

Binomial distribution: 分布函数可以根据定义直接写出。

计算期望与方差时看作 N 个 I.I.d 二项分布的和。

Stirling's approximation: 估计阶乘 $x! \approx x^x e^{-x}$

$\lambda \rightarrow \infty, r \rightarrow \lambda$ 时。

$$e^{-\lambda} \frac{\lambda^r}{r!} \approx \frac{1}{\sqrt{2\pi\lambda}} e^{-\frac{(r-\lambda)^2}{2\lambda}}$$

$$e^{-\lambda} \frac{\lambda^r}{r!} \approx \frac{1}{\sqrt{2\pi\lambda}}$$

$$x! \approx \lambda^{\lambda} e^{\lambda} \sqrt{2\pi\lambda}$$

$$\ln x! \approx x \ln x - x + \frac{1}{2} \ln 2\pi x$$

$$\ln \binom{N}{r} \approx (N-r) \ln \frac{N}{N-r} + r \ln \frac{N}{r} + \frac{1}{2} (\ln 2\pi N - \ln 2\pi r - \ln \binom{m}{N-r})$$

$$\log \binom{N}{r} \approx N H_2(r/N) + \frac{1}{2} \log \frac{N}{2\pi r(N-r)}$$

$$\binom{N}{r} \approx 2^{N H_2(r/N)}$$

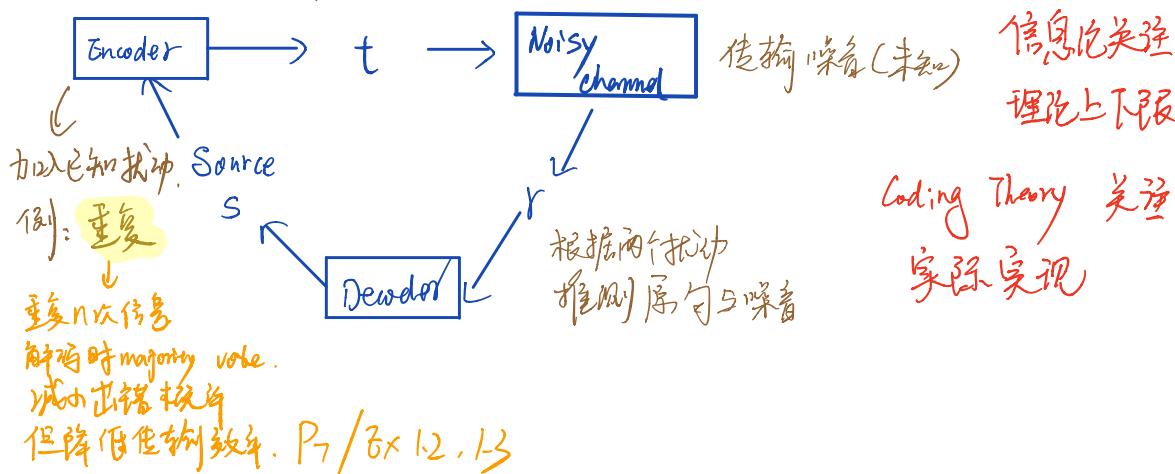
$$\log \binom{N}{r} \approx N H_2(r/N) - \frac{1}{2} \log [2\pi N \frac{N-r}{N} \frac{r}{N}]$$

后面 \sqrt{n} 所比较小，忽略。

在有干扰信道下传输

Channels are always noisy,
every bit has a probability of f to be changed during transmitting.

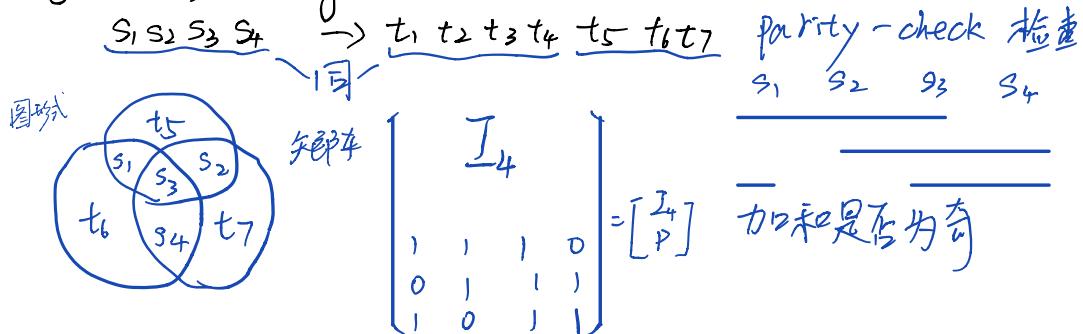
Focus on 'system' solution



例): Block code 对一串数据加干扰

将长度为 k 的序列 s 加干扰为 长度为 N 的序列 ($N \geq k$)

e.g. (7,4) Hamming code



经过信道后解码 r : $s = \arg \min_s \|ts - r\|$

实际操作: 冲突圈 (加和为奇)

反转 [冲突圈内了] [非冲突圈外]

除了画图, 也可用矩阵检查冲突. $H = [P \ I_3]$

问题: 如果信道过程中有

两位出错, 会解码错 3 个

Shannon 证明 decoded bit-error p_b 和信道误码率 p_e 构成的 (R, p_b)
的可走点集和不可走点集的分界线不经过原点.

什么是概率 (两种观点)

- 重复随机试验中的发生频率
- 描述主张的可信度. [Bayesian / Subjective View]
↓ 故可用于描述假说及其推断. P_{obs} Cox axioms.

Likelihood Principle

给定参数 θ , 给定关于数据 d 的模型 $P(d|\theta)$, 观察到数据 d_1 , 则
之后所有的推断和预测都 只和 $P(d_1|\theta)$ 有关

Shannon information content of an outcome x

$$h(x) = \log_2 \frac{1}{p(x)} \quad \text{单位 bits. (衡量了二进制编码文件长度)}$$

Entropy of an ensemble X (集合 X 的熵)

$$H(X) = \sum_{x \in X} P(x) \log \frac{1}{P(x)} \quad [\text{Shannon information}]$$

Another name: uncertainty of X (.. 均匀分布 Entropy 最大, $H(X) \leq \log |A_X|$)

$$H(X) \geq 0 \\ " = " \iff \exists i, \text{s.t. } P_i = 1 \quad \text{Redundancy } 1 - H(X) / \log |A_X|$$

$$\text{联合 Entropy } H(X,Y) = \sum_{x,y} P(x,y) \log \frac{1}{P(x,y)}$$

$$H(X,Y) = H(X) + H(Y) \Leftrightarrow P(x,y) = P(x)P(y)$$

Decomposability of the entropy

$$\begin{aligned} H(\mathbf{p}) &= H[(p_1 + p_2 + \dots + p_m), (p_{m+1} + p_{m+2} + \dots + p_I)] \\ &\quad + (p_1 + \dots + p_m)H\left(\frac{p_1}{(p_1 + \dots + p_m)}, \dots, \frac{p_m}{(p_1 + \dots + p_m)}\right) \\ &\quad + (p_{m+1} + \dots + p_I)H\left(\frac{p_{m+1}}{(p_{m+1} + \dots + p_I)}, \dots, \frac{p_I}{(p_{m+1} + \dots + p_I)}\right) \end{aligned}$$

$$\text{KL 故度 } D_{\text{KL}}(P||Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$$

$$\text{Gibbs' inequality } D_{\text{KL}}(P||Q) \geq 0$$

Jensen's inequality: f 是凸函数，则

$$E[f(x)] \geq f(E[x])$$

凸函数相反

1 byte = 8 bits 可表示 $0 \sim 255$

ASCII 编码，只用 8 bits 中的 7 个 表示从 $0 \sim 127$

Raw bit content of X

$$H_0(X) = \log_2 |A_X|$$

Compression method

lossy: a probability δ of failure. $\delta = P(X \text{ 不在可能取值里})$

lossless: 一一对应

smallest δ -sufficient subset. S_δ is the smallest subset of A_X 满足

$$P(x \in S_\delta) \geq 1 - \delta$$

essential bit content of X

$$H_\delta(X) = \log_2 |S_\delta|$$

Theorem Shannon's source coding theorem

对 X $H(X) = H$ bits. 存在 $\varepsilon > 0$ and $0 < \delta < 1$. $\exists N$

st 对 $N > N_0$, $|\frac{1}{N} H_\delta(X^N) - H| < \varepsilon$

Typical set (β)

在这个集合里的 x 满足 $P(x) \approx p_1^{(P_1 N)} \cdots p_I^{(P_I N)}$

$$\log_2 \frac{1}{p(x)} \approx N \sum p_i \log_2 \frac{1}{p_i} \approx NH.$$

$$\therefore T_{N\beta} = \left\{ x \in \mathcal{A}_x^N : \left| \frac{1}{N} \log_2 \frac{1}{p(x)} - H \right| < \beta \right\}$$

Asymptotic equipartition principle

N i.i.d ensemble r.v. $X^N = (X_1, X_2, \dots, X_N)$,

N 充分大.

outcome $x = (x_1, x_2, \dots, x_N)$ is almost certain to belong to a subset of \mathcal{A}_x^N (have only $2^{NH(x)}$ members) each probability close to $2^{-NH(x)}$

Shannon's source coding theorem.

N iid rv each with entropy $H(X)$ can be compressed into more than $NH(X)$ bits with negligible risk of information loss, as $N \rightarrow \infty$;
conversely if they are compressed into fewer than $NH(X)$ bits it is virtually certain that information will be lost.

也就是说，一个长 N 的序列，里面每个元素是 i.i.d. 的，熵为 $H(X)$ 。

那么它们至少也应该被压缩成 $NH(X)$ 长。

更短的表示必然丢失信息。

Source coding theorem (symbol codes)

对 ensemble X , 存在一个度量 encoding C , 测量 encoded symbol 的平均长度 $L(C, X)$ (即 $L(C, X) \in [H(X), H(X)+1]$) $L(C, X) = \sum_{i=1}^2 p_i l_i$

$L(C, X) = H(X) \Leftrightarrow$ code length for each outcome = Shannon information content

Binary symbol code C of X .

map the range of X , $\mathcal{A}_x = \{a_1, \dots, a_2\}$ to $\{0, 1\}^+$

$c(x)$ denote the codeword corresponding to x
 $l(x)$ denote its length ($i = l(c_{ai})$)

C^+ 是 mapping from $\mathcal{A}_x^+ \rightarrow \{0, 1\}^+$.

$C^+(x_1, x_2, \dots, x_n) = c(x_1)c(x_2)\dots c(x_n)$ (这里是指起来, 不是乘).

A useful symbol code should: no codeword can be prefix of another

Unique decoding ; Easy to decode ; As much compression as possible

Prefix code: no codeword is a prefix of any other codeword.
can be uniquely decoded.

Kraft inequality:

\exists uniquely decodable code C over the binary alphabet $\{0, 1\}$. the codeword length $\sum_{i=1}^I 2^{-l_i} \leq 1$ $I = |\mathcal{A}_C|$

Completeness:

- \nexists uniquely decodable code 使得 Kraft 不等式的等号成立.

Kraft inequality and prefix codes

给定 a set of codeword lengths, 使得 Kraft 不等式 存在

- \exists uniquely decodable **prefix code** with these codeword lengths.

Lower bound on $L[C]$

Uniquely decodable code.

$$L(C, X) = \sum_i p_i l_i \geq H(X) = \sum_i p_i \log \frac{1}{p_i}$$

对每组 $\{l_i\}$, 隐含着一组分布 $\{q_i\}$

$$q_i = 2^{-l_i} / z \quad z = \sum_i 2^{-l_i}$$

Source coding theorem for symbol codes.

\exists ensemble X , 存在 prefix code C

$$H(X) \leq L(C, X) < H(X) + \epsilon$$

If we use a code whose lengths \neq optimal codelengths, Define implicit probabilities

$$L(C, X) = H(X) + \sum_i p_i \log \frac{p_i}{q_i} \quad q_i = 2^{-l_i}$$

Finding Optimal Prefix —— Huffman coding algorithm.

给定每个 symbol 的概率.

排序. 把最小概率的两个 symbol 合并为一个. 重复.

Arithmetic coding

对数据的压缩 entails 数据概率模型.

用一个能产生 context-dependent predictive distribution 的模型替代 Guessing Game 中的 human.

对每个序列 a_1, \dots, a_n , 模型产生概率. 这个概率落在某个二分区间里.

把二分区间用 0..1 表示.