Chapter 8

8.1 $H(X,Y) = H_u + H_v + H_w$

$H(X|Y) = H_u$

$I(X,Y) = H_v$

8.2 Eg X is mixture w $X|y=b_k$ more entropic than the average $X|y=b$

In all cases

$$H(X|Y) = \sum_{x,y} P(x,y) \log \frac{1}{p(x|y)} = \sum P(x,y) \log \frac{p(y)}{p(y|x)p(x)}$$

$$= H(X) + \sum_x p(x) \sum_y p(y|x) \log \frac{p(y)}{p(y|x)}$$

$$= H(X) - \underbrace{\sum_x p(x)\, D_{KL}(p(y|x) \| p(y))}_{I(X;Y)} \leq H(X)$$

= only if $X \perp Y$

8.3 $$H(X,Y) = \sum_{x,y} P(x) P(y|x) \left[\log \frac{1}{p(x)} + \log \frac{1}{p(y|x)}\right]$$

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

8.4 $I(X;Y) := H(X) - H(X|Y)$

$$= \sum_y \sum_x \left[p(x,y) \log \frac{1}{p(x)} - p(x,y) \log \frac{1}{p(x|y)}\right]$$

$$= \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = D_{KL}[p(x,y) \| p(x)p(y)]$$

symm, ≥ 0

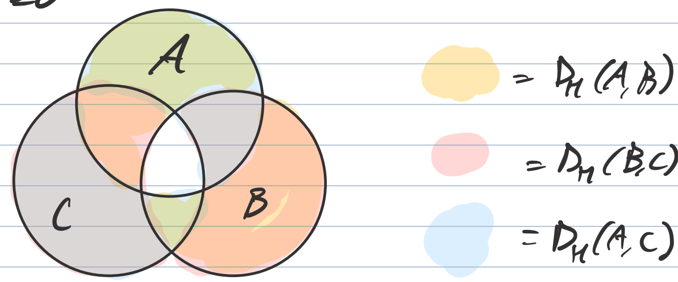
8.5 $D_H := H(X,Y) - I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x)p(y)}{p(x,y)^2}$

$= H(X|Y) + H(Y|X) \geq 0$

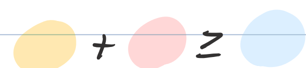
symm

$D_H(X,X) = 0$

Triangle



8.6 Straightfwd calc



⇐ clear from counting

8.7

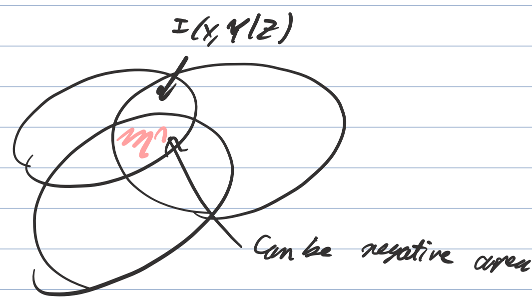
a) $q = 1/2 \Rightarrow P_Z = (1/2, 1/2)$

$\Rightarrow I(X;Z) = H(Z) - H(Z|X)$
$= 0$

b) $P_Z = [pq + (1-p)(1-q), (1-q)p + (1-p)q]$ ← BSC

$H(Z) - H(Z|X) = H_2(pq + (1-p)(1-q)) - H_2(q)$

8.8 For 3 vars:



8.9 $w \to d \to r$

world, data, processed data

$P(w,d,r) = P(w)P(d|w)P(r|d)$

$I(W;D,R) = I(W;D) + \cancel{I(W;R|D)}^{0}$

$= I(W;R) + I(W;D|R)$

$\Rightarrow I(W;R) \leq I(W;D)$

8.10 3 cards

$2/3$ chance black

$H(U) = 1$ $\quad P(U|L)\cdot P(L) = P(U,L)$:

$H(L) = 1$

	0	1
0	$1/3$	$1/6$
1	$1/6$	$1/3$

$H(U|L) = H_2(1/3)$

$\Rightarrow I(U,L) = 1 - H_2(1/3) = 0.08$ bits

Chapter 9

9.1 $P(x=0) = 0.9$ $\quad P(x=1) = 0.1$

$$P(x=1|y=1) = \frac{0.85 \cdot 0.1}{0.15 \cdot 0.9 + 0.85 \cdot 0.1} \approx 0.39$$

9.2 $$P(x=1|y=0) = \frac{0.15 \cdot 0.1}{0.15 \cdot 0.1 + 0.85 \cdot 0.9} \approx 0.02$$

9.3 Z channel:

$$P(x=1|y=1) = \frac{0.85 \cdot 0.1}{0.85 \cdot 0.1 + 0} = 1.0$$

9.4 $$P(x=1|y=0) = \frac{0.15 \cdot 0.1}{0.15 \cdot 0.1 + 0.85 \cdot 0.9} \approx 0.016$$

9.5 BSC $\quad f = 0.15 \quad p(x=0) = 0.9$

$$I(X;Y) = H(Y) - H(Y|X)$$

$$H(Y|X) = \sum_x p(x)\, \underbrace{H(Y|x)}_{H_2(0.15)}$$

$$= H_2(0.15)$$

$$H(Y) = H_2(0.1 \cdot 0.85 + 0.9 \cdot 0.15)$$
$$= H_2(0.22)$$

$$\Rightarrow I = H_2(0.22) - H_2(0.15)$$

$= 0.76 - 0.61 \approx 0.15$

$H_2(X) = 0.47$

9.6 Z-channel:

$H(Y) - H(Y|X)$

$\Downarrow$

$= H_2(0.1 \cdot 0.85) - [0.9 \cdot H(0) + 0.1\ H(0.15)]$

$= 0.42 - 0.1 \cdot 0.61 \approx 0.36$

9.7 For $p_0 = p_1 = 1/2$

$H(Y) = H_2(1/2)$

$H(Y|X) = H_2(0.85)$

$\Rightarrow H_2(0.5) - H_2(0.85) = 0.39$

9.8 $H_2(0.5 \cdot 0.85) - 0.5 \cdot H_2(0.15)$

$0.98 - 0.3 = 0.679$

9.9 $C(Q_{BSC}) = \underbrace{H_2((1-f)p_1 + f(1-p_1))}_{\text{only } p \text{ dependence}} - H_2(f)$

maximized @ $p = 1/2$

9.10 For noisy typewriter, pick the uniform

$\Rightarrow C = \log_2 9$

9.11 Z-channel

$\underbrace{H_2(p_1(1-f)) - p_1 H_2(f)}_{\text{not maximized when } p_1 = 1/2}$

9.12 Did this above

9.13 This time $H(X) - H(X|Y)$

$\Rightarrow \mathcal{H}_2(p_1) - f\, H_2(p_1)$

$P(y=?)$

$= (1-f)\mathcal{H}_2(p_1)$ opt at $p_1 = 1/2 \Rightarrow 1-f$

9.14

$N=1$:

	0	1
0	$1-f$	0
?	f	f
1	0	$1-f$

$N=2$:

	00	10	01	11	
00	$(1-f)^2$	0	0	0	$\Rightarrow 0$
?0	$f(1-f)$	$f(1-f)$	0	0	
10	0	$(1-f)^2$	0	0	
0?	$f(1-f)$	0	$f(1-f)$	0	
??	f^2	f^2	f^2	f^2	$\Leftarrow$ only $\Rightarrow$?
1?	0	$f(1-f)$	0	$f(1-f)$	$\Rightarrow 1$
01	0	0	$(1-f)^2$	0	
?1	0	0	$f(1-f)$	$f(1-f)$	
11	0	0	0	$(1-f)^2$	

Take $\underline{x} \in \mathcal{X}^N$

\# of probable $\underline{y} \simeq 2^{NH(Y)}$

$\simeq 2^{NH(Y|X)}$ probable seqs given $\underline{x}$

$\Rightarrow$ \# of non-confusable inputs $\approx 2^{NH(Y)} / 2^{NH(Y|X)} = 2^{NI(X;Y)}$

Let X maximize $I(X;Y)$

$\Rightarrow$ \# of non-confusable inputs is 2^{NC}

$\Rightarrow$ C bits per N bits $\underline{x}$

$\Rightarrow$ Rate C

9.15 $I(X;Y) = H_2(p_1(1-f)) - p_1 H_2(f)$

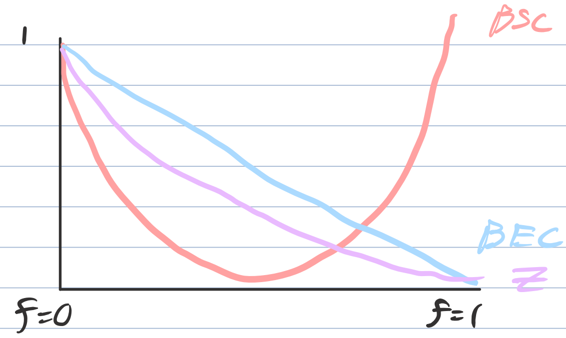
$$\frac{\partial I}{\partial p_1} = (1-f)\log_2 \frac{1-p_1(1-f)}{p_1(1-f)} - H_2(f) = 0$$

$$\Rightarrow p_1(1-f) = \frac{1}{1+2^{H_2(f)/(1-f)}} \quad \Rightarrow p_1^* = \frac{1/(1-f)}{1+2^{H_2(f)/(1-f)}}$$

as $f \to 1$ $p_1^* \to 1/e$ by L'Hopital

When 1 is used, the 'Z' channel injects entropy
not so for 0

9.16



9.17 $H(Y|X) = 2$
$H(Y) = \log 10$ $\Rightarrow C = \log \frac{5}{2}$ bits

9.18 $Q(y|x, \alpha\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-x\alpha)^2}$

$Q(x|y) \propto e^{-\frac{1}{2\sigma^2}(y \mp \alpha)^2}$

$a(y) = \log \frac{P(x=1)}{P(x=-1)} = -\frac{1}{2\sigma^2}[-2y\alpha - 2y\alpha]$

$= \frac{2y\alpha}{\sigma^2}$

$\Rightarrow P(x=1|y) = \frac{1}{1+e^{-2y\alpha/\sigma^2}}$

$P_b = \int_{-\infty}^{0} dy\, Q(y|x=1) = \int_{-\infty}^{-x\alpha} dy \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}y^2} = \Phi(-\frac{x\alpha}{\sigma})$

9.19 see explaination

$H('z') \approx 220$

90% chance of clean z w/ $H \approx 20$
10% of messy w/ $H \approx 220$

$\Rightarrow H(0.1) + 0.1 \cdot 220 + 0.9 \cdot 20 \approx 40$

9.20 $P(\text{distinct}) = \frac{A \cdot (A-1) \cdots (A-S+1)}{A^S}$

not instructive $\rightarrow$ $A = 365 \Rightarrow S \sim 24$

$\#$ pairs $= \dfrac{S\cdot(S-1)}{2}$

$\#$ pairs $\cdot$ P(that pair shares bday) $= \dfrac{S\cdot(S-1)}{2}\dfrac{1}{A}$

$= E(\#$ collisions$)$

small if $S \ll \sqrt{A}$

big if $S \gg \sqrt{A}$

$$1\,(1-1/A)\cdots\left(1-\frac{S-1}{A}\right) \approx \exp\left(-\frac{1}{A}\sum_{i=1}^{S-1} i\right)$$

$$\approx \exp\left(-\frac{S(S-1)}{2A}\right)$$

9.21 Capacity is $H(X) - H(X|Y)$

$\uparrow$ $\log 365$; $H(X|Y) \to 0$

rate is $\log_2 24 \sim 4.6$ bits

$\Rightarrow$ below capacity + 6% chance of error $\approx$

9.22 Select q^k k-tuples of A^k alphabet

$$1-\left(\frac{A^k-1}{A^k}\right)^{q^k-1}$$

$q=364$ $k=1$ $\Rightarrow$ $1-(1-1/A)^{363} \approx 0.63$

$\Rightarrow$ likely failure

As $k \to \infty$ though

$$1-\left(\frac{A^k-1}{A^k}\right)^{q^k-1} \approx \left(\frac{q}{A}\right)^k \to 0 \text{ fast}$$

as $k \to \infty$ this becomes reliable.

Chapter 10

Noisy-Channel Coding

1. $C = \max_{P_X} I(X;Y)$

has $\forall \epsilon > 0$ $R < C$ for N large $\exists$ code of rate R s.t. $p_B^{max} < \epsilon$

2. If bit error p_b is acceptable, can reach rates up to

$$R(p_b) = \frac{C}{1 - H_2(p_b)}$$

3. Higher rates not possible

$\underline{x}$ is typical in $P(\underline{x})$ if $\left| \frac{1}{N} \log \frac{1}{P(\underline{x})} - H(X) \right| < \beta$

similar for $\underline{y}$ in $P(\underline{y})$
$\underline{x}\,\underline{y}$ in $P(\underline{x},\underline{y})$

$\underline{x}, \underline{y}$ jointly typical if above 3 hold "J.T."

$J_{N\beta}$ is set of all jointly typical $\underline{x}\,\underline{y}$

let $\underline{x}, \underline{y} \sim P(\underline{x},\underline{y}) = \prod_{n=1}^{N} p(x_n, y_n)$

1. By LLN the probability that $\underline{x}, \underline{y}$ are jointly typical $\to 1$ as $N \to \infty$

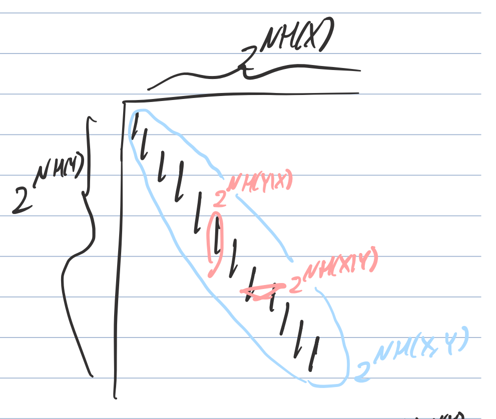
2. $|J_{N\beta}| \simeq 2^{NH(X,Y)}$
   $|J_{N\beta}| \leq 2^{N(H(X,Y)+\beta)}$

3. For $\underline{x} \sim P(\underline{x})$ $\underline{y} \sim P(\underline{y})$ indep

$$P((\underline{x},\underline{y}) \in J_{N\beta}) = \sum_{\underline{x},\underline{y} \in J_{N\beta}} P(\underline{x}) P(\underline{y})$$

$$\leq |J_{N\beta}|\, 2^{-N(H(X)-\beta)}\, 2^{-N(H(Y)-\beta)}$$

$$\leq 2^{-N(I(X;Y)-3\beta)}$$



Mutual info is $\frac{1}{N}\log\frac{1}{p}$ that two random typical sequences x, y are also jointly typical

$\Rightarrow p(\text{you hit a dot}) \sim 2^{N(H(X,Y) - H(X) - H(Y))} = 2^{-NI(X;Y)}$

Shannon's proof:

1. Fix $P(x)$ & generate $S = 2^{NR'}$ codewords of an (N, NR') code
at random according to $\prod_{n=1}^{N} p(x_n)$

2. $P(\underline{y} | \underline{x}^s) = \prod_n P(y_n | x_n^s)$

3. Typical-set decoding (non optimal but good enough):
decode $\underline{y}$ as $\underline{x}$ if $\exists ! \, \underline{x}$ s.t. $\underline{x}, \underline{y}$ are J.T. else error

4. error if $\hat{s} \neq s$

errors

1. $p_B = P(\hat{s} \neq s \,|\, \underset{\text{Fixed code}}{C})$

2. $\langle p_B \rangle = \sum_C P(\hat{s} \neq s | C) \, P(C)$ ← average over codes

3. $p_{BM}(C) = \max_s \, p(\hat{s} \neq s | s, C)$ ← this is what we care about

first focus here

a) By symmetry in C we can assume WLOG $s = 1$

$\delta \geq$ probability that $x, y \notin J_{N\beta}$ → 0 from before
defn
$\delta \to 0$ as $N \to \infty$

b) $P[(x', y) \in J_{N\beta} \text{ for } x' \neq x] \leq 2^{-N(I(X;Y) - 3\beta)}$

$\Rightarrow \langle p_B \rangle \leq \delta + 2^{NR'} \cdot 2^{-N(I(X;Y) - 3\beta)}$

$\nexists ! \, x$ s.t. (x,y) JT $\quad x^s$ not unique

$\Rightarrow \langle p_B \rangle < 2\delta$ as long as $I(X;Y) > R' + 3\beta$

Choose $P(X)$ to maximize $I(X;Y) \Rightarrow C > R' + 3\beta$

c) Since $\langle p_B \rangle < 2\delta$ $\exists C$ with $p_B(C) < 2\delta$ ✓

Focus on that code. Now:

d) For p_B, $\frac{1}{S}\sum_s \overset{\in[0,1]}{p_B(s,C)} < 2\delta \Rightarrow$ Markov: $\mathbb{P}[p_B(s,C)>a] \le \frac{\mathbb{E}_s[p_B(s,C)]}{a} = \frac{2\delta}{a}$

→ error of best half of codewords must all be $< 4\delta$ $\quad a = 4\delta \Rightarrow \frac{1}{2}$

→ Throw away other half

$2^{NR'-1}$ words $\Rightarrow$ rate $R = R' - 1/N$ $\quad R' < C - 3\beta$

10.4 Communication above capacity (part 2 of theorem)

Take noiseless channel (or well-encoded $R<C$ noisy channel)

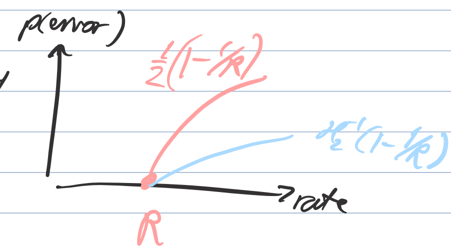
⇒ If we want a $C=1$ bit channel
at $R=2$ ⇒ ignore ½ the bits

→ $1-1/R$ of the bits missing ⇒ guess randomly

guess randomly

⇒ $p_b = \frac{1}{2}(1 - 1/R)$ ← can do better

$= 1/4$



Take (N,K) code

put chunks of N bits in, turn to K bits → "encode" K bits back to N

here, "encoding" $N \to K$ is just error correcting ⇒ qN bits away from length N word (of noiseless channel capacity C)

then $K \to N$ differs by $\sim qN$ bits from original

⇒ $p_b = q$

$\frac{K}{N} = C(q) \Rightarrow \frac{N}{K} = 1/C(q)$

$C_{BSC}(q) = 1 - H_2(q) \Rightarrow$ differs by $p_b = q \Rightarrow N/K = \frac{C}{1-H_2(q)} \Rightarrow \frac{C}{1-H_2(q)}$ error rate For BSC

← bit flip

10.5 Non-achievable part (Part 3 of Theorem)

$P(s, \underline{x}, \underline{y}, \hat{s}) = P(s)P(\underline{x}|s)P(\underline{y}|\underline{x})P(\hat{s}|\underline{y})$

Data processing: $I(s;\hat{s}) \le I(\underline{x};\underline{y}) \le NC$ ← defn of Channel capacity

Rate R & p_b error ⇒ Rate R & bit error probability p_b

not block

10.1 if errors on $\hat{s}$ indep $\quad I(s;\hat{s}) = \underbrace{H(\hat{s})}_{NR} - \underbrace{H(\hat{s}|s)}_{NR\,H_2(p_b)} = NR(1 - H_2(p_b))$ by independence

← NR bits

If there are complex correlations between bits then

key insight ⇒ $H(\hat{s}|s) < NR\,H_2(p_b) \Rightarrow I(s;\hat{s}) \ge NR(1 - H_2(p_b))$

$$\Rightarrow \quad NR(1-\mathcal{H}_2(p_b)) \le I(s;\hat{s}) \le I(x,y) \le NC$$

$$\Rightarrow R \le \frac{C}{1-\mathcal{H}_2(p_b)} \Rightarrow \text{max acheivable } R \text{ is } \frac{C}{1-\mathcal{H}_2(p_b)}$$

10.6 Computing Capacity

10.2 $\quad -\sum_{j,i} Q_{j|i} p_i \log \sum_k Q_{j|k} p_k + \sum_{j,i} Q_{j|i} p_i \log Q_{j|i}$

$$-\sum_y p(y) \log p(y) \qquad H(Y) - H(Y|X) = I(X,Y)$$

$\frac{d}{dp_i}$

$$= \sum_j \left(-1 - \log p_j^y\right) \frac{\partial p_j^y}{\partial p_i^x} + \sum_j Q_{j|i} \log Q_{j|i}$$

$$= -\sum_j \left(1 + \log \sum_k Q_{j|k} p_k\right) Q_{j|i} + \sum_j Q_{j|i} \log Q_{j|i}$$

$*$

10.3

$$\Rightarrow \frac{\partial^2 I}{\partial p_i \partial p_j} = -\sum_\ell \frac{Q_{\ell|j}}{\underbrace{\sum_k Q_{\ell|k} p_k}_{p_\ell^y}} Q_{\ell|i} = -\sum_\ell \frac{\frac{\partial p_\ell^y}{\partial p_i^x}}{\sum_k Q_{\ell|k} p_k} Q_{\ell|i} < 0$$

$$\frac{\partial I}{\partial p} = 0 \Rightarrow \text{global max}$$

$\Rightarrow$ find $\frac{\partial I}{\partial p} = \lambda \quad \forall i \qquad \lambda$ is for $\sum_i p_i = 1$

- misses bdy eg

$$\begin{array}{ccc} 0 & \rightarrow & 0 \\ ? & \nearrow \searrow & \\ 1 & \rightarrow & 1 \end{array} \Rightarrow p = \begin{cases} 1/2 \\ 0 \\ 1/2 \end{cases}$$

10.4 $\quad H(Y) - H(Y|X)$

$p_I \quad -((1-p_I)p_x + \frac{1}{2}p_I)\log(\cdots) - ((1-p_I)(1-p_x) + \frac{1}{2}p_I)\log(\cdots) - p_I \mathcal{H}_2(1/2)$

$(1-p_I)p_x$

$(1-p_I)(1-p_x)$

$$p(y) = \sum p(y|x)p(x)$$

$$\Rightarrow p(0) = 1\cdot(1-p_I)p_x + \tfrac{1}{2}p_I$$
$$p(1) = 1\cdot(1-p_I)(1-p_x) + \tfrac{1}{2}p_I$$

10.5 KKT optimizer

10.6 From $*$: $\sum_j Q_{j|i} \log p_y^j = \underbrace{\sum_j Q_{j|i} \log Q_{j|i} - \lambda - 1}_{> -\infty}$

If $p_y^j = 0$ then LHS $= -\infty$

unless $Q_{j|i}$ $\Rightarrow$ Use all accessible outputs

10.7 p_j^Y is linear in $Q_{j|i}$

$H(Y)$ is concave in $p_j^Y \Rightarrow$ concave in $Q_{j|i}$

$H(Y|X)$ is also concave in $Q_{j|i}$

$p_\lambda(x,y) = p(x)\left(\lambda\, p_1(y|x) + (1-\lambda) p_2(y|x)\right)$ $\Rightarrow$ both $p(x,y)$ $p(x)p(y)$ are convex combinations

$p(x)p(y) = p(x)\left(\lambda p_1(y) + (1-\lambda) p_2(y)\right)$

$I(X_\lambda; Y_\lambda) \le \lambda I(X_1; Y_1) + (1-\lambda) I(X_2; Y_2)$ $\Leftarrow$ I is jointly convex

Because D_{KL} is jointly convex

$\Rightarrow$ I is convex in $Q_{j|i}$

$$D_{KL}(p\|q) = \mathbb{E}_p \log \frac{p}{q} \leftarrow \text{convex in } q$$

$$= \mathbb{E}_q \frac{p}{q} \log \frac{p}{q} \leftarrow \text{convex in } p$$

10.8 Let $p_1(x)$ $p_2(x)$ be optimal

$$\underbrace{I(X_\lambda; Y)}_{} \ge \lambda I(X_1; Y) + (1-\lambda) I(X_2; Y)$$

by optimality

this is an equality

$$p_\lambda(y) = \sum_i Q_{j|i}\, p_{\lambda,i}(x) = \lambda p_1(y) + (1-\lambda) p_2(y)$$

10.8 Let $p_1(x)$ $p_2(x)$ both have $I(Y_i;X_i)=C$

$$p_i(y)=\sum_x \underbrace{p(y|x)}_{\text{const}}\, p_i(x)$$

$$I(\lambda p_1(x)+(1-\lambda)p_2(x);\ \lambda p_1(y)+(1-\lambda)p_2(y)) \geq \lambda I_1 + (1-\lambda) I_2$$

$\Rightarrow$ convex combination also has $I(\lambda p_1+(1-\lambda)p_2)=C$
$\Rightarrow$ above holds with equality

Now $I(X;Y)=H(Y)-H(Y|X)$

$$H(Y|X)=\sum_x p(x) \underbrace{\sum_y p(y|x)\log p(y|x)}_{\text{const}}$$ affine in $p(x)$

$$H(Y_\lambda)=\underbrace{\lambda H(Y_1)+(1-\lambda)H(Y_2)}_{\text{view this as } \mathbb{E}_\lambda H(Y_\lambda) \text{ where } Y_\lambda \sim p_\lambda(y)} \Rightarrow p_\lambda(y) \text{ is } \lambda\text{-indep} \Rightarrow p_1(y)=p_2(y)$$

A discrete memoryless channel is symmetric iff the outputs can be partitioned into subsets s.t.:
For each subset, the matrix $Q^{\text{subset}}_{y\leftarrow x}$
has each row is a perm. of every other
& likewise for columns

Eg 10.9

$P(y=0|x=0)=0.7$ $\quad$ $P(y=0|x=1)=0.1$
$P(y=?|x=0)=0.2$ $\quad$ $P(y=?|x=1)=0.2$
$P(y=1|x=0)=0.1$ $\quad$ $P(y=1|x=1)=0.7$

$$\begin{array}{c|cc}
 & x=0 & x=1 \\
\hline
0 & 0.7 & 0.1 \\
1 & 0.1 & 0.7 \\
\hline
? & 0.2 & 0.2
\end{array}$$

Will later see that
communication @ capacity can
be acheived over symmetric channels
by linear codes

Ex 10.10 Assume partition has only 1 elem

$$I(X;Y)=H(Y)-H(Y|X)$$
$$\sim \sum p_x \underbrace{H(Y|x)}_{x\text{-indep since } p(y|x_1) \text{ is perm of } p(y|x_2)} =: H(r)$$

$$= H(Y) - \underbrace{H(r)}_{p(x)\text{-indep}}$$

$$\le H(\mathrm{Unif}(y)) - H(r)$$

Note $\mathrm{Unif}_x \Rightarrow \mathrm{Unif}_y$ because rows are perms of each other

$\Rightarrow$ For single partition, Unif_x is an optimum

For multiple partitions $p(x) = \mathrm{Unif}(x) \Rightarrow p(y) = \mathrm{Unif}(\ )$ still

Fix x, x' $p(y|x)$ is still a permutation of $p(y|x')$ $\forall x, x'$
regardless of y being partitioned

$$\Rightarrow H(y|x) = \sum_x p(x) H(y|x) = H(y|x_0) =: H(r)$$

$$\Rightarrow I(X;Y) \le H(\mathrm{Unif}(y)) - H(r)$$

with equality for $p(x) = \mathrm{Unif}(x)$

Ex 10.11 For channel we have $I \cdot (J-1)$ d.o.f
for $p(x)$ we have just $I-1$

In the $I(J-1)$-dim space of perturbations about symmetric channel
expect a dimension $I \cdot (J-1) - I - 1 = IJ - 2I - 1$
that leave $p^*(x)$ the same but break symmetry

example:
$$\begin{pmatrix} 0.9585 & 0.0415 & 0.35 & \\ 0.0415 & 0.9585 & & 0.35 \\ & & 0.65 & \\ & & & 0.65 \end{pmatrix}$$

10.7 Reliable communication w/ error ϵ & rate R
at sufficiently large N

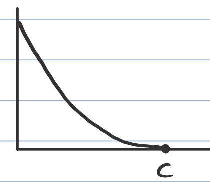
Closer $R \to C$ & smaller ϵ is $\Rightarrow$ larger N

$$P_B \le \exp[-N \underbrace{E_r(R)}_{\text{random coding exponent}}]$$

$P_{B,max}$ also follows this
by expurgation

AKA reliability function

convex in R:



$E_r(R) \to 0$ as $R \to C$

Even for BSC there is no analytic form for E_r

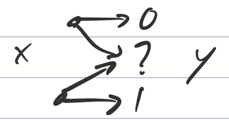
Lower bounds:

$$P_B \geq \exp[-N E_{sp}(R)]$$

sphere packing exponent

$E_{sp}(R)$ also convex decreasing in R

Ex 10.12



$$\Rightarrow Q = \begin{pmatrix} 1-q & 0 \\ q & q \\ 0 & 1-q \end{pmatrix}$$

let $P(x=0)=p \Rightarrow$ $P(y \mid x=0) = (1-q, q, 0)$
$P(y \mid x=1) = (0, q, 1-q)$ $\Rightarrow H(Y|X) = H_2(q)$

p-indep

$$P(y) = [(1-q)p,\ qp+q(1-p),\ (1-q)(1-p)]$$
$$= [(1-q)p,\ q,\ (1-q)(1-p)]$$

$$\Rightarrow H(Y) = -(1-q)p \log (1-q)p - q \log q - (1-q)(1-p) \log (1-q)(1-p)$$

$\psi = -x \log x$
$\psi' = -1 - \log x$

$$\partial_p H(Y) = 0 \Rightarrow p = 1/2$$

$$\Rightarrow H(Y) = -(1-q) \log (1-q)\tfrac{1}{2} - q \log q$$

$$\Rightarrow C = H(Y) - H(Y|X) = 1-q$$

Z channel: Encode second bit as first bit flipped

Either both bits are sent correctly w prob $1-q$
or the bit encoded as 1 flips with prob q

$\tfrac{1}{2}$ 01 $\to$ 01 $(1-q)/2$
00 q
$\tfrac{1}{2}$ 10 $\to$ 10 $(1-q)/2$

$\Rightarrow$ BEC w q

$$\Rightarrow C \geq \frac{1-q}{2}$$

this code just gives an example rate. $C > \frac{1-q}{2}$ is possible

Ex 10.13 Take a set of C connections of N wires

Information content of a partition is $\log \Omega$

$$\Omega = \frac{N!}{\prod_r \underbrace{g_r!}_{\text{permute subsets}} \underbrace{(r!)^{g_r}}_{\text{permute within subset}}} \qquad g_r \text{ subsets of size } r$$

$$\frac{\partial}{\partial g_r}\left[\log \Omega + \lambda \sum r g_r\right] = -\log r! - \log g_r + \lambda r$$

$$\Rightarrow g_r = \frac{e^{\lambda r}}{r!} \Rightarrow \text{optimal is poisson!}$$

$$\sum g_r r = \mu e^{\mu} = N \qquad \mu = e^{\lambda}$$

Chapter 11 : Real Channels

$$P(y|x) = N(y|x, \sigma^2)$$

discrete in time $\Rightarrow$ AGWN channel

$$y(t) = x(t) + n(t)$$

$$n \sim N(0, \sigma^2)$$

Power cost $:= \frac{1}{T}\int_0^T dt\, [x(t)]^2 \leq P$

Transmit N numbers using N basis functions

$$x(t) = \sum_{n=1}^{N} x_n \phi_n(t)$$

$$y_n = \int_0^T dt\, \phi_n(t) y(t) = x_n + \int_0^T dt\, \phi_n(t) n(t)$$

$$= x_n + n_n \qquad n_n \sim N(0, N_0/2)$$

power $< P \Rightarrow \overline{x_n^2} < \frac{PT}{N}$

Bandwidth: $W = \frac{N^{max}}{2T}$

$\Rightarrow N^{max} = 2WT$

By Nyquist sampling theorem if highest freq is W then a signal can be uniquely recovered by sampling at $\Delta t = \frac{1}{2W}$ intervals

$\Rightarrow 2W$ uses /second

If we want to transmit binary x_n have an encoding giving us rate R

power/source bit $E_b = \overline{x_n^2}/R$ vs noise spectral density

$$\frac{E_b}{N_0} = \frac{x_n^2}{2\sigma^2 R}$$

11.2 Inferring the input

$$P(n) = N(0, A^{-1})$$

$$\Rightarrow P(y|s) = N(s, A^{-1})$$

resp. embeddings

$$\frac{P(s=1|y)}{P(s=0|y)} = \frac{P(y|s=1)}{P(y|s=0)}\frac{P(s=1)}{P(s=0)} = \exp\left[y^T A(x_1 - x_0) \underbrace{- \frac{1}{2}x_1^T A x_1 + \frac{1}{2}x_0^T A x_0 + \log\frac{P(s=1)}{P(s=0)}}_{=:\theta}\right]$$

$a(y) = y^T A(x_1 - x_0) + \theta = w^T y + \theta$ } LDA

$a > 0 \Rightarrow s = 1$

$a < 0 \Rightarrow s = 0$

11.3 Capacity of a Gaussian Channel

Ex 11.1 $I(X;Y) = H(Y) - H(Y|X)$

$$\max_{p(x)} I(X;Y) \quad \text{s.t.} \quad \overline{x^2} = v$$

$$\int dx\, P(x)\left[\int dy\, P(y|x) \log \frac{P(y|x)}{P(y)} - \lambda x^2 - \mu\right]$$

$$\Rightarrow \frac{\delta}{\delta P(x)} = \int dy\, P(y|x) \log \frac{P(y|x)}{P(y)} - \lambda x^2 - \mu$$

$$- \int dx'\, P(x') \int dy\, \frac{P(y|x')}{P(y)} \underbrace{\frac{\delta P(y)}{\delta P(x)}}_{P(y|x)}$$

$$\int \cancel{dx'\,dy}\; \cancel{p(x')}\, \frac{p(x',y)}{\cancel{p(y)}\,\cancel{p(x')}}\, \frac{p(x,y)}{\cancel{p(x)}} = 1$$

$$\Rightarrow \forall x: \int dy\; P(y|x) \log P(y) = -\lambda x^2 - \mu'$$

$P(y|x)$ is gaussian w/ mean x $\Rightarrow$ $\log P(y)$ must be quadratic in y

$\Rightarrow$ $P(y)$ is gaussian

Can obtain this using gaussian x

Ex 11.2

$$I = \int dx\,dy\, P(x)\, P(y|x) \log P(y|x) - \int P(y) \log P(y)$$

$$= \frac{1}{2} \log \frac{1}{\sigma^2} - \frac{1}{2} \log \frac{1}{v^2+\sigma^2} \qquad \left(\tfrac{1}{2}\left(1+\log 2\pi\right)\right) \text{ cancel}$$

$$= \frac{1}{2} \log \Big(1 + \underbrace{\frac{v}{\sigma^2}}_{SNR}\Big)$$

Geometric view of noisy-channel coding theorem:

$\underline{x} = (x_1 \cdots x_N)$

Noise power is very close (for large N) to $N\sigma^2$

$\Rightarrow$ $\underline{y}$ is close to lying on the surface of a sphere at x of radius $\sqrt{N\sigma^2}$

If x is generated under $\overline{x^2} = v$

$\Rightarrow$ x is close to the surface of a sphere at 0 of radius $\sqrt{Nv}$

$\Rightarrow$ y is at $\sqrt{N(v+\sigma^2)}$

$$V_S(r, N) = \frac{\pi^{N/2} r^N}{\Gamma(N/2+1)}$$

$$\frac{Vol(S_y)}{Vol(S_{y|x})} \Rightarrow \left[\frac{v+\sigma^2}{\sigma^2}\right]^{N/2} = \left[1 + SNR\right]^{N/2}$$

$$= \exp\left[\frac{N}{2}\log(1+SNR)\right]$$

$$\Rightarrow \quad C \simeq \frac{1}{2}\log(1+SNR)$$

$N/T = 2W$ uses per second

$$\Rightarrow \quad C \cdot \frac{N}{T} = W\log(1+SNR) \qquad \sigma^2 = N_0/2$$
$$v = \overline{x^2} = P/2W$$

$$= W\log\left(1 + \frac{P}{WN_0}\right)$$

$$W_0 \coloneqq P/N_0 \Rightarrow \frac{C}{W_0} = \frac{W}{W_0}\log\left(1 + \frac{W_0}{W}\right)$$

$$C \to W_0 \log e \quad \text{as} \quad \frac{W}{W_0} \to \infty$$

Better to have low SNR large W
than high SNR small W

P, N_0 fixed $\Rightarrow$ W_0 fixed

"Wideband communication" $\to$ 3G

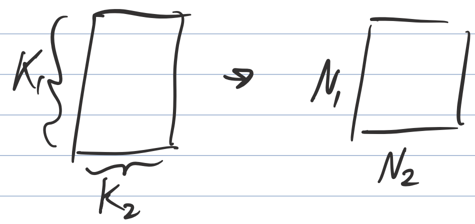
But for social reasons need narrower bands

Concatenation: $\underbrace{\underset{\text{enc}}{C} \to \underset{\text{channel}}{Q} \to \underset{\text{decoder}}{D}}_{\text{super-channel } Q'}$

$C' \to Q' \to D'$ ⇒ Concatenated code

Interleaving: Read in blocks of length > length of C, C'

encode data one way using C
reorder bits → encode another way using C'



Ex 11.3 Finish

11.4 $C = 1 - \underbrace{H(\text{noise})}_{\underbrace{H_2(b)}_{\text{odds of burst}} + \underbrace{Nb}_{\text{flipped bits}}} = 1 - 0.207 \approx 0.793$

Interleaving leads to a bsc w/ $f \sim 0.2 \times 0.5 = 0.1$ iid

$\Rightarrow C = 0.53$

11.5 a) $C = \frac{1}{2} \log\left(1 + \frac{v}{\sigma^2}\right)$

b) C is maximized for $\pm v$ equiprobable

$C = -\int P(y) \log P(y) - \int N(y;0) \log N(y;0)$

annoying - becomes close to $\frac{1}{2} \log 1 + \frac{v}{\sigma^2}$ for $\frac{v}{\sigma^2}$ small

c) Becomes BSC

$$C = 1 - \mathcal{H}_2(\xi) \qquad \xi = \Phi(\sqrt{v}/\sigma)$$