$ bits each time

→ 3x reduction

→ at least 3 weighs



4.2 $H(X,Y) = \sum p(x)p(x) \log p(x) + \log p(y)$

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

when $X \perp Y$

4.3 63 needs 6 qs:

$x \geq 32$?

$x \% 32 \geq 16$?

$\vdots$

$x \% 2 = 1$?

4.4 Reduce by a factor of 7/8 since ASCII doesn't use a byte

4.5 Cannot compress all x uniquely to codes of length $H_0(X)$ because then $|A_X| \leq 2^{\ell} < 2^{H_0} = |A_X|$ $\Leftarrow$

4.6 For $\delta = 1/16$ only compress a, b, c, d

S_δ smallest $\{x$ s.t. $\mathbb{P}(x \notin S_\delta) \leq \delta\}$

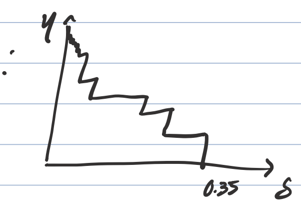
$P(x \in S_\delta) \geq 1-\delta$

$H_\delta = \log S_\delta$

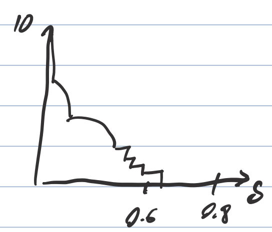
4.7 $\underline{x}$ n flips w $p_0 = 0.9$ $p_1 = 0.1$

$P(\underline{x}) = p_0^{N-r(x)} p_1^{r(x)}$, $r(x) := \#1$s in x

$H_\delta(X^4)$:



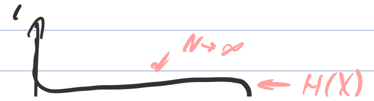
$H_\delta(X^{10})$



Source coding theorem

If we allow even a little error, can compress down to H(X). Regardless of how much we allow, can't do better than H(X)

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



$\forall \epsilon > 0 \; \exists \delta \; \left| \frac{1}{N} H_\delta(X^N) - H \right| < \epsilon$

4.8 Changes in P are equal between cusps
$\Rightarrow$ # elements in H_δ scales linearly with $\log(-\delta)$

Typicality: $r \sim Np_i \pm \sqrt{Np_i(1-p_i)}$

Alphabet of I letters w/ probabilities p_i

$\Rightarrow P(\underline{x})_{typ} = P(x_1) \cdots P(x_N) = p_1^{p_1 N} \cdots p_I^{p_I N}$

$\log_2 \frac{1}{P(\underline{x})_{typ}} \approx N \cdot \sum p_i \log \frac{1}{p_i} = NH(X)$

Typical elements $\underline{x}$ have $P(\underline{x}) \simeq 2^{-NH}$

$T_{N,\beta} : \left\{ \underline{x} \cdot \left| \frac{1}{N} \log \frac{1}{P(\underline{x})} - H \right| < \beta \right\}$

At any fixed β, T_N contains almost all prob as $N \to \infty$

Asymptotic Equipartition:

$X^N = \{\underline{x}\}$ as $N \to \infty$ $\underline{x} \in A_N$ of size 2^{NH_x}
with almost certain probability

Each elem of A_N has $p(\underline{x})$ "close to" 2^{-NH_x}

$H(X) < H_0(X) \Rightarrow 2^{NH(X)} \ll 2^{NH_0(X)}$

Equivalent to source coding

(consider only compressing the 2^{NH} bits in the typical set)

Proofs: Lemma (Chebyshev)

For $t > 0$

$$P(t \geq \alpha) \leq \frac{\bar{t}}{\alpha}$$

Pf: $\sum_{t \geq \alpha} P(t) \leq \frac{1}{\alpha} \sum_{t \geq \alpha} P(t)\, t$

$$\leq \frac{\bar{t}}{\alpha}$$

$\Rightarrow$ Chebyschev 2:

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

Weak LLN:

$x = \frac{1}{N} \sum h_i$

$$P[(x-\bar{h})^2 \geq \alpha] \leq \frac{\sigma_x^2}{N\alpha}$$

Take $\frac{1}{N} \log \frac{1}{P(\underline{x})} = \frac{1}{N} \sum_n h_n$ $\qquad h_n = \log \frac{1}{p(x_n)}$

$\bar{h} = H(X)$

$\sigma = \text{var} \log \frac{1}{p(x_n)}$

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

$$\boxed{P(x) \in T_{N\beta} \geq 1 - \frac{\sigma^2}{\beta^2 N}}$$

pick $\delta = \frac{\sigma^2}{\beta^2 N}$

Next relate $T_{N\beta}$ to $H_\delta(X^N)$

I: $\frac{1}{N} H_\delta(X^N) < H + \epsilon$

The size $T_{N\beta}$ gives upper bound on H_δ since $T_{N\beta}$ is not optimized to minimize size

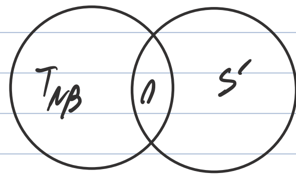
$|T_N| < 2^{N(H+\beta)}$

set $\beta = \epsilon \Rightarrow \delta = \frac{\sigma^2}{\epsilon^2 N} \Rightarrow P(T_{N\beta}) \geq 1-\delta$

$H_\delta(X^N) \leq \log T_{N\beta} \leq N(H+\epsilon)$

II: $\frac{1}{N} H_\delta(X^N) > H - \epsilon$

Assume otherwise. Set $\beta = \epsilon/2$ $\quad S'$ s.t. $|S'| < 2^{N(H-2\beta)}$



$$P(x \in S') = \underbrace{P(x \in S' \cap T_{N\beta})}_{\le 2^{N(H-2\beta)} 2^{-N(H-\beta)}} + \underbrace{P(x \in S' \cap \bar{T}_{N\beta})}_{\le \frac{\sigma^2}{\beta^2 N}}$$

$$\le 2^{-N\beta} + \frac{\sigma^2}{\beta^2 N}$$

set $\beta = \epsilon/2 \Rightarrow P(x \in S') < 1 - \delta \Leftarrow$

$\Rightarrow$ Any subset w/ size $|S| < 2^{N(H-\epsilon)}$ has prob $< 1 - \delta$

$\Rightarrow H_\delta(X^N) > N(H - \epsilon)$

$\Rightarrow \frac{1}{N} H_\delta(X^N)$ concentrates to H

$\log \frac{1}{P(x)}$ are within stdev of $2N\beta$ of each other

as $\beta \to 0$ need N to grow as $\frac{1}{\beta^2}$

to keep $\delta = \frac{\sigma^2}{\beta^2 N}$ fixed

$\Rightarrow \beta \sim \frac{\alpha}{\sqrt{N}}$

$\Rightarrow$ Most probable will be $\sim 2^{2\alpha\sqrt{N}} \times$ the least probable

$\Rightarrow$ equipartition in a weak sense

4.9 Not informative about odd one out, but informative about odd is light/left or heavy/right etc

4.10 $3^4 = 81 > 39 \Rightarrow 4$ weightings

4.11 2 bits of info at each time

4.12 1 3 9 27 $\Rightarrow$ 4 in total

4.13 12 balls labelled by

AAB ABA ABB ABC
BBC BCA BCB BCC
CAA CAB CAC CCA

pan A pan B

Weightings: 1. A * * vs B * *

2. * A * vs * B *

3. * * A vs * * B

Each weighting gives A B C
for each pan above below canonical

$\Rightarrow$ 2 sequences of 3 letters

both CCC $\Rightarrow$ No odd
otherwise for just one of the two pans the sequence is one of the above & names the pan which is Above or Below depending if its in pan A or B

4.14 $4 \cdot \binom{12}{2} = 66 \cdot 4 = 264 \Rightarrow \lceil \log_3 264 \rceil = 6$

$8 \cdot \binom{12}{3} = 8 \cdot 220 = 1760 \Rightarrow \lceil \log_3 1760 \rceil = 7$

this answers b)

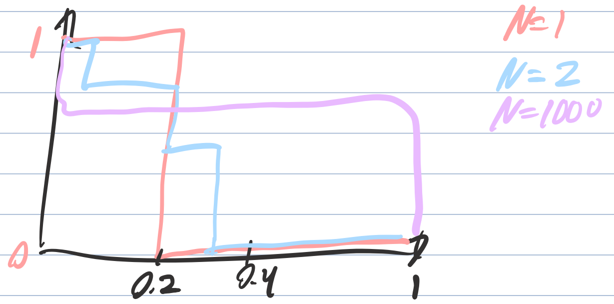
But if there was a ranking of oddness
$\Rightarrow$ extra factor of 3 for 2 balls $\Rightarrow$ 7 weighings

extra factor of $3! + 3 \cdot 2 + 1 \Rightarrow 13$

$\Rightarrow 13 \cdot 1760 \Rightarrow$ 10 weighings

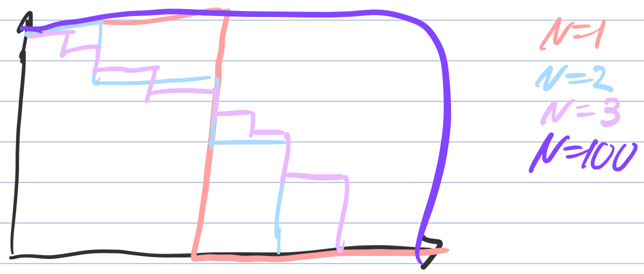
4.15

0.8, 0.2



4.16

0.5, 0.5



4.17 Boltzmann entropy only exists for microcanonical:

$$S_{Boltz} = k_B \log \Omega$$

Gibbs entropy is a Shannon entropy of ensemble:

$$S_{Gibbs} = k_B \sum_i p_i \log \frac{1}{p_i}$$

For $P(x) = \frac{1}{Z} \exp[-\beta E(x)]$

Fixing $E = E_0 \pm \epsilon \Rightarrow P(x) = \frac{e^{-\beta(E \pm \epsilon)}}{Z}$

$\to$ If $E(\vec{x})$ separates into $\sum E(x_i)$ then $S_{micro} \approx S_{Gibbs}$ as $N \to \infty$

"self-averaging"

4.18 $\frac{1}{Z} \frac{1}{x^2+1} \Rightarrow Z = \pi$

Mean & var are undefined

$z = x_1 + x_2$

$$\Rightarrow P(z) = \frac{1}{\pi^2} \int dx \, \frac{1}{x^2+1} \frac{1}{(z-x)^2+1} = \frac{2}{\pi} \frac{1}{4+z^2}$$

$\Rightarrow \frac{x_1+x_2}{2}$ has cauchy dist w/ same width

Alternatively $\hat{P}(\omega) = e^{-|\omega|} \Rightarrow \hat{P}_{\frac{x_1+x_2}{2}} = \sqrt{e^{-2|\omega|}} = e^{-|\omega|}$

4.19 Let $t = \exp(sx)$

$$P(x \geq a) = P(t \geq e^{sa}) \leq \frac{\bar{t}}{e^{sa}} = \frac{\sum P(x) e^{sx}}{e^{sa}} = e^{-sa} g(s) \quad \forall s > 0$$

$t = \exp(sx)$ for $s < 0 \Rightarrow P(x \leq a) = P(t \geq e^{sa})$

4.20 $y = x^{x^{\cdot^{\cdot^{\cdot}}}}$
$= x^y$

$\Rightarrow \log y = y \log x$

$\Rightarrow \frac{1}{y} \log y = \log x$

take $y = 1/p$

Chapter 5

5.1 $\{0,1\}^3 = \{000, \ldots, 111\}$

5.2 $\{0,1\}^+ = \{0, 1, 00, 01, 10, 11, \ldots\}$

5.3 acdbac

a	1	0	0	0
b	0	1	0	0
c	0	0	1	0
d	0	0	0	1

C_0

1000 0010 0001 ...

long!

5.4 $\{0, 101\}$ Is a prefix code C_1

5.5 $\{1, 010\}$ is not C_2

5.6 $\{0, 10, 110, 111\}$ is C_3

5.7 $\{00, 01, 10, 11\}$ is C_4

5.8 C_2 is uniquely decodable nontheless

5.9 yes — $\{1, 101\}$ is uniquely decodable but not prefix

5.10 $A_X: \{a, b, c, d\}$ $\Rightarrow H_X = 1.75$ bits
$P_X: \frac{1}{2}\ \frac{1}{4}\ \frac{1}{8}\ \frac{1}{8}$

$L(C_3, X) = 1.75$ bits

$l_i = \log_2(1/p_i)$ for C_3

5.11 $L(C_4, X)$ is 2 bits

5.12 $C_5: \{0, 1, 00, 11\}$ has $L(C_5, X) = 1.25$
but C_5 not uniquely decodable

5.13 $C_6: \{0\ \ 01\ \ 011\ \ 111\}$ Not prefix — but uniquely decodable
reverse of C_3

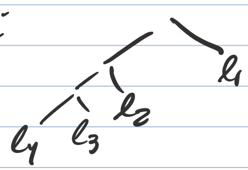
Proof of Kraft

$$z^N = \left(\sum_i 2^{-l_i}\right)^N = \sum_{i_1 \cdots i_N} 2^{-l_{i_1} - \cdots - l_{i_N}} = \sum_{l = N \cdot l_{min}}^{N \cdot l_{max}} 2^{-l} A_l$$ ← # length l words

$$\leq N \cdot l_{max}$$

$\Rightarrow z \leq 1$

5.14 Given l_i satisfying Kraft, can build a tree

eg:  $\Rightarrow$ prefix code

☆ $H(X)$ is lower bound on $L(C, X)$

$q_i = \frac{2^{-l_i}}{z}$

$$\sum_i p_i l_i = \sum_i p_i \log 1/q_i - \log z$$

$\Rightarrow l_i = -\log zq_i$

$= \sum_i p_i \log 1/p_i - \log z$

$\geq H(X)$

Equality iff $l_i = \log_2 1/p_i$

$\Rightarrow l_i$ implicity defines $q_i = \frac{2^{-l_i}}{z}$

☆ $H(X) \leq L(C,X) < H(X)+1$

Set $l_i = \lceil \log 1/p_i \rceil$

$\Rightarrow 2^{-l_i} \leq 1$

$\Rightarrow L(C,X) = \sum_i p_i \lceil \log (1/p_i) \rceil < H(X)+1$

☆ $L(C,X) = H(X) + D_{KL}(p \| q)$

Top-down coding achieves $L(C,X) \leq H(X)+2$ bad!

Huffman: Priority queue: Take two largest probs
append 0,1 to them resp
→ merge and put back

5.16 No better symbol code

By contradiction: take a,b w/ smallest probs
$\Rightarrow$ equal length by Huffman

Assume there is a better (WLOG prefix) code
with $l_a < l_b$

Prefix code never has solo max length leaf so

$\exists$ node c w $p_c > p_a$ $l_c \geq l_b$

Swap a, c $\Rightarrow$ expected length decreases $\Leftarrow$

By contracting the tree over & over you arrive at Huffman

5.17 Can make Huffman out of English

$L \sim 4.15$

$H \sim 4.11$ disparities between l & p_i

(both above & below)

5.18

$A_X = \{a, b, c, d, e, f, g\}$

$P_X = \{.01, .24, .05, .2, .47, .01, .02\}$

Top down .5 .5 perfectly balanced

.25 .25 .47 .03 } not ideal

.01 .24 .03 .2 .01 .02

000 001 010 011 10 110 111

⇒ 2.53 bits

Huffman gives 1.97

Huffman is optimal for an ensemble but

a) p_i can change

b) 1 bit overhead is severe

5.19 No, 11 & 111

5.20 Yes, ternary prefix

5.21 X^2 00 01 10 11 ⇒ $L \sim 1.29$

1 01 000 001 $H \sim 0.938$

X^3 000 001 010 100 011 110 101 111 ⇒ 1.598

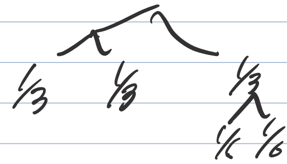
1 011 010 001 00000 00001 00010 00011 1.407

X^4 1 3 3 3 4 6 7 7 7 7 7 9 9 9 10 10

1.9702 1.876

5.22 $\{1/6 \quad 1/6 \quad 1/3 \quad 1/3\}$

$\{1/5 \quad 1/5 \quad 1/5 \quad 2/5\}$



5.23 $p_1 = p_3 + p_4$ (if $p_3 + p_4 = p_2$ then $p_1 = p_2 = 1/3$)

$q^1 = \{1/3 \quad 1/3 \quad 1/6 \quad 1/6\}$

$q^2 = \{2/5 \quad 1/5 \quad 1/5 \quad 1/5\}$

$q^3 = \{1/3 \quad 1/3 \quad 1/3, \ 0\}$

convex hull

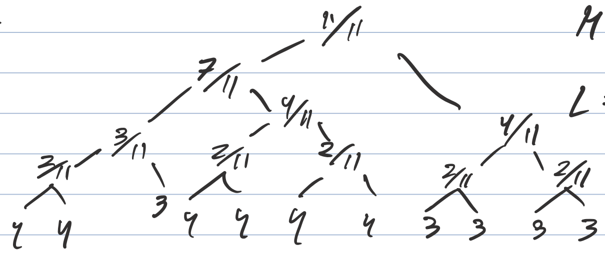
b.c. edges happen from turning 2/3 ineqs to eqs

5.24 Should ask q5 w/ 50% prob

5.25 The lower bound is satisfied w equality

5.26 See next exercise

5.27



$H = \log_2 11 = 3.45$

$L = \frac{5}{11} \cdot 3 + \frac{6}{11} \cdot 4$
$= 3.54$

$\Rightarrow H - L \sim 0.086$

5.28 All symbols $1/I$ $\quad I \neq 2^n$

ξ^+ points assigned value $\lceil \log I \rceil =: l^+$

WLOG $I \nmid 2$

n_- nodes $< 2^{l_-}$ at length l_- $\quad 2(2^{\lfloor \log I \rfloor} - n_-) = n_+$ nodes for l_+

$n_- + n_+ = I \quad \Rightarrow 2(2^{\lfloor \log I \rfloor} - n_-) = I - n_- \Rightarrow 2^{\lceil \log I \rceil} = I + n_-$

minimize $n_+ = I - n_- \Rightarrow$ maximize n_- $\quad \Rightarrow 2^{l_+} = 2I - n_+$

$$\Rightarrow \xi_+ = \frac{n_+}{I} = 2 - \frac{2^{l_+}}{I}$$

$$\Rightarrow L = l^+ - 1 + \xi^+ = l^+ + 1 - \frac{2^{l_+}}{I}$$

$$\frac{\partial}{\partial I} L - H = \frac{2^{\lceil \log I \rceil}}{I^2} - \frac{1}{I \ln 2} = 0 \Rightarrow 2^{\lceil \log I \rceil} \log 2 = I$$

$\lceil \log I \rceil$ N

$\Rightarrow I = 2^{(\log \ldots)} \ln 2 \Rightarrow I \approx 2^{\cdots} \ln 2$

$\Rightarrow l_t = N$

$$N+1 - \frac{2^N}{2^N \ln 2} - \log_2 2^N \ln 2$$

$$= 1 - \frac{1}{\ln 2} - \frac{\ln \ln 2}{\ln 2} \approx 0.086 \quad \checkmark$$

5.29 $N=1$ Huffman gives $L=1$ but $H(X) \approx 0$ $\therefore$ need $N \neq 1$

Need $P[\underbrace{0 \cdots 0}_{N}] \sim 1/2$ for efficient code

$$\Rightarrow N \approx \frac{\log 1/2}{\log .99} \approx 69$$

$\Rightarrow 2^{69}$ entries in the tree

$\approx 5E20$ entries

pretty expensive ~ 100 exabytes

5.30 129 (of the form 2^n+1 between 100 & 200)

Best strategy is huffman tree

$\Rightarrow$ need 7 tests w/ a $2/129$ chance of 8

$\Rightarrow 7 + \frac{2}{129}$

Prof needs $8 \cdot \frac{128}{129} + 7 \cdot \frac{1}{129} = 7 + \frac{122}{129}$

5.31 Wrong way: pick symbol w/ prob p_i & pick random bit from $c(x_i)$

	a_i	$c(a_i)$	p_i	l_i
$C_3 =$	a	0	1/2	1
	b	10	1/4	2
	c	110	1/8	3
	d	111	1/8	3

$$\sum p_i \cdot f_i$$

$$= \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{8} \cdot \frac{2}{3} + \frac{1}{8}$$

$$\sim 1/3$$

Really: $\underline{\sum p_i f_i l_i} = 1/2$

$\sum p_i l_i$ ↖ always the case for symbol codes

Another way: Since $c(x)$ is optimally compressed
if $E \neq 1 \neq 1/2$ we could compress further,
violating the $H(X)$ lower bound

5.32 Same, but now use q-ary tree, build up by picking q least common

5.33 Metacode is incomplete

$x \Rightarrow l_k(x)$ under C_k

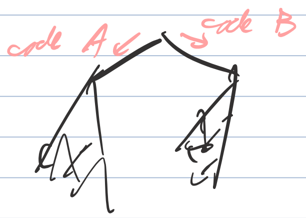
$$l'(x) = \log K + \min_k l_k(x)$$

$$\Rightarrow z = \sum 2^{-l'} = \frac{1}{K} \sum 2^{-\min_k l_k(x)}$$

$$= \frac{1}{K} \sum_k \underbrace{\sum_{x \in A_k} 2^{-l_k(x)}}_{\leq 1}$$

equality only if all $x \in A_k$

$$\leq 1$$



Chapter 6

$P(a) = 0.425$ $P(b) = 0.425$ $P(\square) = 0.15$

$P(a|b) = 0.28$ $P(b|b) = 0.57$ $P(\square|b) = 0.15$

$P(a|bb) = 0.21$ $P(b|bb) = 0.64$ $P(\square|bb) = 0.15$

$P(a|bbb) = 0.17$ $\quad$ $P(b|bbb) = 0.68$ $\quad$ $P(\square|bbb) = 0.15$

$P(a|bbba) = 0.28$ $\quad$ $P(b|bbba) = 0.57$ $\quad$ $P(\square|bbba) = 0.15$

$b \Rightarrow P(\text{string}) \in [0.425, 0.85)$

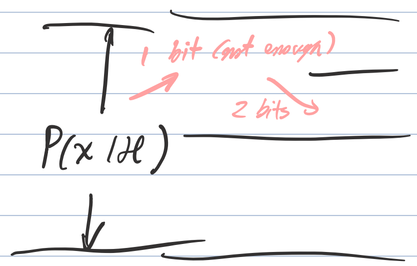
$\Rightarrow$ 01, 10, 11 are first 2

$bb \Rightarrow P(\text{string}) \in [0.544, 0.78)$

$\Rightarrow$ 10 11

$bbb \Rightarrow P(\text{string})$

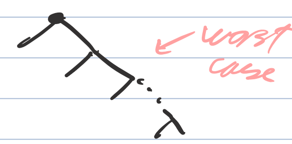
6.1



6.2 ASCII 128

Huffman starts by communicating 128 ints

l_i be as long as 127 or as short as 1



on average they are ~2-17

Say all must be <32 $\Rightarrow$ header of size $5 \times 128 = 640$ bits

Lets say ent/char ~ 4 if IID

For 400 chars ~ 2240

For shorter, header dominates

For Laplace, p_a start ~ 1/2 but this deviates after ~128

$\propto \text{cnt}(p_a)$

For Dirichlet need only ~2
$\alpha = 0.01$

If only a small fraction have high $p_a \Rightarrow$ Dirichlet if nearly uniform $\Rightarrow$ Laplace

If only 2/128 are used equiprobably
$\Rightarrow$ Huffman $H_N \approx \frac{3}{2} N$

0 gets $\ell=1$, 1 gets $\ell=2$

Arithmetic $\approx N$ $\Leftarrow$ appreciate!

If one char is disproportionate:

Huffman $\searrow$ $H=1$

Arithmetic < 1

6.3 1) 32 bits generated / 1 bit output

2) Needs only $H(0.01) \approx 0.081$ bits/symb

$\Rightarrow$ $81 + \underbrace{2}_{\text{term}} = 83$ bits for 1000

Fluctuations in # of 1s produce variation w/ $\sigma \approx 21$

6.4 Otherwise, at fixed length N we'd have a many-to-one issue.

6.5

	1	10	11	100	101	110	111
	1	2	3	4	5	6	7

0,00,000,0000,00010,000,00000,0000,00,

(,0) (1,0),(10,0),(11,0) ,(010,1), (100,0) , (110,0)

6.6

0,01,010,111,0110,0100,1000,1101,01010,00011

0 1 00 001 000 10 0010 101 0000 01

0	λ
1	0
10	1
11	00
100	001
101	000

110	10
111	0010
1000	101
1001	0000
1010	01

6.7 K ones $N-K$ zeros

$p(0) = \frac{N-K}{N}$ $p(1) = N$

$p(0|0) = \frac{N-K-1}{N-1}$ $p(1|0) = \frac{K}{N-1}$

$p(0|1) = \frac{N-K}{N-1}$ $p(1|1) = \frac{K-1}{N-1}$

$p(0|\cdots) = \frac{N-K+\#\text{ones}}{N-n}$

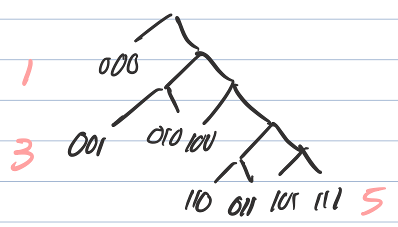
$p(1|\cdots) = \frac{K-\#\text{ones}}{N-n}$

3/5	0	0	0	1	1
			1	0	1
				1	0
		1	0	0	1
				1	0
			1	0	0
2/5	1	0	1	0	1
				1	0
			1	0	0
		1	0	0	0

6.8 $\lceil \log_2 \binom{N}{K} \rceil \approx N\, H_2(K/N)$

Binary string generated by arithmetic code above

6.9 Huffman



$\Rightarrow L \approx 1$

$H(X) \approx 0.24 = 3 \cdot 0.08$

Arithmetic code gives $N \cdot 0.08$

Variance is given by $Var(\#\ 1s) = N \cdot p \cdot (1-p)$
$\approx 0.01 \cdot N$

length is $\ell(r) = r \log\left(\frac{1}{f_1}\right) + (N-r)\log\left(\frac{1}{f_0}\right)$

# of 1s

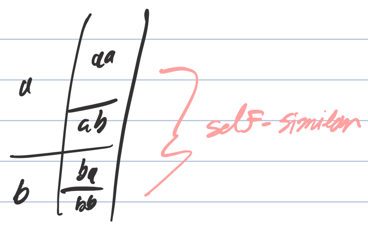
$= r \log \frac{f_0}{f_1} + N \log \frac{1}{f_0}$

$\Rightarrow$ len is $\pm\, 3.14 \cdot \log \frac{f_0}{f_1} \approx 21$ for $N=1000$

$\Rightarrow 80 \pm 21$ bits

$+2$

6.10 Input random bits into arithmetic encoder for sparse source



6.13 long-range correlations w/ intervening junk

2D images

intricate redundancy: (LaTeX file, Postscript)

Mandelbrot set

Ground state of Frustrated Ising model

SBIG would do well

Cellular automata

6.14 $\langle r^2 \rangle \approx N\sigma^2$

$\langle r^4 \rangle = \sum_i \langle (x_i)^4 \rangle + \sum_{i \neq j} \langle (x_i)^2 \rangle \langle (x_j)^2 \rangle$

$$= 3N\sigma^4 + N(N-1)\sigma^4$$

$$\Rightarrow \text{Var}\ r^2 = 2N\sigma^4$$

r is concentrated to be within $\frac{2}{\sqrt{N}}$ % of $\langle r \rangle$

6.15 entropy of P is 2.78

Huffman gives the unique answer w/ $L = 2.81$

6.16 $A_X = \{a, b, c\}$
$\{1/10, 3/10, 6/10\}$

$y = x_1 x_2$ $x_i \sim$ iid from A_X

$H(Y) = 2H(X) = 2 \cdot 1.295 = 2.59$

Huffman on y gives $L = 2.67$

6.17 $470 \pm \sqrt{1000 \cdot 0.1 \cdot 0.9}\ \log_2 \frac{f_0}{f_1}$

$= 470 \pm 30$

6.18 $R = S/L = \dfrac{\sum p_n \log 1/p_n}{\sum p_n l_n}$

$$\Rightarrow \frac{dS}{dp_n} L - \overset{l_n}{\frac{\partial L}{\partial p_n}} \underset{RL}{S} = \mu L^2$$

$$\Rightarrow \frac{dS}{dp_n} = \mu L + l_n \cdot R$$

$$\Rightarrow -1 - \log p_n = \mu L + l_n R \qquad \Rightarrow S = \log Z + RL$$

$$\Rightarrow p_n = \frac{1}{Z} \exp[-R l_n] \qquad \Rightarrow Z = 1$$

$l_n = n \Rightarrow p_n = 2^{-n}$ 1 bit/second

Finish

6.19 $\log_2 52! \approx 226$ bits

6.20

Chapter 7

7.1

a) 8-bit blocks $\Rightarrow$ base 255

$\Rightarrow \frac{1}{q+1}$ belief that current char is final one

$\Rightarrow$ IE # chars $= 256$

$\Rightarrow 256 \times 8$ bits ≈ 2000 bits

b) 100k bytes

$q \cdot \log q = 800k$

$\Rightarrow q \sim 2^{15}$ to 2^{16}

$\Rightarrow$ 16-bit blocks

7.2

$C_\alpha(n) = \underbrace{0 \cdots}_{\text{largest} \sim 100} 01$ (headless binary)

$c_\beta(n) = \underbrace{0000001}_{7}\ 100010$ (headless binary)

$c_\gamma(n) = \underbrace{001}_{3}\ \underbrace{11}_{7}\ 100010$ (headless binary)

$c_\delta(n) = \underbrace{01}_{2}\ \underbrace{1}_{3}\ \underbrace{11}_{7}\ 100010$

C_3 = · ˙ ⁻ · 11

C_7 = ·⁻ ⁻˙ 111

C_{15} = · ⌣ ⁻⌣ 1111 ⇐ shortest

7.3 Encode the # of levels of recursion that C_ω will need to go through

Then encoder uses $C_B(n)$ at each level instead of $C_b(n)$