CHAPTER 2. 信息的度量

独立：$p(x,y)=p(x)\cdot p(y)$（联合分布）

此时 ~~I(X~~ $I(x,y)=I(x)+I(y)$（自信息）

§2.1. 信息的度量 $= f$（复杂度，概率分布）.

§2.2. 香农熵：随机变量不确定性. $*$ 自信息 $I(x)=\log\frac{1}{p(x)}$

$X\sim\begin{pmatrix}x_1\cdots x_a\\ p_1\cdots p_a\end{pmatrix}$（概率分布）. $\mathcal{X}=\{x_1\cdots x_n\}$（取值空间）

$a=\|\mathcal{X}\|$. $p_i=\underline{P_r\{X=x_i\}}$, $x\in\mathcal{X}$.

$P_r\{A\}$：事件A的概率

等价

$p=\{p_1\cdots p_a\}$.

$p(x)=P_r\{X=x\}$, $x\in\mathcal{X}$.

不确定性. $H(X)=H(p)=H(p_1\cdots p_a)$, $p\in\mathcal{P}$ （$\mathcal{P}$为全体概率分布）.

$$H(p)=-\sum_{i=1}^{a} p_i\log_c p_i \quad , c>0.$$

$$=\sum_{i=1}^{a} p_i\log\frac{1}{p_i}=E\left(\log\frac{1}{p(x)}\right).$$

自信息.

(Page. 34. 推导).

c=	单位
2	bit
e	nat
3	tet
10	det ~~nat~~

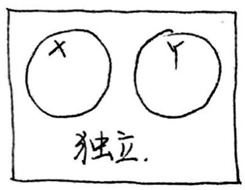
$$\log_c x=\frac{\log_b x}{\log_b c}.$$

性质：对称 & 非负.

联合熵：$H(X,Y)=-\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\cdot\log p(x,y)$. $\rightarrow E(g(X,Y))$.

$p(y)\cdot p(x|y)=p(x,y)$.

X, Y 独立：$H(X,Y)=H(X)+H(Y)$.



条件熵：$H(X|Y)=-\sum_{x}\sum_{y} p(x,y)\cdot\log p(x|y)$.

与联合熵关系：$H(X,Y)=H(X)+H(Y|X)$

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

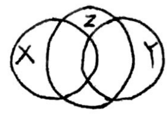
?
什么用

马尔可夫链(Markov)：在n时刻状态与前k个状态有关(记忆性)。

$X^n=(X_1,\cdots X_n)$：有 $(x_1\cdots x_{i-1})\in\mathcal{X}^{i-1}$，$x_i, x_{i+1}\in\mathcal{X}$：

$p((x_1\cdots x_{i-1}), x_{i+1}|x_i)=p((x_1\cdots x_{i-1}|x_i)\cdot p(x_{i+1}|x_i)$ 成立。

此时 $H(X_1\cdots X_n)=\sum_{i=1}^{n}H(X_i|X_{i-1})$.



2-马氏链

" $X\to Z\to Y$ 马尔可夫链. "

§2.3 熵的性质.

1. $H(X)\leq\log a$，均匀分布下最大.（ $1-\frac{1}{x}\leq\ln x\leq x-1$ ）

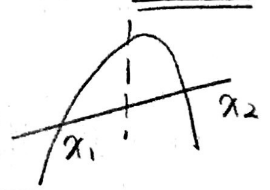
~~2.~~ 可加. $H(Q)=H(p)+\sum_{i=1}^{a}p_i H(q_i')$

$q_i'=(q_{i1}',\cdots q_{ik_i}')$，$i=1,2,\cdots a$.

$-\sum p(x)\log p(x)$

$q_{ij}'=\frac{q_{ij}}{p_i}$

熵 $H(p)$ 是 p 的凹函数



Q

$q_{11}\ \cdots\ q_{1k_1}\longrightarrow p_1$

$q_{21}\ \cdots\ q_{2k_2}\longrightarrow p_2$ ☆ 对数和不等式

$\vdots\qquad\vdots\quad$ (sum). $\quad\vdots$

$q_{a1}\ \cdots\ q_{ak_a}\longrightarrow p_a$

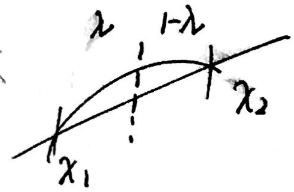
$$\sum a_i\log\frac{a_i}{b_i}\geq\left(\sum a_i\right)\log\frac{\sum a_i}{\sum b_i}$$

当且仅当 $\frac{a_i}{b_i}=C$ 等号成立.

? * Fano 不等式：$X, Y\subset\mathcal{X}$. $P_e=P_r\{X\neq Y\}$，则

二元熵函数

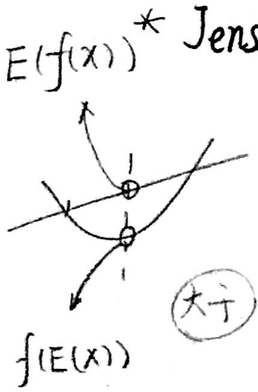
$H(X|Y)\leq H(P_e)+P_e\log(a-1)$.



* Jenson 不等式：上凸：$g(\lambda x_1+(1-\lambda)x_2)\geq\lambda g(x_1)+(1-\lambda)g(x_2)$.

$\lambda\in[0,1]$

若 $g(x)$ 上凸：$E(g(X))\leq g(E(X))$.



(推广) $\longrightarrow$ $g\left(\sum_{i=1}^{k}\lambda_i x_i\right)\geq\sum_{i=1}^{k}\lambda_i g(x_i)$.

§2.4. 互熵：$K(p;q)=\sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)}$.　（又称 散度）.

（相对熵）

（概率分布"差异性"的度量）　　* (p,q) 的凸函数

$D(p\|q)$.

性质：$K(p;q)$：q 固定，则为 p 的 $\boxed{\text{下凸函数}}$.
p 固定，则为 q 的 $\boxed{\text{下凸函数}}$.

互信息：（特殊的互熵）

$$I(X;Y)=\sum_x\sum_y p(x,y)\log\frac{p(x,y)}{p(x)p(y)}.$$

$p(x,y)$ 为与 $p(x)p(y)$ 的互熵

性质：非负、对称、

用图型记
↓

$$I(X;Y)=H(X)+H(Y)-H(X,Y)$$
$$=H(X)-\underline{H(X|Y)}\ \text{疑义度}$$
$$=H(Y)-\underline{H(Y|X)}\ \text{噪音熵}$$
$$I(X;X)=H(X).$$

$I(X;Y)$ 是 $q(y|x)$ 严格下凸函数.

Z 条件下 X,Y 独立：（称为马氏链）$X\to Z\to Y$

$p(x,Y|Z)=p(x|Z)\cdot p(y|Z)$.

（即 $X\perp Y|Z$）

且 $I(X;Y)\leq I(X;Z)$（远小于近）.

+ 条件熵：$I(X;Y|Z)=\sum_{(x,y,z)} p(x,y,z)\cdot\log\frac{p(x,y|z)}{p(x|z)\cdot p(y|z)}$.

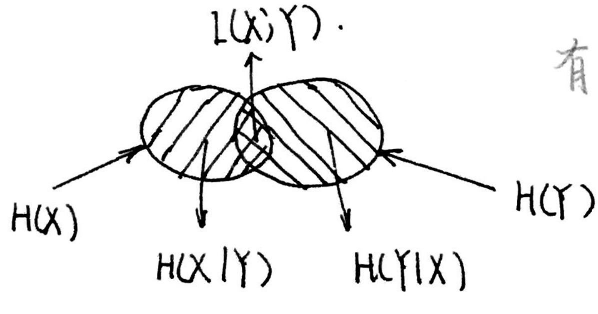
（条件互信息）.

性质：非负，对称：$I(X;Y|Z)=I(Y;X|Z)$.

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

（科尔莫戈罗夫公式）. $\longrightarrow$ $I(X;(Y,Z))=I(X;Y)+I(X;Z|Y)$.

e.g. $X\to Y\to Z$ 马尔可夫链 $\Leftrightarrow$ $I(X;(Y,Z))\ \overset{=}{\neq}\ I(X;Z)$ ~~恒成立~~.



有 $I(X_1\cdots X_n;Y)$
$=\sum_{i=1}^{n} I(X_i;Y|X_{i-1}\cdots X_1)$.

§2.5、连续型随机变量的信息量

更新记号：$p(a',b') = Pr\{a' \leq x \leq b'\}$. $a',b' \in (a,b) = \mathcal{X}$.

叫可微熵. $H(X) = -\int_{\mathcal{X}} p(x) \log \underbrace{p(x)}_{\text{概率密度函数}} .\ dx.$

$$I(X;Y) = \int_{R^2} f(x,y) \cdot \log \frac{f(x,y)}{f(x)f(y)}\, dx dy. \geq 0$$

一些例子. 1. 一维均匀分布：$H(X) = \log a$.

指数分布：$H(X) = 1 - \log\lambda$ ($p(x) = \lambda e^{-\lambda x}$).

正态分布：$H(X) = \frac{1}{2}\log(2\pi e\sigma^2)$.

2. 多维连续分布：$X = (X_1, \cdots X_n)$. , $x = (x_1 \cdots x_n)$.

$$H(X) = -\int_{R^n} p(x) \cdot \log p(x)\, dx$$

(OTHERS)

其他情况均为将 $\Sigma \longrightarrow \int$ 在相应区间积分即可.

积分过程中常用：$\int x \ln x\, dx = \frac{1}{2}x^2 \ln x - \frac{1}{4}x^2 + C$

$$\int x \cdot e^x\, dx = (x-1)e^x + C.$$

§2.6. 最大熵原理：寻找概率分布 ⟶ 熵最大.

$\exists p_0 \in \mathcal{P}$. s.t. $\forall p \in \mathcal{P}$: $-\int_{\mathcal{X}} p(x) \log p_0(x)\, dx = H_0$ H_0 与 p 无关的最大熵.

proof: $-\int_{\mathcal{X}} p(x) \log p(x) dx = -\int_{\mathcal{X}} p(x) \log p_0(x) dx - \int_{\mathcal{X}} p(x) \cdot \log \frac{p(x)}{p_0(x)} dx$

$\leq -\int_{\mathcal{X}} p(x) \log p_0(x) dx$.

最大熵分布：1. 有限区间： $p_0(x) = \frac{1}{b-a}$ $H_0 = \log(b-a)$.

2. 半开区间：✓ $p_0(x) = \frac{1}{\mu} e^{-x/\mu}$ $H_0 = 1 + \log\mu$ $(\mu = \int_0^{\infty} x p(x)\, dx)$

指 $(0,\infty)$. 全直线： $p_0(x) = N(\mu, \sigma^2)$ $H_0 = \frac{1}{2}(1 + \log(2\pi\sigma^2))$.

§2.7. 习题

3.

$y \backslash x$	0	1
0	1/3	0
1	1/3	1/3

$H(X) = -\sum_{x} p(x) \log p(x) = -\left(\frac{2}{3} \cdot \log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right)$

$H(Y) = H(X)$ $= \frac{1}{6}\log 2 + \frac{1}{2}\log 3$.

$H(X|Y) = -\sum\sum p(x,y) \log p(x|Y)$

$= -\left(\cancel{\frac{1}{3}\log 1} + \frac{1}{3}\log\frac{1}{2} + \frac{1}{3}\log\frac{1}{2}\right)$

$= \frac{2}{3}\log 2$

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

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

$= -\left(\frac{1}{3}\log\frac{1}{3} \times 3\right) = \log 3$.

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

$= \frac{1}{3}\log 2$

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

② $= -\sum p(y) \log p(y) + \sum\sum p(x,y) \log \underbrace{p(y|x)}_{= \frac{p(x,y)}{p(x)}}$

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

$= -\sum\sum p(x,y) \log p(x,y) + \sum p(x) \log p(x)$

5. (1) $H(X,Y|Z) = -\sum\sum\sum p(x,y,z) \log p(x,y|z)$.

$H(X|Z) = -\sum\sum p(x,z) \cdot \log p(x|z)$

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

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

12. $H(X) = -\sum p(x) \log p(x)$.　　Lagrange 乘子法？

CHAPTER 3. 通信系统概论.

这一节都在讲各种记号，不必细看

(通信系统)

信源 ⟶ (编)码 ⟶ 信道 ⟶ 译码 ⟶ 信宿. 在作业中理解

番羽

↑

干扰(是信道的成分).

*编码 = 番羽码 + 译码.

消息 ≠ 信号 ⟶ 信息.

↓ 所有情况

码元

消息 ⟶ 入口信号 ⟶ 出口信号 ⟶ 消息

↑ 编码　　↑ 信道　　↑ 译码

§3.2. 信源的概率模型：$S = [\mathcal{X}, p(x)]$.　　, * $\mathcal{Y}$：信宿字母表.

也称"状态集"

信道的概率模型：$C = [\mathcal{U}, \mathcal{V}, p(v|u)]$.

输入字母集　输出字母集　转移概率分布.

番羽码：$f: \mathcal{X} \to \mathcal{U}$

译码：$g: \mathcal{V} \to \mathcal{Y}$.

上述总合称为通信系统：

$$\mathcal{E} = \{S, C\} = \{\mathcal{X}, p(x), \mathcal{U}, p(v|u), \mathcal{V}\}.$$

有编码的通信系统　$\mathcal{E} = \{S, C, (f, g), \mathcal{Y}\}$.

$$= \{\mathcal{X}, p(x), \mathcal{U}, p(v|u), \mathcal{V}, (f, g), \mathcal{Y}\}.$$

以随机变量表示为 (X, U, V, Y).

$$p(x, u, v, y) = p(x)\, p(u|x)\, p(v|u)\, p(y|v).$$

$X \to U \to V \to Y$ 构成马尔可夫链.　　$p(y|x, u, v)$.

§3.3. 序列模型 $\tilde{\Sigma}^n = \{S^n, C^n\}$.

其中: $S^n = [X^n, p(x^n)]$, $C^n = [U^n, p(v^n|u^n), V^n]$.

简记为: $\tilde{S}$, $\tilde{C}$, $\tilde{\Sigma}$. e.g. $X^n = \{(x_1 \cdots x_n) \cdots\}$.

def. $p^n(\underset{\downarrow}{x^n}) = \prod_{i=1}^{n} p(x_i) \longrightarrow$ 无记忆信道序列 $\tilde{S}$

$x^n = (x_1 \cdots x_n)$.

$p^n(v^n|u^n) = \prod_{i=1}^{n} p(v_i|u_i) \longrightarrow$ 无记忆… $\tilde{C}$

信道编码: $(\tilde{f}, \tilde{g}) = \{(f^n, g^n) \dots\}$.

$f^n: x^n \to u^n$ $g^n: v^n \longrightarrow y^n$.

k-阶马尔可夫过程

$Pr\{X_m = x_m \mid X_{m-1} = x_{m-1}, \cdots X_1 = x_1\}$
$= Pr\{X_m = x_m \mid X_{m-1} = x_{m-1}, \cdots X_{m-k} = x_{m-k}\}$. $k \in [0, m)$

k=1时: 马氏过程 (即第 m个码输出时受前一个码影响).

即 $Pr\{X_m = x_m \mid X_{m-1} = x_{m-1}, \cdots X_1 = x_1\}$

$= Pr\{X_m = x_m \mid X_{m-1} = x_{m-1}\} = p(x_m \mid x_{m-1})$

则有 $p(x_1 \cdots x_m) = p(x_1)\, p(x_2|x_1) \cdot p(x_3|x_2) \cdots p(x_m|x_{m-1})$

(可称马氏信源)

二阶 $\begin{array}{c|cccc} & 00 & 01 & 10 & 11 \\ \hline 00 & & & & \\ 01 & & & & \\ 10 & & & & \\ 11 & & & & \end{array}$

e.g. $X = \{0, 1\}$ (一阶)

$p(x_{i+1} = 0 \mid x_i = 0) = 0.25$; $p(x_{i+1} = 0 \mid x_i = 1) = 0.6$
$p(x_{i+1} = 1 \mid x_i = 0) = 0.75$; $p(x_{i+1} = 1 \mid x_i = 1) = 0.4$

(亦可用转移概率矩阵表示)

CHAPTER 4. 信源编码.

§4.1. 种类. ...

等长码, <u>变长码</u> ($\bar{C} = U_0^* = \{C_1 \cdots C_a\}$)

$C_i = (u_{i1}, \cdots u_{ik_i})$ 是一个码元.

Ref. Version 2.

CHAPTER 5 信道编码问题

§5.1. 可达速率 Page. 108.

§5.2. 无记忆信道序列：$p(v^n|u^n) = \prod_{i=1}^{n} p(v_i|u_i)$. 可以用矩阵处理.

u, v ~~分布称为~~ 概率分布分别称为入、出口分布.

$$I(U;V) = \sum_{(u,v)\in U\otimes V} p(u,v)\log\frac{p(u,v)}{p(u)p(v)}. \longrightarrow \text{由} p(u|v) \text{与入口分布决定.}$$

也记作 $I(C,p)$

则信道容量 $C = \max\{I(C;p),\ p\in \mathcal{P}_u\}$.（即找一个入口分布，使互信息最大）

对称信道：转移概率矩阵关于行对称.

无记忆信道：序列容量 $C^n = nC$，且有 $H(Y^n|X^n) = \sum_{i=1}^{n} H(Y_i|X_i)$

§5.3. Reserve.

§5.4. 求信道容量.（$I(p)$ 最大值问题）. $p = (p(u), u\in \mathcal{U})$.

· $I(p)$ 是 p 的上凸函数.

对称

Chapter 5.

$$X^n = (X_1 \cdots X_n) \longrightarrow \boxed{\text{信源Coding}}^{\,f,\ W\in\mathcal{W}} \longrightarrow \boxed{\text{信源译码}}^{\,g} \longrightarrow \hat{X}^n$$

$$\begin{cases} f: \mathcal{X}^n \longmapsto \mathcal{W} \\ g: \mathcal{W} \longmapsto \hat{\mathcal{X}}^n \end{cases} \qquad 码率: R = \frac{1}{n}\log M.$$

$$\cancel{d(x^n, \hat{x}^n) = \begin{cases} \text{other}, & x^n \neq \hat{x}^n \\ 0, & x^n = \hat{x}^n \end{cases}} \qquad d(x, \hat{x}) = \begin{cases} \text{other}, & x \neq \hat{x} \\ 0, & x = \hat{x} \end{cases}$$

(失真测度). $\quad d(x^n, \hat{x}^n) = \frac{1}{n}\sum d(x_i, \hat{x}_i)$.

平均失真: $Ed(X^n, g_n(f_n(X^n))) = \sum p(x^n)\cdot d(x^n, g(f(x^n)))$.

率失真码 (M, n)
率失真对 (R, D) $\Big| \longrightarrow \lim\limits_{n\to\infty} Ed(\sim) \leq D$, 则 (R, D) 可达

率失真函数: $R(D) = \inf\{R : (R, D) \in \mathcal{L}\}$ 最小的码率

失真率函数: $D(R) = \inf\{D : (R, D) \in \mathcal{L}\}$.

$$R^I(D) = \min_{Q(\hat{x}|x),\ Ed \leq D} I(X; \hat{X}).$$

无记忆 + 有界失真: $R(D) = R^I(D)$.

(信息率失真函数).

$\downarrow$ (多维)

$$R_n^I(D) = \cancel{\min}\ \inf I(X; \hat{X}) \Big|_{\mathcal{X}^n \times \hat{\mathcal{X}}^n 上}.$$

且有 $R^I(D) = \lim\limits_{n\to\infty} \frac{1}{n}\cdot R_n^I(D)$.

(平稳信源是成立的).

Chapter 6. 连续

§6.1. $h(X) = -\int_R p(x)\log p(x)\,.\,dx.$ $x \in R.$

$h(X) \leq \log L$ L: length of R.

$\lim_{n\to\infty} (H(X_n) - n\log 2) = h(X)$, The 6.1.1. 条件. Page. 136.

$h(X|Y) = -\iint_{R^2} p(x,y)\log p(x|y)\,.\,dx\,dy$ $\frac{f(x,y)}{f_Y(y)}$ or $f_{X|Y}(x|y)$

$h(X,Y) = -\iint_{R^2} p(x,y)\log \underline{p(x,y)}\cdot dx\,dy$

都用$f(x,y)$代替$p(x,y)$

相对熵(互熵) $D(f\|g) = \int f(x)\cdot\log\frac{f(x)}{g(x)}\,dx.$

互信息: $I(X;Y) = \iint_{R^2} f(x,y)\log\frac{f(x,y)}{f_X(x)f_Y(y)}\,dx\,dy.$

$\lim_{n\to\infty} I(X_n;Y_n) = I(X;Y).$

$0 \leqslant H(X) \leqslant \log N$. $(\log \|\mathcal{X}\|)$

$H(X) \leqslant R$. 否则无法信源编码

信源无错编码 码率 $R = \frac{k}{n} \log_2 D$

(等长). $R \geqslant \log_2 D$.

$0 \leqslant H_\infty(X) \leqslant \log \|\mathcal{X}\|$.

$H_D(X) \leqslant \bar{L}$ 进制为D.

$0 \leqslant C \leqslant \min\{\log \|\mathcal{X}\|, \log \|\mathcal{Y}\|\}$ 信道容量(离散无记忆)

$(1-\varepsilon) 2^{n(H(X,Y)-\varepsilon)} \leqslant \|W_\varepsilon^{(n)}\| \leqslant 2^{n(H(X,Y)-\varepsilon)}$

对称信道：（行列均为置换）

$$C = \max I(X;Y) = \log\|Y\| - h(\underline{r})$$

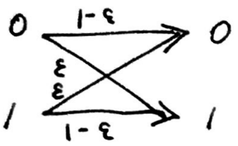
出口均匀分布
入口均匀分布

弱对称信道：（行置换 列和相等）

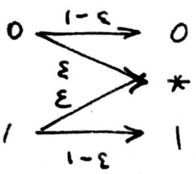
$$C = \max I(X;Y) = \log\|Y\| - h(\underline{r})$$

出口均匀分布
入口不一定.

二进对称：$\boxed{C = 1-h(\varepsilon)}$.

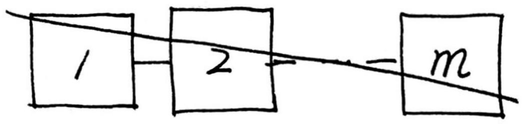


二进对称删除：（单M信道）



$\boxed{C = 1-\varepsilon}$.

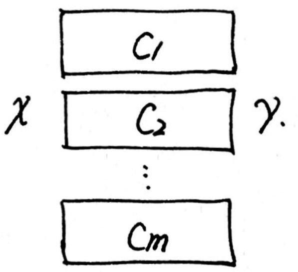
复合信道：



$$\boxed{C = \sum_{i=1}^{m} q_i C_i}$$

调用第i个信道概率为 q_i

并列信道：



$$\boxed{C = \log \sum_{i=1}^{m} 2^{C_i}}$$