Writing on dirty paper

<u>考试通知</u>

时间时间：：7月15日，下午14:30-16:30

地点：D104

联系方式：issjqni@mail.sysu.edu.cn

- 关键技术之一
- 有关几个主要互信息的推导

- 例题4.4.1：二维高斯随机变量集合XY，其中X,Y的均值和方差分别为m_x，m_y和σ_x，σ_y，且相关系数为ρ，求

（1）X,Y的联合分布密度$P_{XY}(xy)$；

（2）$h(XY)$; $h(X)$; $h(Y)$;

（3）$h(Y/X)$; $h(X/Y)$; $I(X;Y)$。

解

（1）设XY的协方差矩阵Σ，则

$$\Sigma=\begin{pmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2\end{pmatrix},\quad \Sigma^{-1}=\frac{1}{\sigma_x^2\sigma_y^2(1-\rho^2)}\begin{pmatrix}\sigma_y^2 & -\rho\sigma_x\sigma_y \\ -\rho\sigma_x\sigma_y & \sigma_x^2\end{pmatrix}$$

利用（4.2.5)式，得

$$P_{XY}(xy)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\{-\frac{1}{2(1-\rho^2)}[\frac{(x-m_x)^2}{\sigma_x^2}-\frac{2\rho(x-m_x)(y-m_y)}{\sigma_x\sigma_y}+\frac{(y-m_y)^2}{\sigma_y^2}]\}$$

续解：

（2）根据高斯变量差熵的公式（4.2.6）、（4.2.2），得

$$h(XY)=\log[2\pi e\sigma_x\sigma_y\sqrt{1-\rho^2}]$$

$$h(X)=\frac{1}{2}\log[2\pi e\sigma_x^2]$$

$$h(Y)=\frac{1}{2}\log[2\pi e\sigma_y^2]$$

（3）根据公式(4.1.15)和（4. 2.22)，得到

$$h(Y/X)=h(XY)-h(X)=\frac{1}{2}\log[2\pi e\sigma_y^2(1-\rho^2)]$$

$$h(X/Y)=h(XY)-h(Y)=\frac{1}{2}\log[2\pi e\sigma_x^2(1-\rho^2)]$$

$$I(X;Y)=h(X)+h(Y)-h(XY)=-\frac{1}{2}\log(1-\rho^2)$$

- 例4.4.2 已知X，S为零均值、互相独立的高斯随机变量集合，方差分别为P、Q；Z为独立于X和S的零均值高斯噪声，方差为N；设 $Y = X + S + Z$， $U = X + \alpha S$，其中 α 为常数。求：（1）$I(U;S)$；（2） $I(U;Y)$
- 解：由已知条件可得 $X \sim N(0,P)$ $S \sim N(0,Q)$ $Z \sim N(0,N)$

$$Y \sim N(0, P+Q+N) \qquad U \sim N(0, P+\alpha^2 Q)$$

（1）

$$\rho_{US} = \frac{E(US)}{\sigma_S \sigma_U} = \frac{E(XS + \alpha S^2)}{\sigma_S \sigma_U} = \frac{\alpha Q}{\sigma_S \sigma_U}$$

$$I(U;S) = H(U) + H(S) - H(US)$$

$$= \frac{1}{2}\log 2\pi e \sigma_U^2 + \frac{1}{2}\log 2\pi e \sigma_S^2 - \log 2\pi e \sigma_U \sigma_S \sqrt{1-\rho^2}$$

$$= \frac{1}{2}\log \frac{\sigma_U^2 \sigma_S^2}{\sigma_S^2 \sigma_Y^2 - (\alpha Q)^2} = \frac{1}{2}\log \frac{P + \alpha^2 Q}{P}$$

- 续解：

（2）$$\rho_{UY}=\frac{E(YU)}{\sigma_Y\sigma_U}=\frac{E(X^2+\alpha XS+XS+\alpha S^2+XZ+\alpha ZS)}{\sigma_Y\sigma_U}=\frac{P+\alpha Q}{\sigma_Y\sigma_U}$$

$$I(U;Y)=H(U)+H(Y)-H(UY)$$

$$=\frac{1}{2}\log 2\pi e\sigma_U^2+\frac{1}{2}\log 2\pi e\sigma_Y^2+\log 2\pi e\sigma_U\sigma_Y\sqrt{1-\rho^2}$$

$$=\frac{1}{2}\log\frac{\sigma_U^2\sigma_Y^2}{\sigma_U^2\sigma_Y^2-(P+\alpha Q)^2}=\frac{1}{2}\log\frac{(P+Q+N)(P+\alpha^2Q)}{PQ(1-\alpha^2)+N(P+\alpha^2Q)}$$

- 关键技术之二
- 有关典型序列和渐近均分性(AEP)

信源序列分组定理

定理

总可以找到N_0

离散无记忆信源

任意给定$\varepsilon > 0, \quad \delta > 0$

使得

长度为$N \geq N_0$的信源序列都可以分成两组

①

序列 $\vec{x}$ 出现的概率 $p(\vec{x})$ 满足：

$$\left| \frac{1}{N} \log p(\vec{x}) + H(X) \right| < \delta$$

(5.2.3)

②

所有符号序列出现概率之和小于ε

总结

- 对离散无记忆信源，给定 $\varepsilon,\delta>0$，令 $N_0=\dfrac{\sigma^2}{\varepsilon\delta^2}$ 取 $N\ge N_0$；那么对长度为N的信源序列，满足下式的为**典型序列**，否则为**非典型序列**。

$$\{\vec{x}:\left|\frac{N_i}{N}-p_i\right|<\varsigma,\ i=1,\cdots,q\}$$

- 定理说明，当N足够大时，典型序列 $\vec{x}$ 的 $\dfrac{-\log p(\vec{x})}{N}$ 的值接近信源的熵
- 对于有记忆的马氏源，定理也成立

渐进均分特性

典型序列的概率估计

设$\vec{x} \in G_1 \quad \Rightarrow -\delta < \dfrac{\log p(\vec{x})}{N} + H(X) < \delta$

$$\Rightarrow -N[H(X)+\delta] < \log p(\vec{x}) < -N[H(X)-\delta]$$

设取2为底 $\Rightarrow 2^{-N[H(X)+\delta]} < p(\vec{x}) < 2^{-N[H(X)-\delta]}$

简记为: $\boxed{p(\vec{x}) = 2^{-N[H(X)\pm\delta]}}$

- 当 δ 足够小时，每个典型序列的概率 $p(\vec{x})$ 接近 $2^{-NH(X)}$ 其偏差不大于 $2^{-N\delta}$ ；
- 此时序列的长度需要很大

典型序列的个数估计

设 N_G 为 G_1 中序列的个数

先估计上界：

利用概率估计的下界 $\Rightarrow N_G \cdot 2^{-N[H(X)+\delta]} < N_G \cdot \min_{\vec{x}} p(\vec{x}) \le 1$

$$N_G < 2^{N(H(X)+\delta)}$$

再估计下界：

利用概率估计的上界 $\Rightarrow 1-\varepsilon \le N_G \cdot \max_{\vec{x}} p(\vec{x}) < N_G \cdot 2^{-N[H(X)-\delta]}$

$$N_G > (1-\varepsilon)2^{N[H(X)-\delta]}$$

$$\boxed{(1-\varepsilon)2^{N[H(X)-\delta]} < N_G < 2^{N[H(X)+\delta]}}$$

✓ 渐近均分特性

当ε, δ 取值很小时（N要求很大），对于典型序列

$$N_G \approx 2^{NH(X)}, \qquad p(\vec{x}) \approx 2^{-NH(X)}$$

当长度N足够大时：

❖典型序列接近等概率 $2^{-NH(X)}$，数目近似于 $2^{NH(X)}$

❖非典型序列出现的概率接近为零

❖ $-\frac{1}{N}\log p(\vec{x}) \to H(X)$ （以概率收敛）

联合典型序列

设离散无记忆平稳信道的转移概率为 p_{ij}，输入与输出序列分别为 $\boldsymbol{x}=(x_1,\ldots,x_n)$ 与 $\boldsymbol{y}=(y_1,\ldots,y_n)$，长度为n；达到信道容量的输入概率为 $P(x_k=a_i)=p_i$，输出概率为: $P(y_k=b_j)=\sum_i p_i p_{ij}$，$1\le k\le n$。

设输入/输出序列对构成序列 $\mathbf{xy}=[x_1y_1,x_2y_2,\cdots;x_ny_n]$。$x_k=a_i$的数目为 n_i，$y_k=b_j$的数目为n_j，

$x_k=a_i,y_k=b_j$ 的数目为 n_{ij}

如果 $n_i = np_i(1\pm\delta)$，对每个 i ，称 $\mathbf{x}$ 为 $\delta-$ 典型序列；如果 $n_j = n\sum_i p_i p_{ij}(1\pm\delta)$，称 $\mathbf{y}$ 为 $\delta-$ 典型序列；如果 $n_{ij} = np_i p_{ij}(1\pm\delta)$ ，对每个 i, j ，则称 $(\mathbf{x}, \mathbf{y})$ 为 $\delta-$ 联合典型序列。

引理：如果（$\mathbf{x}, \mathbf{y}$）为 $\delta-$ 联合典型序列，那么 $\mathbf{x}$ 和 $\mathbf{y}$ 也分别是 $\delta-$ 典型序列。

引理：对于 δ – 联合典型序列 $(\mathbf{x},\mathbf{y})$，有以下关系成立：

$$p(\mathbf{x},\mathbf{y}) = 2^{-nH(XY)(1\pm\delta)}$$

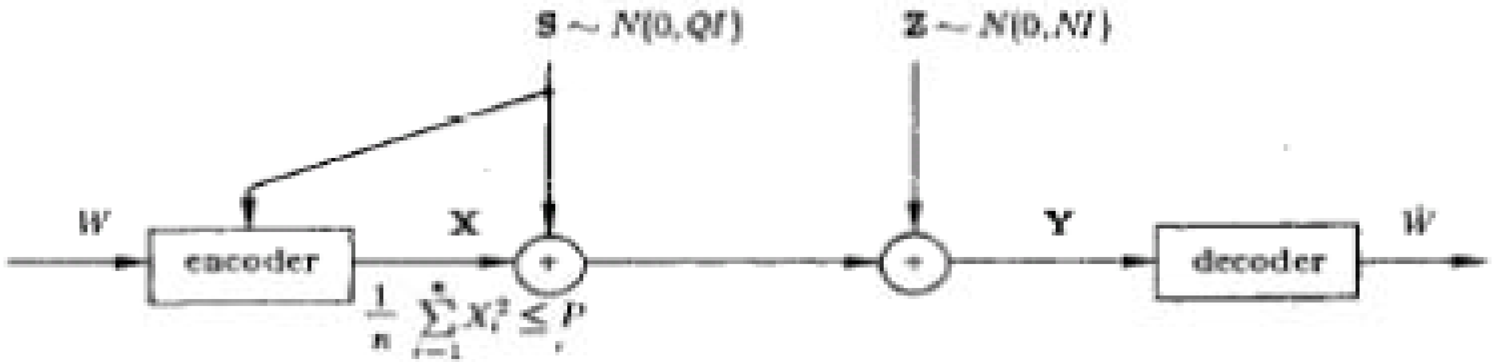
$$p(\mathbf{x}) = 2^{-nH(X)(1\pm\delta)}$$

$$p(\mathbf{y}) = 2^{-nH(Y)(1\pm\delta)}$$

引理： 如果 $\mathbf{y}$ 为 $\delta-$ 典型序列，$\mathbf{x}$ 为与 $\mathbf{y}$ 独立的 $\delta-$ 典型序列，那么与 $\mathbf{y}$ 构成 $\delta-$ 联合典型列的 $\mathbf{x}$ 的个数不大于 $2^{n[H(X/Y)+\delta(H(XY)+H(Y))]}$。

$$\left|F_Y\right| \approx 2^{nH(X|Y)}$$

- 关键技术之三
- 推导



污纸编码模型

M. H. M. Costa, “Writing on dirty paper,” *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439-441, May 1983.

- Gel’fand 和 Pinsker 指出对于离散无记忆信道，如果只有编码器掌握随机状态S的知识，则其信道容量为:

$$C = \max_{p(u,x|s)} \{I(U;Y) - I(U;S)\}.$$

- 含义：

(1) build a channel code over the space of U, with the number of codewords being nearly equal to $2^{nI(U;Y)}$ (where n is the block length of encoding), (2) partition this channel codeword space into cosets of source codes with each coset containing nearly $2^{nI(U;S)}$ codewords, and (3) choose a codeword, U, to represent S from the coset which has an index equal to the message. The size of the index set should be nearly equal to 2^{nC}, where C is given in (3). The signal, X, which is transmitted over the channel is a function $f(U,S)$.

消息个数

说明：

（1）考虑干扰S的编码U，码字个数 $N_u = 2^{nI(U;Y)}$，其中n为码字长度（使用编码U的码率=I(U; Y)=logNu/n ）

（2）将以上码字分为若干个陪集（coset），每个陪集的大小为 $N_{us} = 2^{nI(U;S)}$，这样可以保证每个陪集中有一个码字u可以和状态s联合典型，相当于可以找到一个码字避开s的干扰

（3）按这种方式，实际可以传输的码字个数应该为 $M = 2^{nI(U;Y)} / 2^{nI(U;S)} = 2^{n\{I(U;Y)-I(U;S)\}}$

（4）信道容量为：$I(U;Y) - I(U;S) = \log_2 M / n$

（5）信源的每个信息w可以有 $2^{nI(U;S)}$ 个编码，称为编码陪集

说明（续）

（6）对于状态S，发送码字w的DPC编码过程为：
在索引为w的编码陪集中寻找一个码字u和S联合典型，传输码字u（可以避开S干扰）；接收端收到u，确定其所属陪集索引w，即所传输的符号

- Costa论文中给出的编解码过程如下

First generate $\exp\langle n(I(U;Y)-\epsilon)\rangle$ i.i.d. sequences U, according to the uniform distribution over the set of typical U.[1] Next, distribute these sequences uniformly over e^{nR} bins. For each sequence $\boldsymbol{u}$ let $i(\boldsymbol{u})$ be the index of the bin containing $\boldsymbol{u}$. For encoding, given the state vector $\boldsymbol{S}$ and the message W, look in bin W for a sequence $\boldsymbol{U}$ such that $(\boldsymbol{U}, \boldsymbol{S})$ is jointly typical. Declare an error if no such U can be found. If the number of sequences in bin W is larger than $\exp\langle n(I(U;S)+\delta)\rangle$, the probability of finding no such U decreases to zero exponentially as n increases. Next, choose $\boldsymbol{X}$ such that $(\boldsymbol{X}, \boldsymbol{U}, \boldsymbol{S})$ is jointly typical and send it through the channel. At the decoder look for the unique sequence U such that (U, Y) is jointly typical. Declare an error if more than one or no such sequence exist. Then set the estimate $\hat{W}$ equal to the index of the bin containing the obtained sequence U. If $R < I(U;Y) - I(U;S) - \epsilon - \delta$, the probability of error averaged over all codes decreases exponentially to zero as $n \to \infty$. This shows the existence of a code that achieves rate R with arbitrarily small probability of error.

The problem is reduced to that of finding an appropriate auxiliary variable U. We consider $U = X + \alpha S$, where X and S are independent random variables distributed according to $N(0, P)$ and $N(0, Q)$, respectively, and α is a parameter to be determined. Note that there could be a loss of generality in restricting attention to such U, but as we shall see, the derived answer is clearly optimal. Recalling that $Y = X + S + Z$ with Z distributed according to $N(0, N)$, the relevant mutual informations can be calculated to yield

Let

$$R(\alpha) = I(U;Y) - I(U;S). \tag{5}$$

Then

$$R(\alpha) = \frac{1}{2}\ln\left(\frac{P(P+Q+N)}{PQ(1-\alpha)^2 + N(P+\alpha^2 Q)}\right). \tag{6}$$

Graphs of $I(U;Y)$, $I(U;S)$, and $R(\alpha)$ as functions of α for $P = Q = N = 1$ are presented in Fig. 2.

Maximizing $R(\alpha)$ over α, we get

$$\max_{\alpha} R(\alpha) = R(\alpha^*) = \frac{1}{2}\ln\left(1 + \frac{P}{N}\right) = C^* \tag{7}$$

obtained for $\alpha^* = P/(P+N)$.

高斯信道容量，P-输入信号X功率，N-噪声Z功率，相当于信道只有噪声Z干扰，S可以通过编码有效避开。

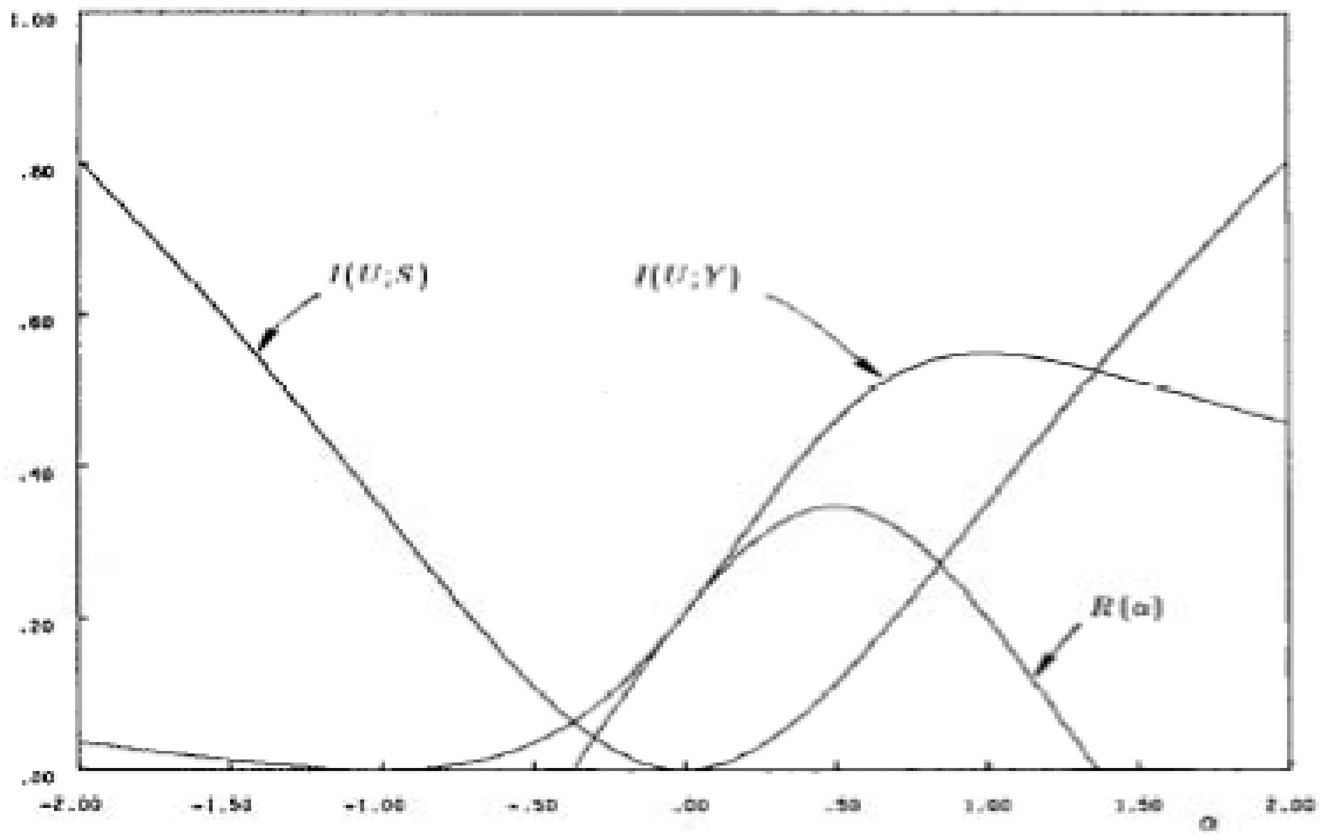


Fig. 2. Graphs of $I(U;Y)$, $I(U;S)$, and $R(\alpha)$ for $P = Q = N = 1$.

相关问题：

Example3: Faulty Registers: Consider a 3-bit memory device, which has one "stuck-at" in any position with a uniform probability of the "stuck-at" being either a one or a zero. When both the encoder and the decoder know the value and the position of the "stuck-at", the encoder can write 2 bits reliably. Now consider the case when only the encoder has access to the "stuck-at". Let $\mathcal{A}$ denote the the

解：

如果只有编码端知道"损坏"比特的位置和取值，是否还是可以通过污纸编码可靠传递2比特信息

步骤1：构造编码U和陪集Coset

Coset-0=(000,111)　　Coset-1=(001,110)

Coset-2=(010,101)　　Coset-3=(011,100)

- 注意：采用以上编码，每个陪集都包含和所有可能错误兼容的码字，例如对于Coset-0，如果bit_1恒错为1，则有码字111；如果bit_2恒错为0，则有码字000。

- 步骤2：根据消息（0-3）编码

  如果要传输消息0，从Coset-0选择码字，相当于（U，S）联合典型

  如果bit_1恒错为1，确定码字111

  如果bit_1恒错为0，确定码字000

- 步骤3：解码

  解码端根据接收码字111或000的陪集索引-0，判定传送的消息为0

复习 CH2 离散信息的度量

1. 自信息、联合自信息、条件自信息、互信息

2. 自信息的平均值为熵 $H(X)=\underset{p(x)}{E}[-\log p(x)]$

3. 条件自信息的平均值为条件熵 $H(X/Y)=\underset{p(xy)}{E}[-\log p(x/y)]$

4. 联合自信息的平均值为联合熵 $H(XY)=\underset{p(xy)}{E}[-\log p(xy)]$

5. 互信息的平均值为平均互信息 $I(X;Y)=\underset{p(xy)}{E}[\log\frac{p(y/x)}{q(y)}]$

6. 条件互信息的平均值为平均条件互信息

$$I(X;Y/Z)=\underset{p(xyz)}{E}[\log\frac{p(y/xz)}{q(y/z)}]$$

复习 CH2

7. 熵的可加性 $H(X_1X_2\cdots X_n)=H(X_1)+H(X_2/X_1)+\cdots+H(X_n/X_1\cdots X_{n-1})$

8. 平均互信息与熵的关系

$$\begin{aligned}I(X;Y)&=H(X)-H(X/Y)\\&=H(Y)-H(Y/X)\\&=H(X)+H(Y)-H(XY)\end{aligned}$$

9. 离散熵与平均互信息都具有非负性

10. 离散最大熵定理 $H(X)=\log n$(n为信源符号数)

11. 平均互信息的凸函数性质

对输入

对信道

平均互信息为输入概率的上凸函数，为条件概率的下凸函数

复习 CH2

12. 凸函数（包含上凸和下凸）的定义和判别

对于 α（$0 \leqslant \alpha \leqslant 1$）及任意两矢量 x_1, x_2，有 **上凸函数（cap）**

$$f[\alpha x_1 + (1-\alpha) x_2] \geqslant \alpha f(x_1) + (1-\alpha) f(x_2)$$

13. Jenson不等式和应用

$$f\left[\sum_{k=1}^{q} \lambda_k x_k\right] \geq \sum_{k=1}^{q} \lambda_k f(x_k)$$

复习 CH3 离散信源

1. 离散无记忆信源：单符号/多维；熵
2. 离散平稳信源：统计特性不随时间改变
3. 离散平稳有记忆信源，理解定理3.2 **离散信源的输出为随机过程，用马氏链**
4. 马氏链定义，齐次马氏链，状态转移矩阵、状态图和网格图
5. Kolmogorov-Chapman 方程，计算多步转移概率
6. 马氏链平稳分布和计算： $\boldsymbol{\pi}^T = \boldsymbol{\pi}^T \mathbf{P}$
7. 马氏源的定义，描述-状态转移概率，m阶马氏链转化为马氏源 **通过合理的定义状态，将m阶马氏链—》一阶马氏信源**
8. 马氏链N次扩展源及熵计算、 N次扩展源的平均符号熵
9. 独立等概信源的熵、独立信源的熵、 m阶马氏源的熵、极限符号熵

复习 CH4 连续信息与连续信源

- 微分熵、条件熵、联合熵
- 连续熵与离散熵的类似性：计算表达式、熵不增加性、可加性；不同点：
- 连续随机变量的信息散度和性质： $D(p // q) = \int p(x) \log \frac{p(x)}{q(x)} dx$
- 一维高斯信源的熵： $\frac{1}{2}\log(2\pi e \sigma^2)$
- 多维独立/相关高斯信源的熵
- 限峰值最大熵定理：均匀分布信源
- 限功率最大熵定理：高斯分布信源
- 连续随机变量的互信息

复习 CH5 无失真信源编码

- 分组码和非分组码；奇异/非奇异码；唯一可译码；即时/非即时码；定长码/变长码；异前置码
- 典型序列、典型序列概率、典型序列个数、渐进均分性AEP

  典型序列概率 $p(\vec{x}) = 2^{-N[H(X)\pm\delta]}$， 个数 $N_G \approx 2^{NH(X)}$，

  $$\left|\frac{1}{N}\log p(\vec{x}) + H(X)\right| < \delta$$

- 香农信源编码定理：码率>=信源熵，存在无失真编码
- 码率（编码中每个信源符号的最大信息量）、编码效率、信息传输速率 - 对于定长/变长码 → **传送一个信源符号平均所需的比特数**
- 变长码存在异前置码的充要条件：Kraft不等式
- 二元Huffman编码，原理和方法

复习 CH6 离散信道及其容量

- 信道容量的定义，包括单符号和矢量离散信道

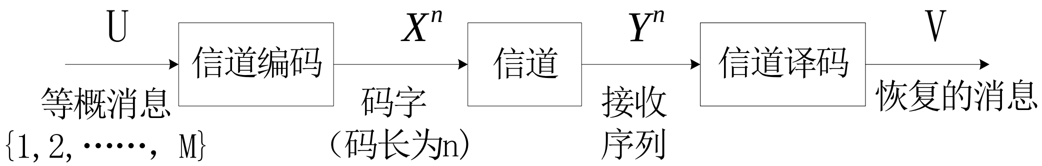
$$C \equiv \max_{p(x)} I(X;Y)$$ **每次使用信道可靠传输的最大信息量**

- 离散信道数学模型：离散无记忆、平稳、离散平稳无记忆可以用一维条件概率描述—转移概率矩阵
- 对称信道、判断方法
- 对称信道容量：当输入等概时达到信道容量

$$C = H(Y) - H(p_1, p_2, ..., p_s)$$

复习 CH7 有噪信道编码

- 信道编码模型：简化通信系统模型（图7-1）



- 信息传输速率（编码码率），(M, n): R= (logM)/n
- 最大后验MAP/最大似然ML译码
- 信道疑义度，费诺不等式

$$H(X\,|\,Y) \le H(p_E) + p_E \log(r-1)$$

- 联合典型序列，个数和概率
- 香农信道编码定理：信道容量C，信息传输R，如果R<C，存在（M, n）码，当n足够长时，错误概率任意小 **（可靠传递）**

复习 CH8 波形信道

Pz（y-x）

- 加性噪声信道：转移概率、条件熵、互信息、信道容量

$$C = \max_{p(x)} h(Y) - h(Z)$$

- 加性高斯噪声信道容量

$$\boxed{\max_{p(x)} I(X;Y) = \frac{1}{2}\log(1+\frac{\sigma_x^2}{\sigma_z^2})}$$

- 一般加性噪声信道的容量界限
- 并联加性高斯噪声信道容量，注水原则及其应用
- 香农信道容量公式及其应用 $C = W\log(1+\frac{P}{N_0 W})$
  - 计算
  - 功率和带宽的互换

复习 CH9 信息率失真函数 （不考）

1、R（D）函数定义：

$$R(D)=\min_{p(y/x)\in P_D} I(X;Y)$$

2、R（D）函数的性质：

（1）定义域：

$$0\le D_{\min}\le D\le D_{\max}$$

$$D_{\min}=\sum_x p(x)\min_y d(x,y)$$

$$D_{\max}=\min_y \sum_x p(x)d(x,y)$$

复习 CH9 信息率失真函数

（2）下凸性：R（D）是D的下凸函数

（3）连续严格递减函数：在（Dmin，Dmax）区间是D的严格递减函数

复习 CH9 信息率失真函数

3、重要的R（D）函数

（1）对称二元信源（汉明失真）

$$R(D)=\begin{cases} H(p)-H(D) & 0\le D\le p \\ 0 & D>p \end{cases}$$

4、限失真信源编码定理

$R>R(D)\Leftrightarrow$ 存在平均失真 $\le D$ 的信源编码

5、限失真信源信道编码定理

$C_{[bps]}>R(D)_{[bps]}\Leftrightarrow$ 存在平均失真 $\le D$ 的信源信道编码