ICC Review

-- Sugar He

Quick Review

Chain Rule

$H\left(U_1, U_2, \cdots, U_N\right)=\sum_{n=1}^N H\left(U_n \mid U_1, U_2, \cdots, U_{n-1}\right)$

Joint MI

$I(X;Y,Z) = I(X;Z) + I(X;Y|Z)$

Fano

$H(X|Y) \le H(P_E) + P_E \log(K-1)$

Power Limit

$H_C(X) \le \frac{1}{2}\log(2\pi e\sigma^2)$

Redundancy

$R = 1 - \frac{H_\infty}{\log K}$

Kraft

$$\sum_{k=1}^{K} D^{n_k} \le 1$$

Src Coding Thm (variable length)

$H(U) \le R = \overline{n}\log D \le H(U) + \frac{\log D}{L}$

AGC (discrete time)

$C = \frac{1}{2}\log(1+\frac{P}{N})$

AGC (continuous time)

$C = W\log(1+\frac{P}{N_0 W})$

$\eta_s = \frac{R}{W},\ \eta_p = \frac{R}{P}$

RDF Support

$$D_{min} = \sum_x p(x)\min_{\hat{x}} d(x,\hat{x})$$

$D_{max} = E[d(X,\hat{x}^*)]$

2 Basic

2.1 self info

$I(X=x_k) \triangleq I(x_k) = -\log p(x_k)$

- 对事件x_k的不确定性
- 确证事件x_k的代价

2.1.1 condition self info

$I(x_k|y_k) = -\log p(x_k|y_j)$

- after y_j happens, info provided by x_k

2.1.2 joint self info

$I(x_k, y_j) = -\log p(x_k, y_j)$

2.1.3 self MI

$I(x_k; y_j) = \log \frac{p(x_k|y_j)}{p(x_k)} = I(x_k) - I(x_k|y_j)$

- info about x_k that contained in y_j

$I(x; y|z) = \log \frac{p(x|y,z)}{p(x|z)} = \log \frac{p(x,y|z)}{p(x|z)p(y|z)}$

$I(x; y, z) = \log \frac{p(x|y,z)}{p(x)} = I(x; y) + I(x; z|y)$

2.2 Entropy

avg self info

$H(X) \triangleq -\sum\limits_x p(x) \log p(x)$

- avg uncertainty of a r.v.
- avg info from observation
- avg cost to verify r.v.

2.2.1 Equivocation

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

- when Y is received, the uncertainty about X, which is the **noise**
- $Y = f(X) \implies H(X|Y) = 0$, no noise

2.2.2 Joint Entropy

$H(X, Y) = -\sum\limits_{x,y} p(x, y) \log p(x, y)$

Chain Rule

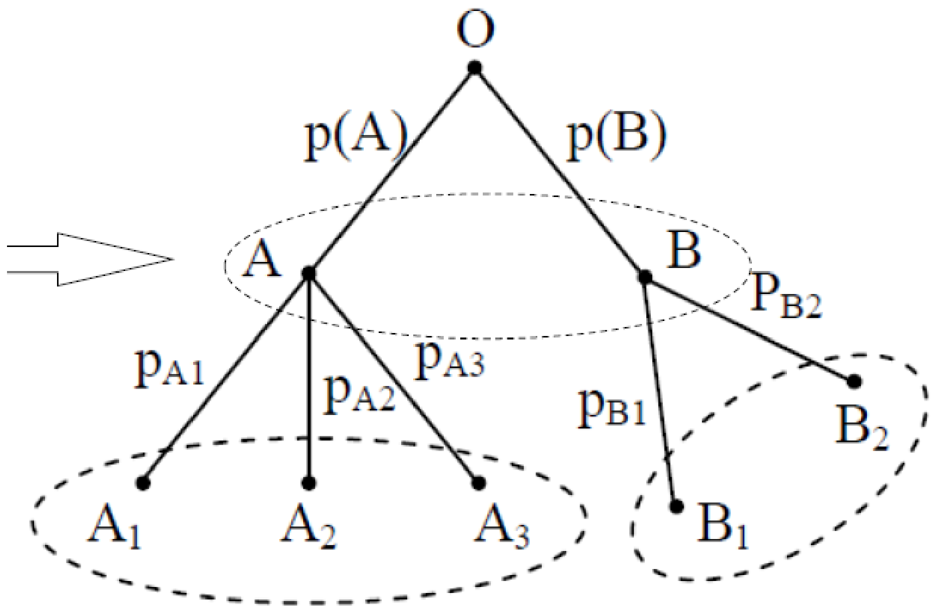
$H(U_1, U_2, \cdots, U_N) = \sum_{n=1}^{N} H(U_n \mid U_1, U_2, \cdots, U_{n-1})$

2.2.3 Property of Entropy

1. $H(X) \geq 0$
2. $\lim\limits_{\epsilon \to 0} H(p_1, \cdots, p_K - \epsilon, \epsilon) = H(p_1, \cdots, p_K)$
3. **additive**

   $H(X_2) = H(X_1) + H(X_2|X_1)$

   $H(X_2|X_1) = \sum\limits_{x_1} p(x_1) H(X_2|x_1)$ is the weighted sum over sub choices



$H(X_1) = H(p(A), p(B))$

$H(X_2) = H(p(A_1), \cdots, p(B_2))$

conditional prob $p_{A_1} = p(A_1|A)$

$H(X_2) = H(X_1) + p(A)H(p_{A_1}, \cdots, p_{A_3}) + p(B)H(p_{B_1}, p_{B_2})$

4. $H(p_1, \cdots, p_K) \leq -\sum\limits_{k=1}^{K} p_k \log q_k,\ \ q_k$ is any other prob. dist.

   proof: $\sum_{k=1}^{K} p_k \log \frac{q_k}{p_k} \leq \log e \cdot \sum_{k=1}^{K} p_k \left(\frac{q_k}{p_k} - 1\right) = 0$

5. $H(p_1, \cdots, p_K) \leq \log K$
6. $H(p_1, \cdots, p_K)$ is concave in p_i

2.3 MI

$I(X;Y) = \sum\limits_{x,y} p(x,y) \log \frac{p(x|y)}{p(x)} = \sum\limits_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$

$I(X;y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y)$

- info of X minus equivocation, is the info of X left in Y
- the effective info that is not polluted by noise
- the info of X that Y provided

2.3.1 Property of MI

- $I(X;Y) \geq 0$
- $H(X;Y) \leq H(X)$
- **convexity** $I(X;Y) = \sum\limits_{x,y} p(x)p(y|x) \log \frac{p(y|x)}{\sum\limits_{x} p(x)p(y|x)} = f(p(x), p(y|x))$
  - given $p(y|x), I(X;Y)$ is concave in $p(x)$
  - given $p(x), I(X;Y)$ is convex in $p(y|x)$

2.3.2 conditional MI

$I(X;Y|Z) = \sum\limits_{x,y,z} p(x,y,z) \log \frac{p(x|y,z)}{p(x|z)}$

2.3.3 joint MI

$I(X;Y,Z) = \sum\limits_{x,y,z} p(x,y,z) \log \frac{p(x|y,z)}{p(x)} = I(X;Z) + I(X;Y|Z)$

- info of X that Y & Z provide is Z provides plus given Z that Y provides

2.3.4 KL Divergence

$D(p\|q) = \sum\limits_{x} p(x) \log \frac{p(x)}{q(x)}$

- $D(p\|q) \geq 0$

2.3.5 Fano Inequality

error prob. $P_E = p(X \neq Y)$

$$H(X|Y) \leq H(P_E) + P_E \log(K-1)$$

proof:
$H(X|Y) = H(p(X=Y), p(X \neq Y)) + P_E H(X|X \neq Y) + (1-P_E) H(X|X=Y) = H(P_E) + P_E H(X|X \neq Y)$

2.4 Markov Chain

$X \to Y \to Z$

$p(x,z|y) = p(x|y)p(z|y) \implies I(X;Z|Y) = 0$

2.4.1 Data Process Inequality

$I(X;Y) \geq I(X;Z)$

$I(X;Y) \geq I(X;Y|Z)$

2.5 Differential Entropy

for continuous r.v.

2.5.1 MI

$I(X;Y) = \iint p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x) p_Y(y)} dxdy$

all the same

2.5.2 Entropy

$H_C(X) = -\int p_X(x) \log p_X(x) dx$

$H_C(X|Y) = -\iint p_{XY}(x,y) \log p_{X|Y}(x|y) dxdy$

$H_C(X,Y) = -\iint p_{XY}(x,y) \log p_{XY}(x,y) dxdy$

$H_C(U^N) = H_C(U_1, \cdots, U_N) = \sum\limits_{n=1}^{N} H_C(U_n | U_1 \cdots U_{n-1}) = \sum\limits_{n=1}^{N} H_C(U_n | U^{n-1})$

2.5.3 Maximum

use Lagrange Multiplier

1. maximum amplitude

   $X \in [-M, M],\ \int_{-M}^{+M} p(x)dx = 1$

   $H_C(X) \le \ln(2M)$

   equal when $X \sim U(-M, M)$

2. maximum power

   $\mathrm{Var}(X) \le \sigma^2$

   $H_C(X) \le \frac{1}{2}\ln(2\pi e\sigma^2)$

   equal when $X \sim \mathcal{N}(m, \sigma^2)$

2.5.4 Entropy Power

$\overline{\sigma}_X^2 = \frac{1}{2\pi e} e^{2H_C(X)}$

$\overline{\sigma}_X^2 \le \sigma_X^2$

equal when $X \sim \mathcal{N}(m, \sigma^2)$

2.6 Stable Discrete Source

- stable: prob. dist. independent of time
- memoryless: X_i are i.i.d.
- markov

entropy of stable source $\boldsymbol{X} = (X_1, \cdots, X_N)$

$H(\boldsymbol{X}) = -\sum p(x_1, \cdots, x_N) \log p(x_1, \cdots, x_N)$

hard to implement & calculate

use **Entropy Rate**

$H_\infty(\boldsymbol{X}) = \lim\limits_{N\to\infty} \frac{1}{N} H(\boldsymbol{X}) = \lim\limits_{N\to\infty} H(X_N | X_1, \cdots, X_{N-1})$

avg entropy per symbol $H_N(\boldsymbol{X}) = \frac{1}{N} H(\boldsymbol{X}) = \frac{1}{N} \sum\limits_{n=1}^{N} F_N(\boldsymbol{X})$

conditional avg entropy $F_N(\boldsymbol{X}) = H(X_N | X_1, \cdots, X_{N-1}) = NH_N(\boldsymbol{X}) - (N-1)H_{N-1}(\boldsymbol{X})$

2.6.1 Property of Stable Src Entropy

- $F_N(\boldsymbol{X}),\ H_N(\boldsymbol{X})$ non-increasing
- $F_N(\boldsymbol{X}) \le H_N(\boldsymbol{X})$

conditional entropy is a more accurate approx to entropy rate

$H_\infty(\boldsymbol{X}) \le H_i(\boldsymbol{X}) \le H(X) \le \log K$

- for memoryless discrete src entropy rate is entropy
- when uniform dist, reach maximum entropy rate

relative rate of entropy $\eta = \frac{H_\infty(\boldsymbol{X})}{\log K}$

冗余度 $R = 1 - \eta$

2.6.2 Markov Src

Markov src is m order $H_{\infty}(\boldsymbol{X}) = H(X_{m+1}|X_1, \cdots, X_m) = H(X|S)$

stable dist. $\boldsymbol{\mu} = \boldsymbol{\mu P}$

$H(X|S) = \sum \mu_i H(X|s_i)$

redundancy $R = 1 - \frac{H_{\infty}}{H_0}$

3 Src Coding

all src are DMS

3.1 Equal Length

- alphabet $\mathcal{A} = \{a_1, \cdots, a_K\}$, with dist. p_i
- code book $\mathcal{B} = \{b_1, \cdots, b_D\}$
- length of message L
- message $u^L = (u_1, \cdots, u_L),\ u_i \in \mathcal{A}$
- length of codeword D
- codeword $c^D = (c_1, \cdots, c_D),\ c_i \in \mathcal{B}$

coding rate

$R \triangleq \frac{N \log D}{L}$

3.1.1 AEP

memoryless src

$p(u^L) = \prod_{l=1}^{L} p(u_l)$

self info of u^L is $H(u^L) = -\log p(u^L)$

avg self info per symbol $I_L = \frac{1}{L} H(u^L) = -\frac{1}{L} \sum_{l=1}^{L} \log p(u_l)$

$\lim_{L \to \infty} I_L = H(U)$

typical set

$A_{\epsilon}^{(L)}(U) = \{u^L; |I_L - H(U)| \le \epsilon\}$

$p(|I_L - H(U)| > \epsilon) \le \frac{\sigma_I^2}{L\epsilon^2}$

Property of typical set

- $p(u^L \in A_{\epsilon}^{(L)}) \ge 1 - \epsilon$
- $p(u^L) \approx 2^{-LH(U)}$
- $|A_{\epsilon}^{(L)}| \approx 2^{LH(U)}$

3.1.2 Src Coding Thm

coding rate

$$R > H(U) + \epsilon$$

then $P_E \to 0$

only encode seq in typical set

$P_E = p(u^L \notin A_\epsilon^{(L)}) < \epsilon$

- reachable $R > H(U)$
- unreachable $R < H(U)$

coding efficiency

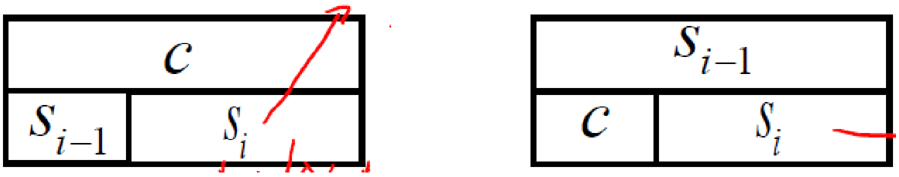
$\eta = \frac{H(U)}{R}$

3.2 Variable Length

avg length $\overline{n} = \sum_{k=1}^{K} p_k n_k$

3.2.1 唯一可译性

后缀分解集



SP method

$S_0 = C$	S_1	S_2	S_3	S_4	S_5	...
0	2	1	0	1	0	...
10			2	2	2	...
12			12		12	...
21			122		122	...
112					1	...
1122						...

112:2|2|12
11:22|210

S_i与连线上一个组成后缀或码字$S_k, k = 0, \cdots, i-1$，把组成码字的连起来会产生ambiguous(组成后缀的跳过)

- 后缀分解集中不包含码字$\Longleftrightarrow$唯一可译码
- $S_1 = \emptyset \Longrightarrow$前缀码

3.2.2 Kraft Inequality

$$\sum_{k=1}^{K} D^{-n_k} \leq 1$$

存在长度为n_i的D元前缀码

唯一可译码满足Kraft Inequality

唯一可译码⟹存在对应长度的前缀码

3.2.3 Src Coding Thm

$\frac{H(U)}{\log D} \leq \overline{n} < \frac{H(U)}{\log D} + \frac{1}{L}$

coding rate $R = \overline{n}\log D,\ H(U) \leq R \leq H(U) + \frac{\log D}{L}$

coding efficiency $\eta = \frac{H(U)}{R}$

3.2.4 Example

二元 Best variable length coding

- the prob. smaller, the length longer
- prob. the least two symbol has same code length, the last bit different

1. Huffman

   D 元要补充到$(D-1)i+1$个symbol，充分利用

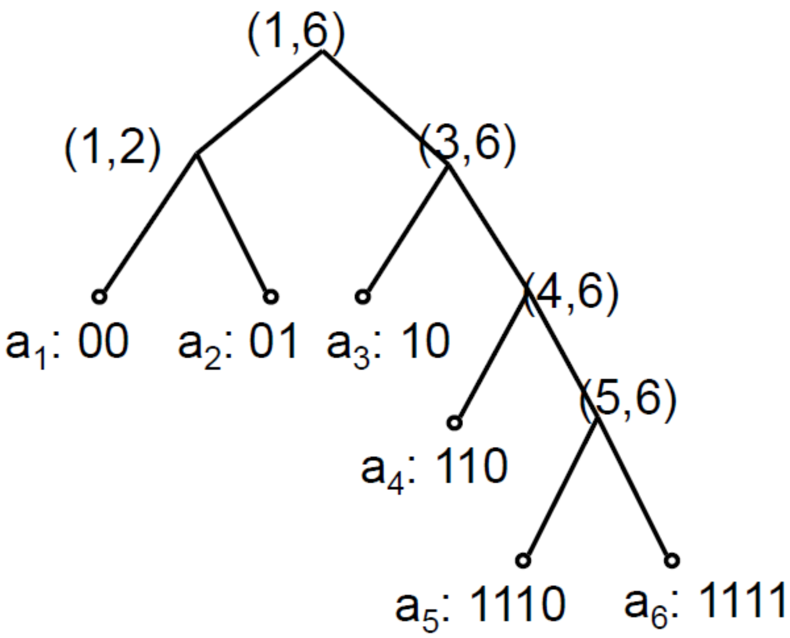
2. Shannon

   - arrange dist. in decreasing order
   - $P_k = \sum_{i=1}^{k-1} p_i$
   - length $l_k = \lceil \log \frac{1}{p_k} \rceil$
   - represent P_k in binary form, take the first l_k bit as codeword
   - $H(U) \leq \overline{n} < H(U) + 1$

3. Fano

   - arrange dist. in decreasing order
   - 概率和对分
   - $H(U) \leq \overline{n} < H(U) + 1 - 2\min p_i$

$$U = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ 0.3 & 0.25 & 0.2 & 0.15 & 0.05 & 0.05 \end{pmatrix}$$



4. SFE

   - no need to arrange dist.

- $\overline{F}(x) = \sum\limits_{i=1}^{x-1} p(i) + \frac{1}{2}p(x)$
- $F(x) = \sum\limits_{i=1}^{x} p(i)$
- $l(x) = \lceil \log \frac{1}{p(x)} \rceil + 1$
- represent $\overline{F}(x)$ in binary form, take the first l_k bit as codeword
- $H(U) \leq \overline{n} < H(U) + 2$

x	$p(x)$	$F(x)$	$\overline{F}(x)$	$\overline{F}(x)$的二进表示	$l(x)$	码字	Huffman 码
1	0.25	0.25	0.125	0.001	3	001	01
2	0.5	0.75	0.5	0.10	2	10	1
3	0.125	0.875	0.8125	0.1101	4	1101	001
4	0.125	1.0	0.9375	0.1111	4	1111	000

3.3 Stable Src Coding

$\frac{H_\infty(\boldsymbol{U})}{\log D} \leq \overline{n} < \frac{H_\infty(\boldsymbol{U})}{\log D} + \frac{1}{L}$

$H_\infty(\boldsymbol{U}) = H(U|S)$

encode for each output of states separately

4 Channel

4.1 DMC

- input $x^n = (x_1, \cdots, x_n)$
- output $y^n = (y_1, \cdots, y_n)$

$p(y^n|x^N) = \prod\limits_{i=1}^{n} p(y_i|x_i)$

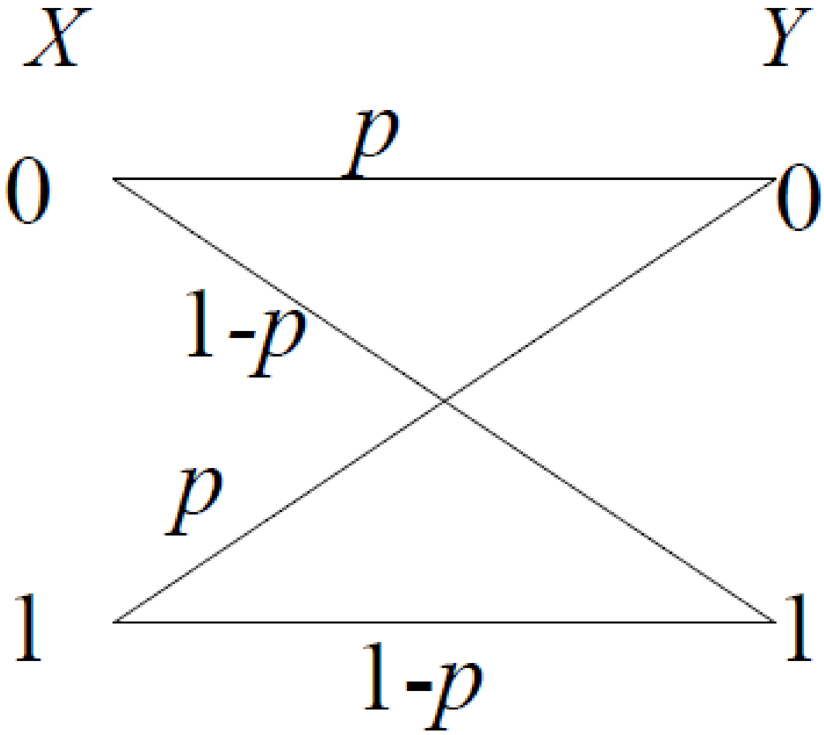
4.1.1 Capacity

$C = \max I(X;Y)$

4.1.2 Examples

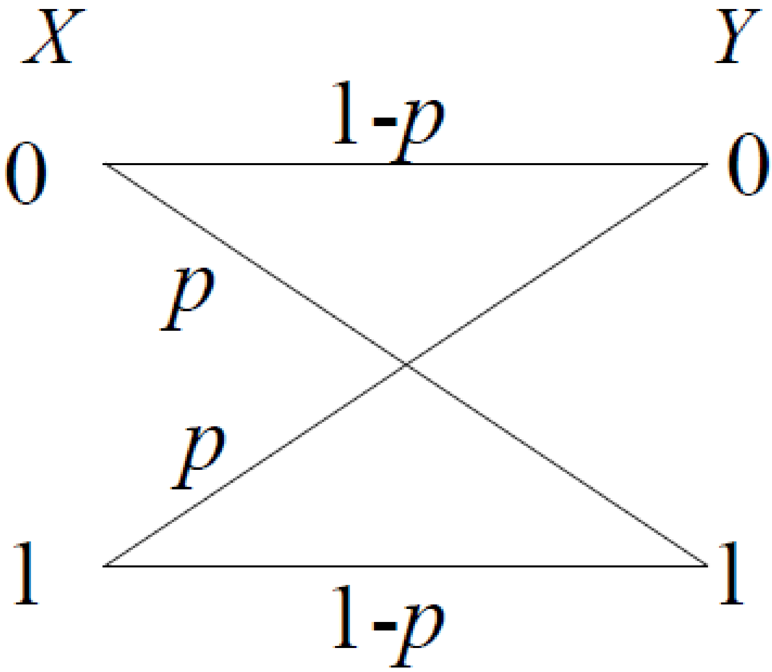
1. Useless

- $H(X|Y) = H(X)$
- $C = 0$



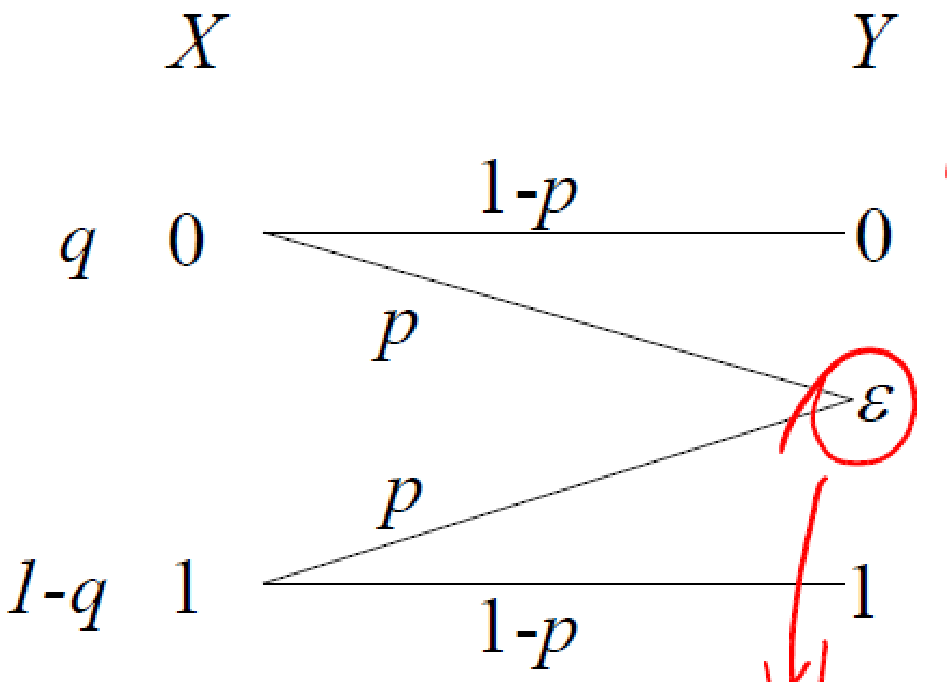
2. BSC

- $I(X;Y) = H(Y) - H(p)$
- $C = 1 - H(p)$ when X is equal dist. so do Y



3. BEC

- $H(Y|X) = H(p)$
- use additive $H(Y) = H(p) + (1-p)H(q)$
- $C = (1-p)$ when X is equal dist.



Input Dist

since $I(X;Y) = f(p(x), p(y|x))$ is concave in $p(x)$ given $p(y|x)$, there must be a maximum

$\max I(X;Y)$

$s.t. \begin{cases} p(x) \geq 0 \\ \sum\limits_x p(x) = 1 \end{cases}$

then $C = \max I(X;Y)$ is the maximum goal

$I(X = x; Y) = C, \ \forall\, x, p(x) > 0$

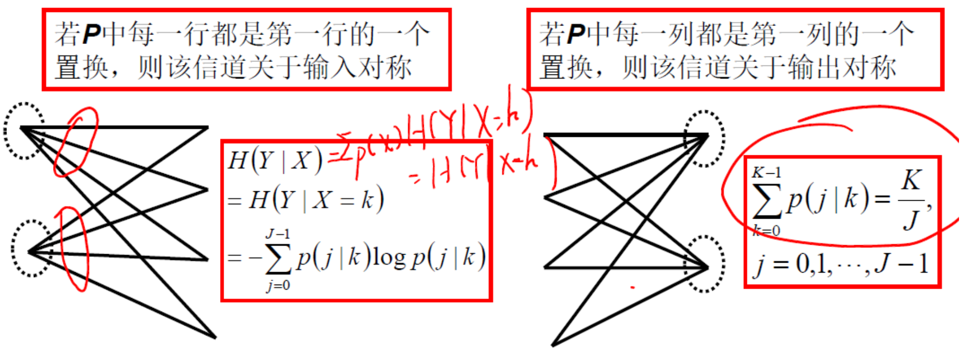
$I(X = x; Y) \leq C, \ \forall\, x, p(x) = 0$

$I(X = x; Y) = \sum\limits_y p(y|x) \log \frac{p(y|x)}{\sum\limits_x p(x)p(y|x)}$

discard ($p(x) = 0$) worst input, make others $I(X = x; Y)$ are equal

4.1.3 Calculation

prob. transfer mat $\boldsymbol{P}$



symmetric in input

$C = H(Y) + \sum\limits_y p(y|x) \log p(y|x)$

symmetric in input & output

$C = \log|\mathcal{Y}| + \sum_y p(y|x)\log p(y|x)$

若一个信道既关于输入对称，又关于输出对称，即***P***中每一行都是第一行的一个置换，每一列都是第一列的一个置换，则该信道是对称的

对一个信道的转移概率矩阵***P*按列**划分，得到若干子信道，若划分出的所有子信道均是对称的，则称该信道是准对称的

pre symmetric chn reach capacity $\Longrightarrow$ X is equal dist.

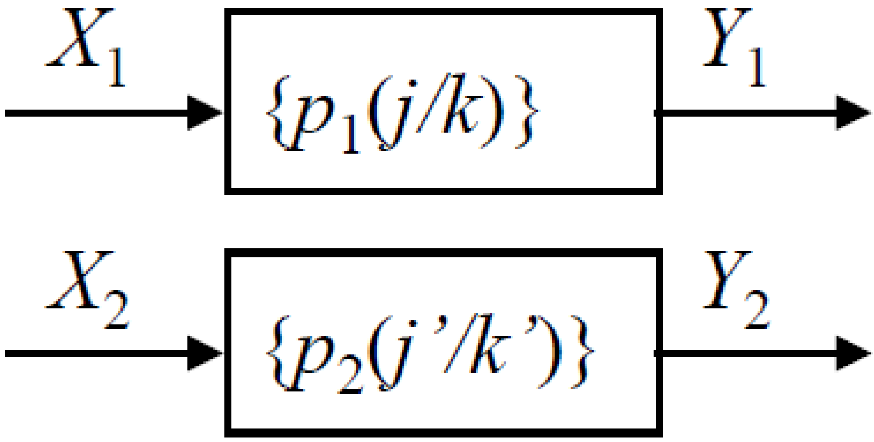
inverse mat

4.1.4 Chn Combination

1. Parallel Chn

   $I(X_1, X_2; Y_1, Y_2) = I(X_1; Y_1) + I(X_2; Y_2)$

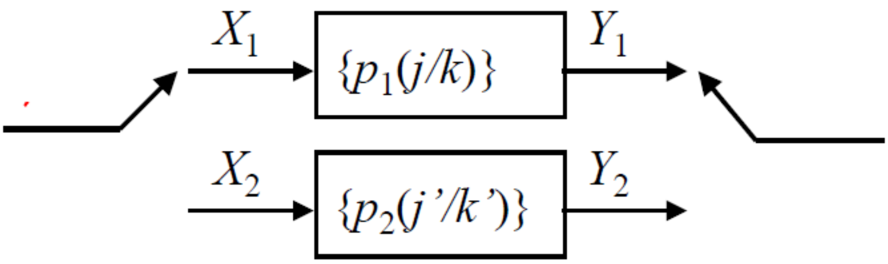
   $C = \log(2^{C_1}2^{C_2})$



2. Switch Chn

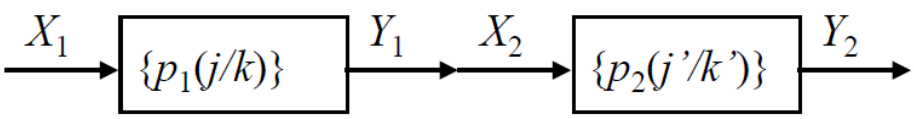
   $I(X;Y) = P_A I(X_1;Y_1) + P_B I(X_2;Y_2) + H(P_A, P_B)$

   $C = \log(2^{C_1} + 2^{C_2}),\ P_A = \frac{2^{C_1}}{2^C},\ P_B = \frac{2^{C_2}}{2^C}$



3. Cascade Chn

   $C \le \min\{C_1, C_2\}$



4.2 Chn Coding Thm

(M, n) code

- # all message M, # bit per message $\log M$
- # bit in codeword n
- encode $X^n(i) : \{1, \cdots, M\} \to \mathcal{X}^n$
- decode $g(y^n) : \mathcal{Y}^n \to \{1, \cdots, M\}$
- error prob. $\lambda_i = p(g(Y^n) \neq i | X^n = X^n(i))$
- max error prob. $\lambda^{(n)} = \max \lambda_i$
- avg error prob. $P_E^{(n)} = \frac{1}{M} \sum\limits_{i=1}^{M} \lambda_i$
- **coding rate** (transmit rate) $R = \frac{\log M}{n}$

4.2.1 Joint Typical Set

$(X^n, Y^n) \sim p(x^n, y^n)$

$\lim\limits_{n\to\infty} I_{nx} = \lim\limits_{n\to\infty} -\frac{1}{n}\log p(x^n) = H(X)$

$A_\epsilon^{(n)} = \{(x^n, y^n); |I_{nx} - H(X)| \leq \epsilon,\ |I_{ny} - H(Y)| \leq \epsilon,\ |I_{nxy} - H(X,Y)| \leq \epsilon\}$

Property

- $p((X^n, Y^n) \in A_\epsilon^{(n)}) \to 1$
- $|A_\epsilon^{(n)}| \approx 2^{nH(X,Y)}$
- if $(\tilde{X}^n, \tilde{Y}^n) \sim p(x^n)p(y^n)$, then $p((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}) \to 0$

only need to encode seq in joint typical set

given x^n, may receive $2^{nH(Y|X)}$ typical seq

$M = 2^{nR} \leq \frac{2^n H(Y)}{2^{nH(Y|X)}} \implies R < H(Y) - H(Y|X) \to C$

4.2.2 Chn Coding Thm

$R < C \iff \exists\ (2^{nR}, n)$ code, $\lim\limits_{n\to\infty} \lambda^{(n)} = 0$

4.3 Discrete Time AGC

noise $Z \sim \mathcal{N}(0, N)$

maximum power $\frac{1}{n} \sum\limits_{i=1}^{n} x_i^2 \leq P$

convert to $X \in \{+\sqrt{P}, -\sqrt{P}\},\ Y \in \{1, 0\}$

4.3.1 Capacity

$I(X;Y) = H_C(Y) - H_C(Z) = H_C(Y) - \frac{1}{2}\log(2\pi e N)$

$C = \frac{1}{2}\log(1 + \frac{P}{N}),\ H(Y)$ reach maximum when $Y \sim \mathcal{N}(0, P+N)$ a.k.a. $X \sim \mathcal{N}(0, P)$

4.3.2 Gaussian Parallel Chn

$Z_i \sim \mathcal{N}(0, N_i)$

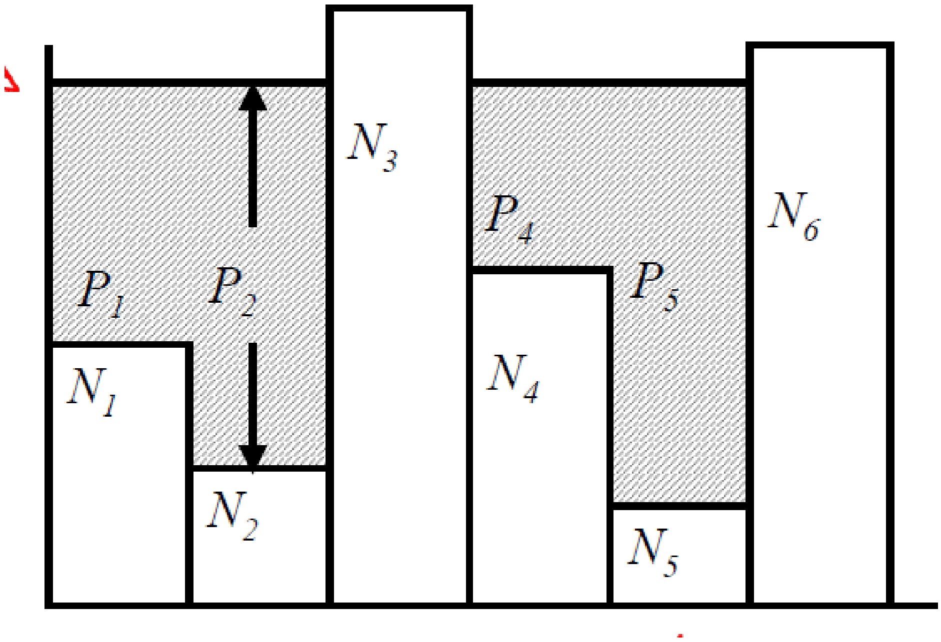
$\sum\limits_{i=1}^{k} P_i \leq P$

solution:

$\overline{P} = \arg_v\{\sum_{i=1}^{k}(v - N_i)^+ = P\}$

$P_i = (\overline{P} - N_i)^+$

$C = \frac{1}{2}\sum_{i=1}^{k}\log(1 + \frac{P_i}{N_i})$



4.4 Continuous Time AGC

- limited band width $[-W, W]$
- a transmission takes T sec
- # samples per transmission $2WT$
- noise spectrum density $\frac{N_0}{2}$

$C_T = \frac{1}{2}\sum_{i=1}^{2WT}\log(1 + \frac{P_i}{N_i}) = WT\log(1 + \frac{P}{N_0 W})$

capacity per sec (bit per sec)

$C = W\log(1 + \frac{P}{N_0 W})$

- $W \nearrow, C \nearrow$
- $\lim_{W\to\infty} C = \frac{P}{N_0}\log e$
- spectrum efficiency $\eta = \frac{R}{W}$
- $\eta_{max} = \log(1 + \frac{P}{N_0 W})$
- power per bit $E_b = \frac{P}{R}$
- power efficiency $\frac{1}{E_b}$
- SNR per bit $\frac{E_b}{N_0} > \frac{2^\eta - 1}{\eta}$
- $\lim_{\eta\to 0}\frac{E_b}{N_0} > \ln 2$

N_0 energy per Hz per sec given by noise

E_b energy per Hz per sec given by signal

5 Rate Distortion Theory

distortion measure $d(x, \hat{x}) : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+$

- hamming distortion $d(x, \hat{x}) = \begin{cases} 0 & x = \hat{x} \\ 1 & x \neq \hat{x} \end{cases}$
- MSE distortion $d(x, \hat{x}) = (x - \hat{x})^2$

regular distortion $\min d(x, \hat{x}) = C_x = 0$

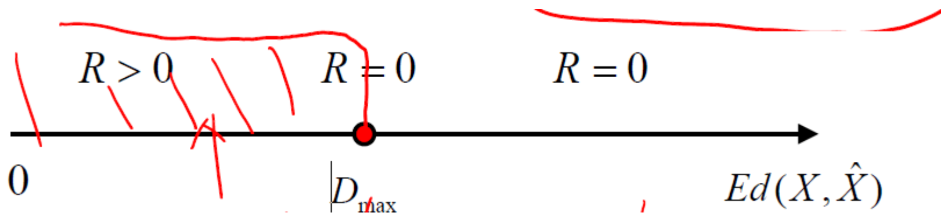
5.1 Min Avg Distortion

when $R = 0$ a.k.a. use a fix point $\hat{x}$ to represent x

$$\hat{x}^* = \arg\min_{\hat{x}} \sum_x p(x) d(x, \hat{x})$$

$$D_{max} = E[d(X, \hat{x}^*)]$$

- when tolerate distortion $D > D_{max}$, no need to code ($R = 0$)
- when use coding, the distortion upper bound $D < D_{max}$



5.1.1 Rate Distortion Pair

reachable RD pair (R, D), $\exists$ a series of $(2^{nR}, n)$ code, that $\lim_{n\to\infty} E[d(X^n, \hat{X}^n)] \leq D$

- rate distortion func $R(D)$, minimum R given D
- distortion rate func $D(R)$, minimum D given R

5.1.2 RDT

info rate distortion func, given src dist. $p(x)$

$$R^I(D) = \min_{E[d(X, \hat{X})] \leq D} I(X; \hat{X})$$

given i.i.d. $p(x)$

$$R(D) = R^I(D)$$

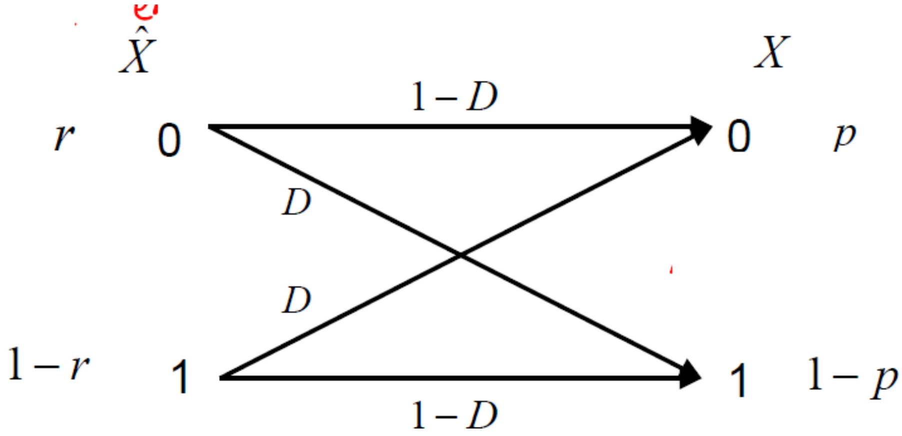
5.2 RDF Calculation

5.2.1 Bernoulli Src under Hamming Distortion

assume $p < \frac{1}{2}$

$$R(D) = \begin{cases} H(p) - H(D) & 0 \leq D \leq p \\ 0 & D > p = \end{cases}$$

reverse check



$r = \frac{p-D}{1-2D}$

5.2.2 Gaussian Src under MSE

$X \sim \mathcal{N}(0, \sigma^2)$

$$R(D) = \begin{cases} \frac{1}{2}\log\frac{\sigma^2}{D} & 0 \le D \le \sigma^2 \\ 0 & D > \sigma^2 \end{cases}$$

$D(R) = \sigma^2 2^{-2R}$

5.2.3 Gaussian Vector Src under MSE

$X_i \sim \mathcal{N}(0, \sigma_i^2)$

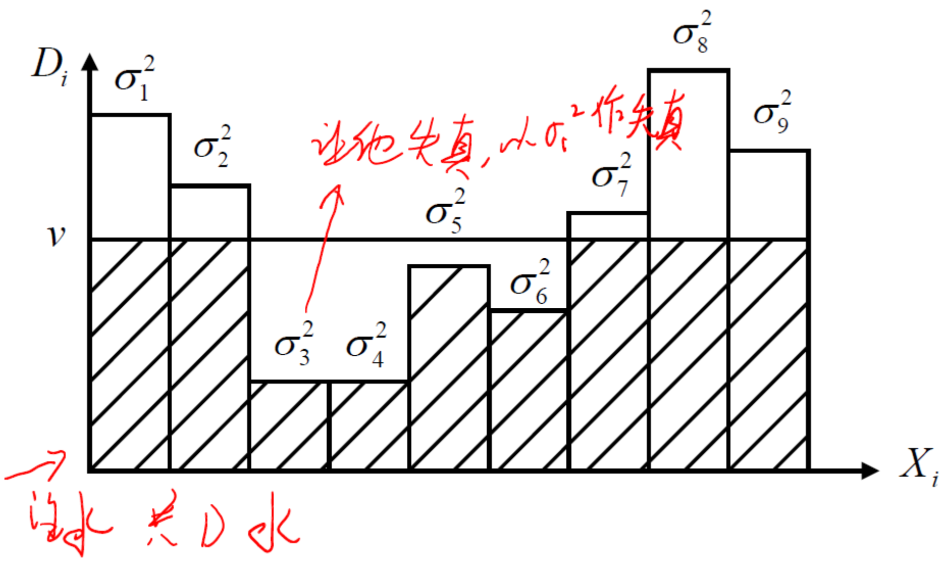
inverse add water

allocate distortion for each component in vector $D = \sum_{i=0}^{m} D_i$

determine a $\overline{D}$

$$D_i = \begin{cases} \overline{D} & \overline{D} < \sigma_i^2 \\ \sigma_i^2 & \overline{D} \ge \sigma_i^2 \end{cases}$$

error too small, no need to tolerate distortion



5.3 Property of RDF

$R(D) = \min I(X; \hat{X})$

5.3.1 Support of RDF

$D_{min} = \sum_x p(x) \min_{\hat{x} \in \hat{\mathcal{X}}} d(x, \hat{x})$

for regular $d(x, \hat{x}),\ D_{min} = 0$

when $D = D_{min}$, x correspond to a $\hat{x},\ H(X|\hat{X}) = 0,\ R(D_{min}) = H(X)$

if continuous and $D \to D_{min},\ R(D) \to \infty$ infinity precision

$D = D_{max}$, then $R = 0$, use one recover point $\hat{x}^*$

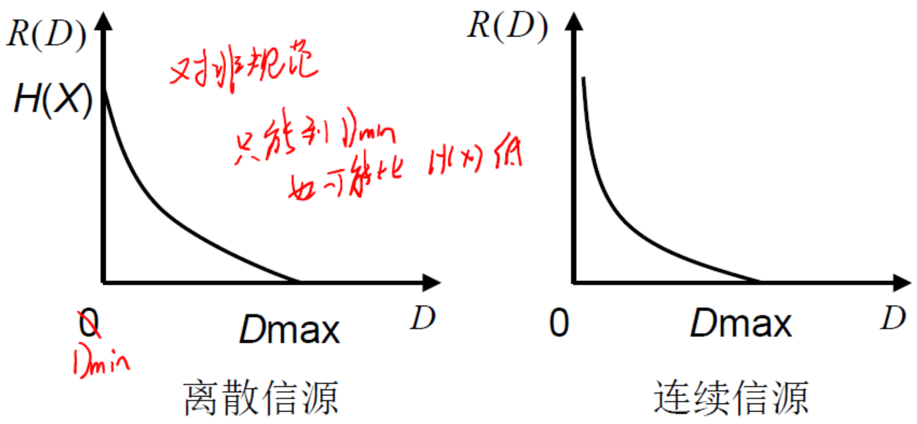
$D_{max} = \min_{\hat{x}} \sum_x p(x) d(x, \hat{x})$

- Bernoulli Src under Hamming: $[0, \min(p, 1-p)]$
- Gaussian Src under MSE: $[0, \sigma^2]$

5.3.2 Convexity

$R(D) = \min I(X; \hat{X})$ is convex in $p(\hat{x}|x)$

5.3.3 Decreasing



5.4 Calculation

- $D_{min} = \sum_x p(x) \min_{\hat{x} \in \hat{\mathcal{X}}} d(x, \hat{x})$
- $D_{max} = \min_{\hat{x}} \sum_x p(x) d(x, \hat{x})$
- if permutation ???, transfer mat $\boldsymbol{P}$ has the symmetric as distortion mat $\boldsymbol{D}$
- calculate $\boldsymbol{P}$
- calculate $D = \sum_{x, \hat{x}} p(x) p(\hat{x}|x) d(x, \hat{x})$
- calculate $I(X; \hat{X})$

一个四元对称信源 $\begin{Bmatrix} X \\ p(x) \end{Bmatrix} = \begin{Bmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{Bmatrix}$，再生字符集为 $\hat{X} = \{0, 1, 2, 3\}$，其失

2019... caches.pdf

真矩阵为， $D = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$

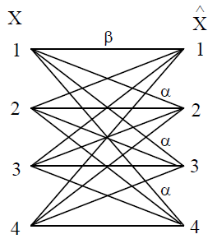
所以由定理 5.3.1，转移概率矩阵具有与失真矩阵相同的对称

$$P = \begin{pmatrix} \beta & \alpha & \alpha & \alpha \\ \alpha & \beta & \alpha & \alpha \\ \alpha & \alpha & \beta & \alpha \\ \alpha & \alpha & \alpha & \beta \end{pmatrix}$$

其中 $\beta + 3\alpha = 1$。设平均失真为 D，则

$$D = \sum_{x,\hat{x}} p(x)p(\hat{x} \mid x)d(x,\hat{x}) = 3\alpha$$

因而 $\alpha = \frac{1}{3}D, \beta = 1 - D$。相应的转移概率图为如下所示



由于 $P(\hat{X} = 1) = P(\hat{X} = 2)$

$$= P(\hat{X} = 3) = P(\hat{X} = 4) = \frac{1}{4}$$

所以 $H(\hat{X}) = 2bit$，

$$H(\hat{X} \mid X) = \sum_{i=1}^{4} P(X = i) H(\hat{X} \mid X = i)$$

$$= -\beta \log \beta - 3 \cdot \alpha \log \alpha$$

$$= -(1-D)\log(1-D) - D \cdot \log \frac{D}{3}$$

于是 $R(D) = \begin{cases} 2 + (1-D)\log(1-D) + D\log\frac{D}{3}, & 0 \le D < \frac{3}{4} \\ 0 & D \ge \frac{3}{4} \end{cases}$

6 Computation Theory

6.1 TM

inst set $(q_i, S_j, S_k, R(LN), q_n)$

- q_i: current state
- S_j: read in symbol
- S_k: write down symbol
- $R(LN)$: movement
- q_n: next state

Von Neumann: store program, ALU centered, binary, divided hw & sw

- controller
- ALU
- memory
- I/O

6.1.2 TM Halting Problem

can not judge transferring to halting within finite step

T: given input program P, if halt, output 1; otherwise, output 0

an outer program S: given input P, using T and a infinite loop, if P halt, S loop; otherwise, S stop

use S as input os S: if S halt, S loop; otherwise, S halt

6.2 K-Complexity

descriptive complexity

$$K(x) = \min_{p:U(p)=x} l(p)$$

reveal

- info of x
- redundancy
- structure

$K(x|l(x)) \leq l(x) + c$

6.2.2 Bounds

$K(x) \leq K(x|l(x)) + 2\log l(x) + c$

$|\{x \in \{0,1\}^* : X(x) < k\}| < 2^k$

6.2.3 K-complexity & Entropy

$\mathrm{E}[\frac{1}{n}K(X^n|n)] \rightarrow H(X)$

incompressable $\lim\limits_{n\to\infty} \frac{K(x_1 \cdots x_n|n)}{n} = 1$

6.3 Bayes

h: hypothesis, D: data

$P(h|D)$: posterior, $p(h)$: prior, $p(D|h)$: likelihood

- MAP: posterior $h_{MAP} = \backslash\mathrm{argmax} P(h|D) = \backslash\mathrm{argmax} P(D|h)P(h)$
- ML: likelihood $h_{ML} = \backslash\mathrm{argmax} P(D|h)$ (when same or unknown $p(h)$)

6.3.1 Bayes Inference

use $P(h|D)$

relay on prior $p(h)$

6.3.2 Bayes Best Classification

further divide h into class v, $h(v_i|h_j)$

\limits is allowed only on operators

$$v^* = \arg\max_{v_j \in V} \sum_{h_i \in H} P(v_j|h_i)P(h_i|D)$$

- 【接上例】：新实例的可能分类集合为$V = \{+, -\}$
  - $P(h_1|D) = 0.4, P(-|h_1) = 0, P(+|h_1) = 1$
  - $P(h_2|D) = 0.3, P(-|h_2) = 1, P(+|h_2) = 0$
  - $P(h_3|D) = 0.3, P(-|h_3) = 1, P(+|h_2) = 0$
  - $\sum_{h_i \in H} P(+|h_i)P(h_i|D) = 0.4$； $\sum_{h_i \in H} P(-|h_i)P(h_i|D) = 0.6$

6.3.3 Naive Bayes Classification

use MAP, prob. calculated from given data

$v^* = \arg\max\limits_{v_j \in \mathcal{V}} P(v_j|D) = \arg\max\limits_{v_j \in \mathcal{V}} P(v_j)P(D|v_j)$

$$v_{NB} = \arg\max_{v_j \in V} P(v_j) * P(\text{Outlook} = \text{Sunny}|v_j) P(\text{Temperature} = \text{Cool}|v_j) * P(\text{Humidity} = \text{High}|v_j) P(\text{Wind} = \text{Strong}|v_j)$$

计算结果：

$P(\text{PlayTennis} = \text{Yes}) = 9/14 = 0.64;$

$P(\text{PlayTennis} = \text{No}) = 5/14 = 0.36$

$P(\text{Wind} = \text{Strong}|\text{PlayTennis} = \text{Yes}) = 3/9 = 0.33$

$P(\text{Wind} = \text{Strong}|\text{PlayTennis} = \text{No}) = 3/5 = 0.60$

$P(\text{Yes})P(\text{Sunny}|\text{Yes})P(\text{Cool}|\text{Yes})P(\text{High}|\text{Yes})P(\text{Strong}|\text{Yes}) = 0.0053$

$P(\text{No})P(\text{Sunny}|\text{No})P(\text{Cool}|\text{No})P(\text{High}|\text{No})P(\text{Strong}|\text{No}) = 0.02$

6.4 Decision Tree

base on given final decision S, for several feature F_i, calculate each conditional entropy $H(F_i|S)$, use the minimum as first level decision, for each choice choose the following features in the same way.

info gain, choose the maximum info gain

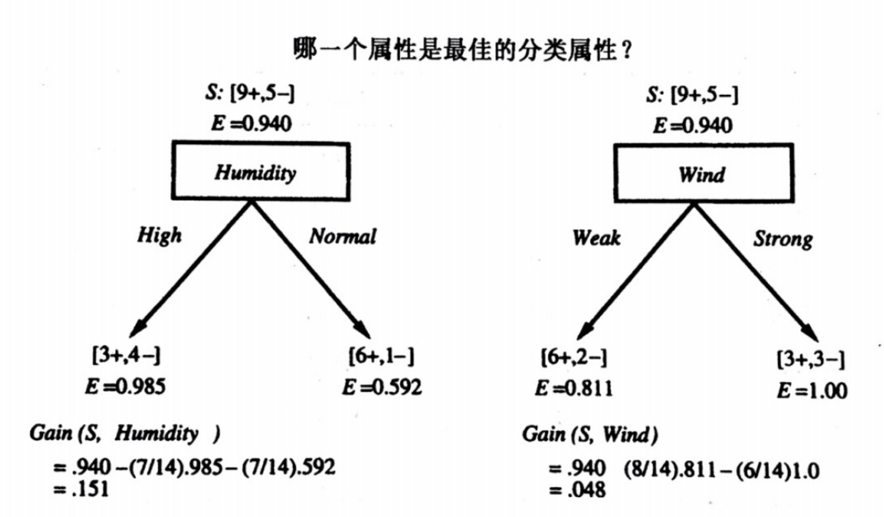
$G(S, F_i) = H(S) - H(S|F_i)$

- 按属性Wind分类14个样例得到的信息增益计算：
  - $\text{Values(Wind)} = \text{Weak, Strong}$
  - $S = [9+, 5-]$
  - $S_{\text{Weak}} = [6+, 2-]$
  - $S_{\text{Strong}} = [3+, 3-]$

$$\text{Gain}(S, \text{Wind}) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= \text{Entropy}(S) - \left(\frac{8}{14}\right) \text{Entropy}(S_{\text{Weak}}) - \left(\frac{8}{14}\right) \text{Entropy}(S_{\text{Strong}})$$

$$= 0.94 - \frac{8}{14} * 0.811 - \frac{6}{14} * 1 = 0.048$$



计算信息增益

所有四个属性的信息增益为:

$$\mathrm{Gain}(S, \mathrm{Outlook}) = 0.246$$
$$\mathrm{Gain}(S, \mathrm{Humidity}) = 0.151$$
$$\mathrm{Gain}(S, \mathrm{Wind}) = 0.048$$
$$\mathrm{Gain}(S, \mathrm{Temperature}) = 0.029$$

6.5 K-means

drawback

- dimensionality
- how to identify k
- how to index each point

7 Control System

7.1 Transfer Function

- input $\boldsymbol{u} \in \mathbb{R}^{p\times 1}$
- state
  $\dot{\boldsymbol{x}} = \boldsymbol{Ax} + \boldsymbol{Bu},\ \boldsymbol{x} \in \mathbb{R}^{n\times 1},\ \boldsymbol{A} \in \mathbb{R}^{n\times n},\ \boldsymbol{B} \in \mathbb{R}^{n\times p}$
- output $\boldsymbol{y} = \boldsymbol{Cx} + \boldsymbol{Du},\ \boldsymbol{y} \in \mathbb{R}^{m\times 1},\ \boldsymbol{C} \in \mathbb{R}^{m\times n},\ \boldsymbol{D} \in \mathbb{R}^{m\times p}$

use LT

$s\boldsymbol{X}(s) = \boldsymbol{AX}(s) + \boldsymbol{BU}(s)$

$\boldsymbol{Y}(s) = \boldsymbol{CX}(s) + \boldsymbol{DU}(s)$

let $(s\boldsymbol{I} - \boldsymbol{A})^{-1} = \boldsymbol{\Phi}(s)$

Transfer Func

$\boldsymbol{G}(s) = \boldsymbol{C\Phi}(s)\boldsymbol{B} + \boldsymbol{D}$

【例】RLC网络

$$\dot{\boldsymbol{x}} = \begin{bmatrix} 0 & \frac{-1}{C} \\ \frac{1}{L} & \frac{-R}{C} \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} \frac{1}{C} \\ 0 \end{bmatrix} u(t)$$

$$\boldsymbol{y} = [0 \quad R]\boldsymbol{x}$$

$$\boldsymbol{A} = \begin{bmatrix} 0 & \frac{-1}{C} \\ \frac{1}{L} & \frac{-R}{C} \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \frac{1}{C} \\ 0 \end{bmatrix}$$

$$\boldsymbol{C} = [0 \quad R]$$

$$[sI - A] = \begin{bmatrix} s & \frac{1}{C} \\ \frac{-1}{L} & s + \frac{R}{C} \end{bmatrix}$$

$$\boldsymbol{\Phi}(\boldsymbol{s}) = [sI - A]^{-1} = \frac{1}{\Delta s} \begin{bmatrix} (s + \frac{R}{L}) & \frac{-1}{C} \\ \frac{1}{L} & s \end{bmatrix}$$

$$\Delta(\boldsymbol{s}) = s^2 + \frac{R}{L}s + \frac{1}{LC}$$

- 传递函数：

$$\boldsymbol{G}(\boldsymbol{s}) = [0 \quad R] \begin{bmatrix} \frac{s + \frac{R}{L}}{\Delta(s)} & \frac{-1}{C\Delta(s)} \\ \frac{1}{L\Delta(s)} & \frac{s}{\Delta(s)} \end{bmatrix} \begin{bmatrix} \frac{1}{C} \\ 0 \end{bmatrix} = \frac{R/LC}{\Delta(s)} = \frac{R/LC}{s^2 + \frac{R}{L}s + \frac{1}{LC}}$$

7.2 Controllability & Observability

7.2.1 Discrete

$\boldsymbol{x}(k+1) = \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{B}\boldsymbol{u}(k)$

Controllability Mat
$\boldsymbol{U_c} = [\boldsymbol{B},\ \boldsymbol{AB}, \cdots, \boldsymbol{A}^{n-1}\boldsymbol{B}] \in \mathbb{R}^{n \times np}$

linear discrete system controllable $\iff \mathrm{rank}\boldsymbol{U_c} = n$

$\boldsymbol{y}(k) = \boldsymbol{C}\boldsymbol{x}(k)$

Observability Mat
$\boldsymbol{U_o} = [\boldsymbol{C};\ \boldsymbol{CA}; \cdots; \boldsymbol{CA}^{n-1}] \in \mathbb{R}^{mn \times n}$

linear discrete system observable $\iff \mathrm{rank}\boldsymbol{U_o} = n$

7.2.2 Continuous

continuous system is similar

- continuous not controllable (observable) $\implies$ discrete not controllable (observable)
- discrete controllable (observable) $\implies$ continuous controllable (observable)

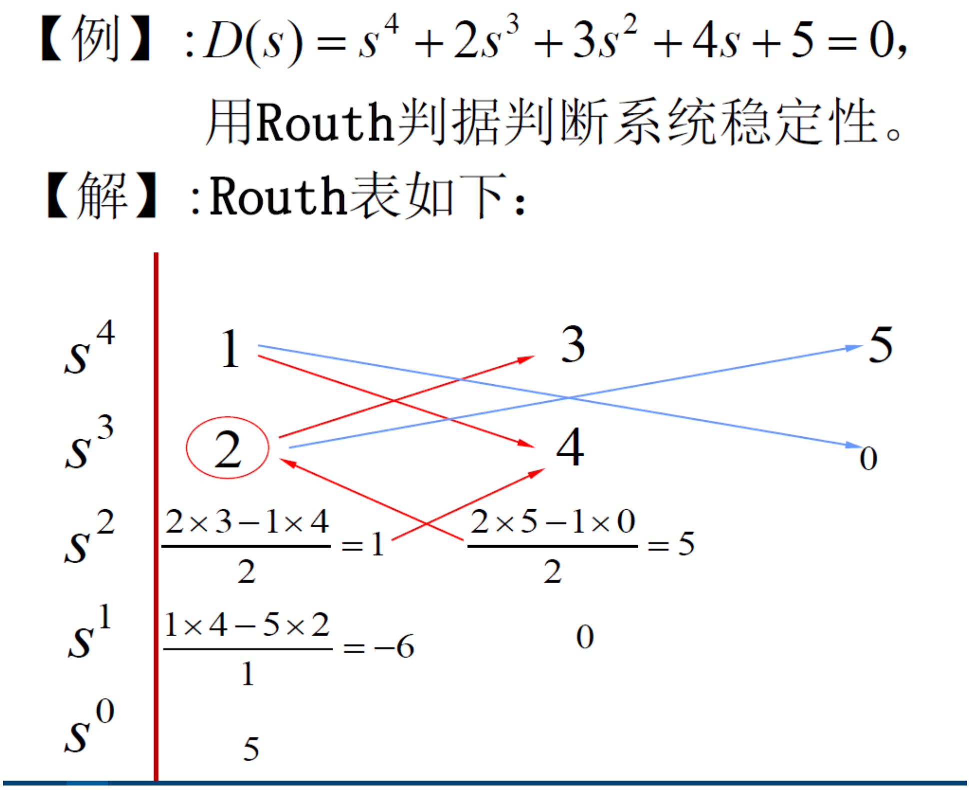
7.3 Stability

Reason: exists delay

7.3.1 Routh

eigen formula $\det[\boldsymbol{\Phi}(s)] = 0$

- stable system $\implies$ all coefficient > 0
- stable system $\iff$ first column of Routh table > 0



close loop system is not stable

7.3.1 李雅普诺夫

1. direct method
   - internal stability
     eigenvalue λ_k
     - 渐进stable: $\forall\ \lambda_k,\ \mathrm{Re}(\lambda_k)<0$
     - 李雅普诺夫stable: $\forall\ \lambda_k,\ \mathrm{Re}(\lambda_k)\le 0,\ \mathrm{Re}(\lambda_k)=0$无重根
     - unstable: $\exists\lambda_k,\ \mathrm{Re}(\lambda_k)>0$ or $\mathrm{Re}(\lambda_k)=0$有重根
   - external stability
     transfer func
     all poles $\mathrm{Re}(p)<0$
2. indirect method

potential $V(x)$ (determined by specific system)

$V(x)>0, \dot{V}(x)<0$, then $V(x)\to 0$