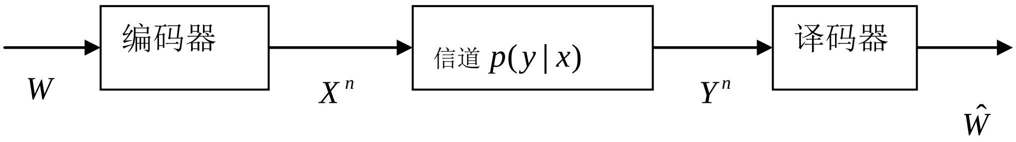
通信信道的数学模型

假设：已知信道转移概率矩阵（可以通过统计试验，以频率代替概率）



符号化表达：

源端的发送行为（向通信信道输入符号 x_i ）的概率：$s_i \triangleq p(x_i)$

输入分布 $S \triangleq \begin{pmatrix} s_1 & s_2 & \dots & s_i & \dots & s_n \end{pmatrix}$

可以定义熵：$H(\mathrm{S}) = -\sum_{i=1}^{n} s_i \log s_i$

转移概率 $p_{ij} \triangleq p(y_j | x_i)$

转移概率矩阵：$P \triangleq (p_{ij})_{n\times m} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nm} \end{pmatrix}$

矩阵的第 i 行记为 $P_i = \begin{pmatrix} p_{i1} & p_{i2} & \dots & p_{ij} & \dots & p_{im} \end{pmatrix}$，是一个概率分布，称为转移矩阵的行分布，简称行分布。

可以定义熵：$H(P_i) = -\sum_{j=1}^{m} p_{ij} \log p_{ij}$，可以称为矩阵的行熵，其本质是源端的某个具体行为所引出的目标端对于状态分布的不确定性度量。

输出概率：目标端检测到某个状态（该状态对应输出符号 $\mathrm{y_j}$ ）的概率 $\mathrm{r_j}$

$r_j \triangleq p(y_j) = \sum_{i=1}^{n} p(x_i) p_{ij} = \sum_{i=1}^{n} s_i p_{ij}$

输出分布 $R \triangleq \begin{pmatrix} r_1 & r_2 & \dots & r_j & \dots & r_m \end{pmatrix}$

可以定义熵：$H(\mathrm{R}) = -\sum_{j=1}^{m} r_j \log r_j$

输入分布、输出分布、转移概率矩阵之间的关系如下：

$$R = SP$$

输入输出联合分布： $p(i,j) \triangleq p(x_i, y_j) = p(x_i)p(y_j \mid x_i) = s_i p_{ij}$

$$U \triangleq (p(i,j))_{n\times m} = \begin{pmatrix} s_1 p_{11} & s_1 p_{12} & \cdots & s_1 p_{1m} \\ s_2 p_{21} & s_2 p_{22} & \cdots & s_2 p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ s_n p_{n1} & s_n p_{n2} & \cdots & s_n p_{nm} \end{pmatrix}$$

$$R = SP$$

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

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

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

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

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

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

命题 1： $I(X,Y) = H(Y) - \sum_{i=1}^{n} s_i H(P_i)$

命题 1 第二项的解释，信道转移概率矩阵的每一行 P_i 构成了一个概率分布，从而有熵 $H(P_i)$，$\sum_{i=1}^{n} s_i H(P_i)$ 是对各行的熵按照输入分布 $S \triangleq \begin{pmatrix} s_1 & s_2 & \dots & s_i & \dots & s_n \end{pmatrix}$ 求平均。

证明：

$$
\begin{aligned}
& I(X;Y)=H(Y)-H(Y\mid X) \\
& =H(Y)-\sum_x p(x)H(Y\mid x) \\
& =H(Y)-\sum_x p(x)\sum_y p(y\mid x)\log\frac{1}{p(y\mid x)} \\
& =H(Y)-\sum_{i=1}^n s_i\sum_{j=1}^m p_{ij}\log\frac{1}{p_{ij}} \\
& =H(Y)-\sum_{i=1}^n s_i H(P_i)
\end{aligned}
$$

对于一般的信道转移概率矩阵，

$$
\begin{aligned}
& I(X,Y)=H(Y)-\sum_{i=1}^n s_i H(P_i) \\
& =\sum_{j=1}^m r_j\log\frac{1}{r_j}-\sum_{i=1}^n s_i H(P_i) \\
& =\sum_{j=1}^m\sum_{i=1}^n s_i p_{ij}\log\frac{1}{\sum_{i=1}^n s_i p_{ij}}-\sum_{i=1}^n s_i H(P_i) \\
& =f(s_1,\ldots,s_i,\ldots,s_n)
\end{aligned}
$$

是关于输入分布 $S\triangleq\begin{pmatrix}s_1 & s_2 & \ldots & s_i & \ldots & s_n\end{pmatrix}$ 的非线性函数，

计算信道容量 $C=\max\limits_{p(x)} I(X;Y)$（其中最大值取自所有可能的输入分布）是**闭凸集上的非线性优化。**

对于一些特殊的矩阵，各行的熵 $H(P_i)$ 相等，问题可以简单化。

命题 2（命题 1 的推论）.

1. 若信道转移概率矩阵的各行互为置换关系，

   则 $I(X,Y)=H(Y)-H(P_i)$，

   且信道容量为 $C=\max\limits_{p(x)} I(X;Y)=\max\limits_{p(x)} H(Y)-H(P_i)$。

2. 进一步，信道转移概率矩阵的各行互为置换关系，

   各列的元素之和相等，则 $C=\max\limits_{p(x)} I(X;Y)=\log m-H(P_i)$，

   在输入分布为等概分布时达到。（证明留作习题）

一些简单特殊的实例

(1) 无噪二元信道（Noiseless Binary Channel）：

$$P=(p_{ij})_{2\times 2}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

$$H(P_i)=0\ (i=1,2)$$

$$C=\max_{p(x)} I(X;Y)=\max_{p(x)} H(Y)-H(P_i)$$
$$=\max_{p(x)} H(Y)=\max_{p(x)} H(R)=\max_{p(x)} H(SP)=\max_{p(x)} H(S)=1$$

(2) 无重叠输出的有噪二元信道（Noisy Channel with Nonoverlapping Outputs）：

$$T=(p_{ij})_{2\times 4}=\begin{pmatrix}\frac{1}{2} & \frac{1}{2} & 0 & 0\\ 0 & 0 & \frac{1}{2} & \frac{1}{2}\end{pmatrix}$$

$$H(P_i)=1\ (i=1,2)$$

$$C=\max_{p(x)} I(X;Y)=\max_{p(x)} H(Y)-H(P_i)$$
$$=\max_{p(x)} H(Y)-1=\max_{p(x)} H(R)-1=\max_{S} H(SP)-1$$
$$=\max_{S} H(\frac{S_1}{2}\ \frac{S_1}{2}\ \frac{S_2}{2}\ \frac{S_2}{2})-1=\log 4-1=2-1$$
$$=1$$

等号的取得条件：输入分布为等概分布。

(3) 有噪打字机(Noisy Typewriter)：

$$P=(p_{ij})_{26\times 26}=\begin{pmatrix}\frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 & 0 & 0\\ 0 & \frac{1}{2} & \frac{1}{2} & \cdots & 0 & 0 & 0\\ 0 & 0 & \frac{1}{2} & \cdots & 0 & 0 & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & \frac{1}{2} & \frac{1}{2} & 0\\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & 0 & 0 & \cdots & 0 & 0 & \frac{1}{2}\end{pmatrix}$$

$H(P_i) = 1\ (i = 1, 2, \ldots, 26)$

由命题 2（2）$C = \max\limits_{p(x)} I(X;Y) = \log m - H(P_i) = \log 26 - 1 = \log 13$

(4) 二元对称信道（Binary Symmetric Channel）

$$P = (p_{ij})_{2\times 2} = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

$H(P_i) = H(p)\quad (i = 1, 2$

由命题 2（2）$C = \max\limits_{p(x)} I(X;Y) = \log m - H(p) = \log 2 - H(p) = 1 - H(p)$

(5) 二元删除信道（Binary Erasure Channel）

$$P = (p_{ij})_{2\times 3} = \begin{pmatrix} 1-\alpha & \alpha & 0 \\ 0 & \alpha & 1-\alpha \end{pmatrix}$$

若 $\alpha = 1/3$，则可以应用命题 2（2）。
在 $\alpha \neq 1/3$ 时，则可以应用命题 2（1）：

$$H(P_i) = (1-\alpha)\log\frac{1}{1-\alpha} + \alpha\log\frac{1}{\alpha} = H(\alpha)\quad (i = 1, 2)$$

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} H(Y) - H(\alpha)$$
$$= \max_{p(x)} H(R) - H(\alpha) = \max_{S} H(S\quad) -PH(\alpha)$$
$$= \max_{S} H((1-\alpha)s_1,\quad \alpha,\quad (1-\alpha)s_2) - H(\alpha)$$
$$= \max_{S}[(1-\alpha)s_1\log\frac{1}{(1-\alpha)s_1} + \alpha\log\frac{1}{\alpha} + (1-\alpha)s_2\log\frac{1}{(1-\alpha)s_2}] - H(\alpha)$$
$$= \max_{S}[(1-\alpha)(s_1\log\frac{1}{s_1} + s_2\log\frac{1}{s_2})]$$
$$= \max_{S}[(1-\alpha)H(S)]$$
$$= 1-\alpha$$

等号的取得条件：输入分布为等概分布。

关于信道容量的几个理论问题

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

1) 数学上的存在性：$I(X;Y)$ 具有哪些性质，足以保证上述定义中的最大值是存在的？
2) 计算上的可行性：如果存在，使用什么算法能够计算出该最大值？

数学上的存在性，可以归结为下面的性质。

信道容量的性质：

1) $0 \le C \le \min(\log n, \log m)$

2) $I(X;Y)$ 是输入分布（一个封闭凸集）的连续凹函数

性质 1 说明了信道容量的非负有限性。性质 2 说明了 $I(X;Y)$ 所取得的局部最大值就是全局最大值。

性质 1 的证明：

$$I(X;Y) \ge 0 \Rightarrow C = \max_{p(x)} I(X;Y) \ge 0$$

$$H(X|Y) \ge 0 \Rightarrow I(X;Y) = H(X) - H(X|Y) \le H(X)$$
$$\Rightarrow C = \max_{p(x)} I(X;Y) \le \max_{p(x)} H(X) = \log n$$

$$H(Y|X) \ge 0 \Rightarrow I(X;Y) = H(Y) - H(Y|X) \le H(Y)$$
$$\Rightarrow C = \max_{p(x)} I(X;Y) \le \max_{p(x)} H(Y) \le \log m$$

结合上述各式下述结论：$0 \le C \le \min(\log n, \log m)$

性质 2 的证明：

输入分布 $S \triangleq \begin{pmatrix} s_1 & s_2 & \dots & s_i & \dots & s_n \end{pmatrix}$

对于一般的信道转移概率矩阵，

$$I(X,Y) = H(Y) - \sum_{i=1}^{n} s_i H(P_i)$$
$$= \sum_{j=1}^{m} r_j \log \frac{1}{r_j} - \sum_{i=1}^{n} s_i H(P_i)$$
$$= -\sum_{j=1}^{m} \sum_{i=1}^{n} s_i p_{ij} \log \sum_{i=1}^{n} s_i p_{ij} - \sum_{i=1}^{n} s_i H(P_i)$$
$$= f(s_1, \dots, s_i, \dots, s_n)$$

上述函数是简单初等函数的复合函数，显然是输入分布的连续函数。

$\sum_{i=1}^{n} s_i H(P_i)$ 是输入分布的线性组合

$\sum_{j=1}^{m} r_j \log \frac{1}{r_j}$ 是输出分布 $R \triangleq \begin{pmatrix} r_1 & r_2 & \ldots & r_j & \ldots & r_m \end{pmatrix}$ 的凹函数，

而 $r_j \triangleq p(y_j) = \sum_{i=1}^{n} p(x_i) p_{ij} = \sum_{i=1}^{n} s_i p_{ij}$ 是输入分布的线性组合，

所以 $I(X,Y)$ 是凹函数

注解： 自变量的线性组合不改变函数的凹凸性。

Kuhn-Tucker 条件

$C=\max\limits_{p(x)} I(X;Y)$

1) 数学上的存在性：$I(X;Y)$ 具有哪些性质，足以保证上述定义中的最大值是存在的？
2) 计算上的可行性：如果存在，使用什么算法能够计算出该最大值？

下面研究计算上的可行性。

由于计算 $C=\max\limits_{p(x)} I(X;Y)$ 可以完整地改写为：

$$C=\max_{p(x)} I(X;Y)=\max_{\mathrm{S}} I(S;SP)=\max_{\mathrm{S}} f(S)$$

$s.t.$

(1) $\sum\limits_{i=1}^{n} s_i=1$

(2) $0\le s_i\le 1,\ i=1,2,\dots,n$

这是有约束的非线性优化问题，它的一般性理论在“非线性最优化”、“非线性规划”研究主题下已经有了相应的结论。但是在本问题中，有一些特别之处可以加以利用：

其一是，约束所形成的集合是封闭凸集；

其二是，约束是线性约束；

其三是，目标函数涉及两个相互制约的部分（信道输入、信道输出）。

其四是，目标函数是凹函数

这些特点决定了可以简化一些定理（例如 Kuhn-Tucker 条件）的表达形式，考虑一些更为有效的算法。

定理（Kuhn-Tucker 条件）：

$f(\alpha)$ 是定义在 R^n 上的凹（concave）函数，$\alpha=(\alpha_1,\alpha_2,\dots,\alpha_i,\dots,\alpha_n)$ 是概率矢量。

假定 $\dfrac{\partial f(\alpha)}{\partial \alpha_i}$（i=1,2,…,n）均存在，且在 R^n 上连续，则 $f(\alpha)$ 在 R^n 上取极大值的充要条件是

$$\begin{cases}\dfrac{\partial f(\alpha)}{\partial \alpha_i}=\lambda, & \forall \alpha_i>0\\[2ex] \dfrac{\partial f(\alpha)}{\partial \alpha_i}\le\lambda, & \forall \alpha_i=0\end{cases}$$

证明：
充分性证明：

设 f 在点 α 处满足 $\begin{cases} \dfrac{\partial f(\alpha)}{\partial \alpha_i} = \lambda, & \forall \alpha_i > 0 \\ \dfrac{\partial f(\alpha)}{\partial \alpha_i} \le \lambda, & \forall \alpha_i = 0 \end{cases}$，下面证明 $f(\alpha)$ 为极大值。

即对任意的概率矢量 $\beta \in R^n$，成立 $f(\beta) - f(\alpha) \le 0$。

$f(\alpha)$ 是定义在 R^n 上的凹函数，所以对任意的 $\theta \in (0,1)$，有

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

其中 $\theta\beta + (1-\theta)\alpha$ 也是概率矢量（注解：全体概率矢量形成闭凸集）。

由于 $\theta \in (0,1)$ 的任意性，可令 $\theta \to 0^+$，得到

$$f(\beta) - f(\alpha) \le \left.\frac{df(\theta\beta + (1-\theta)\alpha)}{d\theta}\right|_{\theta=0}$$
$$= \left.\frac{df(\alpha + \theta(\beta - \alpha))}{d\theta}\right|_{\theta=0}$$
$$= \sum_{i=1}^{n} \frac{\partial f(\alpha)}{\partial \alpha_i}(\beta_i - \alpha_i)$$
$$\le \sum_{i=1}^{n} \lambda(\beta_i - \alpha_i)$$
$$= \lambda(\sum_{i=1}^{n} \beta_i - \sum_{i=1}^{n} \alpha_i)$$
$$= \lambda(1-1)$$
$$= 0$$

充分性获证。

充分性证明中的一个注解：

方向导数的定义 $\dfrac{\partial f(\alpha)}{\partial d} = \lim\limits_{\theta \to 0^+} \dfrac{f(\alpha + \theta d) - f(\alpha)}{\theta \|d\|}$,上面证明中取 $d = \beta - \alpha$ ）

$$\frac{\partial f(\alpha)}{\partial d} = < \nabla f(\alpha), \frac{d}{\|d\|} >$$

$$\lim_{\theta\to 0^+}\frac{f(\alpha+\theta d)-f(\alpha)}{\theta}=\frac{\partial f(\alpha)}{\partial d}\|d\|=<\nabla f(\alpha),\frac{d}{\|d\|}>\|d\|=<\nabla f(\alpha),d>$$

必要性证明：

设 f 在点 α 处取得极大值，且 $\frac{\partial f(\alpha)}{\partial \alpha_i}$（i=1,2,…,n）在点 α 处连续，

则对任意概率矢量 $\beta\in R^n$， $\theta\in(0,1)$ 成立

$$f(\theta\beta+(1-\theta)\alpha)-f(\alpha)\le 0$$
$$\frac{f(\theta\beta+(1-\theta)\alpha)-f(\alpha)}{\theta}\le 0$$

其中 $\theta\beta+(1-\theta)\alpha$ 也是概率矢量。

由于 $\theta\in(0,1)$ 的任意性，可令 $\theta\to 0^+$，得到

$$\left.\frac{df(\theta\beta+(1-\theta)\alpha)}{d\theta}\right|_{\theta=0}$$
$$=\left.\frac{df(\alpha+\theta(\beta-\alpha))}{d\theta}\right|_{\theta=0}$$
$$=\sum_{i=1}^{n}\frac{\partial f(\alpha)}{\partial \alpha_i}(\beta_i-\alpha_i)\le 0$$

由于 α 是概率矢量，所以至少有一个分量是严格正的

（不妨设为 $\alpha_1>0$，且令 $\lambda=\frac{\partial f(\alpha)}{\partial \alpha_1}$），

又由 β 的任意性，不妨设

$$\beta_i=\begin{cases}\alpha_1-\varepsilon & (i=1)\\ \alpha_k+\varepsilon & (i=k\neq 1)\\ \alpha_i & (i\neq 1, i\neq k)\end{cases}$$
$$0<\varepsilon<\alpha_1$$

$$\varepsilon\frac{\partial f(\alpha)}{\partial \alpha_k}-\varepsilon\frac{\partial f(\alpha)}{\partial \alpha_1}\le 0\Rightarrow\frac{\partial f(\alpha)}{\partial \alpha_k}\le\lambda$$

若 $\alpha_k>0$，则还可取 $\varepsilon<0$

$$\varepsilon\frac{\partial f(\alpha)}{\partial \alpha_k}-\varepsilon\frac{\partial f(\alpha)}{\partial \alpha_1}\le 0\Rightarrow\frac{\partial f(\alpha)}{\partial \alpha_k}\ge\lambda$$

从而$\dfrac{\partial f(\alpha)}{\partial \alpha_k} = \lambda$

若$\alpha_k = 0$，则只可取$\varepsilon > 0$，$\dfrac{\partial f(\alpha)}{\partial \alpha_k} \le \lambda$

必要性获证。

下面研究达到信道容量的充要条件。

定理：对信道转移概率矩阵为 P 的离散无记忆信道，

其输入概率分布为$S^*=(s_1^*,s_2^*,...,s_n^*)$，能使$I(X;Y)$取得最大值的充要条件是

$$\begin{cases} I(X=x_i;Y)=C, & \forall i, s_i^*>0 \\ I(X=x_i;Y)\le C, & \forall i, s_i^*=0 \end{cases}$$

$$I(X=x_i;Y)=\sum_{j=1}^{m} p_{ij}\log\frac{p_{ij}}{r_j}=\sum_{j=1}^{m} p_{ij}\log\frac{p_{ij}}{\sum_{i=1}^{n} s_i^* p_{ij}}=D(P_i;R)$$

证明：

$I(X;Y)$是输入分布的凹（concave）函数，C 为极大值，

根据 K-T 条件知，输入概率分布$S^*=(s_1^*,s_2^*,...,s_n^*)$达到 C 的最佳分布的**充要条件**是

存在常数λ，使得

$$\begin{cases} \left.\dfrac{\partial I(X;Y)}{\partial s_i}\right|_{s_i^*}=\lambda, & \forall i, s_i^*>0 \\ \left.\dfrac{\partial I(X;Y)}{\partial s_i}\right|_{s_i^*}\le\lambda, & \forall i, s_i^*=0 \end{cases}$$

$$I(X,Y)=H(Y)-\sum_{i=1}^{n} s_i H(P_i)$$

$$=\sum_{j=1}^{m} r_j\log\frac{1}{r_j}-\sum_{i=1}^{n} s_i H(P_i)$$

$$=-\sum_{j=1}^{m}\sum_{i=1}^{n} s_i p_{ij}\log\sum_{i=1}^{n} s_i p_{ij}-\sum_{i=1}^{n} s_i H(P_i)$$

$$=f(s_1,...,s_i,...,s_n)$$

$$
\begin{aligned}
&\frac{\partial I(X;Y)}{\partial s_i} \\
&= \frac{\partial}{\partial s_i}(H(Y) - \sum_{i=1}^{n} s_i H(P_i)) \\
&= \sum_{j=1}^{m} \frac{\partial H(Y)}{\partial r_j} \frac{\partial r_j}{\partial s_i} - H(P_i) \\
&= \sum_{j=1}^{m} (\log \frac{1}{r_j} - 1) p_{ij} - H(P_i) \\
&= \sum_{j=1}^{m} p_{ij} \log \frac{1}{r_j} - H(P_i) - 1 \\
&= \sum_{j=1}^{m} p_{ij} \log \frac{p_{ij}}{r_j} - 1 \\
&= I(X = x_i; Y) - 1
\end{aligned}
$$

$$
\begin{cases}
\dfrac{\partial I(X;Y)}{\partial s_i} = \lambda, \quad \forall i, s_i^* > 0 \\
\dfrac{\partial I(X;Y)}{\partial s_i} \le \lambda, \quad \forall i, s_i^* = 0
\end{cases}
$$

演变为

$$
\begin{cases}
I(X = x_i; Y) = \lambda + 1, \quad \forall i, s_i^* > 0 \\
I(X = x_i; Y) \le \lambda + 1, \quad \forall i, s_i^* = 0
\end{cases}
$$

令 $C = \lambda + 1$，得命题中的形式

此时 $I(X;Y) = \sum_{i=1}^{n} s_i^* I(X = x_i; Y) = C$

Z 信道的信道容量计算

$$
P = (p_{ij})_{2\times 2} = \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix}
$$

$$
H(P_1) = H(1,0) = 0
$$

$$
H(P_2) = H(1/2, 1/2) = 1
$$

$$
R = (r_1, r_2) = SP = (s_1 + s_2/2, s_2/2)
$$

$$I(X,Y)=H(Y)-\sum_{i=1}^{n}s_iH(P_i)=H(R)-s_2$$

三种解法：其一是转化为单变量的非线性优化问题，其二是利用 K-T 条件，其三是拉格朗日乘子法。

解法一：

$$\frac{s_2}{2}\triangleq p,\ p\in[0,\frac{1}{2}]$$
$$s_1=1-2p$$
$$R=(r_1,r_2)=(1-p,p)$$

$$I(X,Y)=H(R)-s_2=H(1-p,p)-2p=-p\log p-(1-p)\log(1-p)-2p$$

$$\frac{dI(X,Y)}{dp}=\frac{d}{dp}(-p\log p-(1-p)\log(1-p)-2p)$$
$$=-\log p-\log e+\log(1-p)+\log e-2$$
$$=\log\frac{1-p}{p}-2$$

$$0=\frac{dI(X,Y)}{dp}=\log\frac{1-p}{p}-2$$
$$\Rightarrow\frac{1-p}{p}=4$$
$$p=\frac{1}{5}$$
$$(s_1,s_2)=(1-2p,2p)=(\frac{3}{5},\frac{2}{5})$$
$$I(X,Y)=-p\log p-(1-p)\log(1-p)-2p$$
$$=-\frac{1}{5}\log\frac{1}{5}-\frac{4}{5}\log\frac{4}{5}-\frac{2}{5}$$
$$=\log 5-2$$

解法 2

$$\frac{\partial I(X;Y)}{\partial s_i}=\sum_{j=1}^{m}p_{ij}\log\frac{p_{ij}}{r_j}-1=I(X=x_i;Y)-1=\lambda$$

$$\frac{\partial I(X;Y)}{\partial s_1}=\log\frac{1}{r_1}-1=\log\frac{2}{2s_1+s_2}-1=\log\frac{1}{2s_1+s_2}=\lambda$$

$$\frac{\partial I(X;Y)}{\partial s_2}=\frac{1}{2}\log\frac{\frac{1}{2}}{r_1}+\frac{1}{2}\log\frac{\frac{1}{2}}{r_2}-1=\frac{1}{2}\log\frac{1}{2s_1+s_2}+\frac{1}{2}\log\frac{1}{s_2}-1=\lambda$$

$$\Rightarrow 2^{-\lambda}=\frac{8}{5},s_2=\frac{1}{4}2^{-\lambda}=\frac{2}{5},s_1=\frac{3}{8}2^{-\lambda}=\frac{3}{5}$$

$$C=\lambda+1=-\log\frac{8}{5}+1=\log 5-2$$

信道容量的迭代计算算法

1972 年，S.Arimoto 和 R.E.Blahut 分别得到。

$$p_{ij} \triangleq p(y_j \mid x_i)$$

$$P \triangleq (p_{ij})_{n\times m}$$

$$s_i \triangleq p(x_i)$$

$$S \triangleq \begin{pmatrix} s_1 & s_2 & \dots & s_i & \dots & s_n \end{pmatrix}$$

$$r_j \triangleq p(y_j) = \sum_{i=1}^{n} p(x_i) p_{ij} = \sum_{i=1}^{n} s_i p_{ij}$$

$$R \triangleq \begin{pmatrix} r_1 & r_2 & \dots & r_j & \dots & r_m \end{pmatrix}$$

$$R = SP$$

$$\varphi_{ji} \triangleq p(x_i \mid y_j) = \frac{s_i p_{ij}}{r_j} = \frac{s_i p_{ij}}{\sum_{i=1}^{n} s_i p_{ij}}$$

$$\Phi \triangleq (\varphi_{ji})_{m\times n}$$

$$S = R\Phi = SP\Phi$$

$$p(i,j) \triangleq p(x_i, y_j) = p(x_i) p(y_j \mid x_i) = s_i p_{ij} = r_j \varphi_{ji}$$

$S = R\Phi = SP\Phi$ 向我们呈现了 $S = f(S)$ 的形式，该形式提供了迭代计算的可能。

Algorithm for channel capacity:

$$Input: P=(p_{ij})_{n\times m}$$

$$Initialization:$$

$$S^{(0)}=(\frac{1}{n},\frac{1}{n},\dots,\frac{1}{n})$$

$$Iteration:$$

$$\varphi_{ji}^{(k)}=\frac{s_i^{(k)}p_{ij}}{\sum_{i=1}^{n}s_i^{(k)}p_{ij}}$$

$$s_i^{(k+1)}=\frac{\exp(\sum_{j=1}^{m}p_{ij}\log\varphi_{ji}^{(k)})}{\sum_{i=1}^{n}\exp(\sum_{j=1}^{m}p_{ij}\log\varphi_{ji}^{(k)})}$$

$$C^{(k+1)}=\log(\sum_{i=1}^{n}\exp(\sum_{j=1}^{m}p_{ij}\log\varphi_{ji}^{(k)}))$$

关于该算法中的计算公式的由来，可以写成下面的三个定理：

定理 1. 当S和P给定时，对于任何满足$\begin{cases}0\le\varphi_{ji}\le 1\\ \sum_{i=1}^{n}\varphi_{ji}=1\end{cases}$

的矩阵Φ，均有

$$I(X;Y)\ge I(S,\Phi)$$

其中$I(S,\Phi)=\sum_{j=1}^{m}\sum_{i=1}^{n}s_i p_{ij}\log\frac{\varphi_{ji}}{s_i}$

等号成立当且仅当$\varphi_{ji}=\varphi_{ji}^{*}\triangleq\frac{s_i p_{ij}}{r_j}=\frac{s_i p_{ij}}{\sum_{i=1}^{n}s_i p_{ij}}$

证明： 利用简单不等式$\ln x\le x-1$

$$I(S,\Phi)=\sum_{j=1}^{m}\sum_{i=1}^{n}s_i p_{ij}\log\frac{\varphi_{ji}}{s_i}$$

$$I(X;Y)=I(S,\Phi^*)=\sum_{j=1}^{m}\sum_{i=1}^{n}s_i p_{ij}\log\frac{\varphi_{ji}^*}{s_i}$$

$$I(S,\Phi)-I(S,\Phi^*)=\sum_{j=1}^{m}\sum_{i=1}^{n}s_i p_{ij}\log\frac{\varphi_{ji}}{\varphi_{ji}^*}$$

$$\le\sum_{j=1}^{m}\sum_{i=1}^{n}s_i p_{ij}(\frac{\varphi_{ji}}{\varphi_{ji}^*}-1)=0$$

定理 2

当矩阵 Φ 固定时，

$$C=\max_{S} I(S;\Phi)=I(S^*;\Phi)=\log(\sum_{i=1}^{n}\exp(\sum_{j=1}^{m}p_{ij}\log\varphi_{ji}))$$

$$s_i^*=\frac{\exp(\sum_{j=1}^{m}p_{ij}\log\varphi_{ji})}{\sum_{i=1}^{n}\exp(\sum_{j=1}^{m}p_{ij}\log\varphi_{ji})}$$

证明：（提示：利用 K-T 条件）。

定理 3

上述迭代算法是收敛的，即 $\lim_{k\to\infty}C^{(k)}=C$

证明：（提示：考察 $\sum_{i=1}^{n}s_i^*\log\frac{s_i^{(k+1)}}{s_i^{(k)}}$ ）