XI'AN JIAOTONG UNIVERSITY

信息论与编码

第3章 信道与信道容量

张建国

XI'AN JIAOTONG UNIVERSITY

主要内容与基本要求

- **主要内容**
  - 信道的模型与分类；
  - 互信息的定义，计算；
  - 信道容量的定义，计算，对称信道的信道容量；
- **基本要求**
  - 了解信息信道的模型建立方法；理解和掌握互信息的概念、性质和计算方法；
  - 理解信道容量的概念；理解凸函数的含义和性质，了解用凸函数这一数学工具分析信道容量的方法；
  - 理解和掌握非对称信道的信道容量的求解方法、对称信道的信道容量的求解方法。

XI'AN JIAOTONG UNIVERSITY

本章目录

3.0 引言

3.1 信道的模型及分类

3.2 互信息

3.2.1 离散无记忆信道的互信息

3.2.2 连续无记忆信道的互信息

3.3 信道容量

3.3.1 凸函数

3.3.2 信道容量的计算

3.4 本章小结

XI'AN JIAOTONG UNIVERSITY

3.0 引言

信息论定量地研究信息在传输过程中的统计特性。

信道是传输信息的载体，其任务是以信号的形式传输和存储信息。

研究使用信道一次能够传输多少信息（**互信息**）；研究信道使用一次平均能够传输多少信息（平均互信息或**互信息熵**）。

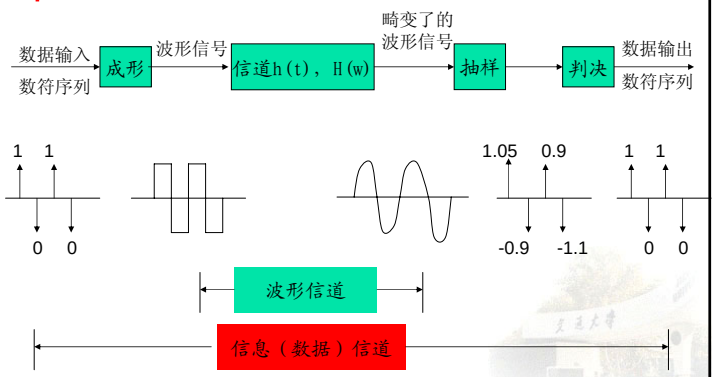
人们总是希望在可靠的前提下，信道能传输的信息越多越好。信道平均所能传输的最大信息量就是**信道容量**。信道传输信息的多少不仅与信道本身的特性有关，还与承载信息的信号形式和信源输出符号的统计特性有关。

本章主要讨论在什么条件下，通过信道的信息量最大。

首先简单介绍信道模型及其分类；然后讨论互信息和互信息熵；最后，讨论**单用户**（无记忆、无反馈、离散）信道的信道容量。

3.1 信道的模型与分类

XI'AN JIAOTONG UNIVERSITY



3.1 信道的模型与分类

XI'AN JIAOTONG UNIVERSITY

一般信道的数学模型：

$$X \longrightarrow \boxed{P(Y|X)} \longrightarrow Y$$

信道的分类

根据输入输出随机信号的特点可以分为**离散信道**、**连续信道**、**半离散**（或半连续）信道以及**波形信道**。

根据用户数量可以分为**单用户信道**和**多用户信道**。单用户信道是指只有一个输入端和一个输出端，信息只朝一个方向传输。

根据信道的记忆特性可分为**无记忆信道**和**有记忆信道**。如果信道的输出信号仅与当前输入信号之间有转移概率关系，而与其它非该时刻的输入、输出信号都无关，就是无记忆信道。

根据信道输入端和输出端的关系可分为**无反馈信道**和**反馈信道**。

根据信道参数与时间的关系可以分为**固定参数信道**和**时变参数信道**。

有干扰信道与**无干扰信道**；**随机差错信道**与**突发差错信道**。

3.1 信道的模型与分类

XI'AN JIAOTONG UNIVERSITY

离散信道的数学模型：输入取值集合$X \in \{a_1, a_2, \cdots a_M\}$，输出取值集合$Y \in \{b_1, b_2, \cdots b_N\}$。前向转移概率$P(y|x)$，它描述了输入信号和输出信号之间的统计依赖关系，显然应有$\sum_y P(y|x) = 1$。

例：二元离散无记忆信道模型

BDMC：Binary Discrete Memoryless Channel。

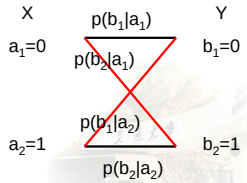
前向转移概率

$P(b_1|a_1)$：输入端发0，信道传输后，输出端收到0的概率；

$P(b_2|a_1)$：发0收1的概率；

$P(b_1|a_2)$：发1收0的概率；

$P(b_2|a_2)$：发1收1的概率。



3.1 信道的模型与分类

XI'AN JIAOTONG UNIVERSITY

特例1：二元对称信道（BSC：Binary Symmetric Channel）

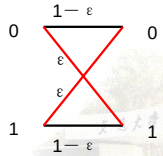
$P(b_1|a_1) = P(b_2|a_2) = 1 - \varepsilon$

$P(b_1|a_2) = P(b_2|a_1) = \varepsilon$

其中ε为传输出错的概率，一般而言：$\varepsilon \le 1/2$

当$\varepsilon = 1/2$时，表示输出和输入无关，强噪信道。强噪信道是不能传递信息的。

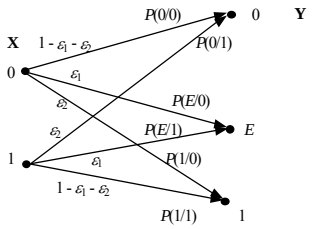
当$\varepsilon = 0$时，理想信道。

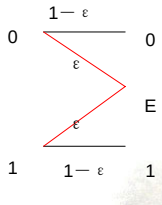


3.1 信道的模型与分类

XI'AN JIAOTONG UNIVERSITY

特例2：二元删错信道





3.2 互信息

XI'AN JIAOTONG UNIVERSITY

互信息的物理意义是信道一次传过的信息量。

3.2.1离散无记忆信道(DMC)的互信息

离散：指信道输入、输出的随机序列的取值在幅度上都是离散的。

无记忆：信道的输出仅与当前的输入有关。

信道的数学描述

设信道输入 $X\in\{a_1,a_2,\cdots,a_K\}$，信道输出 $Y\in\{b_1,b_2,\cdots,b_J\}$，$X$ 与 Y 的联合概率分布为 $P(a_k,b_j)$。由它易求得：

输入概率分布：$P(a_k)=\sum_j P(a_k,b_j)$ 输出概率分布：$P(b_j)=\sum_k P(a_k,b_j)$

前向转移概率分布：

$$P(b_j\mid a_k)=\frac{P(a_k,b_j)}{P(a_k)}$$

后向转移概率分布：

$$P(a_k\mid b_j)=\frac{P(a_k,b_j)}{P(b_j)}$$

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

直观来看，信道一次传多少信息与两个因素有关，一是输入的分布，二是前向转移分布，也就是要知道输入和信道的特性。

如信道输入 a_k，则输入信道的信息量为：

$$I(a_k)=\log\frac{1}{p(a_k)}$$

传过信道后，收方收到 b_j，则接收后 a_k 所含的信息量变为：

$$I(a_k\mid b_j)=\log\frac{1}{p(a_k\mid b_j)}$$

传输前后 a_k 的信息量发生了变化，所以传过的信息量为：

$$I(a_k;b_j)\triangleq I(a_k)-I(a_k\mid b_j)=\log\frac{1}{p(a_k)}-\log\frac{1}{p(a_k\mid b_j)}=\log\frac{p(a_k\mid b_j)}{p(a_k)}$$

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

可见，经过信道传输后，不确定性发生了变化，这个变化的得来是因为信道传输，也就是信道传输消除了一些不确定性，即传输了一些信息量。这个传过的信息量就是前后不确定性的变化，即传输前后 a_k 的信息量之差。

稍做变换，可得：

$$I(a_k;b_j)=\log\frac{p(a_k\mid b_j)}{p(a_k)}=\log\frac{p(a_k\mid b_j)p(b_j)}{p(a_k)p(b_j)}=\log\frac{p(a_k,b_j)}{p(a_k)p(b_j)}=\log\frac{p(b_j\mid a_k)}{p(b_j)}=I(b_j;a_k)$$

此式说明互信息具有对称性，也可以合说成是 a_k 和 b_j 间的相互信息，英文中称为mutual information。

当信道没有干扰时，信源发 a_k，信宿获取其全部信息量，即信源信息通过信道全部流到信宿，此时有：

$$I(a_k;b_j)=I(a_k)$$

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

上面的互信息是某次传输通过信道的信息量，它是个随机量。因此有必要研究一次传输平均通过的信息量，这就是平均互信息，即互信息熵。

$$I(X;Y)\equiv \underset{XY}{E}\left[I(x;y)\right]=\sum_{k,j}p(a_k,b_j)I(a_k;b_j)=\sum_{k,j}p(a_k,b_j)\log\frac{p(a_k\mid b_j)}{p(a_k)}$$

一般写作：

$$I(X;Y)=I(Y;X)=\sum_{x,y}p(x,y)I(x;y)$$

$$=\sum_{x,y}p(x,y)\log\frac{p(x\mid y)}{p(x)}=\sum_{x,y}p(x,y)\log\frac{p(y\mid x)}{p(y)}$$

$I(x;y)$ 可正、可负、可为0；而 $I(X;Y)$ 具有非负性。

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

互信息熵变形1：

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

$$=\sum_{x,y}p(x,y)\log\frac{1}{p(x)}-\sum_{x,y}p(x,y)\log\frac{1}{p(x\mid y)}$$

$$=\sum_{x}p(x)\log\frac{1}{p(x)}-\sum_{x,y}p(x,y)\log\frac{1}{p(x\mid y)}$$

$$=H(X)-H(X\mid Y)$$

$$\sum_{x,y}p(x,y)\log\frac{1}{p(x)}=\sum_{x}\sum_{y}p(x)p(y\mid x)\log\frac{1}{p(x)}$$

$$=\sum_{x}p(x)\log\frac{1}{p(x)}\sum_{y}p(y\mid x)=\sum_{x}p(x)\log\frac{1}{p(x)}$$

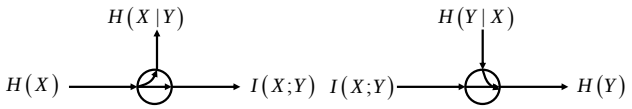
3.2 互信息

XI'AN JIAOTONG UNIVERSITY

$H(X)$ 输入熵：X 的先验不确定度

$H(X\mid Y)$ 条件疑义熵：表示收到变量 Y 后，对随机变量 X 仍然存在的不确定度，代表了在信道中损失的信息，因此又叫损失熵。

意义：输入熵减去因噪声干扰丢掉的疑义熵，就是传过的熵。



互信息熵变形2 ：

由互信息的对称性，可得 $I(X;Y)=H(Y)-H(Y\mid X)$

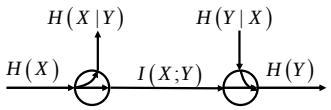
$H(Y)$ 信宿熵

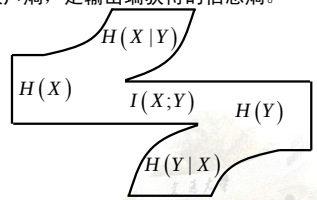
3.2 互信息

XI'AN JIAOTONG UNIVERSITY

$H(Y\mid X)$噪声熵：表示发出随机变量 X 后，对随机变量 Y 仍然存在的平均不确定度。如果信道中不存在任何噪声，发送端和接收端必存在确定的对应关系，发出 X 后必能确定对应的 Y，而现在不能完全确定对应的 Y，这显然是由信道噪声所引起的。

意义：传过的熵加上干扰引入的噪声熵，是输出端获得的信息熵。





互信息熵变形3：

由联合熵与条件熵之间的关系：

$$H(X,Y)=H(X)+H(Y\mid X)=H(Y)+H(X\mid Y)$$

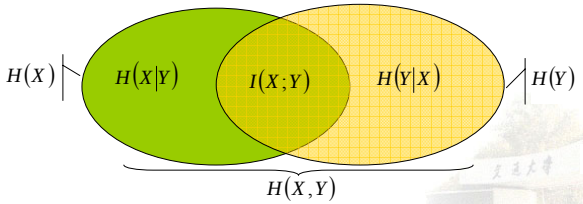
3.2 互信息

XI'AN JIAOTONG UNIVERSITY

将其代入 $I(X;Y)=H(X)-H(X|Y)$，有：

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

各种熵之间的关系可用文氏图表示



3.2 互信息

XI'AN JIAOTONG UNIVERSITY

变形总结

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

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

下面考察理想信道和强噪信道两种极限情况下的互信息熵。

$I(X;Y)$ 表示从 Y 中提取的关于 X 的信息量。

$I(Y;X)$ 表示从 X 中提取的关于 Y 的信息量。

当 X 和 Y 统计独立时，就不可能从一个随机变量获得关于另一个随机变量的信息，因为统计独立的随机变量间不提供任何启示，互相毫无关系，所以 $I(X;Y)=I(Y;X)=0$。而强噪信道的输入和输出相互独立，所以它们不可能相互得到信息。结论：**强噪信道不能用来传信息。**

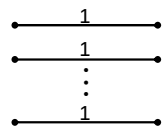
当两个随机变量 X 和 Y 一一对应时，从一个变量就可以充分获得关于另一个变量的信息，即 $I(X;Y)=I(Y;X)=H(X)=H(Y)$，也就是说从 Y 中可以完全获得 X 的信息，从 X 中也可以完全获得 Y 的信息。而理想信道的输入和输出就是一一对应的。结论：**理想信道可传输所有输入的信息。**

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

例1：理想（无噪无损）信道

信道模型，前向转移概率：$p(b_j|a_k)=\begin{cases}1 & k=j\\ 0 & k\neq j\end{cases}$



联合概率

$$p(a_k,b_j)=p(a_k)p(b_j|a_k)=\begin{cases}p(a_k) & k=j\\ 0 & k\neq j\end{cases}$$

互信息熵为：

$$I(X;Y)=\sum_{x,y}p(x,y)\log\frac{p(x|y)}{p(x)}$$
$$=\sum_{k,j}p(a_k,b_j)\log\frac{p(a_k|b_j)}{p(a_k)}$$
$$=\sum_{k}p(a_k)\log\frac{1}{p(a_k)}$$
$$=H(X)$$

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

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

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

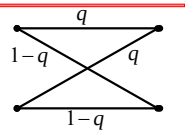
例2. 强噪信道

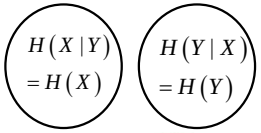
信道模型，前向转移概率：$p(y|x)=p(y)$

联合概率：$p(x,y)=p(x)p(y)$

互信息熵：

$$I(X;Y)=\sum_{x,y}p(x,y)\log\frac{p(y|x)}{p(y)}$$
$$=\sum_{x,y}p(x,y)\log\frac{p(y)}{p(y)}$$
$$=\sum_{x,y}p(x,y)\log 1$$
$$=0$$





$$I(X;Y)=H(X)-H(X|Y)$$
$$=H(Y)-H(Y|X)$$
$$=0$$

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

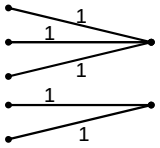
例3：无噪有损信道

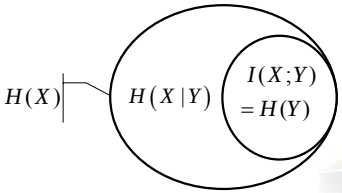
损失熵 $H(X|Y)\neq 0$

噪声熵 $H(Y|X)=0$

所以：$I(X;Y)=H(X)-H(X|Y)$

$=H(Y)$





3.2 互信息

XI'AN JIAOTONG UNIVERSITY

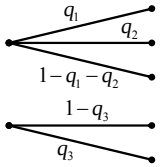
例4：有噪无损信道

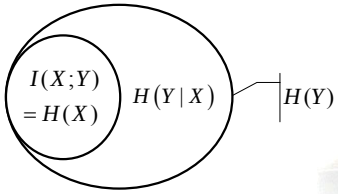
损失熵 $H(X|Y)=0$

噪声熵 $H(Y|X)\neq 0$

所以：$I(X;Y)=H(X)$

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





Example

3.2 互信息

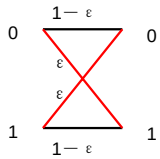
XI'AN JIAOTONG UNIVERSITY

例5：对于二元对称信道BSC，证明

$H(Y|X)=\Omega(\varepsilon)$

$I(X;Y)=\Omega(\varepsilon+p-2\varepsilon p)-\Omega(\varepsilon)$

其中 $\Omega(x)=x\log\frac{1}{x}+(1-x)\log\frac{1}{1-x}$



证：

$$H(\mathbf{Y}|\mathbf{X})=-\sum_{\mathbf{XY}}P(xy)\log P(y|x)$$
$$=-P(0)P(0|0)\log P(0|0)-P(0)P(1|0)\log P(1|0)$$
$$-P(1)P(0|1)\log P(0|1)-P(1)P(1|1)\log P(1|1)$$
$$=-(1-\varepsilon)\log(1-\varepsilon)-\varepsilon\log\varepsilon=\Omega(\varepsilon)$$

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

$H(\mathbf{Y})=\Omega[p(b_1)]$ $\quad p(b_1)=\sum_k p(a_k,b_1)=p(a_1,b_1)+p(a_2,b_1)$

$$=p(a_1)p(b_1|a_1)+p(a_2)p(b_1|a_2)$$
$$=p(1-\varepsilon)+(1-p)\varepsilon$$
$$=p+\varepsilon-2\varepsilon p$$

$H(\mathbf{Y})=\Omega[p(b_1)]=\Omega(p+\varepsilon-2\varepsilon p)$

$\therefore\ I(X;Y)=\Omega(\varepsilon+p-2\varepsilon p)-\Omega(\varepsilon)$

特殊情况

- 强噪信道 $I(X;Y)=\Omega(1/2+p-p)-\Omega(1/2)=0$
- 理想信道 $I(X;Y)=\Omega(0+p-0)-\Omega(0)=H(X)$

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

互信息熵的性质：对称、非负。

定理3.2.1：令(X,Y)为一个离散联合事件集合，则X和Y之间的互信息熵必满足$I(X;Y)\ge 0$，当且仅当X和Y相互统计独立时取等号。

证明：改证 $-I(X;Y)\le 0$

$$-I(X;Y)=\sum_{x,y}p(x,y)\log\frac{p(x)}{p(x|y)}\qquad p(x,y)\ne 0$$
$$=(\log e)\sum_{x,y}p(x,y)\ln\frac{p(x)}{p(x|y)}\le(\log e)\sum_{x,y}p(x,y)\left[\frac{p(x)}{p(x|y)}-1\right]$$
$$=(\log e)\left[\sum_{x,y}p(x)p(y)-\sum_{x,y}p(x,y)\right]=(\log e)[1-1]=0$$

$\therefore\ I(X;Y)\ge 0$

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

当且仅当 $p(x)/p(x|y)=1$，即 $p(x)=p(x|y)$ 时，也即X、Y独立时上式取等号。 #

由 $I(X;Y)=H(X)-H(X|Y)=H(Y)-H(Y|X)\ge 0$

立即可以得到两个推论：

- 输入熵大于等于疑义熵 $H(X)\ge H(X|Y)$
- 信宿熵大于等于噪声熵 $H(Y)\ge H(Y|X)$

无条件的熵大于等于有条件的熵。

换个角度，即 $I(X;Y)\le H(X)$ $\quad I(X;Y)\le H(Y)$

两种特殊信道：

- 理想信道 $I(X;Y)=I(Y;X)=H(X)=H(Y)$
- 强噪信道 $I(X;Y)=I(Y;X)=0$

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

3.2.2 连续无记忆信道的互信息

实际中常见的加性高斯信道就是一种连续无记忆信道。

连续无记忆信源的相对熵。

平均互信息量**具有对称性**并**可用信息熵和条件熵表示：**

$$I(X;Y)=H_{abs}(X)-H_{abs}(X|Y)$$
$$=-\int_{-\infty}^{\infty}p(x)\log p(x)\,\mathrm{d}x-\lim_{\Delta x\to 0}\log\Delta x\int_{-\infty}^{\infty}p(x)\,\mathrm{d}x$$
$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\log p(x|y)\,\mathrm{d}x\,\mathrm{d}y+\lim_{\Delta x\to 0}\log\Delta x\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\,\mathrm{d}x\,\mathrm{d}y$$
$$=-\int_{-\infty}^{\infty}p(x)\log p(x)\,\mathrm{d}x+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\log p(x|y)\,\mathrm{d}x\,\mathrm{d}y$$
$$=H(X)-H(X|Y)\quad \text{信息单位/样值}$$

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

可见，在计算连续消息的平均互信息量时，信源熵和条件熵中的绝对熵中的无穷大项抵消了。

$$
\begin{aligned}
I(X;Y) &= I(Y;X) \\
&= \iint p(x,y)I(x;y)dxdy = \iint p(x,y)I(y;x)dxdy \\
&= \iint p(x,y)\log\frac{p(x|y)}{p(x)}dxdy = \iint p(x,y)\log\frac{p(y|x)}{p(y)}dxdy \\
&= \iint p(x,y)\log\frac{p(x,y)}{p(x)p(y)}dxdy \\
&= H(X)-H(X|Y) \\
&= H(Y)-H(Y|X) \\
&= H(X)+H(Y)-H(X,Y)
\end{aligned}
$$

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

离散时的互信息熵是求和，总可以求出结果。而连续时的互信息熵是积分，并不一定能把积分求出来，这与概率密度的形式有关。

常见的加性高斯信道可以求出互信息熵。

例3.2.2：设一个信道具有加性高斯噪声N。令信道输入为一零均值高斯随机变量X，具有概率密度函数 $p_x(x)=\frac{1}{\sqrt{2\pi E}}\exp\left(-\frac{x^2}{2E}\right)$

其中$E=\delta_x^2$是输入信号的"能量" 或均方值， N是零均值和方差为σ_n^2的高斯变量，即 $p_N(n)=\frac{1}{\sqrt{2\pi\sigma_n^2}}EXP\left(-\frac{x^2}{2\sigma_n^2}\right)$

信道输出 $y=x+n$ ，求互信息熵。

（提示：先求$H(Y|X)$，再求$H(Y)$，最后求$I(X;Y)$）

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

定理3.2.2：若一连续随机变量X有高斯概率分布，且均值为0，方差（或能量）为σ^2，则其自信息熵为

$$H(X)=\frac{1}{2}\log 2\pi e\sigma^2$$

证明：
$$
\begin{aligned}
H(X) &= -\int_{-\infty}^{\infty}p(x)\log p(x)dx \\
&= -\int_{-\infty}^{\infty}p(x)\log\left\{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{x^2}{2\sigma^2}\right]\right\}dx \\
&= -\int_{-\infty}^{\infty}p(x)\log\frac{1}{\sqrt{2\pi\sigma^2}}dx+\frac{\log e}{2\sigma^2}\int_{-\infty}^{\infty}p(x)x^2dx \\
&= \log\sqrt{2\pi\sigma^2}+\frac{\log e}{2\sigma^2}\sigma^2=\frac{1}{2}\log(2\pi\sigma^2)+\frac{1}{2}\log e \\
&= \frac{1}{2}\log(2\pi e\sigma^2) \qquad \#
\end{aligned}
$$

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

$$p(y|x)=p(x+n|x)=p_N(n)=p_N(y-x)=\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left[-\frac{(y-x)^2}{2\sigma_n^2}\right]$$

$$
\begin{aligned}
H(Y|X) &= \iint p(x,y)\log\frac{1}{p(y|x)}dxdy=\int p(x)\left[\int p(y|x)\log\frac{1}{p(y|x)}dy\right]dx \\
&= \int p(x)\left[\int p(y|x)\log\left(\sqrt{2\pi\sigma_n^2}\exp\left[\frac{(y-x)^2}{2\sigma_n^2}\right]\right)dy\right]dx \\
&= \int p(x)\left[\int\log\sqrt{2\pi\sigma_n^2}\,p(y|x)dy+\int p(y|x)\frac{(y-x)^2\log e}{2\sigma_n^2}dy\right]dx \\
&= \int p(x)\left[\log\sqrt{2\pi\sigma_n^2}+\frac{\log e}{2\sigma_n^2}\int p(y|x)(y-x)^2dy\right]dx \\
&= \int p(x)\left[\log\sqrt{2\pi\sigma_n^2}+\frac{\log e}{2\sigma_n^2}\bullet\sigma_n^2\right]dx=\frac{1}{2}\log 2\pi e\sigma_n^2
\end{aligned}
$$

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

再求 $H(Y)$，需要先知道 Y 的概率分布 $p(y)$。由于 $y=x+n$，X 是高斯变量，N 是高斯变量，且 X、N 是相互独立的，所以 Y 也是高斯随机变量，其均值为0，方差为 $\sigma_y^2=\sigma_x^2+\sigma_n^2=E+\sigma_n^2$

$$p(y)=\frac{1}{\sqrt{2\pi\left(E+\sigma_n^2\right)}}\exp\left[-\frac{y^2}{2\left(E+\sigma_n^2\right)}\right] \qquad H(Y)=\frac{1}{2}\log 2\pi e\left(E+\sigma_n^2\right)$$

所以，有：$I(X;Y)=H(Y)-H(Y|X)$

$$=\frac{1}{2}\log 2\pi e\left(E+\sigma_n^2\right)-\frac{1}{2}\log 2\pi e\sigma_n^2$$

$$=\frac{1}{2}\log\frac{E+\sigma_n^2}{\sigma_n^2}$$

$$=\frac{1}{2}\log\left(1+\frac{E}{\sigma_n^2}\right)$$

3.2 互信息

XI'AN JIAOTONG UNIVERSITY

通信原理或其它课程中常见的形式为：$C=W\log(1+S/N)$ bits/s

前面的互信息熵是一次平均传输的信息量。二者通过信息率联系。

在一段时间 T 内的信道容量为：$C=WT\log(1+S/N)$

它表明**一个信道可靠传输的最大信息量完全由** W、T、S/N **所确定**。一旦这三个物理量给定，理想通信系统的极限信息传输率确定了。由此可见，对一定的信息传输率来说，带宽 W，传输时间 T 和信噪功率比 S/N 三者之间可以相互转换。

若传输时间 T 固定，则扩展信道的带宽 W 就可以降低信噪比的要求；反之，带宽变窄，就要增加信噪比。

如果信噪比固定，则增加信道的带宽 W 就可以缩短传送时间 T，换取传输时间的节省；或者花费较长的传输时间来换取频带的节省。

如果频带宽度不变，我们可以采用增加时间 T 来改善信噪比。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

定义3.3.1： 一离散无记忆信道的容量是使用信道一次平均所能传输的最大互信息量，求最大值应对输入概率分布进行。

$$X \longrightarrow \boxed{P(Y|X)} \longrightarrow Y$$

$$C \triangleq \max_{p(x)} I(X;Y)$$

$$=\max_{p(x)}\sum_{x,y}p(x,y)\log\frac{p(y|x)}{p(y)}$$

$$=\max_{p(x)}\sum_{x,y}p(x,y)\log\frac{p(y|x)}{\sum_x p(x)p(y|x)}$$

$$=\max_{p(x)}\sum_{x,y}p(x)p(y|x)\log\frac{p(y|x)}{\sum_x p(x)p(y|x)}$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

互信息熵既跟输入分布有关，又与信道的前向转移分布有关。

应该对谁求最大？ 显然是对输入分布。

- 前向转移概率反映了信道的特性，它的选择往往受实际的物理条件限制，信道一旦选定，信道所具有的转移特性就确定了。
- 从另一角度看，实际中常用同一信道传输具有不同统计特性的消息，我们不可能为每种消息都设计一个最佳信道。
- 对给定信源和信道统计特性，总可以通过在信源与信道输入间插入信道编码器使信道输入具有期望的分布。

所以要看输入分布如何变，才会与给出的信道前向转移分布最为匹配，从而充分利用信道。注意，**在给出信道容量时，一定要同时给出能达到该信道容量的输入分布**，只有这两个同时给出，才是有意义的。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

接下来要做的是：**如何求这个最大值？**

所求式是关于多变量非线性函数在等式 $\sum_x p(x)=1$ 和不等式 $p(x)\ge 0$ 的约束下的求极值问题。

可利用凸函数求极值的办法，共分三步：

1）引入凸函数概念，导出使凸函数取最大值的充要条件；

2）证明 $I(X;Y)$ 是 $p(x)$ 的凸函数，并求 $I(X;Y)$ 最大时 $p(x)$ 应满足的充要条件；

3）由充要条件求解 $p(x)$，求出信道容量。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

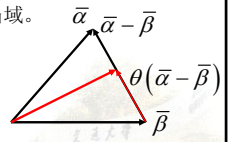
3.3.1 凸域与凸函数

3.3.1.1 凸（区）域

凸函数一定是定义在凸域上的。

定义3.3.2：设 R 是定义在实 K 维空间中的一个域，$\bar{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_K)$ 与 $\bar{\beta}=(\beta_1,\beta_2,\cdots,\beta_K)$ 是域中的任意两个向量。如果对任意的 $0\le\theta\le 1$，向量 $\theta\bar{\alpha}+(1-\theta)\bar{\beta}$ 也在 R 域中，则把 R 域称为凸域。

$$\theta\bar{\alpha}+(1-\theta)\bar{\beta}=\bar{\beta}+\theta(\bar{\alpha}-\bar{\beta})$$
$$=\bar{\alpha}+(1-\theta)(\bar{\beta}-\bar{\alpha})$$

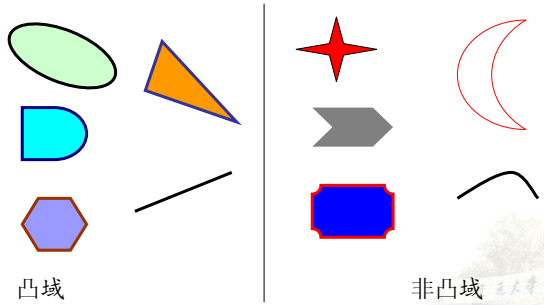


凸域的几何意义：**凸域中任意两点连线必在该域中。**

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

凸域中任意两点连线必在该域中。



一维和二维凸域的例子

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

互信息熵是定义在输入概率分布上的。

定义3.3.3：若一向量所有分量非负且和为1，则该向量称为概率向量。所有概率向量所构成的域（或空间），称为**概率向量域**。

定理3.3.1：概率向量域是凸域。

证明：令 $\bar{\alpha}$ 和 $\bar{\beta}$ 是概率向量域 R 中任意两个向量，根据凸域定义，需证：当 $0\le\theta\le 1$ 时，向量 $\bar{\gamma}=\theta\bar{\alpha}+(1-\theta)\bar{\beta}$ 也是一个概率向量，即各分量非负，且分量和为1。

对 $\bar{\gamma}$ 的任一分量，有：$\gamma_k=\theta\alpha_k+(1-\theta)\beta_k\ge 0$

$\bar{\gamma}$ 的所有分量之和为：$\sum_k\gamma_k=\sum_k[\theta\alpha_k+(1-\theta)\beta_k]$

$$=\theta\sum_k\alpha_k+(1-\theta)\sum_k\beta_k$$
$$=\theta+(1-\theta)=1$$

3.3 信道容量

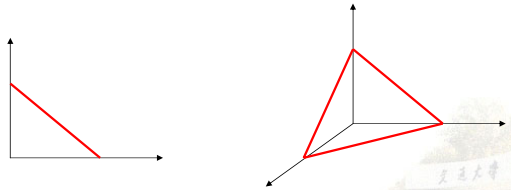
XI'AN JIAOTONG UNIVERSITY

即 $\bar{\gamma}$ 的所有分量非负且分量和为1。

根据定义3.3.3，$\bar{\gamma}$ 是一概率向量，即 $\bar{\gamma}$ 仍在R中。

根据定义3.3.2，概率向量域是凸域。 #

（二维、三维概率向量图）



3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

3.3.1.2 凸函数定义及性质

上凸（Convex ∩）、下凸（Convex ∪，Concave或称凹）

定义3.3.4 凸函数：设一向量的实函数 f 定义于向量空间的某凸域 R，如对于 R 中所有的 $\bar{\alpha}$，$\bar{\beta}$，以及 θ，$0<\theta<1$，函数 f 满足

$$\theta f(\bar{\alpha})+(1-\theta)f(\bar{\beta})\le f\left[\theta\bar{\alpha}+(1-\theta)\bar{\beta}\right]$$

则称 f 为凸函数。如式中不等式反向，则 f 称为凹函数。如式中 ≤代之以<或>，则称 f 严格凸或严格凹。

当 $\bar{\alpha}$ 和 $\bar{\beta}$ 取定时，不等式两边只与 θ 有关。

左边：$g_l(\theta)=\theta f(\bar{\alpha})+(1-\theta)f(\bar{\beta})=\theta\left(f(\bar{\alpha})-f(\bar{\beta})\right)+f(\bar{\beta})$

右边：$g_r(\theta)=f\left[\theta\bar{\alpha}+(1-\theta)\bar{\beta}\right]$

显然：$g_l(0)=g_r(0)=f(\bar{\beta})\quad g_l(1)=g_r(1)=f(\bar{\alpha})$

3.3 信道容量

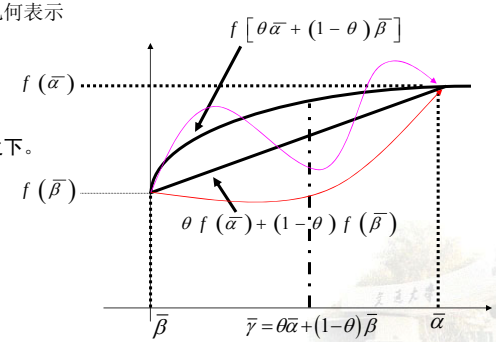
XI'AN JIAOTONG UNIVERSITY

$$\theta f(\bar{\alpha})+(1-\theta)f(\bar{\beta})\le f\left[\theta\bar{\alpha}+(1-\theta)\bar{\beta}\right]$$

一元凸函数的几何表示

意义：

任意弦在曲线之下。



3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

凸函数的性质1：任意凸函数的正线性组合仍是凸函数。

如果 $f_1(\bar{\alpha}), f_2(\bar{\alpha}),\cdots f_L(\bar{\alpha})$ 均为凸函数，且 $c_1, c_2,\cdots c_L$ 是正数，则

$$f_s(\bar{\alpha})\triangleq\sum_{l=1}^{L}c_l f_l(\bar{\alpha})$$ 是凸函数。

如 $f_l(\bar{\alpha})$ 中任一个严格凸，则 $f_s(\bar{\alpha})$ 严格凸。

证明：$\theta f_s(\bar{\alpha})+(1-\theta)f_s(\bar{\beta})=\theta\sum_{l=1}^{L}c_l f_l(\bar{\alpha})+(1-\theta)\sum_{l=1}^{L}c_l f_l(\bar{\beta})$

$$=\sum_{l=1}^{L}c_l[\theta f_l(\bar{\alpha})+(1-\theta)f_l(\bar{\beta})]$$

$$\le\sum_{l=1}^{L}c_l f_l[\theta\bar{\alpha}+(1-\theta)\bar{\beta}]$$

$$=f_s[\theta\bar{\alpha}+(1-\theta)\bar{\beta}]\qquad\#$$

显然，当任意一个严格凸时，小于等于号变为小于号，即组合函数严格凸。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

凸函数的性质2：对于一维向量（一元函数），如在某一区间处处有$d^2 f(\alpha)/d\alpha^2 \le 0$，则$f(\alpha)$在该区间为凸函数。如≤变为<，则$f(\alpha)$是严格凸函数。

凸函数的性质3：如$\bar{\alpha}_1, \bar{\alpha}_2, \cdots \bar{\alpha}_L$是某凸域中一组向量，在该域中$f(\bar{\alpha})$为凸函数，而$(\theta_1, \theta_2, \cdots \theta_L)$是一个概率集合，即分量非负且和为1。则有

$$\sum_{l=1}^{L} \theta_l f(\bar{\alpha}_l) \le f\left[\sum_{l=1}^{L} \theta_l \bar{\alpha}_l\right]。$$

如果将$\bar{\alpha}$看作是一个离散的随机向量，且任一向量$\bar{\alpha}_l$有概率θ_l，则上式等价于$E[f(\bar{\alpha})] \le f[E(\bar{\alpha})]$，即函数的均值不超过均值的函数。

可以通过归纳法证明。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

定理3.3.2：任一随机变量X的信息熵是严格凸函数。

证明：令X的概率分布为$\bar{P} = \{p_1, p_2, \cdots p_K\}$，其信息熵为

$$H(X) = -\sum_{i=1}^{K} p_i \log p_i$$

定义函数$f(p) = -p \log p \quad 0 < p \le 1$，换底，有$f(p) = -p \cdot \log(e) \cdot \ln(p)$

显然，它是关于p的一维函数，对其求导有

$$f'(p) = (-\log e)\left(\ln p + p \cdot \frac{1}{p}\right) = (-\log e)(\ln p + 1); \quad f''(p) = (-\log e) \cdot \frac{1}{p} < 0$$

由性质2，$f(p)$严格凸。

而$H(X) = -\sum_{i=1}^{K} p_i \log p_i = \sum_{i=1}^{K} 1 \cdot f(p_i)$

由性质1，可得$H(X)$是严格凸的。 #

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

3.3.1.3 凸函数求最大值

在介绍了凸域、凸函数的定义、几何意义和性质后，下面看凸函数如何求最大的问题。

求最大值 → 求极大值（边界点） → 求驻点 → 求导

设$f(\bar{\alpha})$是定义在某凸域R上的凸函数，其中$\alpha_k \ge 0, \quad 1 \le k \le K$，求最大值。

求驻点：

$$\frac{\partial f(\bar{\alpha})}{\partial \alpha_k} = 0 \qquad 1 \le k \le K$$

可能出现以下结果，存在驻点且满足约束、存在驻点但不满足约束、不存在驻点。下面分别讨论。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

1） **存在驻点且满足约束**

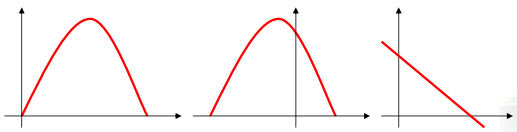
用反证法很容易证明，函数在此点取得最大值。

2） **存在驻点但不满足约束**

即驻点在定义域之外，此时的**最大值在边界**，不能通过求驻点的方法求最大值。

3） **不存在驻点**

说明函数（曲面）单调，此时的**最大值也在边界**。



$$\frac{\partial f(\bar{\alpha})}{\partial \alpha_k} = 0 \quad 对\alpha_k > 0$$

$$\frac{\partial f(\bar{\alpha})}{\partial \alpha_k} \le 0 \quad 对\alpha_k = 0$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

进一步，如果$f(\bar{\alpha})$定义在一个概率向量域上，则约束变为：

$$0 \le \alpha_k < 1, \quad \sum_k \alpha_k = 1, \quad 1 \le k \le K$$

由于每个分量都在0、1之间取值，该约束条件有两个边界。在此只考虑最大点在边界0上的情况。因为如果一个分量为1，就意味着其它分量都为0，也就是变成确定事件了。

利用拉格朗日乘数法，该有约束条件的极值可以等价于求**凸函数**

$$f(\bar{\alpha}) + \lambda\left[1 - \sum_k \alpha_k\right]$$

在给定域的最大值。

根据上面的结论，**定义于概率向量域上的凸函数的最大点应满足：**

$$\frac{\partial f(\bar{\alpha})}{\partial \alpha_k} = \lambda \quad 对于\alpha_k > 0; \quad \frac{\partial f(\bar{\alpha})}{\partial \alpha_k} \le \lambda \quad 对于\alpha_k = 0$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

我们首先使用广义拉格朗日乘数法进行说明。

前述问题等价于在等式$\sum_k \alpha_k = 1$和不等式$\alpha_k \ge 0, \quad 1 \le k \le K$的约束下求$-f(\bar{\alpha})$的最小值。其广义拉格朗日函数为：

$$L(\alpha_k, \lambda, \rho_k) = -f(\bar{\alpha}) + \lambda\left(\sum_{k=1}^{K} \alpha_k - 1\right) + \sum_{k=1}^{K} \rho_k(-\alpha_k)$$

根据Kuhn-Tucker定理，$(\alpha_k^*, \lambda^*, \rho_k^*)$是最优解的充要条件是：

$$\left.\frac{\partial L(\alpha_k, \lambda, \rho_k)}{\partial \alpha_k}\right|_{(\alpha_k^*, \lambda^*, \rho_k^*)} = 0 \qquad \left.\frac{\partial L(\alpha_k, \lambda, \rho_k)}{\partial \lambda}\right|_{(\alpha_k^*, \lambda^*, \rho_k^*)} = 0$$

$$\rho_k^* \alpha_k^* = 0 \qquad -\alpha_k^* \le 0 \qquad \rho_k^* \ge 0 \qquad k = 1, \ldots, K$$

右边称为KKT互补条件

$\rho_k > 0$时，为紧约束

$\rho_k = 0$时，非紧约束

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

对各变量求偏导可得如下方程组：

$$\frac{\partial L}{\partial \alpha_k} = -\frac{\partial f(\bar{\alpha})}{\partial \alpha_k} + \lambda - \rho_k = 0, \quad k = 1, \ldots, K$$

整理可得：

$$\frac{\partial f(\bar{\alpha})}{\partial \alpha_k} = \lambda - \rho_k, \quad k = 1, \ldots, K$$

由KKT条件：

当约束为紧约束时，解在边界上，对应的系数$\rho_k > 0$，因此$\frac{\partial f(\bar{\alpha})}{\partial \alpha_k} \le \lambda$

当约束为非紧约束，对应的系数$\rho_k = 0$，此时$\frac{\partial f(\bar{\alpha})}{\partial \alpha_k} = \lambda$。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

定理3.3.3：令$f(\bar{\alpha})$是定义于区域R的凸函数，$\bar{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_k)$是一概率向量。设偏微分$\partial f(\bar{\alpha}) / \partial \alpha_k$除在个别方向可能有

$$\lim_{\alpha_k \to 0} \frac{\partial f(\bar{\alpha})}{\partial \alpha_k} = \infty$$

外，在R内有定义且连续。则$f(\bar{\alpha})$在R内达到最大的充要条件是：

$$\frac{\partial f(\bar{\alpha})}{\partial \alpha_k} = \lambda \quad 对所有\alpha_k > 0的k$$

$$\frac{\partial f(\bar{\alpha})}{\partial \alpha_k} \le \lambda \quad 对所有\alpha_k = 0的k$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

充分性：即要证明满足两式时，$f(\bar{\alpha})$在R区域内最大。

设$\bar{\beta}$是概率向量域内任意一点，则相当于要证明：$f(\bar{\beta})-f(\bar{\alpha})\le 0$

证明前先总结一下已知条件：

- ✓ $f(\bar{\alpha})$是定义在概率向量域上的凸函数；
- ✓ 一阶偏微分存在且连续；
- ✓ 对于某λ和某一点$\bar{\alpha}$，有下两式成立

$$\frac{\partial f(\bar{\alpha})}{\partial \alpha_k}=\lambda \quad 对于\alpha_k>0;$$

$$\frac{\partial f(\bar{\alpha})}{\partial \alpha_k}\le\lambda \quad 对于\alpha_k=0$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

由凸函数定义，对$0<\theta<1$，有：

$$\theta f(\bar{\beta})+(1-\theta)f(\bar{\alpha})\le f\left[\theta\bar{\beta}+(1-\theta)\bar{\alpha}\right]$$

整理、移项，可得：

$$f(\bar{\beta})-f(\bar{\alpha})\le\frac{f[\theta\bar{\beta}+(1-\theta)\bar{\alpha}]-f(\bar{\alpha})}{\theta}$$

$$=\frac{f[\bar{\alpha}+\theta(\bar{\beta}-\bar{\alpha})]-f(\bar{\alpha})}{\theta}$$

$$=\frac{f(\bar{\alpha}+\Delta\bar{\alpha})-f(\bar{\alpha})}{\theta} \qquad 式中\Delta\bar{\alpha}=\theta(\bar{\beta}-\bar{\alpha})$$

对θ趋于0，上式仍成立。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

由于一阶偏微分存在且连续，使用多元函数的一阶泰勒级数展开，得：

$$f\left[\bar{\alpha}+\theta\left(\bar{\beta}-\bar{\alpha}\right)\right]=f(\bar{\alpha})+\sum_k\theta(\beta_k-\alpha_k)\frac{\partial f(\bar{\alpha})}{\partial\alpha_k}+O\left(\Delta\alpha_k^2\right)$$

代入上式，并令θ趋于0

$$f(\bar{\beta})-f(\bar{\alpha})\le\lim_{\theta\to0}\frac{f[\bar{\alpha}+\theta(\bar{\beta}-\bar{\alpha})]-f(\bar{\alpha})}{\theta}$$

$$=\lim_{\theta\to0}\sum_k(\beta_k-\alpha_k)\frac{\partial f(\bar{\alpha})}{\partial\alpha_k}+O(\theta)$$

$$=\sum_k(\beta_k-\alpha_k)\frac{\partial f(\bar{\alpha})}{\partial\alpha_k}$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

按$\alpha_k=0$和$\alpha_k>0$将上式拆成两部分

$$f(\bar{\beta})-f(\bar{\alpha})\le\sum_k(\beta_k-\alpha_k)\frac{\partial f(\bar{\alpha})}{\partial\alpha_k}$$

$$=\sum_{\substack{K_1\\ \alpha_k=0}}\beta_k\frac{\partial f(\bar{\alpha})}{\partial\alpha_k}+\sum_{\substack{K_2\\ \alpha_k>0}}(\beta_k-\alpha_k)\frac{\partial f(\bar{\alpha})}{\partial\alpha_k}$$

$$\le\sum_{K_1}\beta_k\lambda+\sum_{K_2}(\beta_k-\alpha_k)\lambda$$

$$=\lambda\left(\sum_k\beta_k-\sum_{K_2}\alpha_k\right)=\lambda(1-1)=0$$

充分性得证。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

必要性：若函数 $f(\bar{\alpha})$ 在 $\bar{\alpha}$ 达最大，则 $\bar{\alpha}$ 一定满足定理中二式。

已知条件

- ✓ $f(\bar{\alpha})$ 是定义在概率向量域上的凸函数；
- ✓ $f(\bar{\alpha})$ 在 $\bar{\alpha}$ 处取得最大值。
- ✓ 一阶偏微分存在且连续；

最终证明目标:

$$\frac{\partial f(\bar{\alpha})}{\partial \alpha_k}=\lambda \quad 对于\ \alpha_k>0;$$

$$\frac{\partial f(\bar{\alpha})}{\partial \alpha_k}\le\lambda \quad 对于\ \alpha_k=0$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

由于在 $\bar{\alpha}$ 处达最大值，任取区域内另一点 $\bar{\alpha}+\theta(\bar{\beta}-\bar{\alpha})$，$0<\theta<1$ 有：

$$f\left[\bar{\alpha}+\theta(\bar{\beta}-\bar{\alpha})\right]-f(\bar{\alpha})\le 0$$

两边除以 θ，并代入泰勒级数展开结果，可得：

$$\sum_i(\beta_i-\alpha_i)\frac{\partial f(\bar{\alpha})}{\partial \alpha_i}\le 0$$

由于 $\bar{\alpha}$ 是概率向量，其至少有一个分量大于0，不失一般性，设 $\alpha_1>0$，然后选择另一向量

$$\bar{\beta}=\bar{\alpha}+\varepsilon\bar{I}_k-\varepsilon\bar{I}_1,\quad 其中\ 0<\varepsilon\le\alpha_1,\quad \bar{I}_k=[0,0,\cdots,i_k=1,0,\cdots,0]$$

显然 $\bar{\beta}$ 仍是一概率向量，且

$$\bar{\beta}-\bar{\alpha}=[-\varepsilon,0,\cdots,\varepsilon,\cdots 0]$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

将差向量代入到上式，得

$$\varepsilon\cdot\frac{\partial f(\bar{\alpha})}{\partial \alpha_k}-\varepsilon\cdot\frac{\partial f(\bar{\alpha})}{\partial \alpha_1}\le 0 \qquad \therefore\ \frac{\partial f(\bar{\alpha})}{\partial \alpha_k}\le\frac{\partial f(\bar{\alpha})}{\partial \alpha_1}$$

如果同时还有 $\alpha_k>0$，选 $\bar{\beta}=\bar{\alpha}-\delta\bar{I}_k+\delta\bar{I}_1$，其中 $0<\delta\le\alpha_k$，可得：

$$\delta\cdot\frac{\partial f(\bar{\alpha})}{\partial \alpha_1}-\delta\cdot\frac{\partial f(\bar{\alpha})}{\partial \alpha_k}\le 0 \qquad \therefore\ \frac{\partial f(\bar{\alpha})}{\partial \alpha_k}\ge\frac{\partial f(\bar{\alpha})}{\partial \alpha_1}$$

综上，有

$$\begin{cases}\dfrac{\partial f(\bar{\alpha})}{\partial \alpha_k}=\dfrac{\partial f(\bar{\alpha})}{\partial \alpha_1} & \alpha_k>0\\[2ex] \dfrac{\partial f(\bar{\alpha})}{\partial \alpha_k}\le\dfrac{\partial f(\bar{\alpha})}{\partial \alpha_1} & \alpha_k=0\end{cases}$$

令 $\lambda=\dfrac{\partial f(\bar{\alpha})}{\partial \alpha_1}$，定理得证。#

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

3.3.2 信道容量的计算

首先要证明互信息熵是定义在输入概率分布上的凸函数，这样才能利用凸函数取最大值的充要条件，得到互信息熵达最大值的充要条件。

定理3.3.4：一离散无记忆信道有 K 个输入和 J 个输出，其转移概率为 $P(j|k)$，$1\le j\le J$，$1\le k\le K$。令 $\bar{Q}=[Q(1),\cdots,Q(K)]$ 为信道输入的概率分布。

则

$$I(X;Y)=\sum_{j,k}Q(k)P(j|k)\log\frac{P(j|k)}{\sum_i Q(i)P(j|i)}$$

是 $\bar{Q}$ 的凸函数。

证明中记 $I(X;Y)=f(\bar{Q})$。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

证明1：由于前向转移概率$P(y|x)$确定，任取两个概率向量$\bar{Q}_1$和$\bar{Q}_2$，对应的联合概率分别记为$P_1(x,y)=Q_1(x)P(y|x)$和$P_2(x,y)=Q_2(x)P(y|x)$

输出分布分别记为$P_1(y)=\sum_x P_1(x,y)$和$P_2(y)=\sum_x P_2(x,y)$。

令 $\bar{Q}=\theta\bar{Q}_1+(1-\theta)\bar{Q}_2, 0<\theta<1$，对应的联合概率分布为：

$$P(x,y)=Q(x)P(y|x)=\theta P_1(x,y)+(1-\theta)P_2(x,y)$$

由凸函数定义，需证明：$\theta f(\bar{Q}_1)+(1-\theta)f(\bar{Q}_2)\le f\left[\theta\bar{Q}_1+(1-\theta)\bar{Q}_2\right]$

$$\theta f(\bar{Q}_1)+(1-\theta)f(\bar{Q}_2)-f\left[\bar{Q}\right]$$

$$=\sum_{x,y}\theta P_1(x,y)\log\frac{P(y|x)}{P_1(y)}+\sum_{x,y}(1-\theta)P_2(x,y)\log\frac{P(y|x)}{P_2(y)}$$

$$-\sum_{x,y}\left[\theta P_1(x,y)+(1-\theta)P_2(x,y)\right]\log\frac{P(y|x)}{P(y)}$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

$$=\theta\sum_{x,y}P_1(x,y)\log\frac{P(y)}{P_1(y)}+(1-\theta)\sum_{x,y}P_2(x,y)\log\frac{P(y)}{P_2(y)}$$

因为$\log x$是上凸函数，根据凸函数的性质3，有：

$$\sum_{x,y}P_1(x,y)\log\frac{P(y)}{P_1(y)}\le\log\left[\sum_{x,y}P_1(x,y)\frac{P(y)}{P_1(y)}\right]$$

$$=\log\left[\sum_y\frac{P(y)}{P_1(y)}\sum_x P_1(x,y)\right]=\log\sum_y P(y)=0$$

同理：

$$\sum_{x,y}P_2(x,y)\log\frac{P(y)}{P_2(y)}\le 0$$

由于 $0<\theta,1-\theta<1$，所以有：

$$\theta f(\bar{Q}_1)+(1-\theta)f(\bar{Q}_2)-f\left[\theta\bar{Q}_1+(1-\theta)\bar{Q}_2\right]\le 0$$

即：$\theta f(\bar{Q}_1)+(1-\theta)f(\bar{Q}_2)\le f\left[\theta\bar{Q}_1+(1-\theta)\bar{Q}_2\right]$ 证完#

3.3 信道容量

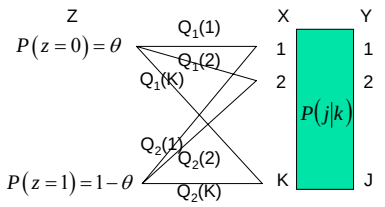
XI'AN JIAOTONG UNIVERSITY

证明2：任取两个概率向量$\bar{Q}_1$和$\bar{Q}_2$，并令

$$\bar{Q}=\theta\bar{Q}_1+(1-\theta)\bar{Q}_2, 0<\theta<1 \quad 。$$

根据凸函数的定义，需要证明：

$$\theta f(\bar{Q}_1)+(1-\theta)f(\bar{Q}_2)\le f\left[\theta\bar{Q}_1+(1-\theta)\bar{Q}_2\right] \qquad *$$



$$\bar{Q}_1=\bar{P}(x|z=0)$$

$$\bar{Q}_2=\bar{P}(x|z=1)$$

$$\bar{Q}=\theta\bar{Q}_1+(1-\theta)\bar{Q}_2$$

$$=p(z=0)\cdot\bar{P}(x|z=0)$$

$$+p(z=1)\cdot\bar{P}(x|z=1)$$

$$=\bar{P}(x,z=0)+\bar{P}(x,z=1)=\bar{P}(x)$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

因此，原不等式（*）的右边可以写作：

$$f\left[\theta\bar{Q}_1+(1-\theta)\bar{Q}_2\right]=f[\bar{Q}]=f[\bar{P}(x)]=I(X;Y)$$

而左边可以写作：

$$f(\bar{Q}_1)=f[\bar{P}(x|z=0)]=I(X;Y|z=0)$$

$$f(\bar{Q}_2)=f[\bar{P}(x|z=1)]=I(X;Y|z=1)$$

$$\therefore\theta f(\bar{Q}_1)+(1-\theta)f(\bar{Q}_2)$$

$$=p(z=0)I(X;Y|z=0)+p(z=1)I(X;Y|z=1)=I(X;Y|Z)$$

于是，证明原不等式（*）可以改为证明：

$$I(X;Y|Z)\le I(X;Y) \qquad **$$

下面倒一个关系式，

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

$$I(Y;ZX)=I(Y;U)=\sum_{y,u}p(y,u)\log\frac{p(y,u)}{p(y)p(u)}=\sum_{x,y,z}p(x,y,z)\log\frac{p(x,y,z)}{p(y)p(x,z)}$$

$$I(Y;Z)+I(Y;X\mid Z)$$

$$=\sum_{y,z}p(y,z)\log\frac{p(y\mid z)}{p(y)}+\sum_{z}p(z)\sum_{x,y}p(x,y\mid z)\log\frac{p(y,x\mid z)}{p(x\mid z)p(y\mid z)}$$

$$=\sum_{y,z}\sum_{x}p(x,y,z)\log\frac{p(y\mid z)}{p(y)}+\sum_{x,y,z}p(x,y,z)\log\frac{p(y,x\mid z)}{p(x\mid z)p(y\mid z)}$$

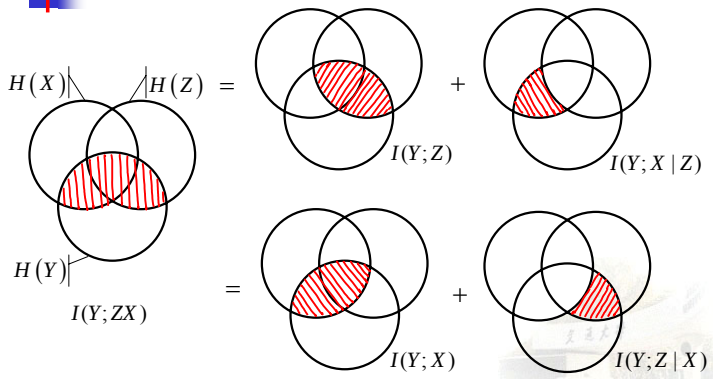
$$=\sum_{x,y,z}p(x,y,z)\log\frac{p(y,x\mid z)p(z)}{p(y)p(x\mid z)p(z)}=\sum_{x,y,z}p(x,y,z)\log\frac{p(x,y,z)}{p(y)p(x,z)}$$

所以：$I(Y;ZX)=I(Y;Z)+I(Y;X\mid Z)$

由于X和Z对称，所以二者互换有：$I(Y;ZX)=I(Y;X)+I(Y;Z\mid X)$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY



3.3 信道容量

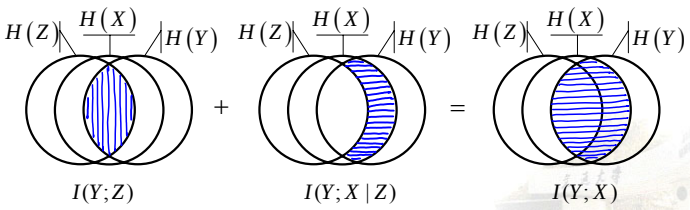
XI'AN JIAOTONG UNIVERSITY

结合以上两式，有：

$$I(Y;Z)+I(Y;X\mid Z)=I(Y;X)+I(Y;Z\mid X)$$

对前述的信道模型，当X给定时，Y和Z之间的互信息等于零，即

$$I(Y;Z\mid X)=0$$



3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

又 $I(Y;Z)\ge 0$

所以可得：$I(Y;X\mid Z)\le I(Y;X)$ 或 $I(X;Y\mid Z)\le I(X;Y)$ 。 证完#

结论：**互信息熵对信道输入概率分布是个凸函数**，因此可用凸函数求极值的方法求互信息熵的最大值。

意义：对给定的信道转移概率分布，总可以找到某一个先验概率分布的信源，使平均互信息量即传送的信息量达到相应的最大值，这时称这个信源为该信道的**匹配信源**。

$$\frac{\partial I(X;Y)}{\partial Q(x_n)}=\lambda \quad 对于 Q(x_n)>0;$$

$$\frac{\partial I(X;Y)}{\partial Q(x_n)}\le\lambda \quad 对于 Q(x_n)=0$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

接下来求互信息熵 $I(X;Y)$ 对概率分量 $Q(x_n)$ 的偏导数。

$$\frac{\partial I(X;Y)}{\partial Q(x_n)}=\frac{\partial}{\partial Q(x_n)}\left[\sum_k\sum_j Q(x_k)p(y_j|x_k)\log\frac{p(y_j|x_k)}{p(y_j)}\right]$$

$$=\sum_k\sum_j\frac{\partial}{\partial Q(x_n)}\left[Q(x_k)p(y_j|x_k)\log\frac{p(y_j|x_k)}{\sum_i Q(x_i)p(y_j|x_i)}\right]$$

$$=\sum_j p(y_j|x_n)\log\frac{p(y_j|x_n)}{p(y_j)}+\sum_k\sum_j Q(x_k)p(y_j|x_k)\frac{\partial}{\partial Q(x_n)}\log\frac{p(y_j|x_k)}{p(y_j)} \quad *$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

对*式中第二项的求偏导部分，有：

$$\frac{\partial}{\partial Q(x_n)}\log\frac{p(y_j|x_k)}{p(y_j)}=\log e\bullet\frac{p(y_j)}{p(y_j|x_k)}\bullet p(y_j|x_k)\bullet\left(-\frac{1}{p^2(y_j)}\right)\bullet\frac{\partial}{\partial Q(x_n)}p(y_j)$$

$$=-\frac{\log e}{p(y_j)}\bullet\frac{\partial}{\partial Q(x_n)}\sum_i Q(x_i)p(y_j|x_i)$$

$$=-\frac{\log e}{p(y_j)}p(y_j|x_n)$$

将结果代入*式的第二项，可得：

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

$$\sum_k\sum_j Q(x_k)p(y_j|x_k)\left[-\frac{\log e}{p(y_j)}p(y_j|x_n)\right]$$

$$=-\log e\sum_k\sum_j p(x_k,y_j)\frac{p(y_j|x_n)}{p(y_j)}$$

$$=-\log e\sum_j\frac{p(y_j|x_n)}{p(y_j)}\sum_k p(x_k,y_j)$$

$$=-\log e\sum_j\frac{p(y_j|x_n)}{p(y_j)}p(y_j)$$

$$=-\log e\sum_j p(y_j|x_n)$$

$$=-\log e$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

所以，*式变为

$$\frac{\partial I(X;Y)}{\partial Q(x_n)}=\sum_j p(y_j|x_n)\log\frac{p(y_j|x_n)}{p(y_j)}-\log e$$

$$=I(X=x_n;Y)-\log e$$

其中，$I(X=x_n;Y)$ 是输入为 x_n 时的互信息熵，称为部分互信息熵。它与互信息熵之间的关系如下：

$$\sum_{x_n}p(x_n)I(X=x_n;Y)=\sum_{x_n}p(x_n)\sum_{y_j}p(y_j|x_n)\log\frac{p(y_j|x_n)}{p(y_j)}$$

$$=\sum_{x_n}\sum_{y_j}p(x_n)p(y_j|x_n)\log\frac{p(y_j|x_n)}{p(y_j)}$$

$$=\sum_{x_n}\sum_{y_j}p(x_n,y_j)\log\frac{p(y_j|x_n)}{p(y_j)}=I(X;Y)$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

定理3.3.5：离散无记忆信道的互信息熵 $I(X;Y)$ 达最大时，输入概率向量 $\bar{Q}=[Q(x_1),Q(x_2),\cdots,Q(x_K)]$ 应满足的充要条件是：

$$\frac{\partial I(X;Y)}{\partial Q(x_n)}=I(X=x_n;Y)-\log e=\lambda \qquad Q(x_n)>0\text{时},$$

$$\frac{\partial I(X;Y)}{\partial Q(x_n)}=I(X=x_n;Y)-\log e\le\lambda \qquad Q(x_n)=0\text{时}。$$

至此，完成了三大步中的前两步，即凸函数求极值的充要条件，互信息熵是输入分布的凸函数且得到了其达到最大时输入分布应满足的充要条件。下面要用这个充要条件求出互信息熵达到最大时的输入分布。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

基于定理3.3.5，可以按下面的步骤求信道容量。

1）对所有概率大于0的输入，求出其部分互信息熵。

2）令所有的部分互信息熵相等，即

$$I(X=x_1;Y)=I(X=x_2;Y)=\cdots=I(X=x_K;Y)$$

总共可以得到 $K-1$ 个独立方程。

再加上约束

$$\sum Q(x_n)=1$$

组成 K 个方程，解出输入概率分布。

3）将解出的概率分布代入互信息熵公式计算信道容量。

除了非常简单的信道，对一般信道而言，这个过程很复杂。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

定理3.3.6：为使一个具有转移概率 $P(y_j|x_k)$ 的离散无记忆信道(DMC)达到其信道容量 C（Capacity），输入概率向量

$$\bar{Q}(x)=[Q(x_1),Q(x_2),\cdots,Q(x_K)]$$

应满足的充要条件是：

$$I(X=x_n;Y)=C \qquad Q(x_n)>0,$$

$$I(X=x_n;Y)\le C \qquad Q(x_n)=0.$$

定理3.3.6与定理3.3.5的区别是，这里终于提到了信道容量 C。这个定理的实质内容是说部分互信息熵小于等于的那个常数就是信道容量。

定理3.3.5中说 $I(X=x_n;Y)-\log e\le\lambda$，$\lambda$ 和 $\log e$ 都是常数，而定理3.3.6说明这个常数 $\lambda+\log e$ 就是信道容量 C。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

令 $\lambda+\log e=C'$

由信道容量定义：$C=\max I(X;Y)=\max\sum_{x_n}Q(x_n)I(X=x_n;Y)$

$$=\sum_{x_n}Q(x_n)C'$$

$$=C'\sum_{x_n}Q(x_n)=C'\bullet 1=C'$$

所以互信息熵小于等于的那个常数就是信道容量 C。

直观解释：如果某个输入有更大的互信息，可以更多使用它，这又会降低该输入的互信息；经过充分变化，所有输入获得相等的信息。

另一方面，只有每个输入都达到互信息最大，才能达到信道容量。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

这个定理很精确、简练（elegance）。它说明，要达到信道容量，要求所有的部分互信息熵都相等，从这些等式方程中就可以求出输入概率分布，相当于解方程组，这个过程在实际信道中其实是非常困难的，对于大多数普通信道都做不到这点。而对于对称信道，可以很容易地求出达到信道容量的输入分布。

定义3.3.5：对称无记忆信道定义

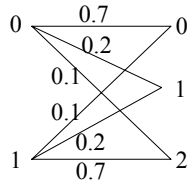
将信道的前向转移概率**以输入为行，输出为列**表示成矩阵形式。如果该矩阵可按列（**输出**）划分为若干个子集，每一子集具有下列特性：每一行都是其它任一行的排列，每一列（如果列数大于1）都是其它任一列的排列，则该信道称为对称信道。

有些参考书，把只有一个子集的称为（**严格**）**对称信道**；需要两个或两个以上子集的称为**准对称信道**。**强对称信道**或均匀信道。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

举例

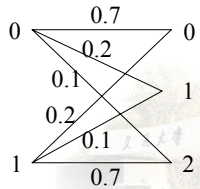


$$\begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & 0.7 & 0.2 & 0.1 \\ 1 & 0.1 & 0.2 & 0.7 \end{array}$$

对称

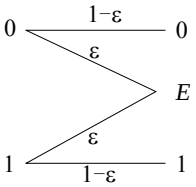
$$\begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & 0.7 & 0.2 & 0.1 \\ 1 & 0.2 & 0.1 & 0.7 \end{array}$$

不对称



3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

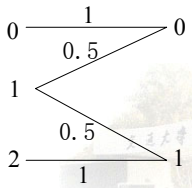


$$\begin{array}{c|ccc} & 0 & E & 1 \\ \hline 0 & 1-\varepsilon & \varepsilon & 0 \\ 1 & 0 & \varepsilon & 1-\varepsilon \end{array}$$

对称

$$\begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1 & 0 \\ 1 & 0.5 & 0.5 \\ 2 & 0 & 1 \end{array}$$

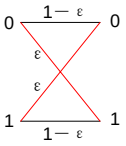
不对称



3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

二元对称信道（BSC）



$$\begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1-\varepsilon & \varepsilon \\ 1 & \varepsilon & 1-\varepsilon \end{array}$$

定理3.3.7：对称离散无记忆信道，当输入等概分布时达到其信道容量。

对于对称信道，该定理由充要条件直接给出了需要求解的输入分布结果，即输入等概。从而省去了求解方程组获得输入分布的过程，使得求信道容量的过程大大简化。知道输入等概后，代入任何一个部分互信息熵的式子即可求出信道容量。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

证明：设信道转移概率矩阵（$K\times J$维）可划分为两个子集，其中的N列为一个子集（I_1），其余$J-N$列为另一子集（I_2）。

写各个输入的部分互信息熵并比较

$$I(x=k;Y)=I_1(x=k;Y)+I_2(x=k;Y)$$

$$I(x=1;Y)=\sum_{j=1}^{N}p(j|1)\log\frac{p(j|1)}{p(j)}+\sum_{j=N+1}^{J}p(j|1)\log\frac{p(j|1)}{p(j)}$$

$$=\sum_{j=1}^{N}p(j|1)\log\frac{p(j|1)}{\sum_i Q(i)p(j|i)}+\sum_{j=N+1}^{J}p(j|1)\log\frac{p(j|1)}{\sum_i Q(i)p(j|i)}$$

输入等概

$Q(i)=1/K$

$i=1,\cdots,K$

$$=\sum_{j=1}^{N}p(j|1)\log\frac{p(j|1)}{\frac{1}{K}\sum_i p(j|i)}+\sum_{j=N+1}^{J}p(j|1)\log\frac{p(j|1)}{\frac{1}{K}\sum_i p(j|i)}$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

所以有：

$$I(x=1;Y)=\sum_{j=1}^{N}p(j|1)\log\frac{p(j|1)}{\frac{1}{K}\sum_i p(j|i)}+\sum_{j=N+1}^{J}p(j|1)\log\frac{p(j|1)}{\frac{1}{K}\sum_i p(j|i)}$$

$$I(x=2;Y)=\sum_{j=1}^{N}p(j|2)\log\frac{p(j|2)}{\frac{1}{K}\sum_i p(j|i)}+\sum_{j=N+1}^{J}p(j|2)\log\frac{p(j|2)}{\frac{1}{K}\sum_i p(j|i)}$$

$$\vdots$$

$$I(x=K;Y)=\sum_{j=1}^{N}p(j|K)\log\frac{p(j|K)}{\frac{1}{K}\sum_i p(j|i)}+\sum_{j=N+1}^{J}p(j|K)\log\frac{p(j|K)}{\frac{1}{K}\sum_i p(j|i)}$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

考虑到对称信道的列排列性质，各式第一个和式中的，

$$\frac{1}{K}\sum_i p(j|i)=C_1$$

再由行排列性质， $I_1(x=k;Y)=C_2$

同理， $I_2(x=k;Y)=C_3$

则 $I(x=k;Y)=C_2+C_3=C \qquad k=1,2,\cdots,K$

由定理3.3.6，信道达到信道容量，且

$$C=I(x=k;Y) \qquad k=1,2,\cdots,K$$

例：计算二元对称信道（BSC）的信道容量。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

方法1：按定理3.3.7求部分互信息熵

$$C=I(x=0;Y)=\sum_j p\left(y_j|x=0\right)\log\frac{p\left(y_j|x=0\right)}{\frac{1}{2}\sum_i p\left(y_j|x_i\right)}$$

$$=p(0|0)\log\frac{p(0|0)}{\frac{1}{2}\left(p(0|0)+p(0|1)\right)}+p(1|0)\log\frac{p(1|0)}{\frac{1}{2}\left(p(1|0)+p(1|1)\right)}$$

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

$$=(1-\varepsilon)\log\left[2(1-\varepsilon)\right]+\varepsilon\log(2\varepsilon)=\log 2-\Omega(\varepsilon)$$

$$=1-\Omega(\varepsilon)\quad \text{bit}$$

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

方法2：按原始定义求互信息熵。在3.2节，我们曾经得到BSC的互信息熵

$$I(X;Y)=H(Y)-H(Y|X)=\Omega(\varepsilon+p-2\varepsilon p)-\Omega(\varepsilon)$$

当信道一定时，第二项是个常数，因此当第一项达到最大时，信道容量达到最大。而第一项是二元变量信息熵，当等概时取最大值1bit。

所以信道容量为：$1-\Omega(\varepsilon)$ bit

也可以：

$$C=\max I(X;Y)=I(X;Y)|_{p=1/2}$$
$$=\Omega(\varepsilon+p-2\varepsilon p)-\Omega(\varepsilon)=\Omega(1/2)-\Omega(\varepsilon)$$
$$=\log 2-\Omega(\varepsilon)=1-\Omega(\varepsilon)\quad \text{bit}$$

方法3

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

（严格）对称信道容量的简化计算

输入、输出均是对称的，不用划分子集（或者说只有一个子集）。

$$C=\max I(X;Y)=\max[H(Y)-H(Y|X)]$$
$$=\max H(Y)-H(Y|X)=\max H(Y)-H(Y|X=x_i)$$

显然，信宿熵最大时达到信道容量。由列排列性质（即列和相等）可以知道：只要输入符号等概，输出符号就是等概的。

所以，（严格）对称信道在输入等概时达到信道容量，且可以用下式简化计算。

$$C=\max H(Y)-H(Y|X=x_i)=\log N_{out}-H(Y|X=x_i)$$

其中，N_{out} 为输出符号的个数。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

准对称信道容量的简化计算

按输出划分子集，每个子集符合对称性。

可以证明，当输入分布等概时，达到信道容量，且可以用下式计算：

$$C=\log N_{in}-H(Y|X=x_i)-\sum_{k=1}^{K}L_k\log C_k$$

Proof

其中：N_{in} 为输入符号的个数；

K是互不相交的子集个数；

L_k是第 k 个子集（子矩阵）的行元素之和；

C_k是第 k 个子集（子矩阵）的列元素之和；

准对称信道的输出不等概。

3.3 信道容量

XI'AN JIAOTONG UNIVERSITY

（严格）对称信道是准对称信道的特例

可以看作只划分了一个子集 $K=1$。

显然有，行和 $L_1=1$；而列和 $C_1=N_{in}\bullet L_1/N_{out}=N_{in}/N_{out}$，代入准对称信道的计算公式有，

$$C=\log N_{in}-H(Y|X=x_i)-\sum_{k=1}^{K}L_k\log C_k$$
$$=\log N_{in}-H(Y|X=x_i)-1\bullet\log N_{in}/N_{out}$$
$$=\log N_{out}-H(Y|X=x_i)$$

3.4 本章小结

XI'AN JIAOTONG UNIVERSITY

- 信道模型的建立；
- DMC互信息量的定义、意义及计算；
- DMC互信息熵的定义、意义、变形、计算、性质；
- 连续无记忆信道互信息熵的定义、计算；
- DMC信道容量的定义、计算；
- 凸域、凸函数的定义、几何意义，凸函数性质、求最大值的充要条件；（定理3.3.1-3.3.3）
- DMC的互信息熵是输入概率分布的凸函数；（定理3.3.4）
- DMC信道部分互信息熵最大的充要条件；（定理3.3.5-3.3.6）
- DMC对称信道的定义及信道容量的计算；（定理3.3.7）
- 一般信道的信道容量的计算（仅限于二元信道）