**自信息：I(xi)=**

**互信息：I(xi;yj)=**

**信息熵：H(X)=**

联合熵：H(XY)=

**条件熵：H(X|Y)=**

**平均互信息：I(X;Y)=**

信息：对事物运动状态或存在方式的不确定性的描述。

**信息流通系统：信源->编码器->信道->译码器->信宿**

信源：产生消息的物理实体

编码器：把消息转变为信号

信道：传播信号的物理实体(噪声源产生干扰)

译码器：把信号转变为消息

信宿：接受信号的物理实体

数据处理定理：若X,Y,Z 构成马尔科夫链，则I(X;Z)<=I(X;Y),I(X;Z)<=I(Y;Z)，即信息量不会越传越多。

信源分类：离散信源，连续信源，波形信源

离散信源：离散时间产生离散信息

连续信源：离散时间产生连续信息

波形信源：连续时间产生连续信息

概率分布是否随时间推移改变：平稳、非平稳

随机变量之间是否统计独立：有记忆，无记忆

离散单符号信源：输出离散数值的单个符号的信源

离散多符号信源：输出符号序列

平均符号熵HN(X)：随机变量序列中前N个随机变量的联合熵求平均

**熵率/极限熵H∞：N趋向无穷时的平均符号熵**

**m阶马尔科夫信源H∞=Hm+1=**

**信源剩余度：γ=1-H∞/log q** (q:信源符号数)

信道分类：离散/连续/波形；有/无记忆；有/无噪声；两/多端口；恒参(平稳)/随参(非平稳)

输入分布P(X)：

**信道矩阵P：**

**信道容量C：存在一种可以使信息传输率R=I(X;Y)最大的输入分布P(X)，在这种分布下的信息传输率即信道容量**

通俗理解：信道能最高效传输信息的传输率

**求C：**

1. **列方程组 Σp(yj|xi)log p(yj|xi)=Σp(yj|xi)βj 解出βj**
2. **C=logΣ2^βj**
3. **验证：p(yj)=2^(βj-C) 求出输出分布**
4. **再根据输出分布，列方程组Σp(xi)p(yj|xi)=p(yj)求出输入分布**

**若输入分布均非负，则结果无误**

信道容量定理：达到信道容量后，所有p(xi)>0的输入的信息量I(xi;Y)都相同，均为C

信源编码：降低信源符号之间的相关性，增加信息传输的有效性

编码器：把信源符号S={s1……sq}编成码符号C={w1……wq}

wi由li个码元组成，例如wi=010，li=3，码元为0和1

所有码符号的码长相同：固定长度码。不同：可变长度码

分组码：一个信源符号si对应一个固定的码符号wi。

非奇异码：分组码中的码符号各不相同。

**唯一可译码：任意有限长非奇异码序列只能唯一分割成一个个码子，不存在多种分法。**

**即时码：唯一可译码收到一个码字后能立刻译码，不需等待下一个码。**

码->分组码->非奇异码->唯一可译码->即时码

**判断是否是唯一可译码：**

1. **构造C的尾随后缀集合F1(例如：C={0,01}，0是01的前缀，则1是尾随后缀)**
2. **在Fi和C中寻找新的未出现过的尾随后缀，构造Fi+1。**
3. **如果在F中存在C中的元素，则非唯一可译码。如果不存在未出现过的元素了，则是唯一可译码。**

**判断是否是即时码：C中每个元素都不是另一个元素的前缀。**

定长信源编码定理：l logr> NH(S) (l长码符号携带信息量大于N长信源序列)

编码效率η=H(S)/(l logr/N) H(S):信源的平均信息量(编码前) l logr/N: 平均每个信源符号携带信息量(编码后)

**变长码 即时码、唯一可译码存在的充要条件：Kraft-McMillan不等式 Σr^(-li) i=1,2……q**

**满足Kraft不等式可构造即时码，满足McMillan不等式可构造唯一可译码**

条件是一样的，名字不一样。也就是唯一可译码一定能构造即时码

平均码长L\_=Σp(si)li 码符号/信源符号。即码长在信源概率分布下的平均值

编码后的信息传输率R=H(S)/L\_ 比特/码符号。即平均每个码元的信息量

信源每秒提供的信息量Rt=R/t 比特/秒

紧致码/最佳码：平均码长最小的唯一可译码。

紧致码的平均码长满足H(S)/logr <= L\_< (H(S)/logr)+1

无失真变长信源编码定理(香农第一定理)

对信源进行N阶拓展的紧致码满足H(S)/logr <= LN\_/N < (H(S)/logr)+1/N

当N趋近于无穷时，LN\_/N=H(S)/logr

LN\_表示N阶拓展信源编码的平均码长

编码效率：η = [H(S)/logr]/L\_ <=1

码的剩余度 γ=1-η

**香农编码：**

1. **把q个信源符号按概率递减排序**
2. **Li=[-log(p(si))]向上取整**
3. **计算累加概率F(si)=Σp(sk) k=1,2……i-1，变换成二进制小数，取小数点后li位做码字**

**费诺编码：跟香农编码很像**

**不用重排。**

**码长Li=[-log(p(si))]向上取整+1**

**累加概率为F(si)=[Σp(sk),k=1,2……i-1]+1/2\*p(si) ，**

**二元霍夫曼码：**

1. **递减排列**
2. **合并最末位两项，分别标上0、1，把合并的概率求和，再重新排序 (为保证码长的方差最小，尽量使新出现的项靠前)**
3. **直到只有最后两项，标上0、1**
4. **从后往前读出码字**

r元霍夫曼码：方法一样。需满足q=(r-1)θ+r，q为码数，θ为合并次数。如果不满足，补几个0概率的码，补到满足，最后再删除。

信道编码：根据相关性检测和纠正传输中的差错

译码规则：根据f(yj)=xi 把输出yj翻译回输入xi。r个输入，s个输出，则有r^s种译码规则

平均错误概率PE：Σp(yj)p(e|yj) j=1,2……s

**最大后验概率译码规则：把信道矩阵每一行都乘以输入概率，再在每一列中找到最大值进行译码。平均错误概率为除了找出项之外的和**

**极大似然译码规则：直接在信道矩阵每一列中找到最大值进行译码。平均错误概率为除了找出项之外的每一项，再乘以该行的输入概率，再求和。**

**简单重复编码：把一个符号重复n次。例如0->00000,1->11111**

**译码时0多译成0，1多译成1**

**错误概率为C(n,n)\*p^n+C(n,n-1)\*p^(n-1)\*(1-p)+……+C(n,(n+1)/2)\*p^(n+1)/2\*(1-p)^(n-1)/2**

**信息传输率R=(logM)/n (M:信源符号数 n：重复次数)**

(5,2)线性码：按照x4=x1,x3=x5=x1异或x2构造出00000,01101,10111,11010

译码时译为相差最小的码。

错误概率1-[(1-p)^5+5\*p\*(1-p)^4+2\*p^2\*(1-p)^3]

汉明距离：两个序列相同位置不同码的个数。如xi=0000和yj=0011的距离d(xi,yj)=2

码的最小距离dmin：任意两个码字之间的汉明距离的最小值。

选择编码码字时，应该选择最小距离最大的码。译码时，将码译成距离最小的码。

有噪信道编码定理(香农第二定理) 当信息传输率R<C时，只要码长n足够长，总能找到一种编码可以使错误概率PE任意小。否则无论n取多大都无法找到。

限失真信源编码：保证一定质量的同时，在信宿近似还原信源输出的信息。

**失真函数d(xi,yj)表示输入xi输出yj引起的失真。d(xi,yj)=0表示无失真。**

**失真矩阵D**

汉明失真D=对角线全0，其余全1

平方误差失真D=[] d(xi,yj)=(xi-yj)^2

平均失真D\_即d(xi,yj)在XY的联合概率分布下求平均

N维拓展：D\_(N)=ND\_

**D失真许可的实验信道：满足平均失真D\_<=D的信道**

所有D失真许可的信道的集合：BD

信息率失真函数R(D)=min I(X;Y) D失真许可的实验信道的集合中平均互信息最小的那个。信息传输率最小，平均失真D\_最大，为D

研究信息率失真函数的目的：用最少的码传递在能接受失真范围内的准确度的信息，提高有效性。

**R(D)的定义域：(Dmin,Dmax)**

**Dmin=Σp(xi)minj{d(xi,yj)}，D矩阵每行最小值乘以输入概率分布再求和。**

**通常Dmin=0，此时不存在任何失真，R(Dmin)=H(X)**

**Dmax=minj{Σp(xi)d(xi,yj)}，D矩阵乘以输入概率分布，每列求和再找出最小值。**

**R(Dmax)=0，不能传输任何信息。**

**当Dmin<D<Dmax时，H(X)>R(D)>0**

限失真信源编码定理(香农第三定理)：若要求D\_<D，只要N足够大，一定存在满足信息传输率R>R(D)的编码，不存在R<R(D)的编码能使D\_<D。

限失真信源编码的目的：希望信息传输率R尽可能小，最好趋近于R(D)。