信息论

何扬槊 3180102687

信息与电子工程学院

浙江大学

December 18, 2021

Outline

通信系统

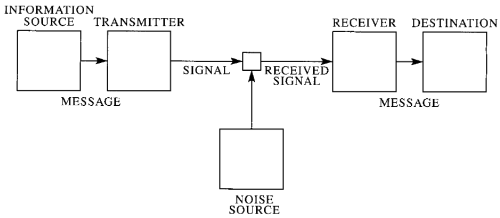

图. 1. 通用的通信系统模型 [1]

- 信源: 产生消息序列
- 发送器: 对消息进行处理得到信号
- 信道: 信号传输的媒介，一般有噪声
- 接收器: 从信号重构消息
- 信宿: 消息接收者

Basic Questions

1. 如何衡量信息?
2. 如何定义信源?
3. 如何定义信道?
4. 如何描述传输过程?
5. 如何应对噪声?

Outline

信源

- 如何用数学语言描述信源?

信源

- 如何用数学语言描述信源?
- 离散信源产生多少信息?

信源

- 如何用数学语言描述信源?
- 离散信源产生多少信息?
- 如何度量信源产生消息的量?

熵

希望有一个度量 $H(p_1, \cdots, p_n)$ 满足 3 个性质

1. H 是 p 的连续函数
2. $p_i = \frac{1}{n}$ 时，若 $n \nearrow$, $H \nearrow$
3. 一个选择被分为多个子选择，则原 H 为各个 H 的加权和

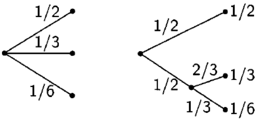

图. 2. 熵的可加性要求

$$H(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) = H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2}H(\frac{1}{3}, \frac{2}{3})$$

熵

满足 3 个性质的函数有如下形式

$$H = -K \sum_{i=1}^{n} p_i \log p_i \tag{1}$$

K 根据单位选取，bit 一般取 $K = 1$

Jump to Appendix

熵的性质

1. $H(X) = 0 \iff p_i = 0$ or 1
2. $H(X) \leq \log n$
3. $H(X)$ is concave in $p(x)$

熵的性质

- 联合熵: $H(X,Y) = -\sum_{x,y} p(x,y)\log p(x,y)$

  $H(X) = -\sum_{x,y} \log \sum_{y} p(x,y)$
- $H(X,Y) \leq H(X) + H(Y)$
- 条件熵: $H(Y|X) = -\sum_{x} p(y|x)\log p(y|x) = -\sum_{x,y} p(x,y)\log p(y|x)$

  $H(Y|X) = H(X,Y) - H(X)$
- $H(Y) \geq H(Y|X)$

熵的性质

- 联合熵: $H(X,Y) = -\sum_{x,y} p(x,y)\log p(x,y)$

  $H(X) = -\sum_{x,y}\log\sum_{y} p(x,y)$
- $H(X,Y) \le H(X) + H(Y)$
- 条件熵: $H(Y|X) = -\sum_{x} p(y|x)\log p(y|x) = -\sum_{x,y} p(x,y)\log p(y|x)$

  $H(Y|X) = H(X,Y) - H(X)$
- $H(Y) \ge H(Y|X)$
- $H(X,Y) = H(X) + H(Y|X)$

  $H(X_1,\cdots,X_N) = \sum_{n=1}^{N} H(X_n|X_1,\cdots,X_{n-1})$

其他度量

- 互信息: $I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$

其他度量

- 互信息: $I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$
- 性质

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
I(X;Y) &= H(Y) - H(Y|X) \\
I(X;Y) &= H(X) + H(Y) - H(X,Y)
\end{aligned}
\tag{2}
$$

其他度量

- 互信息: $I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$
- 性质

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
I(X;Y) &= H(Y) - H(Y|X) \\
I(X;Y) &= H(X) + H(Y) - H(X,Y)
\end{aligned}
\tag{2}
$$

其他度量

- 互信息: $I(X;Y)=\sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$
- 性质

$$\begin{aligned} I(X;Y) &= H(X)-H(X|Y) \\ I(X;Y) &= H(Y)-H(Y|X) \\ I(X;Y) &= H(X)+H(Y)-H(X,Y) \end{aligned} \tag{2}$$

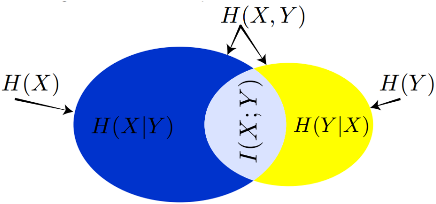

图. 3. 熵的关系

其他度量

- KL 散度: $D(p\|q) = \sum_{x_1,\cdots,x_N} p(x_1,\cdots,x_N)\log\frac{p(x_1,\cdots,x_N)}{q(x_1,\cdots,x_N)}$

其他度量

- KL 散度: $D(p\|q) = \sum_{x_1,\cdots,x_N} p(x_1,\cdots,x_N)\log\frac{p(x_1,\cdots,x_N)}{q(x_1,\cdots,x_N)}$
- 描述两个分布的"距离"

$$D(p(x,y)\|p(x)p(y)) = I(X;Y) \tag{3}$$

信源的熵

极限情况下可以由消息序列的统计量得到信源的熵

$P(B_i)$ 为一个消息序列 B_i 的概率

$$G_N = -\frac{1}{N}\sum_i P(B_i)\log P(B_i) \tag{4}$$

消息序列 B_i 后面的符号为 s_j

$$F_N = -\sum_{i,j} P(B_i, s_j)\log P(s_j|B_i) - NG_N - (N-1)G_N \tag{5}$$

信源的熵

- $N \nearrow$, $F_N \searrow$ (conditioning decrease entropy)
- $G_N = \frac{1}{N}\sum F_N \geq F_N$
- $\lim_{N\to\infty} G_N = \lim_{N\to\infty} F_N = H(\mathcal{S})$

定义 (熵速率)

$$H(\mathcal{S}) = \lim_{N\to\infty} \frac{H(s^N)}{N} \tag{6}$$

信源的熵

考虑 N 长序列，为 1 阶 Markov 过程

$$H(\mathcal{S}) = -\sum_{i,j} \pi_i p(j|i) \log p(j|i) \tag{7}$$

进一步若符号 $\{s_i\}$ 相互独立，信源的熵退化为符号的熵

$$H(\mathcal{S}) = H(S) = -\sum_{i} P(s_i) \log P(s_i) \tag{8}$$

信源的熵

考虑 N 长序列，符号 $\{s_i\}$ 相互独立，出现符号 s_i 的次数的期望为 p_iN

$$\mathsf{E}[P(s^N)] = \mathsf{E}[p] = \prod_{i=1}^{N} p_i^{p_iN}$$

$$\mathsf{E}[-\frac{\log P(s^N)}{N}] = -\frac{1}{N}\sum_i Np_i \log p_i = H(\mathcal{S})$$

信源的熵

考虑 N 长序列，符号 $\{s_i\}$ 相互独立，出现符号 s_i 的次数的期望为 $p_i N$

$$\mathsf{E}[P(s^N)] = \mathsf{E}[p] = \prod_{i=1}^{N} p_i^{p_i N}$$

$$\mathsf{E}[-\frac{\log P(s^N)}{N}] = -\frac{1}{N}\sum_i N p_i \log p_i = H(\mathcal{S})$$

定义 (典型列)

$$A_\epsilon^{(N)} = \{s^N : \mid -\frac{\log p}{N} - H(\mathcal{S}) \mid \leq \epsilon\} \tag{9}$$

Jump to Appendix

信源编码定理

定理 (离散无记忆信源编码定理)

当编码速率 R 与信源的熵 $H(\mathcal{S})$ 满足 10，错误概率 $P_e^{(N)} \longrightarrow 0$

$$R < H(\mathcal{S}) \tag{10}$$

其中 $R = \frac{\log M}{N}$，M 为码字的数量

Outline

信道

两个问题

- 如何用数学语言描述信号在信道传输的过程?

信道

两个问题

- 如何用数学语言描述信号在信道传输的过程?
- 如何应对信号传输过程中引入的噪声?

信道

- 符号集 $\{a, b\}$，发送器以 100 symbol/s 的速度发送消息，每个符号错误概率为 $P_e = 0.01$，有效的传输速度是多少 symbol/s?

信道

- 符号集 $\{a, b\}$，发送器以 100 symbol/s 的速度发送消息，每个符号错误概率为 $P_e = 0.01$，有效的传输速度是多少 symbol/s?
- $R = 90$ symbol/s?

信道

- 符号集 $\{a, b\}$，发送器以 100 symbol/s 的速度发送消息，每个符号错误概率为 $P_e = 0.01$，有效的传输速度是多少 symbol/s?
- $R = 90$ symbol/s?
- 如果 $P_e = 0.5$，$R = 50$ symbol/s?

信道传输速率

- 噪声是接收信号中丢失的部分
- 噪声 (疑义度)$\Longrightarrow H(X|Y)$
- 信道传输速率

$$R = H(X) - H(X|Y) \tag{11}$$

Jump to Fano

信道容量

- 可能实现的最大传输速率是信道的容量 C

$$
\begin{aligned}
C &= \max_{P(X)}\{H(X) - H(X|Y)\} \\
&= \max_{P(X)} I(X;Y)
\end{aligned}
\tag{12}
$$

信道容量

- 可能实现的最大传输速率是信道的容量 C

$$C = \max_{P(X)} \{H(X) - H(X|Y)\}$$

- 更一般的信道容量

$$C = \lim_{N \to \infty} \frac{1}{N} \max_{P(X^N)} I(X_1 \cdots X_N; Y_1 \cdots Y_N) \tag{13}$$

信道编码定理

定理 (离散无记忆信道编码定理)

离散信道容量为 C，信号传输速率 R，满足

- $\forall\, R < C,\ \exists\, (2^{nR}, n)$ code, s.t. $\lambda^{(n)} \to 0$
- $\forall\, (2^{nR}, n)$ code with $\lambda^{(n)} \to 0 \Rightarrow R \leq C$

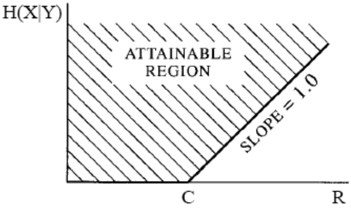

图. 4. 疑义度与给定传输速率的关系

Jump to Appendix

典型列 intuition

- N 长的序列, 传输速率为 R, 即传输 2^{NR} 条序列

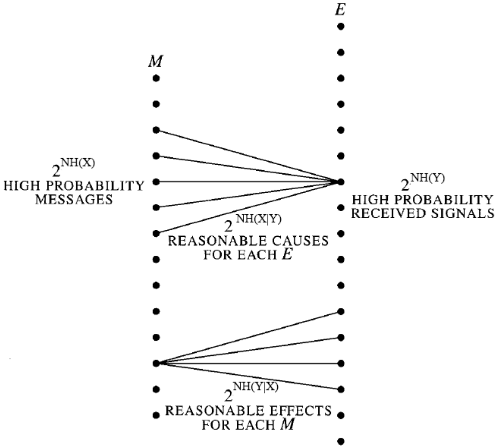

图. 5. 典型列示意

典型列 intuition

- N 长的序列, 传输速率为 R, 即传输 2^{NR} 条序列
- 输入中约 $2^{NH(X)}$ 条典型列, 输出中约 $2^{NH(Y)}$ 条典型列

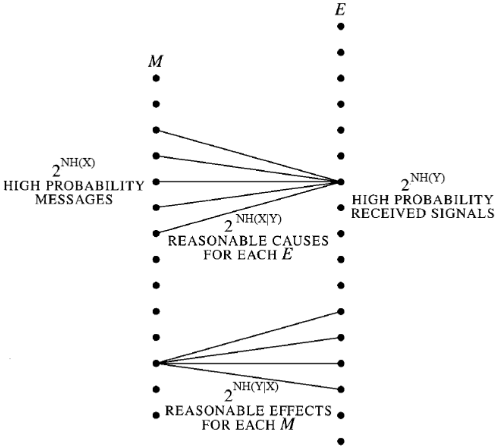

图. 5. 典型列示意

典型列 intuition

- N 长的序列, 传输速率为 R, 即传输 2^{NR} 条序列
- 输入中约 $2^{NH(X)}$ 条典型列, 输出中约 $2^{NH(Y)}$ 条典型列
- 一条输入序列可能对应 $2^{NH(Y|X)}$ 条输出序列 (疑义度)

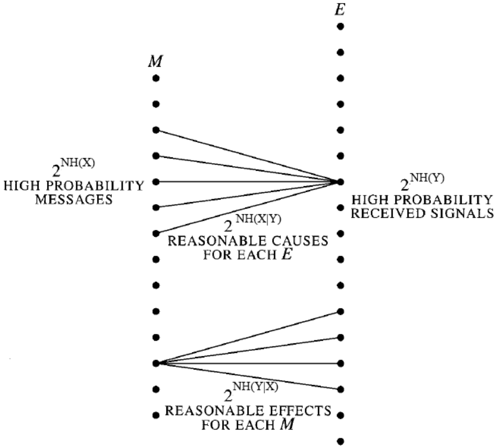

图. 5. 典型列示意

典型列 intuition

- N 长的序列, 传输速率为 R, 即传输 2^{NR} 条序列
- 输入中约 $2^{NH(X)}$ 条典型列, 输出中约 $2^{NH(Y)}$ 条典型列
- 一条输入序列可能对应 $2^{NH(Y|X)}$ 条输出序列 (疑义度)
- 输出序列之间不发生重合

$$2^{NR}2^{NH(Y|X)} < 2^{NH(Y)} \quad (14)$$

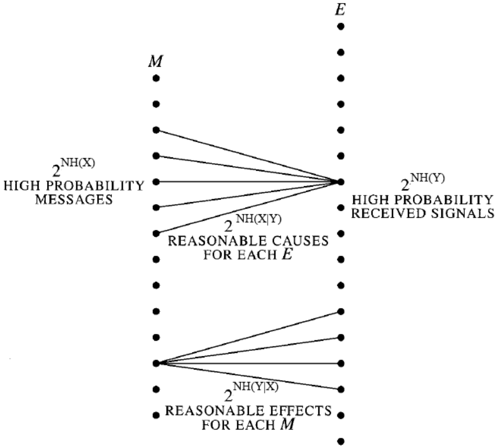

图. 5. 典型列示意

联合典型列 intuition

- 输入中约 $2^{NH(X)}$ 条典型列，
  输出中约 $2^{NH(Y)}$ 条典型列

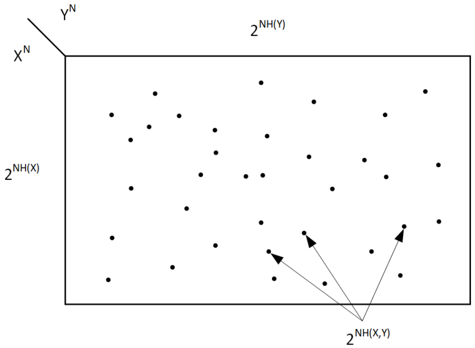

图. 6. 联合典型列示意 [2]

联合典型列 intuition

- 输入中约 $2^{NH(X)}$ 条典型列，
  输出中约 $2^{NH(Y)}$ 条典型列
- 共 $2^{NH(X,Y)}$ 条联合典型列，
  概率 $2^{-NI(X;Y)}$

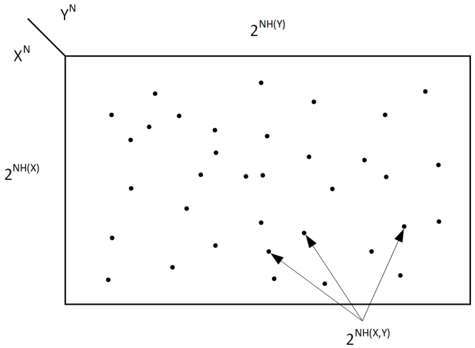

图. 6. 联合典型列示意 [2]

联合典型列 intuition

- 输入中约 $2^{NH(X)}$ 条典型列，输出中约 $2^{NH(Y)}$ 条典型列
- 共 $2^{NH(X,Y)}$ 条联合典型列，概率 $2^{-NI(X;Y)}$
- 给定 Y^N，约 $2^{NI(X;Y)}$ 条 X^N 形成典型列

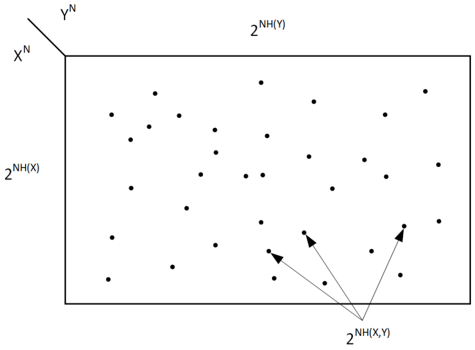

图. 6. 联合典型列示意 [2]

Outline

Chap 2 基本概念

- 各种定义：熵，互信息，微分熵，熵速率

Chap 2 基本概念

- 各种定义：熵，互信息，微分熵，熵速率
- 重要定理：链式法则，Fano 不等式

Chap 2 基本概念

- 各种定义：熵，互信息，微分熵，熵速率
- 重要定理：链式法则，Fano 不等式
- 计算：Markov 信源的冗余度

Chap 3 信源编码

- 重要定义：AEP 性质（直观理解就好了吧）

Chap 3 信源编码

- 重要定义：AEP 性质（直观理解就好了吧）
- 重要定理：信源编码定理，Kraft 不等式

Chap 3 信源编码

- 重要定义：AEP 性质（直观理解就好了吧）
- 重要定理：信源编码定理，Kraft 不等式
- 码的定义：non-singular，uniquely decodable，instantaneous

Chap 3 信源编码

- 重要定义：AEP 性质（直观理解就好了吧）
- 重要定理：信源编码定理，Kraft 不等式
- 码的定义：non-singular，uniquely decodable，instantaneous
- 计算：SP 方法，构造 Huffman 码

Chap 3 信源编码

- 重要定义：AEP 性质（直观理解就好了吧）
- 重要定理：信源编码定理，Kraft 不等式
- 码的定义：non-singular，uniquely decodable，instantaneous
- 计算：SP 方法，构造 Huffman 码
- 细节：多元 Huffman，最佳二元码要求

Chap 4 信道编码

- 重要定义：joint AEP，信道容量

Chap 4 信道编码

- 重要定义：joint AEP，信道容量
- 重要定理：信道编码定理，AWGN 信道容量

Chap 4 信道编码

- 重要定义：joint AEP，信道容量
- 重要定理：信道编码定理，AWGN 信道容量
- 信道：BSC，BEC，组合信道

Chap 4 信道编码

- 重要定义：joint AEP，信道容量
- 重要定理：信道编码定理，AWGN 信道容量
- 信道：BSC，BEC，组合信道
- 计算：对称 DMC 的容量计算，注水法则

Chap 5 率失真理论

- 重要定义：rate-distortion pair

Chap 5 率失真理论

- 重要定义：rate-distortion pair
- 计算：不同失真下的率失真方程，逆注水法则

Chap 6 计算理论

- 重要定义：图灵机，K 复杂度，先验，后验

Chap 6 计算理论

- 重要定义：图灵机，K 复杂度，先验，后验
- 计算：贝叶斯分类，决策树，K-means

Chap 6 计算理论

- 重要定义：图灵机，K 复杂度，先验，后验
- 计算：贝叶斯分类，决策树，K-means
- 细节：图灵停机问题，K 复杂度上下界

Chap 6 计算理论

- 重要定义：传递函数，能控性，能观性，稳定性（Routh，李雅普诺夫）

Chap 6 计算理论

- 重要定义：传递函数，能控性，能观性，稳定性（Routh，李雅普诺夫）
- 计算：判断能控能观，判断稳定性

Chap 6 计算理论

- 重要定义：传递函数，能控性，能观性，稳定性（Routh，李雅普诺夫）
- 计算：判断能控能观，判断稳定性
- 细节：外部稳定性，内部稳定性

Outline

远程声控系统

实现一个远程声音控制系统

- 首先采集不同的语音指示信号，进行适当压缩
- 然后通过噪声信道实现远程传输
- 远端接收后再通过适当计算识别出是何指示
- 最后送入一个处于未知状态、但能控/能观的控制系统，完成不同的控制动作

远程声控系统

实现一个远程声音控制系统

- 信息采集，信源编码 ⇒ 采样 + 量化，Huffman 编码
- 信道编码，（信号调制）⇒ Hamming 码 +BSC，Polar+BPSK+AGWN
- 解码，（解调），信息识别 ⇒ 语音识别
- 信号控制 ⇒ PID 控制系统

推荐

- A Mathematical Theory of Communication, C. S. Shannon
- Elements of Information Theory, T. M. Cover
- ECE 563 - Information Theory `https://courses.engr.illinois.edu/ece563/fa2021`
- EE514A Information Theory `https://www.bilibili.com/video/BV1Vb411G7ZD`
- CC98 `https://www.cc98.org/topic/5031994`

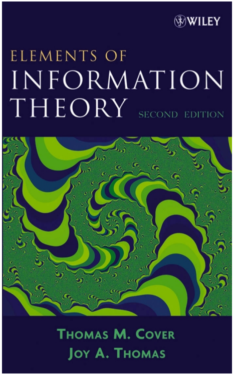

图. 7. 信息论基础 [2]

Outline

Reference

C. E. Shannon, “A mathematical theory of communication,” *ACM SIGMOBILE mobile computing and communications review*, vol. 5, no. 1, pp. 3–55, 2001.

T. M. Cover, *Elements of information theory*.
John Wiley & Sons, 1999.

谢谢

谢谢

Outline

熵

对等概率的熵 $H(\frac{1}{n},\cdots,\frac{1}{n}) \triangleq A(n)$

为满足 3(可加性)，对于 m 长序列有 $A(s^m) = mA(s)$，同样的有 $A(t^n) = nA(t)$

选取 s, m, t, n 使得

$$\begin{aligned} s^m &\le t^n \le s^{m+1} \\ \frac{m}{n} &\le \frac{\log t}{\log s} \le \frac{m}{n} + \frac{1}{n} \end{aligned} \tag{15}$$

熵

为满足 2(单调递增)，$A(s^m) \leq A(t^n) \leq A(s^{m+1})$

即有

$$\frac{m}{n} \leq \frac{A(t)}{A(s)} \leq \frac{m}{n} + \frac{1}{n} \tag{16}$$

结合 15和 16，

$$\left| \frac{\log t}{\log s} - \frac{A(t)}{A(s)} \right| < \epsilon \Longleftrightarrow A(t) = K \log t \tag{17}$$

熵

推广到非等概情况，将等概率的 n 种情况分为 $n = \sum n_i$，对应概率 $p_i = \frac{n_i}{n}$

利用 3，

$$
\begin{aligned}
K \log n &= H(p_1, \cdots) + K \sum p_i \log n_i \\
&\Longrightarrow \\
H(p_1, \cdots) &= K \sum p_i \log \frac{1}{p_i}
\end{aligned} \tag{18}
$$

Jump Back

典型列

考虑 N 长序列，符号 $\{s_i\}$ 相互独立，理想的出现符号 s_i 的次数为 $p_i N$

设理想的序列分布为 $Q(s^N) = \prod_{i=1}^{N} p_i^{p_i N}$

考察两个分布的差异

$$\begin{aligned} D(P\|Q) &= \sum P(s^N)\log\frac{P(s^N)}{Q(s^N)} \\ &= \sum P(s^N)[\log P(s^N) - \log Q(s^N)] \\ &= \mathsf{E}[-\frac{\log P(s^N)}{N} - H(\mathcal{S})] \end{aligned} \tag{19}$$

定义 (典型列 (KL 散度))

$$A_\epsilon^{(N)} = \{s^N : D(P\|Q) \leq \epsilon\} \tag{20}$$

Jump Back

Fano 不等式

定理 (Fano 不等式)

$\mathcal{X}$ 为符号集，Y 为对 X 的估计

$$H(X|Y) \leq H(P_e) + P_e \log(|\mathcal{X}| - 1) \tag{21}$$

其中，$P_e \triangleq P(X \neq Y)$

Proof.

$$\begin{aligned} H(X|Y) &= H(P_e) + P_e H(X|X \neq Y) \\ &\leq H(P_e) + P_e \log(|\mathcal{X}| - 1) \end{aligned}$$

□

Jump Back

Codebook

- generate $(2^{nR}, n)$ randomly $X \sim p(x)$
- 2^{nR} codewords $X^n \sim \prod_{i=1}^{n} p(x_i)$
- a message $W,\ \Pr\{W = w\} = 2^{-nR}$
- code $\mathcal{C} = \begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_n(1) \\ \vdots & \vdots & \ddots & \vdots \\ x_1\left(2^{nR}\right) & x_2\left(2^{nR}\right) & \cdots & x_n\left(2^{nR}\right) \end{bmatrix}$
    - $X^n(w)$ corresponds to w^{th} row
    - $\Pr\{\mathcal{C}\} = \prod_{w=1}^{2^{nR}} \prod_{i=1}^{n} p(x_i(w))$

Decoder

- received sequence $Y^n \sim p(y^n|x^n(w)) = \prod_{i=1}^{n} p(y_i|x_i(w))$
- receiver minimizing error probability (maximum likelihood) is hard, use jointly typical decoding
  - $\exists$ only one $\hat{W},\ s.t.\ (X^n(\hat{W}), Y^n) \in A_\epsilon^{(n)}$
  - declare $\hat{W}$ was sent
  - otherwise assume an error
- decoding error $\mathcal{E} = \{\hat{W} \neq W\}$

Error Analysis

- $\Pr\{\mathcal{E}\} = \sum_{\mathcal{C}} \Pr\{\mathcal{C}\} P_e^{(n)}(\mathcal{C})$
  - for each codebook, average error $P_e^{(n)}(\mathcal{C}) = \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \lambda_w(\mathcal{C})$
  - $\lambda_w(\mathcal{C})$: error prob. for message w using codebok $\lambda_w(\mathcal{C})$
- $\Pr\{\mathcal{E}\} = \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \sum_{C} \Pr\{C\} \lambda_w(C)$
  - average over all codebook is symmetric & independent of w
  - $\sum_{C} \Pr\{C\} \lambda_w(C) \to \sum_{C} \Pr\{C\} \lambda_1(C)$
- $\Pr\{\mathcal{E}\} = \sum_{C} \Pr\{C\} \lambda_1(C) = \Pr\{\mathcal{E} | W = 1\}$

Achievablity

- i^{th} codeword & Y^n jointly typical $E_i = \{(X^n(i), Y^n) \in A_\epsilon^{(n)}\}$
- $\Pr\{\mathcal{E}\} = \Pr\{E_1^c \cup \bigcup_{i=2}^{2^{nR}} E_i | W = 1\} \leq \Pr\{E_1^c | W = 1\} + \sum_{i=2}^{2^{nR}} \Pr\{E_i | W = 1\}$
  - joint AEP$\Rightarrow \Pr\{E_1^c | W = 1\} \leq \epsilon$
  - $X^n(1) \perp\!\!\!\perp X^n(i) \Rightarrow \Pr\{E_i, i \geq 2\} \leq 2^{-n(I(X;Y)-3\epsilon)}$
- $\Pr\{\mathcal{E}\} \leq \epsilon + 2^{-n(I-R-3\epsilon)}$
- $R < I - 3\epsilon \Rightarrow \Pr\{\mathcal{E}\} \leq 2\epsilon$

Jump Back