CC

Noiseless coding - entropy

1. DMC
2. BSC, BEC
3. information rate $\rho(C) = \frac{1}{n}log_2 m$; error rate = $e(\hat{C}) = \max_{x \in M} P(error|xsent)$
4. **transmit reliably** at rate R if there exists $(C_n)_{n=1}^{\infty}$ with each C_n a code of length n such that $lim_{n\to\infty}\rho(C_n) = R$ & $lim_{n\to\infty}\hat{e}(C_n) = 0$.
5. A code is a function $c : A \to B^*$, $c(a)$ are codewords; $c^* : A^* \to B^*$
6. decipherable: induced map c^* is injective
7. block code: all words same length; comma code;
8. prefix-free code: is a code where no codeword is a prefix of any other distinct word
9. **Kraft's inequality**: $|A| = m, |B| = a$, $c : A \to B^*$ has word lengths $l_1, \dots, l_m$. Then $\sum_{i=1}^{m} a^{-l_i} \le 1$
10. A prefix-free code exists if and only if Kraft's inequality holds
11. (McMillan). Any decipherable code satisfies Kraft's inequality
12. Cor: A decipherable code with prescribed word lengths exists if and only if a prefix-free code with the same word lengths exists

1. $H(X) = -\sum_{i=1}^{b} p_i \log p_i$
2. note: $H(p)' = \log \frac{1-p}{p}$, $p = \frac{1}{2}$ giving entropy 1
3. **Gibb's inequality**: $-\sum_{i=1}^{n} p_i \log p_i \le -\sum_{i=1}^{n} p_i \log q_i$. [hint: $\ln q_i/p_i \le q_i/p_i - 1$]
4. Cor: $H(p_1, p_2, \dots, p_n) \le \log n$
5. **Shannon's Noiseless Coding Theorem**: $H(X)/\log a \le E[S] < H(X)/\log a + 1$ [left: Gibb's, $q_i = a^{-l_i}/D$; right: $l_i = lower[-\log_a p_i] + 1$]
6. Shannon-Fano Coding
7. Huffman Coding is optimal [lemma: $p_i p_j$, $l_i l_j$, ; maximal length differ only one last]
8. $H(X, Y)$
9. $H(X, Y) \le H(X) + H(Y)$ [Gibb's, p_{ij} replace by $p_i q_j$]

Error correcting codes - noisy channels

1. binary $[m, n]$-code, Hamming distance
2. ideal observer, maximal likelihood(maximising $P(x$ received $|\, c$ sent)), munimum distance [later two equivalent if $p < 1/2$]
3. d-error detecting: changing up to d digits in each codeword cannot produce another; e-error correcting if knowing that $x \in 0, 1^n$ differs fom a codeword in at most e places we can deduce the codeword.
4. Repetition Code: $[n, 2]$-code, info rate $1/n$
5. Simple parity check: $[n, 2^{n-1}]$, info rate $\frac{n-1}{n}$
6. Hamming code; [7,16,3]-code, 1-error-correcting
7. $[n, m, d]$-code. Minimum distance d, $(d-1)$-error-detecting, $[\frac{d-1}{2}]$-error-correcting

BCH codes

Fuck it

Shift Registers

1. Def
2. Berlekamp-Massey

Cryptography

1. plaintext M, ciphertext C, key K; $e : M \times K \to C$; $d : C \times K \to M$
2. $M = C = \{A, B, \ldots, Z\}^* = \Sigma^*$; Simple substitution: K ={permutations of Σ}; Vigenere cipher: $K = \Sigma^d$ for some D: write out below, sum, mod 26; Caesar cipher: d = 1
3. perfect secrecy: Say (M, K, C) has perfect secrecy if $H(M|C) = H(M)$, i.e M and C are independent
4. perfect secrecy implies $|K| \geq |M|$.
5. message equivocation is $H(M|C)$; key equivocation is $H(K|C)$
6. $H(M|C) \leq H(K|C)$
7. unicity distance: the least n such that $H(K|C^{(n)}) = 0$, i.e the smallest number of encrypted messages required to uniquely determine the key.
8. $U := \frac{log|K|}{log|A|-H}$; $R = 1 - H/log|A|$ redundancy, where $M = C = A$
9. one-time-pad: perfect secrecy
10. key: private key for decryption, public key for encryption

1. Let $p = 4k - 1$ be prime. If the equation $x^2 \equiv d(modp)$ has a solution then $x \equiv d^k(modp)$ is a solution
2. Rabin cryptosystem: private key: p,q =3mod4; Public key: N = pq. $M = C = 1, \ldots, N-1 = Z_n^\star$. Encrypt $m \in M$ as $c = m^2(modN)$. The ciphertext is c
3. Breaking Rabin as difficult as factorizing N
4. RSA
5. Finding the RSA private key (N, d) from the public key (N, e) is essentially as difficult as factoring N
6. Authenticity using RSA; problem: homo attack, Existential forgery (($s^e(modN), s$) valid signed message)
7. Homo attack

CC Quick Run.

Noiseless coding - entropy.

1. DMC: Discrete message chanel.

$p_{ij} = \mathbb{P}(c_j \text{ received} \mid x_i \text{ sent})$

fixed and indep.

2. BSC: $\Sigma_A = \Sigma_B = \{0,1\}$,

3. Info rate $\rho(C) = \frac{1}{n}\log m$.

$\hat{e}(C) = \max_{x \in M} \mathbb{P}(\text{error} \mid x \text{ sent})$

4. Transmit reliably at rate R:

~~$\exists (x_1, \dots, x_n)$~~ s.t. $\exists (C_n)_{n=1}^{\infty}$ with C_n codewords with length n.

s.t. $\lim_{n\to\infty} \rho(C_n) = R$ $\lim_{n\to\infty} \hat{e}(C_n) = 0$.

5. Info capacity = max R.

6. Decipherable: Induced map $C^*: A^* \to B^*$ is injective

prefix-free code: No codeword is a prefix of any other distinct word.

Kraft's inequality: $|A|=m, |B|=a$, $C: A \to B^*$ has word length $l_1, \dots, l_m$

Then $\sum_{i=1}^{m} a^{-l_i} \leq 1$

Claim: Prefix free code exists $\Leftrightarrow$ Kraft's hold.

Proof: ($\Rightarrow$). Rewrite ~~it as~~ $\sum_{i=1}^{m} a^{-l_i}$ as

$\cancel{\sum_{i=1}^{s} n_s a^{-s}}$, where $s = \max\{l_i\}$,

$\sum_{i=1}^{s} n_i a^{-i}$

$n_s =$ ~~# codewords with~~ ~~# l_i with value~~

$n_i =$ # codewords with length i.

Then $\sum_{i=1}^{s} n_i a^{-i} \leq 1$

$\Leftrightarrow \sum_{i=1}^{s} n_i a^{s-i} \leq a^s$

Consider all ~~code with~~ code of length s:

It has max a^s possible combinations.

For each codeword of length i, ~~it takes~~

all other codewords can't start with it,

so ~~has a~~ takes a^{s-i} possible combo.

Then we.

($\Leftarrow$). We create ~~us~~ codewords by length:

Start from $i=1$, ~~we can use~~

one filled, left with

$(\Rightarrow)$ $n_1 a^{s-1} + n_2 a^{s-2} + \cdots + n_s \leq a^s$.

LHS is number of strings of length s in B with some codewords of C as a prefix,

RHS is ~~n~~ number of strings of length s.

So we have it true.

$(\Leftarrow)$. Given $n_1, \ldots, n_s$ satisfying it,

proceed by induction.

exists $\hat{C}$ with n_l codewords of length l for all $l \leq s-1$. Then we can add $\geq n_s$ new codewords of length s to $\hat{C}$, maintain prefix-free property.

McMillan: Any decipherable code satisfies Kraft's inequality.

Proof: Say if $|A| = m$, $|B| = a$, $c: A \to B^*$ is decipherable. with word length $l_1, \ldots, l_m$.

Consider $\left(\sum_{i=1}^{m} a^{-l_i}\right)^R$

$$= \left(\sum_{i=1}^{s} n_i a^{-i}\right)^R$$

$$= \sum_{i=1}^{Rs} b_i a^{-i} \quad \text{after expansion}$$

~~No~~ Since it's decipherable,
~~we~~ each b_i is at most a^i.
as code of length i can only ~~to~~ be deciphered in a unique way, corresponds to 1 sequence of codewords

$$\left(\sum_{i=1}^{m} a^{-l_i}\right)^R \le \sum_{i=1}^{Rs} a^{-i} = \frac{1}{a}\left[1+\frac{1}{a}+\cdots+\frac{1}{a^{Rs-1}}\right]$$

$$= \frac{1}{a}\cdot\frac{1-\frac{1}{a^{Rs}}}{1-\frac{1}{a}}$$

~~Let R~~

$$\sum_{i=1}^{m} a^{-l_i} \le \left(\frac{1}{a}\cdot\frac{1-\frac{1}{a^{Rs}}}{1-\frac{1}{a}}\right)^{\frac{1}{R}}$$

$\to 1$ as $R \to \infty$.

Done!

$$\left(\sum_{i=1}^{m} a^{-l_i}\right)^R \le \sum_{i=1}^{Rs} a^{-i} a^{i} = Rs$$

$$\sum_{i=1}^{m} a^{-l_i} \le (Rs)^{\frac{1}{R}} \to 1 \quad \text{as } R \to \infty$$

So ~.

1. $H(X) = -\sum p_i \lg p_i$

$H(P)' = \frac{d}{dp}\left(-p\lg p - (1-p)\lg(1-p)\right)$

$= -\log p - 1 + 1 + \log(1-p)$.

$= \lg \frac{1-p}{p}$.

$\Rightarrow p = \frac{1}{2}$ max, $H(P) = 1$.

Gibb's inequality: $-\sum p_i \log p_i \le -\sum p_i \log q_i$

where $\sum q_i = 1$.

Proof: $\cancel{\log\frac{q_i}{p_i} \le \frac{q_i}{p_i} - 1}$

$\ln\frac{q_i}{p_i} \le \frac{q_i}{p_i} - 1$

$\Rightarrow -\sum p_i \lg\frac{p_i}{q_i} = \sum p_i \lg\left(\frac{q_i}{p_i}\right)$

$\le \frac{\sum p_i\left(\frac{q_i}{p_i} - 1\right)}{\ln 2}$

$= \frac{\sum q_i - \sum p_i}{\ln 2} = 0$.

Claim: $\frac{H(X)}{\log a} < \mathbb{E}[S] < \frac{H(X)}{\log a} + 1$

LHS: $H(X) = -\sum p_i \log p_i$ Let $q_i = \frac{a^{-l_i}}{\sum a^{-l_i}}$ ✓

RHS: Let ~~$l_i = \lceil -\log_a p_i \rceil$~~

$l_i = \lfloor -\log_a p_i \rfloor + 1$ ✓

Huffman coding is optimal:

Lemma: 1. If $p_i > p_j$, then $l_i \le l_j$.

2. Among codewords with max length, have two differ only by the last digits.

Proof: 1. ✓ o.w. swap.

2. ✓. o.w. delete last digit.

Thm: By induction. Say $\mathbb{E}(C_n) = \mathbb{E}(C_{n-1}) + p_{n-1} + p_n$.

If C_n' optimal, take two that differ only by last digits.

$$C_{n-1}'(\mu_i) = \begin{cases} C_n'(\mu_i), & 1 \le i \le n-2 \\ C_{n-1}'(v) = y. \end{cases}$$

where $C_n'(\mu_{n}) = y0$, $C_n'(\mu_{n-1}) = y1$.
$y \in \{0,1\}^*$.

Then ✓

8. $H(X,Y) = -\sum_{x\in A}\sum_{y\in B} \mathbb{P}(X=x, Y=y) \log(\mathbb{P}(\quad))$.

9. $H(X,Y) \le H(X) + H(Y)$:

Let $\cancel{P_{ij}}\ P_{xy} = \sim$

$P_x = \sim$

$P_y = \sim$.

Note: $\sum\sum P_{xy} = 1$

$\sum\sum P_x P_y = 1$.

Then: $\cancel{-\sum_{x\in A}\sum_{y\in B} \mathbb{P}(X=x, P_{ij} \log P_{ij}}$

$\cancel{-\sum_{x\in A}\sum_{y\in B} P_{ij} \log P_{ij}}$.

$-\sum_{x\in A}\sum_{y\in B} P_{xy} \log P_{xy}$

$\le -\sum_{x\in A,\, y\in B} P_{xy} \log P_x P_y$.

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

Ideal observer: $\max \mathbb{P}(c \text{ sent} \mid \underset{\sim}{x} \text{ received})$

ML maximizer: $\max \mathbb{P}(x \text{ received} \mid c \text{ sent})$

m-d decoding:

Equivalent: if $p < \frac{1}{2}$, then

$\mathbb{P}(x \text{ re} \mid c \text{ sent}) \to n$ digits.

$= p^{d(x,c)} (1-p)^{n-d(x,c)}$

$= (1-p)^n \cdot \left(\frac{p}{1-p}\right)^{d(x,c)}$.

$V(n,r) = |B(x,r)| = \sum_{i=1}^{r} \binom{n}{i}$.

Hamming's bound: $|C| \le \frac{2^n}{V(n,e)}$.

Hamming is perfect:

Hamming (n,d)

$n = 2^d - 1$.

Parity-check H ~~has~~ is $(d \times 2^d - 1)$ matrix;

~~ect~~ each column is non-zero element of $\mathbb{F}_2^d$.

$m = 2^{n-d}$.

$V(n,e) = V(n,1) = 2^d - 1 + 1 = 2^d$.

$$2^{n-d} = \frac{2^n}{2^d}.$$

$A(n,d+1) \le A(n,d)$:

$A(n,d)$: ~~maximal number of codewords s.t.~~ ~~each co~~ $\max\{m : \exists [n,m,d]\text{-code}\}$

Say C is $[n,m,d+1]$-code.

for x,y s.t. $d(x,y) = d+1$.

Let c be that on a certain jth digit c same as y, opposite to x, and other digits same as x.

Then ~~$d(c,m)$~~

for any other z, $\cancel{d(c,z) \le d(c,x) + d(x,z) \le 1 + d(c,x)}$.

$$d+1 \le d(z,x) \le d(z,c) + d(c,x)$$
$$= d(z,c) + 1$$
so $d(z,c) \geq d$.

GSV bound: Note that there does not exist $x \in \mathbb{F}_2^n$ with $d(x,c) \geq d$ for all $c \in C$, otherwise # to ~~m is~~ $|A(n,d)|$ is maximized.

So $\mathbb{F}_2^n \subseteq \bigcup_{c \in C} \bar{B}(c, d-1)$. $\Rightarrow$

C^+ : Parity check $[n+1, m, d/d+1]$

C^- : Puncture $[n-1, m, d/d-1]$

C' : shortened. $[n-1, m', d']$

$d' \geq d$

$\cancel{\frac{1}{4}\log 4 + \frac{3}{4}\log\frac{4}{3} = \frac{1}{2} + \frac{3}{4}\cdot(2-\log_2 3)}$

$\cancel{2-\log}$

---

3. $H(X|Y) = H(X,Y) - H(Y)$:

$$H(X|Y) = \sum_{y\in B} P_y H(X|Y=y)$$

$$= \sum_{y\in B} P_y \sum_{x\in A} \frac{P_{xy}}{P_y} \log \frac{P_{xy}}{P_y}$$

$$= \sum_{y\in B}\sum_{x\in A} P_{xy}\log P_{xy} - \sum_{y\in B} P_y \log P_y$$

4. $H(X|Y) = H(X,Y) - H(Y)$

$$\le H(X) + H(Y) - H(Y) \le H(X)$$

5. $H(X,Y,Z) = H(Z|X,Y) + H(X|Y) + H(Y)$

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

$\cancel{H(X|Y,Z) = H(Z|X,Y) + H(X|Y) - H(Z|Y)}$

$\cancel{\ge H(Z|X,Y) + H(X|Y) - H(Z)}$

$$H(X|Y) = H(X|Y,Z) + H(Z|Y) - H(Z|X,Y)$$

$$\le H(X|Y,Z) + H(Z)$$

6. Fano's inequality: X, Y take values in A, $|A| = m$.

Let $p = \mathbb{P}(X \neq Y)$. Then

$$H(X|Y) \le \underbrace{H(X|Y,Z)}_{p\log(m-1)} + \underbrace{H(Z)}_{H(p)}.$$

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

2nd coding: For DMC,

operation capacity = info capacity.

2017. P2, S$\overline{II}$,

11G. $H(X) = -\sum p_i \log p_i$,

Gibb's: $-\sum p_i \log p_i \leq -\sum p_i \log q_i$.

Proof: $\ln \frac{p_i}{q_i} \leq \frac{p_i}{q_i} - 1$.

$\Rightarrow$ $\cancel{\sum p_i \log \frac{p_i}{q_i}}$

$\cancel{\leq -\sum p_i}$ $\sum p_i \left(\log \frac{q_i}{p_i}\right)$

$\leq \sim$.

$\leq \sim \ldots$ ✓

DMC.

$$\begin{array}{c|ccc} & 0 & 1 & * \\ \hline 0 & 1-\alpha-\beta & \alpha & \beta \\ 1 & \alpha & 1-\alpha-\beta & \beta \end{array}$$

$H(Y|X) = -\left[\alpha \log \alpha + \beta \log \beta + (1-\alpha-\beta)\log(1-\alpha-\beta)\right]$.

$I(X,Y) = H(X) + H(Y) - H(X,Y)$

$= \cancel{H(X)} \cancel{H(Y) - H(Y|X)}$

$\cancel{H(X) - H(Y|X)}$

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

$H(Y|X) = H(X,Y) - H(X)$.

$$I(X,Y) = H(Y) - H(Y|X)$$
$$= H(Y) - [\qquad]$$

$$\begin{matrix} p \\ 1-p \end{matrix} \rightarrow \begin{pmatrix} (1-\alpha-\beta)p+\alpha(1-p) \\ \alpha p+(1-\alpha-\beta)(1-p) \\ \beta \end{pmatrix}$$

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

$p, 1-p$

$$\begin{pmatrix} p-2\alpha p-\beta p+\alpha \\ p(\end{pmatrix} \begin{pmatrix} p(1-2\alpha-\beta)+\alpha \\ p(2\alpha-1-\beta)+(1-\alpha-\beta) \\ \beta \end{pmatrix}$$

$$\beta = p(2\alpha-1-\beta)+(1-\alpha-\beta)$$

$$p = \frac{2\beta+\alpha-1}{2\alpha-1-\beta}.$$

$$\begin{pmatrix} 1-2\beta \\ \beta \\ \beta \end{pmatrix}$$

$$-[(1-2\beta)\log(1-2\beta)+2\beta\log 2\beta]$$

$\cancel{\beta\log\beta - 2\beta\log 2\beta - (1-2\beta)\log(1-2\beta) + 2\beta}$

$$\beta\log\beta + \theta = [\beta\log\beta + (1-\beta)\log(1-\beta) \cancel{+1-\beta}] + (1-\beta).$$

Yes.

let $P_1=\beta$.

$$P_2=P_3=\frac{1-\beta}{2}$$

$W_C(1,0)=1$: $A_n=1$.

$11\cdots 1\in C$.

Plainte

$e: M\times K\to C$

$d: C\times K\to M$.

$|K|=26!$

$|K|=26^d$.

Caesar: $|K|=26$.

Perfect secrecy: $H(M|C)=H(M)$.

~~$e_k(m)=c_0$~~.

Fix m_0 & k_0, with >0 prob, Then $c_0=e_{k_0}(m_0)$ >0 prob.

$\mathbb{P}(C=c_0)=\mathbb{P}(C=c_0|M=m_0)$.

So must $\exists\, k\in K$ with $c_0=e_k(m)$.

If m_1, m_2 same key, then $e_k(m_1)=c_0=e_k(m_2)$

So $m_1=m_2$. So $m\mapsto k$ injective.

One-time-pad:

$\mathbb{P}(M=m, C=c)$

$= \mathbb{P}(M=m, K=c-m)$

$= \mathbb{P}(M=m)\mathbb{P}(K=c-m) = \mathbb{P}(M=m)\frac{1}{q^N}$.

so M, C indep.

$H(M|C) \leq H(K|C)$:

~~$H(M|C) = H$~~

[Note $H(M, C, K) = H(C, K)$
$H(M|C, K) = 0$.

$$H(K|C) = H(K, C) - H(C)$$
$$= H(M, C, K) - \overset{0}{H(M|K, C)} - H(C)$$
$$= H(M, K, C) - H(C)$$
$$= H(K|M, C) + H(M, C) - H(C)$$
$$= H(K|M, C) + H(M|C)$$
$$\leq H(M|C).$$

Unicity: Least n s.t. $H(K|C^{(n)}) = 0$.

$$U := \frac{\log|K|}{\log|A| - H}.$$

$$H(K|C^{(n)}) = H(K, C^{(n)}) - H(C^{(n)})$$
$$= H(K, M^{(n)}, C^{(n)}) - H(C^{(n)})$$
$$= H(K, M^{(n)}) - H(C^{(n)})$$
$$= H(K) + H(M^{(n)}) - H(C^{(n)})$$

as $K, M^{(n)}$ indep.

CC. Noiseless. Day 1.

1. DMC: $P_{ij} = \mathbb{P}(y_j \text{ received} \mid x_i \text{ sent})$

~~the same~~

The same for each channel use and indep of all past and future uses

2. BSC. $A = B = \{0,1\}$.

channel matrix $\begin{array}{c} 0 \\ 1 \end{array} \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$

$$\begin{pmatrix} 1-p & 0 & p \\ 0 & 1-p & p \end{pmatrix}.$$

Code: $c: A \to B^*$.

Decipherable: $c^*: A^* \to B^*$ is injective.

Block code: all words same length

Comma code:

Prefix-free code: No codeword is a prefix of any other distinct word.

$|A| = m, |B| = a.$

$c: A \to B^*.$

$$\begin{matrix} l_1, & l_2, & \cdots, & l_m \\ \| & \| & & \| \\ |c(a_1)| & |c(a_2)| & \cdots & |c(a_m)| \end{matrix}$$

Claim: $\sum_{i=1}^{m} a^{-l_i} \leq 1 \iff$ prefix free exists.

proof: $\cancel{\sum n_s a^{-t}}$ If c is prefix free.

$(\Leftarrow)$ $\sum_{i=1}^{s} n_i a^{-i} \leq 1$

$$\sum_{i=1}^{s} n_i a^{s-i} \leq a^s.$$

total number of strings of length s

number of strings of length s in B with some codeword of c as prefix.

$(\Rightarrow)$ Induction

$$n_1 a^{s-1} + n_2 a^{s-2} + \cdots + n_{s-1} a + n_s \leq a^s.$$

First $s-1$ terms of LHS sum to # strings of length s with a codeword of $\hat{c}$ as a prefix.

McMillan: Any decipherable code satisfy Kraft.

proof: Decipherable: $C^*: A^* \to B^*$ is injective.

~~$c(a) \neq c(b)$~~ ~~$c(a), c(b), \ldots$~~

$s = \max_{1 \le i \le m} l_i$.

$$\left(\sum_{i=1}^{m} a^{-l_i}\right)^R = \sum_{l=1}^{Rs} b_l a^{-l}.$$

b_l: # ways choosing R codewords of total length l.

Any string correspond to ≤ 1

C decipherable $\Rightarrow$ $\Rightarrow b_l \le a^l$.

So $\left(\sum_{i=1}^{m} a^{-l_i}\right)^R \le Rs$

$$\sum_{i=1}^{m} a^{-l_i} \le (Rs)^{1/R} \to 1 \text{ as } R \to \infty.$$

Cor: $H(p_1, p_2, \dots, p_n)$

$$= -\sum p_i \lg p_i$$

$$\leq -\sum p_i \lg \frac{1}{n}$$

$$= \log n.$$

Gibb's inequality: $-\sum p_i \lg p_i < -\sum p_i \lg q_i.$

$$e^{-x} < 1-x$$

$$\ln(1-x) < -x.$$

$$\ln(x) < x-1.$$

$$\ln\frac{q_i}{p_i} \leq \frac{q_i}{p_i} - 1.$$

$$-\sum p_i \lg \frac{p_i}{q_i}$$

$$\leq -\sum p_i \left(\frac{q_i}{p_i} - 1\right)$$

$$= 0.$$

Shannon's Noiseless Coding Thm:

$$\frac{H(X)}{\log a} \le E[S] < \frac{H(X)}{\log a} + 1$$

LHS: $\cancel{H(x) \le}$

$\cancel{P_i\, C\, C}$

$E[S] = P_i l_i$

Let $q_i = \frac{a^{-l_i}}{D}$ $\Rightarrow \sum q_i = 1$

$\cancel{\Rightarrow l_i = Dq_i}$

$\cancel{e^{q_i D} = a^{-l_i}}$

$\cancel{P_i \log}$ $\cancel{P_i l_i = P_i\left(-\frac{Dq_i}{\log a}\right)}$ $\cancel{-l_i \log a = Dq_i \quad l_i = -\frac{Dq_i}{\log a}}$

$$\log(a)\left[\sum P_i l_i\right] = \sum P_i(\log q_i + \log D)$$

$$\ge -\sum P_i \log P_i - \log D$$

$$\ge -\sum P_i \log P_i \quad \checkmark \text{ as } D \le 1$$

$$= H(X)$$

$\log q_i D = -l_i \log a$

$l_i = -\frac{\log q_i D}{\log a}$

RHS: Take $l_i = \lfloor \log_a P_i \rfloor + 1 = \lfloor \frac{\log P_i}{\log a} \rfloor + 1$

Then $\sum P_i l_i \le \sum P_i\left(\frac{\log P_i}{\log a} + 1\right)$

$$= \frac{H(X)}{\log a} + 1 \quad \checkmark$$

Huffman coding is optimal

Optimal code:

Lemma: ① If $p_i > p_j$, then $l_i \le l_j$.

True, as otherwise swap ~~$c(m_i)$~~, ~~$c(m_j)$~~ $c(a_i)$, $c(a_j)$.

② For codewords with maximal length exists two that differ only by the last digits.

True. Otherwise can cut the last digit for some. still legit code.

Then: By induction.

Smallest 2 to be ~~p_i, p_j~~ p_n, p_{n+1}

~~Differ by last digit.~~ ~~cons~~

$\mathbb{E}[S_{n+1}] = p_n + p_{n+1} + \mathbb{E}[S_n]$ by using Huffman

If not optimal S_{n+1} ~~is~~ optimal (with c'_n)

Then ~~①~~ Wlog last two codewords max length, differ by last digit. in final position.

Let c'_{n-1} ~~(M_i)~~ be s.t. $c'_{n-1}(M_i) = \begin{cases} c'_n(M_i) & 1 \le i \le m-2 \\ c'_n(V) = y \end{cases}$

$\mathbb{E}[S'_{n+1}] = \mathbb{E}[S'_n] + p_n + p_{n+1}$. $\mathbb{E}[S'_n] \le \mathbb{E}[S_n]$. Done

$\cancel{H(X,Y) = \sum}$

$H(X) = -\sum p_i \log p_i$

$H(Y) = -\sum_{B} q_j \log q_j .$

$A = \{x_1, \dots, x_m\}$

$B = \{y_1, \dots, y_n\}$

Note: Let $p_{ij} = \mathbb{P}(X = x_i, Y = y_j)$

$\cancel{H(X,Y) = -\sum_X \sum_Y}$

$$H(X,Y) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \log p_{ij}$$

$$\le -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \log p_i q_j$$

$$\le -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \log p_i - \sum_{j=1}^{n} \left(\sum_{i=1}^{m} p_{ij} \log q_j\right)$$

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

CC 2021.

P1, SII, 11K.

$$H(X) = -\sum P_i \log P_i.$$

$$E(S) = \sum_{i=1}^{N} P_i \, c(M_i)$$

Decipherable binary code.

~~wlog~~ $\exists$ Prefix free code with word length $s_1, \ldots, s_N$.

~~$s_1 + \cdots + s_N$~~

$$\frac{1}{N} \sum s_i \geq H(X)$$

$$= \log N. \checkmark$$

2019. P2, SII.

P2, SII, 12G.

(i) X. $\frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6}$.

(ii) X. About $\frac{1}{4}$.

(iii) ~~V. Induction.~~ X.

CC 2017.

P2, SI. 3G. $l_1, l_2, \dots, l_n$ satisfy

① Prefix-tree exists iff Kraft's. [Re-write]

② Decipherable code satisfy Kraft.

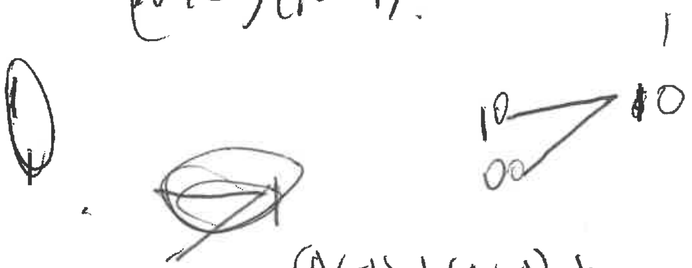
$[C(\ )^R = \sim \ ]$

$\le Rs.$

2016. P1, SI. 3G.

$l_1 + \dots + l_N \le \frac{1}{2}(N^2+N-2)$:

$(N+2)(N-1)$.



$(N-1)+(N-1)+\dots$

$= \frac{N(N-1)}{2} + (N-1)$

$= \frac{(N+2)(N-1)}{2}$ ✓.

CC Day 2.

Noisy Channels.

1. Binary $[m,n]$-code:
   size $m = |C| \subseteq \{0,1\}^n$
   n: length of code

2. Hamming distance: $d(x,y) = |\{i : 1 \le i \le n, x_i \neq y_i\}|$

3. Ideal observer: decode $\max \mathbb{P}(c \text{ sent} \mid x \text{ received})$
   ML decoding: $\max \mathbb{P}(x \text{ received} \mid c \text{ sent})$
   min din: minimize $d(x,c)$.

   If $p < \frac{1}{2}$ then (ii)(iii) equivalent:

Hamming's bound: e-error correcting code of length n has

$$m \le \frac{2^n}{\sum_{i=0}^{e} \binom{n}{e}} = \frac{2^n}{V(n,e)}.$$

perfect: $2^n = m V(n,e)$.

~~Hamming:~~

3. Hamming:

Parity ~~ch~~ check H: $d \times (2^d - 1)$ matrix. $\nearrow n$

columns are elements of ~~basis~~ $(\mathbb{F}_2^d)^*$

Code-length: $n = 2^d - 1$

$m = |C| = 2^{n-d}$ as $(n-d)$ dimensions are free

$e = 1$ as $d = 3$ [any 3 are L.D, 2 are not.]

$V(n,e) = n+1 = 2^d - 1 + 1 = 2^d$.

$\therefore 2^{n-d} = \frac{2^n}{2^d}$

$m = \frac{2^n}{V(n,e)}$ ✓.

So Hamming code is perfect. ✓

4. $A(n, d+1) \le A(n,d)$:

$A(n,d) = \max\{m : \exists [n,m,d]\text{-code}\}$.

If C is $[n,m,d+1]$-code, then

for x, y s.t. $d(x,y) = d+1$,

~~for~~ ~~change~~ change x to z s.t. differ in a digit i which $x_i \ne y_i$, $z_i = y_i$.

Then ~~$d(x,$~~ $d(y,z) = d$.

$d(z,t) \ge d(x,t) - 1 \ge d$.

So done.

Let x be z.

C' is $[n,m,d]$-code.

5. $\underbrace{\frac{2^n}{V(n,d-1)} \le}\; |A(n,d)| \; \underbrace{\le \frac{2^n}{V(n, \lfloor \frac{d-1}{2} \rfloor)}}_{\text{Hamming's bound.}}$

~~$B(x, d-1)$~~

Union of ~~any~~ ball of a codeword cover whole.

$\mathbb{F}_2^n \subseteq \bigcup_{c \in C} \bar{B}(c, d-1)$

$\Rightarrow 2^n \le \sum_{c \in C} |\bar{B}(c, d-1)| = mV(n, d-1)$.

---

6. C^+: Parity check.

$[n+1, m, d']$ code.

$\{(c_1, c_2, \ldots, c_n, \sum_{i=1}^{n} c_i) : (c_1, c_2, \ldots, c_n) \in C\}$

C^-: Puncture

$[n-1, m, d']$ code, $d-1 \le d' \le d$.

$\sim$

C': Shortened cod.

$[n-1, m', d']$, $d' \ge d$, $[m' \ge \frac{m}{2}$ for some choice of $\alpha)$

$\{(c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_n) : (c_1, \ldots, c_{i-1}, \alpha, c_{i+1}, \ldots, c_n) \in C\}$,

1. Transmit reliably at rate R,

$\exists$ ~~$c_1, c_2, \dots$~~ $(C_n)_{i=1}^{\infty}$ s.t.

each C_i of length n,

$$\lim_{n\to\infty} \rho(C_n) = R,$$

$$\lim_{n\to\infty} \hat{e}(C_n) = 0.$$

$$\rho(C) = \frac{1}{n}\log_2(M).$$

Operational capacity $=$ sup{reliable transmition rate}.

3) ~~$H(X,Y) = \sum_{i=1}$~~

$$H(X|Y) = \cancel{\sum_{i=1}^{m}} \sum_{j=1}^{m} q_j \cancel{\sum} \cancel{H(X|Y=y_j)}$$

$$= \sum_{j=1}^{m} q_j$$

3 -

Claim: $H(X,Y) = H(X|Y) + H(Y)$:

$$H(X|Y) = \sum_{y\in B} \mathbb{P}(Y=y)\, H(X|Y=y)$$

$$= \sum_{y\in B} \mathbb{P}(Y=y)\left[-\sum_{x\in A} \mathbb{P}(X=x|Y=y)\log \mathbb{P}(X=x|Y=y)\right]$$

$$= -\sum_{y\in B} \mathbb{P}(X=y) \sum_{x\in A} \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)} \log \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)}$$

$$= -\sum_{y\in B}\sum_{x\in A} \mathbb{P}(X=x, Y=y)\log \mathbb{P}(X=x, Y=y)$$

$$+ \sum_{y\in B}\sum_{x\in A} \mathbb{P}(X=x, Y=y)\log \mathbb{P}(Y=y)$$

$$= H(X,Y) - H(Y).$$

Thus, since we also have

$$H(X,Y) \le H(X) + H(Y) \quad [\text{Proof by Gibbs}],$$

we have $H(X|Y) \le H(X)$

Claim: $H(X,Y) \le H(X|Y,Z) + H(Z)$.

Proof: $H(X,Y,Z) = H(Z|X,Y) + H(X|Y) + H(Y)$

$$\| $$

$$H(X|Y,Z) + H(Z|Y) + H(Y)$$

$$\Rightarrow H(X|Y) = H(X|Y,Z) + H(Z|Y) - H(Z|X,Y)$$

$$\le H(X|Y,Z) + H(Z)$$

Fano's inequality. X, Y r.v. takes values in A, $|A| = m$.

$$H(X|Y) \le H(p) + p\log(m-1).$$

Let $Z = \begin{cases} 1, & X \ne Y \\ 0, & X = Y. \end{cases}$ $p = \mathbb{P}(X != Y)$.

$$H(X|Y) \le H(X|Y,Z) + \underbrace{H(Z)}_{= H(p).}$$

$$\le p\log(m-1) + H(p)$$

when $Z = 0$, X determined

$Z = 1$, X has $m-1$ choises,

so $H(X|Y=y, Z=1) \le \log(m-1)$.

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

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

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

$= I(Y;X)$

≥ 0, $=$ iff X, Y indep

Inter capa $= \max_{x} I(X;Y)$

CC 2022

P2.55. DMC: P_{ij} ~~$= \mathbb{P}(y_i \text{ recev} | x_i \text{ sent})$~~

~~fixed~~

$P_{ij} = \mathbb{P}(b_j \text{ received} | a_i \text{ sent})$

Two DMCs. Product channel

$\max_{X \times X \in A \times A} I(Y \times Y ; X \times X)$

$= \max H(X \times X) + H(Y \times Y) - H(X \times X, Y \times Y)$

$= -\sum_{x_1, x_2 \in A^2} P_{x_1} P_{x_2} \log P_{x_1} P_{x_2} - \sum_{y_1, y_2 \in B^2} \cdots + \sum P_{x_1 y_1} P_{x_2 y_2} \log P_{x_1 y_1} P_{x_2 y_2}$

$=$ ~~$\max_{X} I(Y|X)$~~ ~~$I(X,Y)$~~

$C_1 + C_2$.

2020 P2, SI.

3I. CC.

(a) Info capasity $= \cancel{I}\max_X I(X,Y)$.

where Y follows distribution of a DMC after X.

(b)
$$\begin{array}{c|ccc} & A & B & * \\ \hline A & \frac{1}{2} & 0 & \frac{1}{2} \\ B & 0 & \frac{1}{2} & \frac{1}{2} \end{array}$$

$I(Y|X) = H(\tfrac{1}{2})$.

$I(X,Y) = \cancel{I(Y|X) = H(}\ H(Y) - H(Y|X)$

$= H(Y) - H(\tfrac{1}{2})$.

$\cancel{Y \le \log 2}$ $H(Y)$: $\frac{\alpha}{2}, \frac{1-\alpha}{2}, \frac{1}{2}$.

$H(Y) = [\frac{1}{2}\log\frac{1}{2} + \frac{\alpha}{2}\log\frac{\alpha}{2} + \frac{1-\alpha}{2}\log\frac{1-\alpha}{2}]$

$= \cancel{1\log}\ \frac{1}{2}\log 2 - \frac{\alpha}{2}\log\alpha$
$- \frac{1-\alpha}{2}\log(1-\alpha)$
$+ \frac{\alpha}{2}\log 2 + \frac{1-\alpha}{2}\log 2$.

$= 1 + \frac{H(\alpha)}{2}$

$\geq 1 + \frac{1}{2}$ when $\alpha = \frac{1}{2}$.

Thus $I(X,Y) \le \frac{1}{2}$.

$$P = \begin{array}{c c} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \end{array}$$

$X : \alpha$, $Y : 1-\alpha$.

$$I(X,Y) = H(Y) - H(Y|X)$$
$$= H\left(\alpha + \frac{1-\alpha}{2}\right) - \alpha \cdot 0 - (1-\alpha) \cdot H\left(\frac{1}{2}\right)$$
$$= H\left(\frac{1+\alpha}{2}\right) - (1-\alpha)$$

$$\cancel{\frac{d\, I(X,Y)}{d\alpha} = \frac{d}{d\alpha}\left[-\frac{1+\alpha}{2}\log\frac{1+\alpha}{2} + \frac{1-\alpha}{2}\log\frac{1-\alpha}{2} + (1-\alpha)\right]}$$
$$\cancel{=}$$

Let $p = \frac{1+\alpha}{2}$ Then $\alpha = 2p - 1$

$\frac{d\alpha}{dp} = 2.$

$$\frac{d}{dp} I(X,Y) \cancel{= \log\frac{p}{1-p} \cdot 2 + 2}$$

$$= 2\log_2\frac{p}{1-p} + 2.$$

$$2\log\frac{p}{1-p} + 2 = 0$$
$$\log\frac{p}{1-p} = -1.$$

$$\frac{p}{1-p} = \frac{1}{2}.$$
$$2p = 1-p.$$
$$p = \frac{1}{3}.$$

So $I(X;Y) = H\left(\frac{2}{3}\right) - \frac{2}{3}.$

$$= \log 3 - \frac{4}{3}.$$

$$H(P_1, P_2, P_3) \le H(P_1, 1-P_1) + (1-P_1)$$

Proof: $P_1 \log P_1 + P_2 \log P_2 + P_3 \log P_3$

$$\le P_1 \log P_1 + P_2 \log \frac{1-P_1}{2} + P_3 \log \frac{1-P_1}{2}.$$ ✓

CC 2021

P3, S1. 3K.

Haming code of length 2^d-1:

$n=2^d-1$.

H 是 $d \times (2^d-1)$ matrix.

s.t. each column is distinct non-zero elements in $\mathbb{F}_2^d$.

① Perfect: $V(n,e) \cdot m = 2^n$.

$m = 2^{n-d}$.

$V(n,1) = 2^d-1+1 = 2^d$.

② 1-error detecting as $d=3$:

Any 2 column LI,

3 LD.

$111\cdots1 \in C$, as

all rows of H sum to 0

(even number of 1's)

$\frac{2^d-1+1}{2}$ 1's ~~actually~~

CC 2021

P2, SII, 12k.

$M: \mathbb{F}_2^d \to \mathbb{F}_2^d$

$$\begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_{d-1} \end{pmatrix} \to \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{d-1} \\ f(c_0, c_1, \ldots, c_{d-1}) = c_d \end{pmatrix}$$

$$\begin{pmatrix} c_0 & c_1 & \cdots & c_{k-1} \\ c_1 & c_2 & & c_k \\ \vdots & \vdots & & \vdots \\ c_{d-1} & c_d & & c_{k+d-2} \end{pmatrix} = \begin{pmatrix} c_0 & c_1 & & c_{k-1} \\ c_1 & c_2 & & c_0 \\ \vdots & \vdots & & c_1 \\ & & & \vdots \\ c_{d-1} & c_d & & c_{d-2} \end{pmatrix}$$

If ~~c_0~~ $c_0 a_0 + c_1 a_1 + \cdots + c_{k-1} a_{k-1} = 0$

$c_1 a_0 + c_2 a_1 + \cdots + c_0 a_{k-1} = 0$

Apply ~~M to rows of H.~~

f to rows of H.

obtain $(c_d \; c_{d+1} \cdots c_{d-1})$

Then ...

(c) If x odd weight,

$\cancel{X^n - 1}$

$$(X^{n-1} + X^{n-2} + \dots + 1)(X-1) = 0$$

$\cancel{\frac{X^n - 1}{X - 1}}$ $\cancel{(X-1)}$

$X-1 \mid g \rightarrow$ generator poly.

CC 2018.

P1, S II. 11H.

1. $\mathrm{rank}(C_1|C_2) = \mathrm{rank}(C_1) + \mathrm{rank}(C_2)$

$\{(x|x)\}\ \{(0,y)\}$ basis.

2. $d(C_1, C_2) = \min\{2d(C_1), d(C_2)\}$.

3. $C_1 : P_1 \quad k_1 \times n$

$C_2 : P_2. \quad k_2 \times n.$

$C_1 | C_2 :$ $\cancel{\begin{pmatrix} P_1 & P_2 \end{pmatrix}}$.

$\cancel{\begin{pmatrix} P_1 & P_1 P_2 \end{pmatrix}}$. $\begin{pmatrix} P_1 & P_1 P_2^T \end{pmatrix}$

$RM(d,r) = (R(d-1,r)) \mid RM(d-1,r-1)$.

$\underset{RM(d,r)}{\mathrm{rank}} = \sum_{s=0}^{r} \binom{d}{s}$

CC 2017

P1, SII, 10G.

$W_C(s,t) = \sum_{j=0}^{n} A_j t^n \left(\frac{s}{t}\right)^j$.

$W_C(1,1) = \sum_{j=0}^{n} A_j = |C| = 2^k$.

Claim: $W_C(s,t) = W_C(t,s) \Leftrightarrow W_C(1,0) = 1$.

proof: $(\Leftarrow)$ $W_C(1,0) = 1 \Rightarrow A_n = 1$.

$\Rightarrow 11\ldots1 \in C$.

Claim: $A_j = A_{n-j}$.

Proof: $\cancel{11\ldots1+x}$

If $x \in A_j$, $11\ldots1+x \in A_{n-j}$.

Bijection. $\square$.

So $W_C(s,t) = W_C(t,s)$.

$(\Rightarrow)$ If $W_C(s,t) = W_C(t,s)$, then $W_C(1,0) = W_C(0,1) = 1$. ✓

Dual code $C^\perp$ of C: $\{y : x \cdot y = 0 \ \forall x \in C\}$

(i) $\underline{y} \in \mathbb{F}_2^n$. If $\underline{y} \in C^\perp$, then $\sum_{x \in C} (-1)^{x \cdot y} = |C| = 2^k$.

Else, $\cancel{x \cdot y}$ Take $\cancel{x \cdot y = 1}$ $\underline{s} : \underline{s} \cdot \underline{y} = 1$.

Then $C \to C$

$\underline{x} \mapsto \underline{x} + \underline{s}$ bijection.

So same number of $\underline{x} \cdot \underline{y} = 0$ or 1 ✓.

Extend def of weight:

$w(\underline{y})$ for $y \in \mathbb{F}_2^n$.

$$\sum_{y \in \mathbb{F}_2^n} t^{w(y)} (-1)^{x \cdot y} = (1-t)^{w(x)} (1+t)^{n-w(x)}$$

① $$\sum_{x \in C} \left(\sum_{y \in \mathbb{F}_2^n} (-1)^{x \cdot y} \left(\frac{s}{t}\right)^{w(y)} \right)$$

$$= \sum_{y \in \mathbb{F}_2^n} \left(\frac{s}{t}\right)^{w(y)} \left(\sum_{x \in C} (-1)^{xy} \right)$$

$$= \sum_{y \in C^\perp} \left(\frac{s}{t}\right)^{w(y)} \cdot 2^k$$

$$= 2^k \sum_{y \in C^\perp} \left(\frac{s}{t}\right)^{wy}.$$

② $$\sum_{x \in C} \left(\sum_{y \in \mathbb{F}_2^n} (-1)^{x \cdot y} \left(\frac{s}{t}\right)^{w(y)} \right)$$

$$= \sum_{x \in C} \left(1 - \left(\frac{s}{t}\right)\right)^{w(x)} \left(1 + \left(\frac{s}{t}\right)\right)^{n-w(x)}$$

$$2^k W_{C^\perp}(s,t) = W_C(t-s, t+s) \checkmark.$$

CC Day 3.

Hamg $(7,4)$ - code
$\uparrow$ rank

$= \text{Hamg } [7, 16, 3]$ - code

$n = 2^d - 1$, $\quad 16 = 2^4$

$d = 7 - 4 = 3$.

a.s.

$\mathbb{P}(\|X_n - X\|_2 \to 0 \ \forall n) \to 1$

prob.

$\mathbb{P}(\|X_n - X\|_2 \geq \varepsilon) < \delta$.

$\forall \varepsilon > 0 \ \exists M(\varepsilon) < \infty$ s.t. $\qquad \theta \xrightarrow{d} z$

$\mathbb{P}(\|\theta\|_2 > M(\varepsilon)) < \varepsilon$.

Feed back shift registe

FSR is a map $f: \mathbb{F}_2^d \to \mathbb{F}_2^d$ given by

$f(x_0, \dots, x_{d-1}) = (x_1, x_2, \dots, x_{d-1}, C(x_0, \dots, x_{d-1}))$,

where $C: \mathbb{F}_2^d \to \mathbb{F}_2$.

Say register has length d.

Stream associated to an initial fill $(y_0, \dots, y_{d-1})$

is infinite sequence $(y_0, y_1, \dots, y_j, \dots)$,

with $y_n = C(y_{n-d}, y_{n-d+1}, \dots, y_{n-1})$ for all $n \geq d$.

If $C(x_0, x_1, \dots, x_{d-1}) = \sum_{i=0}^{d-1} a_i x_i$,

Say it's LFSR.

$$y_n = \sum_{i=0}^{d-1} a_i y_{n-d+i}$$

Auxillary Poly $X^d + a_{d-1}X^{d-1} + \dots + a_1 X + a_0$

Feedback poly $\check{P}(X) = a_0 X^d + \dots + a_{d-1} X + 1$

CC Day 4

3. Perfect secrecy:

$$H(M|C)=H(M).$$

4. Perfect $\Rightarrow |K| \geq |M|$.

$\forall\, m \in M$ have $\mathbb{P}(C=c_0)=\mathbb{P}(C=c_0 \mid M=m)$.

So $\exists$ key $k \in K$ with $c_0=e_k(m)$.

If m_1, m_2 give same key k,

then $e_k(m_1)=c_0=e_k(m_2)$, so $m_1=m_2$.

So $m \mapsto k$ injective.

5. $H(M|C) \leq H(K|C)$

Note $M=d(C,K)$, so

$H(M|C,K)=0$, $H(C,K)=H(M,C,K)$

$$H(K|C)=H(K,C)-H(C)$$

$\cancel{\leq H(M,C,K) \pm H(M|K,C) \pm H(C)}$

$= H(M,K,C)-H(C).$

$= H(K|M,C)+H(M,C)-H(C)$

$= H(K|M,C)+H(M|C)$

$\leq H(M|C)$

$$U = \frac{\log|K|}{\log|A| - H}.$$

Rabin:

① Private key: $p, q \equiv 3 (4)$ primes

② Public key: $N (= pq)$.

Encoding: $c \mapsto c^2 \pmod N$.

Decoding: Receive c.

$x^2 \equiv c \ (pq)$

$\Rightarrow x^2 \equiv c \ (p)$

$x^2 \equiv c \ (q)$.

$p = 4k_1 - 1$.

$\cancel{c^k \equiv x^{2k}}$

Claim. $(c^k)^2 \equiv c \ (p)$

Proof. $c^{2k} \equiv x^{4k} \equiv x^2 \equiv c \ (pq)$. ✓.

So $x \equiv \pm c^k \ (p)$

$x \equiv \pm c^k \ (q)$.

By CRT, Done.

Thm: Breaking Rabin as difficult as factoring N.

Proof: $(\Leftarrow)$ ✓

$(\Rightarrow)$ Say we have algo to compute $f(x)$ square root mod N.

Pick $x \pmod N$ randomly.

$y = f(x)$

$y^2 \equiv x^2$.

with $P = \frac{1}{2}$, $x \not\equiv \pm y (N)$,

so $(N, x-y)$ is a non-trivial factor of N.

Fails, do again. ✓

RSA:

Private: d.

Public: N, e.

$N = pq \to$ large p, q.

e random, $(e, \varphi(N)) = 1$.

$de \equiv 1 (\varphi(N))$.

$e: m \mapsto m^e$

$d: c \mapsto c^d$.

CC. 2018.

P2, 5II, 12H.

$O_p(x)$: $\mathrm{ord}(x)$ in $\mathbb{F}_p^*$.

$\phi(N) \mid 2^a b$. $(p-1)(q-1) \mid 2^a b$.

If $O_p(x^b) \neq O_q(x^b)$,

$$x^{2^a b} \equiv 1 \ (p)$$
$$x^{2^a b} \equiv 1 \ (q).$$

< Since ~~1 or~~ only 1 & $(-1)^2 = 1$.

$\exists\ 2^t b$, $x^{2^t b} \equiv 1 \ (p)$
$x^{2^t b} \not\equiv 1 \ (q)$

$\Rightarrow p-1 \mid 2^t b$, $q-1 \nmid 2^t b$

Claim: number of x satisfying $O_p(x^b) \neq O_q(x^b) \geq \frac{\phi(N)}{2}$.

Proof: S.T.S. For each possible value of $O_p(x^b) = k$

~~number of x satisfying O~~

number of x ~~of this~~ s.t. $O_p(x^b) = k$

$\leq \frac{p-1}{2}$.

Let g be primitive root of P.

$g^{P-1} \equiv 1 (P)$, $O_P(g^b)$ is of form 2^k

suppose $O_P(g^b) = 2^t$ $(0 \le t \le a)$

Let $x = g^k$, then $x^b = g^{bk}$ for odd b

~~$x^b = g^{bk}$, $O_P(x^b) = \frac{2^t}{(2^t, k)}$~~ $O_P(x^b) = \frac{2^t}{(2^t, k)}$

~~So $O_P(x^t) = 2^t$ iff~~

So ~~$O_P(x^b) =$~~ $O_P(x^b) = 2^t$ iff k is odd

$$O_P(x^b) = O_P(g^{bk}) \begin{cases} = 2^t & k \text{ odd} \\ < 2^t & k \text{ even} \end{cases}$$

So: for all $x = g^k$, k odd ~~even~~,

$O_P(x^b) = 2^t$.

It has size $= \frac{P-1}{2}$.

Others have size $\le \frac{P-1}{2}$.

Done.

Thus By CRT, $\ge \frac{P-1}{2}(q-1)$

$= \frac{1}{2}\varphi(N)$.

CC 2016 P3

Unicity distance:

least n s.t. $H(K|C^{(n)})=0$.

i.e. smallest n to uniquely determine key

Assume:

i) All messages ~~indep~~ same prob

ii) all ~~do~~ codewords same prob

(iii) $H(M^{(n)}) \sim nH$ for some H const, n large.

Then: $H(K|C^{(n)})=0$

$$\begin{aligned} &\;\;\| \\ &H(K,C^{(n)}) - H(C^{(n)}) \\ &= \cancel{H(K)}\; H(K, M^{(n)}, C^{(n)}) - H(C^{(n)}) \\ &= H(K, M^{(n)}) - H(C^{(n)}) \\ &= H(K) + H(M^{(n)}) - H(C^{(n)}) \\ &= \log|K| + \cancel{n\log}\; nH - n\log|\Sigma| \; \cancel{-n\log|A|} \end{aligned}$$

Since $K, M^{(n)}$ indep.

$\Rightarrow$ $\cancel{n = \frac{}{\log|A| - H}}$

$$n = \frac{\log|K|}{\log|\Sigma| - H}$$