第1章 信息论基础

1.3 离散信道

第8讲 2007.10.18.

本节的主要内容

- 信道的数学模型
- 信道有关的信息熵
- 平均互信息
- 信道容量

教学目的与要求

1. 掌握离散信道的数学模型及其分类。

2. 熟练掌握与信道有关的各种信息熵的定义及计算方法。

3. 掌握平均互信息的定义、性质及计算方法。（重点）

4. 深刻理解信道容量的概念（难点），掌握对称信道容量的计算方法。

- 计划学时：4学时
- 外语关键词：

信道: **Channel**　　对称信道: **Symmetric channel**

传输矩阵: **Transmit Matrix**

噪声熵: **Noise** Entropy　　损失熵: **Loss** Entropy

信道容量: **Channel Capacity**　　联合熵: **Joint** Entropy

平均互信息: Average Mutual Information

参考文献

1.傅祖芸：**信息论—基础理论与应用**

电子工业出版社（2001年8月第一版）

2.傅祖芸：**信息理论与与编码学习辅导及精选题解**

电子工业出版社（2004年7月第一版）

3.吴伟陵：**信息处理与编码**

人民邮电出版社（1999年7月第一版）

4.曹雪虹：**信息论与编码**

北京邮电大学出版社 (2001年8月第一版)

[温旧引新]

- 通信系统的基本组成

信源 → 信道 → 信宿

发出信息　　传输信息　　接收信息

- 信 息 熵 的 定 义：

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

1.3.1 信道的数学描述

- 信道是信息的传输媒体。这里存在噪声干扰，也包含通信设备对信号的作用，是通信系统最为复杂的部分。

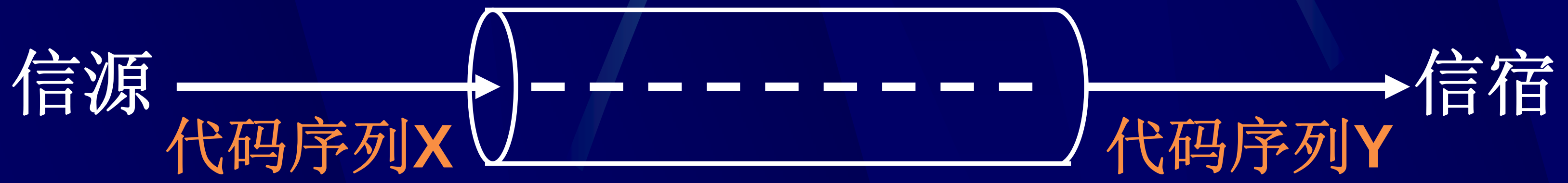
- 在讨论信道中的信息传输规律这个主要问题时，对实际信道抽象出三个简化模型：。

  1. 广义信道模型：
  2. 符号映射模型：
  3. 传输概率模型：

（1）广义信道模型

- 把码型变换、调制与解调、复用与交换等信号处理过程，统统概括在广义信道之中。

- 广义信道对于信源，就是代码序列的入口；对信宿，是代码序列的出口；它只是一个不透明的通道，所有的传输与处理过程都被屏蔽其中。



（2）符号映射模型

- 信源发出的是序列X，信宿收到的是代码Y 。
- 代码序列X取自符号集合$A=\{a_1, a_2,\ldots\ldots, a_m\}$

代码序列Y出现的符号构成集合$B=\{b_1, b_2,\ldots\ldots, b_s\}$

- 为了模型具有普遍性，假设两个集合不同，以便讨论信道中由于噪声、损耗等因素造成误码。
- 通信不是实物的传递，不要求发a_1收a_1，只要X与Y有一定的对应关系，即可传输信息。

（3）传输概率模型

- 当进入广义信道的符号为a_i 时，由于噪声与损耗，从广义信道出来的符号有可能是集合B 中的任意一个，其中是b_j的概率为$p_{ij}=p(b_j|a_i)$，叫前向概率。
- 由于 i=1, 2, ……, m； j=1, 2, ……, s；所以上述条件概率共有ms个。
- 将它们排列成矩阵，叫做信道传输矩阵。

$$P=\begin{pmatrix} p_{11} & p_{12} & \cdots\cdots & p_{1s} \\ p_{21} & p_{22} & \cdots\cdots & p_{2s} \\ \cdots & \cdots & \cdots\cdots & \cdots \\ p_{m1} & p_{m2} & \cdots\cdots & p_{ms} \end{pmatrix}$$

信道的基本类型

- 根据三个简化模型，把信道分为以下类型：

（1）无噪无损信道：

$a_i \leftrightarrow b_j$ 是一一对应的， $p_{ij}=p(b_j|a_i)=\begin{cases} 0 & 当 i \neq j \\ 1 & 当 i = j \end{cases}$

传输矩阵为单位方阵：

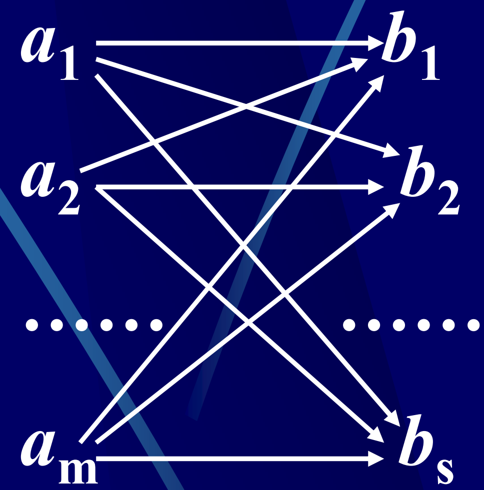
$$P = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

$$\begin{array}{ccc} a_1 & \longleftrightarrow & b_1 \\ a_2 & \longleftrightarrow & b_2 \\ & \cdots\cdots\cdots\cdots & \\ a_m & \longleftrightarrow & b_s \end{array}$$

（2）有噪有损信道:

$a_i \leftrightarrow b_j$ 是多----多对应的， $p_{ij}=p(b_j|a_i) \neq \delta_{ij}$

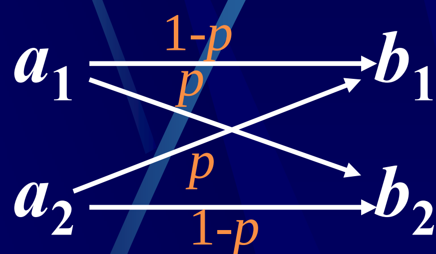
传输矩阵为 $m\times s$ 矩阵。



其中两个典型实例是:

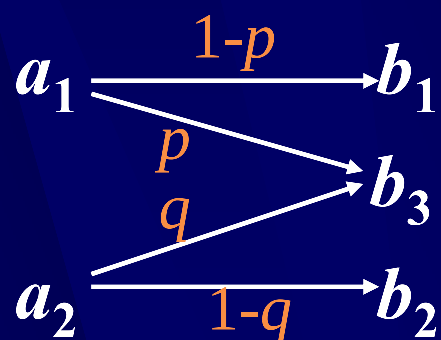
- 二元对称信道（BSC)

$$P=\begin{pmatrix}1-p & p\\ p & 1-p\end{pmatrix}$$



- 二元删除信道（BEC)

$$P=\begin{pmatrix}1-p & 0 & p\\ 0 & 1-q & q\end{pmatrix}$$



（3）有噪无损信道与无噪有损信道：

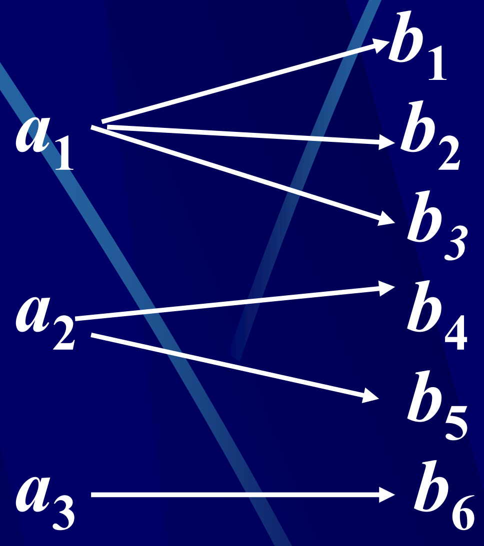
- 噪声与损耗都会造成误码，原因很难区分，而且往往综合发生作用。但是作为数学模型把二者加以区分，可以简化问题，有利于梳理概念。

- 假设噪声是加性的，作用只是增加接收符号的不确定性，让原来的一个接收符号对应一组(多个)接收符号，而不会改变原来的发送符号。我们把这种分组一多对应关系称之为“弥散”。

- 假设损耗是减性的，它只会使若干发送符号因为损失了信息而变得模糊难辨，而不会创生出新的接收符号。其结果是让一组发送符号对应一个接收符号，我们把这种分组多一对应关系称之为“归并”。

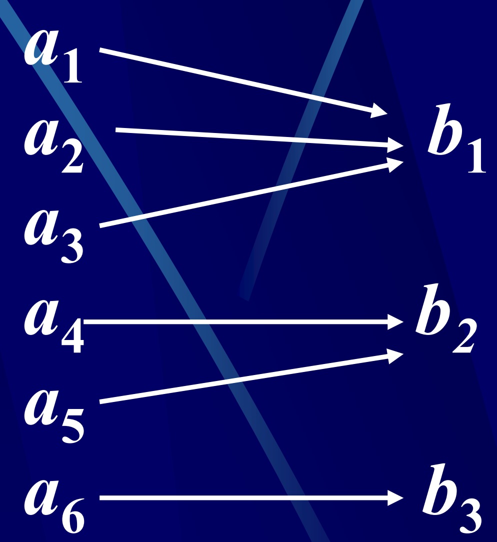
有噪无损信道是只存在弥散不存在归并的传输过程，其结果只能是分组一多对应。

$$P=\begin{pmatrix} p_{11} & p_{12} & p_{13} & 0 & 0 & 0 \\ 0 & 0 & 0 & p_{24} & p_{25} & 0 \\ 0 & 0 & 0 & 0 & 0 & p_{36} \end{pmatrix}$$



有损无噪信道是只存在归并不存在弥散的传输过程，其结果只能是分组多一对应。

$$P=\begin{pmatrix} p_{11} & 0 & 0 \\ p_{21} & 0 & 0 \\ p_{31} & 0 & 0 \\ 0 & p_{42} & 0 \\ 0 & p_{43} & 0 \\ 0 & 0 & p_{63} \end{pmatrix}=\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$



（4）对称信道：

如果传输矩阵各行各列都是一些相同元素的重排，则成为对称矩阵。如：

$$P=\begin{pmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{3} \end{pmatrix}$$

$$P=\begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{pmatrix}$$

如果仅行是重排或仅列是重排，则只能算是准对称矩阵。如：

$$P=\begin{pmatrix} 1-p-\varepsilon & p-\varepsilon & 2\varepsilon \\ p-\varepsilon & 1-p-\varepsilon & 2\varepsilon \end{pmatrix}$$

$$P=\begin{pmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \\ 0.4 & 0.6 \end{pmatrix}$$

1.3.2 信息在信道中的传输特性

1. 通信过程所涉及的各种概率

- 信源符号概率(先验概率)：$p(a_i)$ $i=1, 2, \ldots\ldots m$
- 信道传输概率(前向概率)：$p(b_j|a_i)$ $j=1, 2, \ldots\ldots s$
- 联合概率: $p(a_i b_j) = p(a_i) \cdot p(b_j|a_i)$
- 接收符号概率: $p(b_j) = \sum_{i=1}^{m} p(a_i b_j)$
- 后验概率: $p(a_i \mid b_j) = p(a_i b_j) / p(b_j)$

有关数学知识

- 条件概率：P(A|B), P(B|A)
- 联合概率：P(AB)=P(A)P(B|A)

  或：P(AB)=P(B)p(A|B)
- 归一化条件：Σ_A P(A)=1；Σ_A P(A|B)=1；

  Σ_AP(AB)= P(B)；Σ_BP(AB)= P(A)；

  Σ_A Σ_B P(AB)=1；
- 贝叶斯公式：

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{P(A)P(B|A)}{\sum_A P(AB)}=\frac{P(A)P(B|A)}{\sum_A P(A)P(B|A)}$$

2. 通信过程所涉及的各种信息熵

每种概率都反映一个方面的不确定性，用 **(*-log p*)** 将其表达为自信息，再求出统计平均，就是该方面的信息熵。

（1）信源熵 (先验熵):

它是信源符号概率的统计平均值，代表信源发出符号的平均不确定度。

$$H(X) = -\sum_{i=1}^{m} p(a_i) \log p(a_i)$$

（2）噪声熵 (散布度):

$$H(Y|X) = -\sum_{i=1}^{m}\sum_{j=1}^{s} p(a_i b_j)\log p(b_j|a_i)$$

它是前向传输概率的统计平均值，反映了信道中噪声的有无与大小。理由如下：

- 只要是无噪，就不会有“弥散”发生。不论是无损无噪信道还是有损无噪信道，$p_{ij}= p(b_j|a_i)$都是非1即0（同一分组为1，不同分组为0），必然使$H(Y|X)=0$;
- 若噪声存在，则有“弥散”发生，p_{ij}必为小数，使$H(Y|X) \neq 0$;
- 噪声越强，“弥散”越严重，p_{ij}就越小，$H(Y|X)$越大

（3）联合熵：

$$H(XY) = -\sum_{i=1}^{m}\sum_{j=1}^{s} p(a_i b_j)\log p(a_i b_j)$$

它是联合概率的统计平均值，代表每发送并接收一个符号（无论正确还是误码）的平均不确定程度。

（4）接收符号熵：

它是接收符号概率的统计平均值，代表信宿接收一个符号的平均不确定程度。

$$H(Y) = -\sum_{i=1}^{m} p(b_{j)} \log p(b_j)$$

（5）损失熵（后验熵）：

$$H(X|Y) = -\sum_{i=1}^{m}\sum_{j=1}^{s} p(a_i b_j)\log p(a_i|b_j)$$

它是后验概率的统计平均值，代表信道中损失掉的平均不确定程度，它反映了损失的有无与大小。理由如下：

- 只要是无损，不论是无噪无损还是有噪无损信道，都不会有“归并”发生，从接收符号能唯一指认所发符号，其后验概率 $p(a_i|b_j|)$ 非1即0（同一分组为1，不同分组为0），必然使 $H(X|Y)=0$;
- 若损失存在，后验概率则为小数，使$H(X|Y) \neq 0$;
- 损失越多，“归并”越严重，后验概率就越小，$H(X|Y)$越大。

3. 平均互信息

- 系统完成收发一个符号的通信过程后，关于符号a_i的不确定度的变化为：

$$I(a_i;b_j)=[-\log p(a_i)]-[-\log p(a_i|b_j)]=\log\frac{p(a_i|b_j)}{p(a_i)}$$

式中$p(a_i|b_j)$是收到符号b_j后，关于发送符号为a_i的后验概率。

- 统计平均而言，平均每收发一对符号信宿所获得的信息量为：

$$I(X;Y)=E[I(a_i;b_j)]=\sum_{i=1}^{m}\sum_{j=1}^{s}p(a_ib_j)\log\frac{p(a_i|b_j)}{p(a_i)}$$

- 利用信源熵与损失熵的定义，可推知：

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

- $I\ (X;Y)$叫做平均互信息。代表系统完成收发一个符号的通信过程后，所消除掉的平均不确定度，即信宿从每个符号中平均所获得的信息量，是扣除了损失之后的净信息量。

- 从概念上讲，信息熵与信息量是不同的，信息熵代表平均不确定度，它是一个相对量。信息量是两熵之差，反映通信后中不确定度的变化，它是一个绝对量。

4. 各种信息熵及平均互信息之间的关系

利用：$p(a_i b_j) = p(a_i)\, p(b_j | a_i) = p(b_j)\, p(a_i | b_j)$

有： $\log p(a_i b_j) = \log p(a_i) + \log p(b_j | a_i)$

$= \log p(b_j) + \log p(a_i | b_j)$，

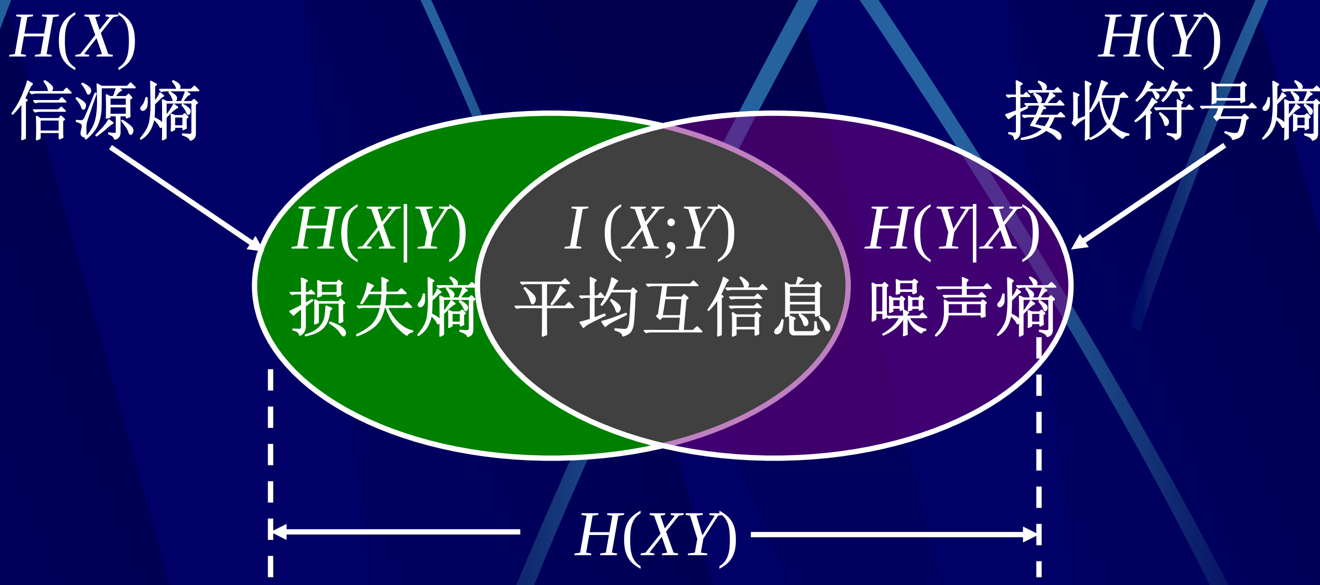
两边求统计平均则不难得到：

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

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

两式相加可得： $I(X;Y) = H(X) + H(Y) - H(XY)$

下图示出信源熵、接收符号熵、损失熵、噪声熵以及平均互信息之间的关系：



从信源熵中扣除损失熵，或从接收符号熵中扣除噪声熵，都会得到平均互信息。

[例2]已知信源先验概率$p(x)$={0.7,0.3}，信道传输矩阵$P=\begin{pmatrix}0.3 & 0.2 & 0.5\\0.4 & 0.3 & 0.3\end{pmatrix}$；试计算各信息熵和互信息。

解:(1)先验熵： $H(X) = -0.7\log_2 0.7 - 0.3\log_2 0.3$

$= (-0.7\lg 0.7 - 0.3\lg 0.3)/\lg 2 = 0.881$ bit/符号

(2)联合熵：

$$P(XY)=\begin{pmatrix}0.7\times0.3 & 0.7\times0.2 & 0.7\times0.5\\0.3\times0.4 & 0.3\times0.3 & 0.3\times0.3\end{pmatrix}=\begin{pmatrix}0.21 & 0.14 & 0.35\\0.12 & 0.09 & 0.09\end{pmatrix}$$

$H(XY) = -0.21\log 0.21 - 0.14\log 0.14 - 0.35\log 0.35$

$-0.12\log 0.12 - 0.09\log 0.09 - 0.09\log 0.09$

$=2.3924$ bit/符号

(3)噪声熵：

由 $P=\begin{pmatrix}0.3 & 0.2 & 0.5\\ 0.4 & 0.3 & 0.3\end{pmatrix}$ 和 $P(XY)=\begin{pmatrix}0.21 & 0.14 & 0.35\\ 0.12 & 0.09 & 0.09\end{pmatrix}$

$H(Y|X)= -0.21\log 0.3 -0.14\log 0.2 -0.35\log 0.5$

$-0.12\log 0.4 -0.09\log 0.3 -0.09\log 0.3$

$= 1.5114$ bit/符号

(4)接收符号熵： 由 $p(y_j)=\sum_{i=1}^{m} p(x_i y_j)$

$P(Y)=$（0.21+0.12，0.14+0.09，0.35+0.09）

=（0.33, 0.23, 0.44）

$H(Y)= -0.33\log 0.33 -0.23\log 0.23 -0.44\log 0.44$

$=1.5366$ bit/符号

(5)损失熵：

$$p(x_i \mid y_j) = \frac{p(a_i b_j)}{p(b_j)} \qquad P(XY) = \begin{pmatrix} 0.21 & 0.14 & 0.35 \\ 0.12 & 0.09 & 0.09 \end{pmatrix}$$

$$P(X \mid Y) = \begin{pmatrix} \frac{0.21}{0.33} & \frac{0.14}{0.23} & \frac{0.35}{0.44} \\ \frac{0.12}{0.33} & \frac{0.09}{0.23} & \frac{0.09}{0.44} \end{pmatrix} = \begin{pmatrix} \frac{7}{11} & \frac{14}{23} & \frac{35}{44} \\ \frac{4}{11} & \frac{9}{23} & \frac{9}{44} \end{pmatrix}$$

$H(X|Y)$= -0.21log(7/11) - ……0.09log(9/44)=0.8558 bit/符号

或：$H(X|Y)=H(XY)-H(Y)$=2.3924-1.5266=0.8558 bit/符号

(6)平均互信息：

$I(X;Y)=H(X)-H(X|Y)$= 0.881 –0.8558=0.0252 bit/符号

1.3.3 平均互信息的性质

1. 非负性

平均互信息不会是负值：$I(X;Y)\geqslant 0$

实际上，因为条件熵不大于无条件熵：

$$H(X)\geqslant H(X|Y)$$

必然有：$I(X;Y)=H(X)-H(X|Y)\geqslant 0$

表明通信总可以获得一些信息，至少是零信息，不会是负信息。

- 值得注意的是平均互信息作为互信息统计平均值:

$$I(X;Y) = E[\,I(a_i\,;\,b_j)\,]$$

是恒为正的，但并不意味着每对互信息 $I(a_i\,;\,b_j)$ 也都恒为正值。对于具体问题中指定的一对a_i 和 b_j，为正为负都有可能。

- 但只要统计平均值为正，就能保证通信总能获得非负的净信息。

2. 有界性

根据定义式 $I(X;Y)=H(X)-H(X|Y)$ 和 $H(X|Y)$ 的非负性即可证明：

$$I(X;Y) \leqslant H(X)$$

表明平均互信息是有界的，最大不会超过信源熵。

这也提示我们，不要奢望获取多于信源所能提供的净信息量。

3. 对称性

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

证明如下：

$$I(a_i;b_j)=\frac{p(a_i|b_j)}{p(a_i)}=\frac{p(b_j)p(a_i|b_j)}{p(b_j)p(a_i)}=\frac{p(a_ib_j)}{p(a_i)p(b_j)}$$

$$=\frac{p(a_i)p(b_j|a_i)}{p(a_i)p(b_j)}=\frac{p(b_j|a_i)}{p(b_j)}=I(b_j;a_i)$$

所以 $I(X;Y)=E[\,I(a_i;b_j)\,]=E[\,I(b_j;a_i)\,]=I(Y;X)$

- 对称性表明信道是可逆的。

4. 极值性

$$I(X;Y)=\sum_{i=1}^{m}\sum_{j=1}^{s}p(a_ib_j)\log\frac{p(b_j|a_i)}{p(b_j)}=\sum_{i=1}^{m}\sum_{j=1}^{s}p(a_i)p(b_j|a_i)\log\frac{p(b_j|a_i)}{p(b_j)}$$

式中： $p(b_j)=\sum_{i=1}^{m}p(a_ib_j)=\sum_{i=1}^{m}p(a_i)p(b_j|a_i)$

- 归根结底，平均互信息$I(X;Y)$是信源发信概率$p(a_i)$和信道传输概率$p(b_j|a_i)$的泛函，信源和信道的统计性质共同决定了互信息的大小。

- 定理1：**当信道给定时，平均互信息仅由信源性质决定，总存在一个信源能使互信息取极大值；**

信源是信息的提供者，对互信息作正贡献。能使信宿得到最大净信息的信源是最佳匹配信源。

- 定理2：**当信源给定时，平均互信息仅由信道性质决定，总存在一个信道能使互信息取极小值。**

信道是信息被损失与受干扰的地方，对互信息起负贡献。能使信宿得到最小净信息的信道是最差不匹配信道。

[例3] 无记忆信源发出0和1两符号的概率分别为ω和1−ω，信道为二元对称信道，错传概率为p；验证极值性。

•**解**：已知信源概率矢量$\boldsymbol{p(X)}$=(ω，1−ω)

信道传输矩阵：$P(Y|X)=\begin{pmatrix}1-p & p\\ p & 1-p\end{pmatrix}$

于是联合概率矩阵：$P(XY)=\begin{pmatrix}(1-p)\omega & p\,\omega\\ p(1-\omega) & (1-p)(1-\omega)\end{pmatrix}$

接收符号概率：

$p(\boldsymbol{Y})=\{(1-p)\,\omega+p(1-\omega)，\ p\omega+(1-p)\,(1-\omega)\}$

- 平均互信息：$I(X;Y)=H(Y)-H(Y|X)$

$$H(Y) = -[(1-p)\omega+p(1-\omega)]\log[(1-p)\omega+p(1-\omega)]$$

$$-[p\omega+(1-p)(1-\omega)]\log[p\omega+(1-p)(1-\omega)]$$

$$H(Y|X)=-(1-p)\ \omega\cdot\log(1-p)-p(1-\omega)\cdot\log p$$

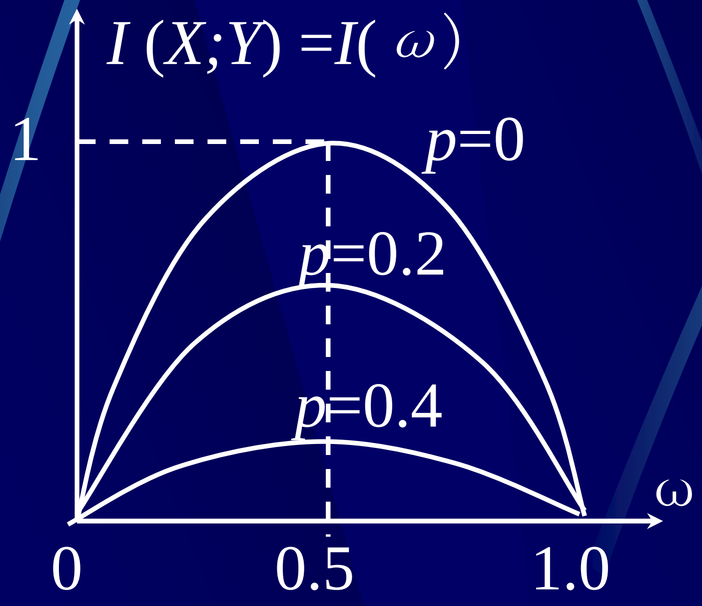
$$-p\ \omega\cdot\log p-(1-p)(1-\omega)\cdot\log(1-p)$$

- 代入整理化简得：$I(X;Y)=H(w)-H(p)$
- 式中：$H(p)=-p\log p-(1-p)\log(1-p)$

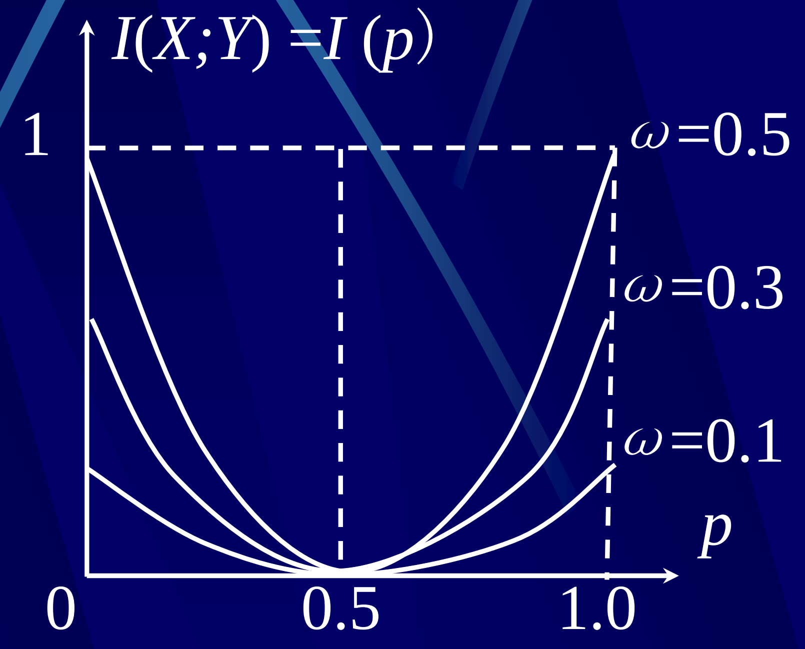
$$H(w)=-w\log w-(1-w)\log(1-w)$$

$$w=(1-p)\omega+p(1-\omega)$$

● 信道p给定时，平均互信息随信源概率ω的变化情况：



● 信源ω给定时，平均互信息随信道传输概率p的变化情况：



1.3.4 信道容量

1. 传码率与传信率

传码率R_B：每秒信道所传输的码元数。单位是波特(B)，所以传码率也叫波特率。一般说来信道所传输的码元就是信源所发出的码元，传码率=发码率。

发信率R_b：每秒信源所发送的信息数。单位叫比特/秒(b/s)，所以发信率也叫比特率，$R_b=R_B \cdot H(X)$ 。

传信率R_t：每秒信道所传输的净信息数。单位也是比特/秒(b/s)，$R_t=R_B \cdot I(X;Y)$。

2. 信道容量的定义

❖ 对于给定的信道，传信率最大可达到多大呢？

❖由上节定理1知，$I(X;Y)$存在一个极大值，这就是该信道所能达到的最大传信率。

❖定义 $$C=\underset{\{p(x)\}}{Max}\, I(X;Y)$$

叫做信道容量，它反映给定信道在通信中平均每符号所能传递的最大净信息量。

❖有时，也把单位时间的最大传信率定义为信道容量，记做 $C_t=C\,R_B$;

- 如何找到这个极大值？由定理1知，给定信道传输概率后，平均互信息只是信源概率的函数，通过改变信源概率，来寻找能使平均互信息取极大的那个信源，然后就能求出该极大值。

- 信道容量的大小，是给定信道的属性，完全由信道本身的性质决定。

- 不要因为是通过寻找最佳信源的方法来求信道容量的就无认为信道容量是与信源有关。信道给定，信道容量就给定了。 不论你用什么方法去寻找它，甚至你不去寻找，它也是存在的。

3. 简单信道的信道容量

（1）无损信道：

因损失熵$H(X|Y)=0$，故互信息：

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

其最大值出现在m个信源符号等概的情况下：

$$C=\max\{I(X;Y)\}=\max\{H(X)\}=\log m\ ;$$

等概信源是最佳信源。

（2）无噪信道：

因噪声熵$H(Y|X)=0$，故：

$I(X;Y)=H(Y)-H(Y|X)=H(Y)$；

其最大值出现在S个输出符号等概的情况下：

$$C=\max\{I(X;Y)\}=\max\{H(Y)\}=\log S;$$

❖须注意，信源发送符号等概未必保证输出符号也等概，这里的最佳信源不一定是等概信源，而是某个能使输出符号等概的信源。

（3）对称信道：

设对称信道的传输矩阵某行元素为：$p_1, p_2, \ldots\ldots, p_S$，这里的$p_i$是相应收、发符号的前向传输概率$p(b_j|a_i)$。则该行元素的平均不确定度为：

$$H(p_1, p_2, \ldots\ldots, p_S) = -p_1 \log p_1 - p_2 \log p_2 \ldots\ldots - p_S \log p_S$$

既然对称信道的传输矩阵各行都是相同元素的重排，那么$H(p_1, p_2, \ldots\ldots, p_S)$就是与行无关的常数，于是噪声熵：

$$H(Y \mid X) = -\sum_{i=1}^{m} p(a_i) \sum_{j=1}^{s} p(b_j \mid a_i) \log p(b_j | a_i)$$

$$= \sum_{i=1}^{m} p(a_i) H(p_1, p_{2_2} \cdots, p_s) = H(p_1, p_2, \cdots, p_s)$$

因此，对称信道的容量为：

$$C=\max\{I(X;Y)\}=\max\{H(Y)-H(Y|X)\}$$

$$=\log S-H(p_1,p_2,\ldots\ldots,p_S);$$

[例3]求对称信道 $P=\begin{pmatrix}0.2 & 0.3 & 0.2 & 0.3\\ 0.3 & 0.2 & 0.3 & 0.2\end{pmatrix}$ 的信道容量

解： $C=\log 4-H(0.2,0.3,0.2,0.3)$

$=2+2\times(0.2\log 0.2+0.3\log 0.3)=0.03$ bit/符号；

4. 信道容量的泛函求解方法

当$m=s$且传输矩阵非奇异时，利用拉格朗日乘子法求解泛函极值，得到以下计算信道容量与最佳信源的方法：

① 列 方 程：$$\sum_{j=1}^{s} p(b_j|a_i)\beta_j = \sum_{j=1}^{s} p(b_j|a_i)\log p(b_j|a_i)$$

例 如 信 道：$$P = \begin{pmatrix} 0.9 & 0.1 \\ .02 & 0.8 \end{pmatrix}$$

方 程 为：$$\begin{cases} 0.9\beta_1 + 0.1\beta_2 = 0.9\log 0.9 + 0.1\log 0.1 \\ 0.2\beta_1 + 0.8\beta_2 = 0.2\log 0.2 + 0.8\log 0.8 \end{cases}$$

求解得到：β_1=−0.433；β_2=−0.794；

②由公式 $C=\log\sum_{j=1}^{s}2^{\beta_j}$ 可求出信道容量：

例中信道容量：$C=\log(2^{-0.433}+2^{-0.794})=0.398$

③接收符号概率 $p(b_j)=2^{\beta_j-C}$

例中 $p(b_1)=2^{-0.433-0.398}=0.562$；同理 $p(b_2)=0.438$

④由 $p(b_j)=\sum_{k=1}^{m}p(a_k)p(b_j\mid a_k)$ 列出关于 $p(a_i)$ 的方程：

例中：

$$\begin{cases}0.9p(a_1)+0.2p(a_2)=0.562\\0.1p(a_1)+0.8p(a_2)=0.438\end{cases}$$

解得最佳信源概率为：$p(a_1)=0.517$；$p(a_2)=0.483$

小结

- 信道的描述：

  信道传输概率，信道的分类。

- 信道有关是信息熵：

  噪声熵，损失熵，信源符号熵，接收符号熵，联合熵。

- 平均互信息：定义，计算方法。
- 信道容量：定义，含义，计算。

课后复习题

- 思考题：

 1. 为什么说平均互信息是通信的净信息？

 2. 如何理解信道容量？

- 作业题：

 教材第32页习题一第11、14、23题；

第1章 信息论基础

1.4 连续信源和波形信道

第16讲 2007.12.4.

本节的主要内容

- 连续信源的数字化
- 连续信源的信息熵
- 波形信道
- Shannon公式

教学目的与要求

1. 正确理解连续信源的信息熵的定义。

2. 掌握连续信源信息熵的特点及计算方法。

3. 掌握波形信道的特点及信道容量的定义、性质。 （难点）

4. 深刻理解Shannon公式的概念和重要意义。 （重点）

- 计划学时：2学时
- 外语关键词：

随机过程：random process

抽样定理：sampling theorem

连续信源：continuous information source

波形信道：wave channel

香农公式：Shannon formula

[温旧引新]

- 离散信源信息熵的定义：

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

- 离散信道平均互信息的定义：

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

- 离散信道信道容量的定义：

$$C=\underset{\{p(x)\}}{Max}\, I(X;Y)$$

1.4.1 连续信源的信息熵

- 连续信号是应用极广的一类信号。
- 发送连续信号的信源叫连续信源。
- 传输连续信号的信道叫波形信道。
- 如何将离散信源与信道的信源熵的理论推广到连续信号的有关问题中呢？
- 首先是连续信号信息熵定义问题。

1. 连续信号的数学描述

- 通信中的连续信号是用随机过程来描述的。
- 随机过程有两个要素，一是时间，二是样本。
- 从头到尾地观测一次随机信号，则得到它的一个样本函数$x_i(t)$, 它随时间无规变化，并且每次观测都不会相同。全体样本的集合构成随机过程。样本函数是随机过程的一次实现。
- 每给定一个时间s，各个样本函数都会有一个不同的取值，它们随机分布，其全体构成一个随机变量Xs。随机变量是随机过程在给定时刻的抽样。

- 连续信号通过抽样、量化两步变为离散信号。
- 抽样就是抽取随机过程在给定时刻的随机变量。
- 量化则是将随机变量离散化的过程。
- 为了将离散信号的信源熵的理论推广到连续信号，也应当分两步来实现：

（1）实现量化的逆过程：将离散随机变量的信息熵推广到连续的随机变量；

（2）实现抽样的逆过程：将某时刻的一个样本的信息熵推广到等间隔的各个时刻。

2. 连续信源的信息熵的定义

（1）从离散到连续的推广：

- 高等数学定积分中曾多次用等份区间的方法，化连续函数为分段不连续变量来求面积、弧长……，当区间无穷小时，其极限便是相应的结果。
- 现在用这个方法来计算连续信源的信息熵。
- 设连续信号取值范围在 $[a, b]$ 的有限区间，其概率密度函数为 $p(x)$；
- 将 $[a,b]$ 均分为 M 个量化台阶，宽度 $\Delta=(b-a)/M$

- 样本落在第*i*个量化台阶的概率：

$$P_i = \int_{a+(i-1)\Delta}^{a+i\Delta} p(x)dx$$

- 根据中值定理，在该台阶内总有一点x_i能使：

$$P_i = p(x_i)\cdot\Delta$$

- 于是离散分布的量化值的信息熵为：

$$H(X_M) = -\sum_{i=1}^{M} P_i \log P_i = -\sum_{i=1}^{M} [p(x_i)\Delta]\log[p(x_i)\Delta] =$$

$$= -\sum_{i=1}^{M} p(x_i)\Delta \log p(x_i) - \sum_{i=1}^{M} p(x_i)\Delta \log \Delta$$

- 当$M\to\infty$时 $\Delta\to 0$，求和就变成了积分：

$$H(x)=\lim_{\Delta\to 0}H(X_M)=-\int_{a}^{a}p(x)\log p(x)dx+\infty$$

- 信息熵代表信源发信的不确定度，连续信源取值无穷多，其信息熵无穷大是合理的。
- 求熵差时两个无穷大互相抵消，并不影响信息量的大小
- 定义连续信源的**相对熵**为：

$$h(x)=-\int_{-\infty}^{\infty}p(x)\log p(x)dx$$

（2）从单个样本到序列的推广：

- 抽样定理指出，对于频率有限的连续信号，只要抽样频率不小于信号最高频率F的两倍，就能无失真地恢复原信号。

- 如果信号存在的时间为T，那么抽样信号就是一个具有$N=2FT$个样值的离散时间函数。

- 从信息角度看，它就是长度为N的随机符号序列。

- 如果各个样值彼此无关，则序列的信息熵等于：

$$h(\vec{x}) = N \cdot h(x)$$

- 如果各个样值相互关联（有记忆信源），则序列的信息熵：$h(\vec{x}) \le N \cdot h(x)$

[例1]一维连续信号x(t)在[*a*,*b*]取间均匀分布，其概率密度函数为：

$$p(x)=\begin{cases}0 & \text{当} x<a \text{ 和 } x>b \\ \frac{1}{b-a} & \text{当} a\le x\le b\end{cases}$$

求相对熵和序列信息熵。

解：

$$h(x)=-\int_{-\infty}^{\infty} p(x)\log p(x)dx$$

$$=\frac{1}{b-a}\int_a^b \log(b-a)dx=\log(b-a)$$

$$h(\vec{x})=N\cdot h(x)=2FT\log(b-a)$$

[例2]一维高斯分布的连续信号，其概率密度函数为：
$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-m)^2}{2\sigma^2} \right)$$

求相对熵和序列信息熵。

解：因 $-\log p(x) = \frac{1}{2}\log(2\pi\sigma^2) + \frac{(x-m)^2}{2\sigma^2}\log e$

利用 $\int_{-\infty}^{\infty} p(x)dx = 1$ 和 $\int (x-m)^2 p(x)dx = \sigma^2$

$$h(X) = -\int_{-\infty}^{\infty} p(x)\log p(x)dx = \frac{1}{2}\log(2\pi e\sigma^2)$$

$$h(\vec{X}) = \frac{N}{2}\log(2\pi e\sigma^2)$$

1.4.2 具有最大熵的连续信源

1、峰值功率受限条件下的信源最大熵:

- 峰值功率受限等价于信号幅度（即随机变量的取值）受限。
- 可以证明：若输出波幅度限定在[a, b]之内，则均匀分布的连续信源具有最大熵，熵值为 $h_{max}(X)=\log(b-a)$;
- 进而推知，当随机向量幅度受限时，只有各分量统计独立且均匀分布时，信源具有最大熵。

2、平均功率受限条件下的信源最大熵：

- 对均值为零的信号而言，平均功率受限等价于方差受限。
- 可以证明：当输出波方差为有限值时，正态分布的连续信源具有最大熵，其熵值为：

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

- 均值为零时，方差 σ^2 就等于平均功率P，所以这个最大熵又可写为：

$$h_{\max}(X)=\frac{1}{2}\log(2\pi eP)$$

3、熵功率：

- 高斯白噪声是均值为零、平均功率为$n_0/2$的正态随机过程，它具有最大熵：$h_{\max}(X)=\frac{1}{2}\log(\pi e n_0)$
- 其它分布方式的连续信源，若平均功率相同，熵值必小于此值。
- 反之，在相同熵值的条件下高斯信源应具有最小的平均功率：$P_0=\frac{2^{2h}}{2\pi e}$
- 其它分布形式的信源，若熵值也等于h，则平均功率一定大于P_0，**P_0是与该信源有相同熵值的高斯信源的平均功率，叫做熵功率。**

1.4.3 波形信道的信道容量

1、波形信道的互信息：

- 原理：

- 把离散信道理论推广到波形信道，同样也经历两步过渡：
- 第一步，从单符号离散信道过渡到传输单个连续随机变量的**基本连续信道**；
- 第二步，从传输单个连续随机变量的基本连续信道过渡到传输多个连续随机变量的**多维连续信道**；当维数趋于无穷时，多维连续信道就回到了波形信道。

- 对于基本连续信道，需要讨论输入和输出两个随机变量的X与Y的问题。

- 仿照对连续信源信息熵的推广办法，不难推广得到连续信道的联合熵、条件熵、以及互信息：

- 联合熵： $h(XY) = -\iint p(xy)\log p(xy)dxdy$

- 噪声熵 $h(Y|X) = -\iint p(xy)\log p(y|x)dxdy$

- 损失熵 $h(X|Y) = -\iint p(xy)\log p(x|y)dxdy$

- 平均互信息： $I(X;Y) = h(X) - h(X|Y)$

$$= h(Y)-h(Y|X)= h(X)+h(Y)- h(XY)$$

- 推广到多维连续信道时，输入、输出的随机变量X和Y过渡为随机向量$X=(x_1x_2\ldots\ldots x_N)$和 $Y=(y_1y_2\ldots\ldots y_N)$
- 联合熵、条件熵以及互信息过渡为：
- 联合熵：$h(\vec{X}\vec{Y}) = -\iint p(\vec{x}\vec{y})\log p(\vec{x}\vec{y})d\vec{x}d\vec{y}$
- 噪声熵：$h(\vec{Y}\,|\,\vec{X}) = -\iint p(\vec{x}\vec{y})\log p(\vec{y}\,|\,\vec{x})d\vec{x}d\vec{y}$
- 损失熵：$h(\vec{X}\,|\,\vec{Y}) = -\iint p(\vec{x}\vec{y})\log p(\vec{x}\,|\,\vec{y})d\vec{x}d\vec{y}$
- 平均互信息：$I(\vec{X};\vec{Y}) = h(\vec{X}) - h(\vec{X}\,|\,\vec{Y}) = h(\vec{Y}) - h(\vec{Y}\,|\,\vec{X})$

$$= h(\vec{X}) + h(\vec{Y}) - h(\vec{X}\vec{Y})$$

2、波形信道的信道容量：

- 先考虑传输单个随机变量的高斯加性信道。输入、输出信道的随机变量为X和Y，**高斯信道**指混入信道中的噪声的概率密度分布呈高斯型，**加性**信道指噪声n与输入变量X的关系是统计无关的相加关系：$Y = X + n$
- 因联合概率密度：$p(xy)=p(xn)=p(x)p(n)$
- 故信道传输概率：$p(y|x)=p(xy)/p(x)=p(n)$

$$h(Y \mid X) = -\iint p(xy)\log p(y \mid x)dxdy = -\int p(n)\log p(n)dn$$

- 互信息：$I(X;Y)=h(Y)-h(Y|X)= h(Y)- h(n)$
- 信道容量： $C = \max\,[h(Y)- h(n)]$
- 实际通信系统中，大量存在的热噪声和环境噪声都具有高斯白噪声的特点，设它的平均功率为P_n，则：

$$h(n)=\frac{1}{2}\log(2\pi e\sigma^2)=\frac{1}{2}\log(2\pi eP_n)$$

- 另一方面，在平均功率为有限值时，$h(Y)$的最大值应出现在Y的均值为零且为高斯分布的的情况下，
- 于是, 平均功率受限的高斯加性信道的信道容量为

$$C=\frac{1}{2}\log(2\pi eP_Y)-\frac{1}{2}\log(2\pi eP_n)=\frac{1}{2}\log\frac{P_Y}{P_n}$$

- 习惯上，用发送端信号功率来写公式，由

$$\sigma_Y^2 = \sigma_X^2 + \sigma_n^2$$

得到：$P_Y=P_X+P_n$；

所以 $C = \frac{1}{2}\log(1+\frac{P_X}{P_n})$

- 推广到多维无记忆高斯加性连续信道，当各抽样值(i=1, 2, ……, N)方差都相同时

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

1.4.4 香农公式及其意义

- 当抽样频率为2F时，注意N=2FT就有：

$$C = FT\log(1+P_X/P_n)$$

- 香农公式，适用于高斯加性信道。但是它也可作为非高斯加性信道的信道容量上限值。
- 一般认为，最高频率就是带宽，取单位时间的最大净信息为信道容量，香农公式也常写为：

$$C = B\log(1+P_X/P_n)$$

- 由公式可见，信道容量的大小与带宽**B**和信噪比P_X/P_n有关，它随着带宽的增大和信噪比的增大而变大。
- 但是因为噪声功率$P_n=n_0B$，式中n_0为噪声的功率谱密度，所以随着带宽的增大，噪声功率会变大，信噪比随之减小，又会使信道容量变小。
- 综合起来分析，如果把香农公式写为：

$C=B\log(1+P_X/n_0B)$，并令$x=P_X/(n_0B)$，

则：

$$C=\frac{P_X}{n_0x}\log(1+x)=\frac{P_X}{n_0}\log(1+x)^{\frac{1}{x}}$$

- 当$B\to\infty$时，$x\to 0$，$(1+x)^{1/x}\to e$，于是：
$$\lim_{B\to\infty} C == \frac{P_X}{n_0}\log e = 1.4427\frac{P_X}{n_0}$$

- 随着带宽的增大，信道容量的增大会越来越慢，最后不再改变，趋于一个理论极限值$1.4427P_X/n_0$

- 定义$B_0=P_X/n_0$叫做临界带宽，由$n_0B_0=P_X$知，临界带宽B_0的含义是噪声与信号功率相等时的带宽。于是，无限带宽所对应的信道容量理论极限值为临界带宽的1.4427 倍。

香农公式的意义

- 香农公式给出了带宽F、时间T、和信噪比P_X/P_n三者之间的制约关系。
- 对于给定的信道（即信道容量不变）的情况下，可以牺牲一些通信效率，用扩展带宽（如CDMA）或延长时间(如积累法接收弱信号）的办法赢得信噪比（通信的质量）的改善；
- 也可以牺牲一些信噪比来换取更高的信道传输率。（如IP电话以及可视电话）

[例3]某图片含2.25×10^6个像素，采用12级量化电平传输。假定各电平等概出现，信道中信噪比为30dB，若要求3分钟完成传输，需要多大的带宽？

解：

传信率：$R=\dfrac{2.25\times10^6\times\log12}{3\times60}=4.48\times10^4\ bit/s\le C$

信噪比： $10\log(P_x/P_n)$=30dB；即P_x/P_n =10^3

根据香农公式：

$$B=\frac{C}{\log_2(1+P_x/P_n)}=\frac{4.48\times10^4}{\log_2 1001}=4.49kHz$$

[例4]某通信系统采用调制指数 $\beta=5$的调频方式发送时，接收端信噪比为20dB；如果信道不变，采用单边带调制，理论上接收端信噪比应为多少分贝才能使通信质量保持不变？

解：用脚标1表示调频，脚标2表示调幅。应有：

$$B_1\log_2(1+S_1/N_1)= B_2\log_2(1+S_2/N_2)$$

换底：$B_1\lg(1+S_1/N_1)= B_2\lg (1+S_2/N_2)$

在信噪比比1大很多时：$B_1\lg(S_1/N_1)= B_2\lg (S_2/N_2)$

现在 $B_1=2(\beta+1)B_m = 2(\beta+1) B_2 =12 B_2$

所以 $\lg (S_2/N_2)=12 \lg(S_1/N_1)$

即：$(S_2/N_2)_{dB}= 12 (S_1/N_1)_{dB} =240dB$

- 香农公式在C=常数条件下给出的是出了带宽B与信噪比P_X/P_n的等效**搭配关系**：较大的带宽搭配较小的信噪比与较大的带宽搭配较小的信噪比均能得到同样的信道容量，达到相同的通信效果。
- 决不要以为这个公式给出的是带宽B与信噪比P_X/P_n之间的**因果关系**：误认为当带宽较大时信噪就会比较小，尤其是不能得出“较大的带宽得到了较小的**输出**信噪比”这样的错误结论。

小结

- 连续信源信息熵的定义：

$$h(x) = -\int_{-\infty}^{\infty} p(x)\log p(x)dx$$

- 波形信道有关的信息熵：

联合熵，噪声熵，损失熵，平均互信息

- 波形信道的信道容量
- 香农公式极其意义

课后复习题

- 思考题：

1. 如何理解连续信源信息熵无穷？

2. 举例说明香农公式对通信的指导意义。

- 作业题：

教材第32页习题一第17、20题；