○1

2
熵和互信息量的基本性质

2. 熵和互信息量的基本性质

- 离散随机变量
  - 一维随机变量X，取值离散

    即 $X \in \{x_1, x_2, \ldots, x_r\}$
  - N维随机矢量**X**，时间离散、取值离散

    即 $\mathbf{X}=(X_1, X_2, \ldots, X_N)$

    $X_i \in \{x_1, x_2, \ldots, x_r\}$

    $i=1,2,\ldots,N$

目录

定理2.1:设离散随机变量X在定义域

$R=\{x_1, x_2, \ldots, x_r\}$内取值,

则 $0 \le H(X) \le \log r$

$H(X) = 0$,

当且仅当对于某个i， $p_i=1$

$H(X) = \log r$,

当且仅当对于所有i， $p_i=1/r$

目录

2.2 互信息量的性质

定理2.2: 对于任意一对随机变量X和Y,

$$I\,(X;Y) \geq 0$$

当且仅当X, Y相互独立时,

$$I\,(X;\,Y) = 0$$

目录

- 熵的性质
- 互信息量的性质
- 条件熵及Fano不等式
- 复合观测的互信息量
- 数据处理定理
- 互信息量与信道输入概率分布的关系
- 互信息量与信道转移概率分布的关系
- 互信息量与信道输入符号相关性的关系
- 互信息量与信道记忆特性的关系

定理2.3: 设X, Y, Z是离散随机变量,

对于每个z，定义

$$A(z) = \sum_{x,y} p(y)\, p(z/x, y)$$

则 H(X/Y) ≤ H(Z) + E[log A(z)]

证明:

$$H(X/Y)=E\left[\log\frac{1}{p(x/y)}\right]=\sum_{x,y,z}p(x,y,z)\log\frac{1}{p(x/y)}$$

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

$$\le\sum_{z}p(z)\log\left[\frac{1}{p(z)}\bullet\sum_{x,y}\frac{p(x,y,z)}{p(x/y)}\right]$$

$$A(z)=\sum_{x,y}p(y)p(z/x,y)$$

$$=\sum_{z}p(z)\log\frac{1}{p(z)}+\sum_{z}p(z)\log\sum_{x,y}\frac{p(x,y,z)}{p(x/y)}$$

(其中$p(x,y,z)/p(x/y)=p(x,y,z)p(y)/p(x,y)=p(y)p(z/x,y)$)

$$=H(z)+E[\log A(z)]$$

证毕!

例2.1：设X和Y是两个离散随机变量，

定义其和为Z=X+Y（普通加法），

若X和Y相互独立，

求证：$H(X) \leq H(Z)$

证明1：设X取值于$(x_1, x_2, \ldots, x_m)$.

Y取值于$(y_1, y_2, \ldots, y_n)$.

对于Z的任一取值z_i

设$z_i = x_{i1} + y_{i1} = x_{i2} + y_{i2} = \ldots = x_{ik} + y_{ik}$

定义$A(z) = \sum_{x,y} p(y) p(z/x,y)$

则$A(z = z_i) = \sum_{y} p(y) \sum_{x} p(z_i / x, y) = \sum_{l=1}^{k} p(y_{il}) \leq 1$

因此$A(z) \leq 1$

证明1（续）：利用定理2.3，有

$$H(X/Y) \le H(Z)+E[\log A(z)] \le H(Z)$$

又因为X和Y相互独立，

即$I(X;Y)=0$

则 $H(X/Y)=H(X)-I(X;Y)=H(X)$

因此 $H(X) \le H(Z)$

证明2：因为 $Z=X+Y$

所以对于任意y_i和k，

有 $p(z_k/y_i)=p(x_k)$

因此 $H(Z/Y=y_i)=H(X)$

$$H(Z/Y)=\sum_{yi}p(y_i)H(Z/Y=y_i)=H(X)$$

而 $H(Z)-H(Z/Y)=I(Z;Y)\geq 0$

因此 $H(Z)\geq H(Z/Y)=H(X)$

推论: (Fano Inequality)

设X,Y是随机变量,

取值于符号集(x_1, x_2, … , x_r).

定义$P_e = P\{X \neq Y\}$.

则 $H(X/Y) \leq H(P_e) + P_e \log(r-1)$

证明：定义z = 0, 若x = y; z = 1, 若$x \neq y$.

则 $P_r(z = 0) = 1- P_e$, $P_r(z = 1) = P_e$

$H(Z) = H(P_e, 1-P_e) = H(P_e)$

已知$A(z)=\sum_{x,y} p(y)p(z/x,y)$

其中$A(z=0)=\sum_{x,y} p(y)p(x=y/x,y)$

$$=\sum_{y} p(y)\left[\sum_{x} p(x=y/x,y)\right]=\sum_{y} p(y)=1$$

$$A(z=1)=\sum_{x,y} p(y)p(x\neq y/x,y)$$

$$=\sum_{y} p(y)\left[\sum_{x} p(x\neq y/x,y)\right]=\sum_{y} p(y)(r-1)=r-1$$

证明（续）：

根据定理2. 3

$$H(X/Y) \le H(Z) + E[\log A(z)]$$

有 $H(X/Y) \le H(P_e) + P_e \log(r-1) + (1-P_e) \log 1$

$$= H(P_e) + P_e \log(r-1)$$

分析：

(1) H(X/Y)是Y已知后，仍需通过观测X才能获得的信息量；

(2) 观测X的一种方法是首先确定X?=Y:

如果X=Y，则达到目的；

如果X≠Y，则X有r-1种可能的取值.

(3) 检测X?=Y等同于观测Z，

而$H(Z)=H(P_e)$

(4) 如果X≠Y（概率为P_e ）

保留的不确定量最多为log(r-1)

因此 $H(X/Y) \le H(P_e) + P_e \log(r-1)$

目录

定理2.4: $I(X, Y; Z) \geq I(Y; Z)$.

当且仅当$p(z/x, y) = p(z/y)$时,

等式成立.

$I(X, Y; Z) \geq I(X; Z)$.

当且仅当$p(z/x, y) = p(z/x)$时,

等式成立.

2.4 复合观测的互信息量

证明：$I(Y;Z) - I(X,Y;Z) = E\left[\log\frac{p(z/y)}{p(z)} - \log\frac{p(z/x,y)}{p(z)}\right]$

$$= E\left[\log\frac{p(z/y)}{p(z/x,y)}\right]$$

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

$$\le \log\sum_{x,y,z} p(x,y,z)\frac{p(z/y)}{p(x,y,z)/p(x,y)}$$

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

$$= \log 1 = 0$$

证毕！

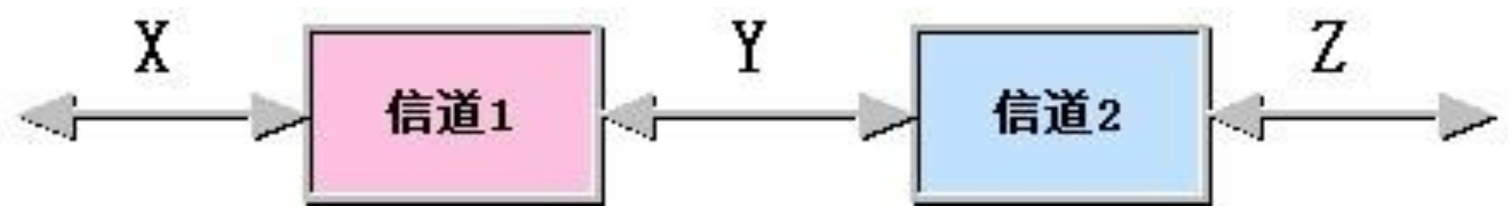
目录

○2
1

- 熵的性质
- 互信息量的性质
- 条件熵及Fano不等式
- 复合观测的互信息量
- <u>数据处理定理</u>
- 互信息量与信道输入概率分布的关系
- 互信息量与信道转移概率分布的关系
- 互信息量与信道输入符号相关性的关系
- 互信息量与信道记忆特性的关系

2.5 数据处理定理

Markov信道链:



(X, Y, Z)是Markov链 $\Rightarrow$ (Z, Y, X)也是Markov链

即 $p(z/x,y) = p(z/y) \Rightarrow p(x/y) = p(x/y,z)$

2.5 数据处理定理

定理2.5: 设 (X, Y, Z) 是一个Markov链, 则有

$$I(X;Z) \le \begin{cases} I(X;Y) \\ I(Y;Z) \end{cases}$$

2.5 数据处理定理

证明：由定理2.4, 有 $I(X;Z) \le I(X,Y;Z)$

因为(X, Y, Z)是Markov链，有

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

所以 $I(X;Z) \le I(Y;Z)$

又因为(X, Y, Z)是Markov链

$\Rightarrow$ (Z, Y, X)也是Markov链

有 $I(Z,Y;X)=I(Y;X)=I(X;Y)$

而 $I(Z,Y;X) \ge I(Z;X)=I(X;Z)$

所以 $I(X;Z) \le I(X;Y)$

2.5 数据处理定理

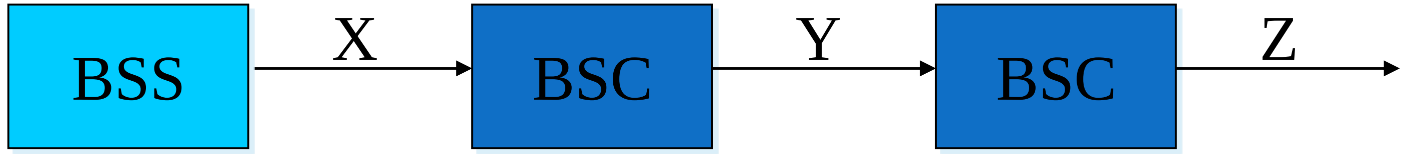
物理意义：数据处理**不能增加**信息！

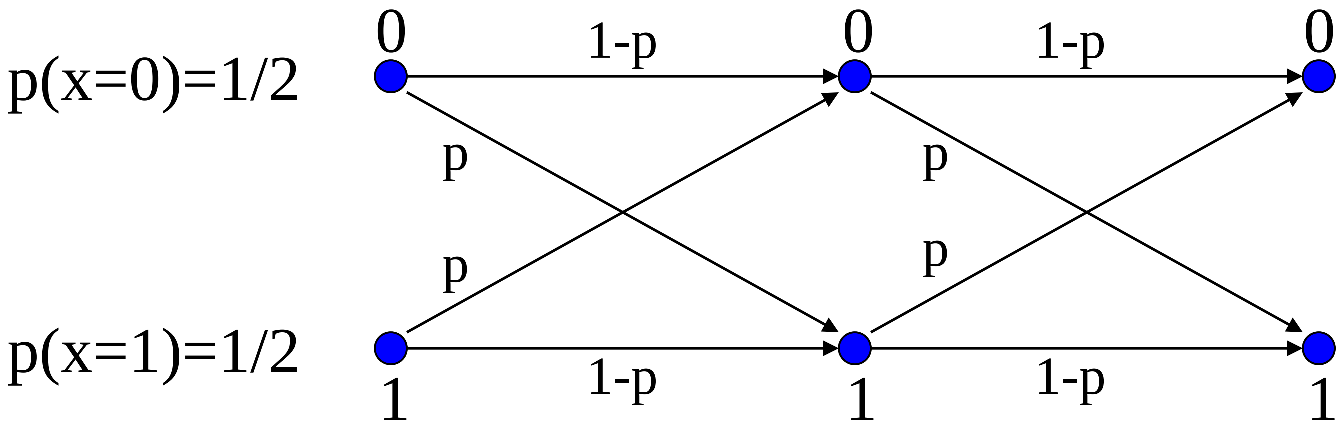
例2.2: X1, X2, X3是独立随机变量，

因此(X1, X1+X2, X1+X2+X3)是Markov 链

有I(X1; X1+X2+X3)≤I(X1; X1+X2)

例题2.3：



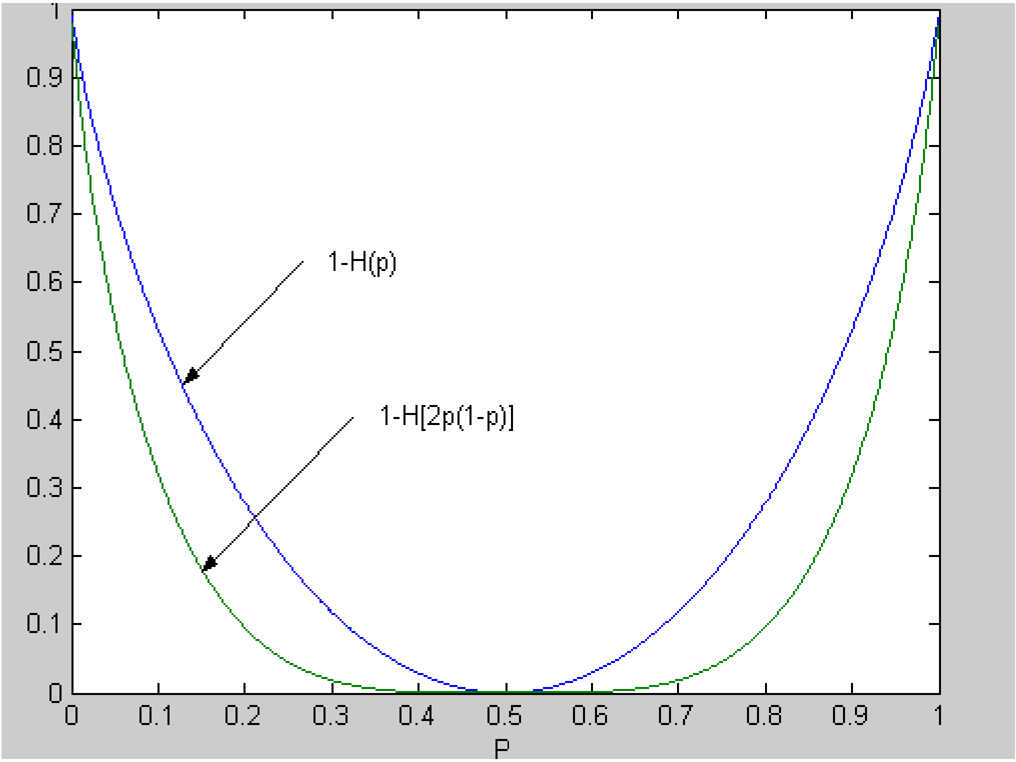


例题2.2（续）：

$$I(X;Y)=1-H_2(p)$$

$$I(X;Z)=1-H_2[2p(1-p)]$$

例题2.2（续）：



TRANSMITTER

The Comms. Research Group deep in thought

CHANNEL

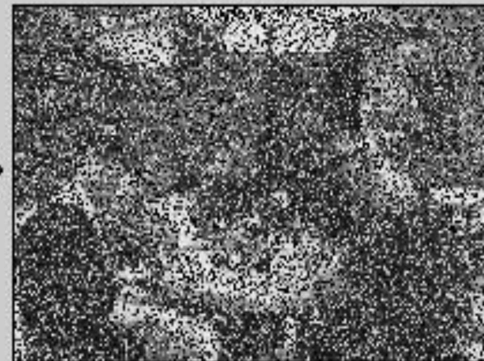


Image Corrupted by Channel Noise (1.0 dB)

RECEIVER

After 1 Turbo Decoding Step

After 2 Iterations

Turbo Decoding of a Bit-Map Image

Turbo Decoding is a powerful error correcting tool, that can be used to protect digital signals transmitted over radio channels. The term 'Turbo' refers to the iterative (repeated) processing of the received signal. Turbo codes tend to perform better than any other code at low signal-to-noise ratios (noisy signals). Here we can see a simulation of a digitised image being transmitted over a noisy radio channel.

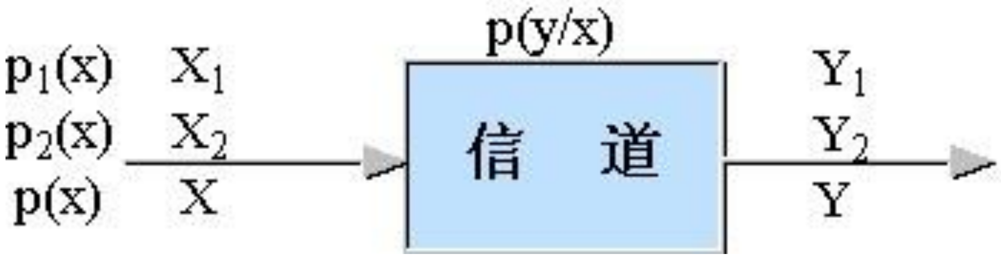
After 10 Iterations

目录

- 熵的性质
- 互信息量的性质
- 条件熵及Fano不等式
- 复合观测的信息量
- 数据处理定理
- <u>互信息量与信道输入概率分布的关系</u>
- 互信息量与信道转移概率分布的关系
- 互信息量与信道输入符号相关性的关系
- 互信息量与信道记忆特性的关系

2.6 互信息量与信道输入概率分布的关系

定理2.6: I(X; Y)是信道输入概率分布p(x)的**上凸函数**.



说明: $p(x) = \alpha\, p_1(x) + \beta\, p_2(x). (\alpha, \beta \geq 0 \quad \alpha + \beta = 1)$

定理成立, 即 $\alpha\, I(X_1; Y_1) + \beta\, I(X_2; Y_2) \leq I(X;Y)$

证明: $\alpha I(X_1;Y_1)+\beta I(X_2;Y_2)-I(X;Y)$

$$=\sum_{x,y}\alpha p_1(x,y)\log\frac{p(y/x)}{p_1(y)}+\sum_{x,y}\beta p_2(x,y)\log\frac{p(y/x)}{p_2(y)}$$

$$-\sum_{x,y}\left[\alpha p_1(x,y)+\beta p_2(x,y)\right]\log\frac{p(y/x)}{p(y)}$$

$$=\alpha\sum_{x,y}p_1(x,y)\log\frac{p(y)}{p_1(y)}+\beta\sum_{x,y}p_2(x,y)\log\frac{p(y)}{p_2(y)}$$

2.6 互信息量与信道输入概率分布的关系

证明：$$\leq \alpha \log \sum_{x,y} p_1(x,y)\frac{p(y)}{p_1(y)} + \beta \log \sum_{x,y} p_2(x,y)\frac{p(y)}{p_2(y)}$$

$$= \alpha \log \sum_{y} \frac{p(y)}{p_1(y)} \sum_{x} p_1(x,y) + \beta \log \sum_{y} \frac{p(y)}{p_2(y)} \sum_{x} p_2(x,y)$$

$$= \alpha \log \sum_{y} \frac{p(y)}{p_1(y)} p_1(y) + \beta \log \sum_{y} \frac{p(y)}{p_2(y)} p_2(y)$$

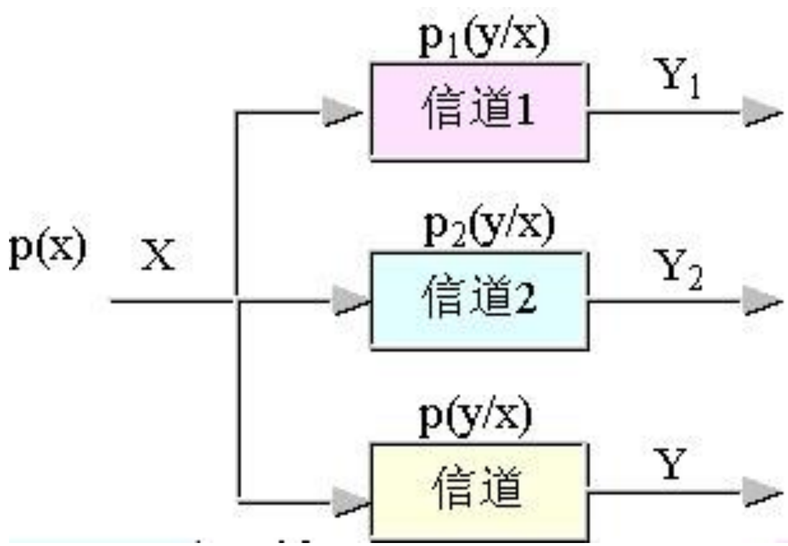
$$= \alpha \log 1 + \beta \log 1 = 0$$

证毕！

目录

- 熵的性质
- 互信息量的性质
- 条件熵及Fano不等式
- 复合观测的互信息量
- 数据处理定理
- 互信息量与信道输入概率分布的关系
- 互信息量与信道转移概率分布的关系
- 互信息量与信道输入符号相关性的关系
- 互信息量与信道记忆特性的关系

定理2.7: I(X; Y)是信道转移概率分布p(y/x)的**下凹函数**.



说明： $p(y/x)=\alpha p_1(y/x)+\beta p_2(y/x)$. $(\alpha,\beta \geq 0 \quad \alpha+\beta=1)$

定理成立, 即 $I(X;Y) \leq \alpha\, I(X;Y_1)+\beta\, I(X;Y_2)$

目录

定理2.8：信道的输入是离散无记忆的

即 $$p(\mathbf{x}) = p(x_1, x_2, \cdots, x_N) = p(x_1)p(x_2)\cdots p(x_N) = \prod_{i=1}^{N} p(x_i)$$

则 $$I(\mathbf{X};\mathbf{Y}) \geq \sum_{i=1}^{N} I(X_i;Y_i)$$

2.8 互信息量与信道输入符号相关性的关系

证明：

$$I(\mathbf{X};\mathbf{Y}) = E\left[\log\frac{p(\mathbf{x}/\mathbf{y})}{p(\mathbf{x})}\right]$$

$$= E\left[\log\frac{p(\mathbf{x}/\mathbf{y})}{p(x_1)p(x_2)\cdots p(x_N)}\right]$$

$$\sum_{i=1}^{N} I(X_i;Y_i) = \sum_{i=1}^{N} E\left[\log\frac{p(x_i/y_i)}{p(x_i)}\right]$$

$$= E\left[\log\frac{p(x_1/y_1)\cdots p(x_N/y_N)}{p(x_1)\cdots p(x_N)}\right]$$

2.8 互信息量与信道输入符号相关性的关系

证明：$\sum_{i=1}^{N} I(X_i;Y_i) - I(\mathbf{X};\mathbf{Y})$

$$= E\left[\log \frac{p(x_1/y_1)\cdots p(x_N/y_N)}{p(\mathbf{x}/\mathbf{y})}\right]$$

$$\le \log E\left[\frac{p(x_1/y_1)\cdots p(x_N/y_N)}{p(\mathbf{x}/\mathbf{y})}\right]$$

$$= \log\left[\sum_{\mathbf{x},\mathbf{y}} p(\mathbf{x}/\mathbf{y})p(\mathbf{y})\frac{p(x_1/y_1)\cdots p(x_N/y_N)}{p(\mathbf{x}/\mathbf{y})}\right]$$

$$= \log\left[\sum_{\mathbf{y}} p(\mathbf{y})\left(\sum_{x_1} p(x_1/y_1)\cdots\sum_{x_N} p(x_N/y_N)\right)\right] = \log 1 = 0$$

说明：

- $(X_1, X_2, \ldots, X_N)$作为一个噪声信道的输入
- $(Y_1, Y_2, \ldots, Y_N)$作为其输出
- 定理2.8说明：如果输入是相互独立的，$\mathbf{Y}$所提供的关于$\mathbf{X}$的信息量，比每个Y_i所提供的关于相应X_i的信息量的总和要多.

2.8 互信息量与信道输入符号相关性的关系

例题：

- 设$X_1, X_2, \ldots, X_n$是独立同分布的离散随机变量
  它们的熵都为H
- 设π是集合{1，2，…，n}的一个置换，并且令$Y_i=X_{\pi(i)}$
- 则 $I(\mathbf{X}; \mathbf{Y}) = nH$
- 但 $\sum I(X_i; Y_i) = kH$
  这里k是π中固定点的数目，即满足π(i)=i的整数i的数目
- 特别是如果π没有固定点
  例如 $\pi(i) \equiv i+1 (\mathrm{mod}\ n)$，则$\sum I(X_i; Y_i)=0$

目录

- 熵的性质
- 互信息量的性质
- 条件熵及Fano不等式
- 复合观测的互信息量
- 数据处理定理
- 互信息量与信道输入概率分布的关系
- 互信息量与信道转移概率分布的关系
- 互信息量与信道输入符号相关性的关系
- 互信息量与信道记忆特性的关系

2.9 互信息量与信道记忆特性的关系

定理2.9：信道是离散无记忆的

即 $p(\mathbf{y}/\mathbf{x})=p(y_1,y_2,\cdots,y_N/x_1,x_2,\cdots,x_N)$

$$=p(y_1/x_1)p(y_2/x_2)\cdots p(y_N/x_N)=\prod_{i=1}^{N}p(y_i/x_i)$$

则 $I(\mathbf{X};\mathbf{Y})\leq\sum_{i=1}^{N}I(X_i;Y_i)$

例题：

- 设X是熵为H的随机变量
- 令$X_1=X_2=\dots=X_n=Y_1=\dots=Y_n=X$
- 满足定理2.9的条件

有 $I(\mathbf{X};\mathbf{Y})=H$,

$\sum I(X_i;Y_i)=nH$

定理2.8和2.9的推论

推论：信道的输入和信道本身都是离散无记忆的，

则 $I(\mathbf{X};\mathbf{Y})=\sum_{i=1}^{N}I(X_i;Y_i)$

习题

2.1 证明定理2.4.

2.2 证明定理2.7.

2.3 证明定理2.9.

2.6 如果X, Y和Z是同一取值空间上的三个离散随机变量, 定义条件互信息量

I (X;Y / Z)为:

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

证明: (a) I (X; Y / Z) = I (Y; X / Z)

(b) I (X; Y, Z) = I (X; Z) + I (X; Y / Z)

(c) I (X; Y / Z) ≥ 0, 当且仅当（X, Z, Y）是Markov链时等式成立.