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 $x! \simeq x^x e^{-x}$

$\lambda \to \infty$ $r \to \lambda$ 时.

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

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

$$\lambda! \simeq \lambda^\lambda e^{-\lambda}\sqrt{2\pi\lambda}$$

后面项所比较小，忽略.

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

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

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

$$H_2(x) \equiv x\log\frac{1}{x} + (1-x)\log\frac{1}{(1-x)}$$

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

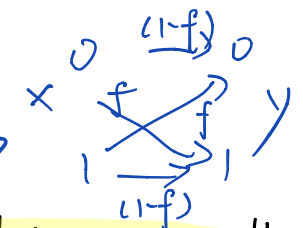
$$\log\binom{N}{r} \simeq N H_2(r/N) - \frac{1}{2}\log\left[2\pi N \frac{N-r}{N}\frac{r}{N}\right]$$

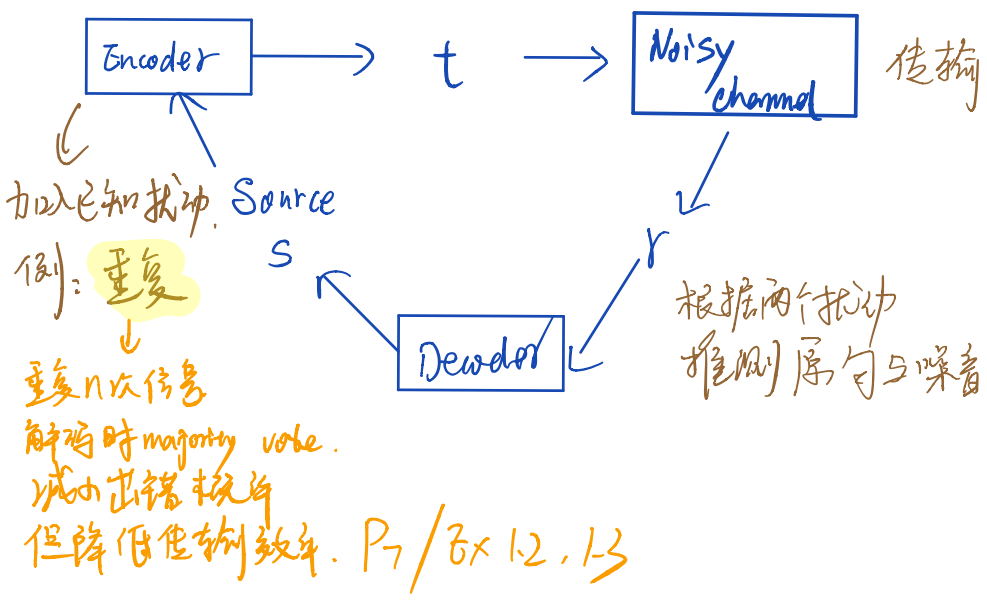
在有干扰信道下传输

Channels are always noisy,

every bit has a probability of f to be changed during transmitting

Focus on 'system' solution





信息论关注理论上下限

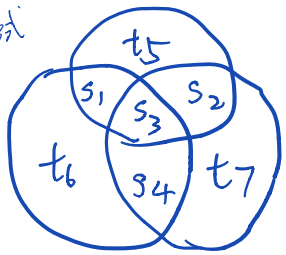
Coding Theory 关注实际实现

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 ($N \geq K$)

eg. (7,4) Hamming code

$s_1 s_2 s_3 s_4 \rightarrow t_1 t_2 t_3 t_4 t_5 t_6 t_7$　parity-check 检查



同矩阵

$$\begin{bmatrix} I_4 \\ \\ \\ \end{bmatrix} = \begin{bmatrix} I_4 \\ P \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

s_1	s_2	s_3	s_4

加和是否为奇

经过传输后解码 r：$s = \arg\min_s \|t(s) - r\|$

实际操作：冲突圈（加和为奇）

反映 [冲突圈内了 ∩ [非冲突圈外了

除了画圈也可用矩阵检查冲突　$H = [P\ I_3]$

问题：如果传输过程中有两位出错，会解码错3个

Shannon 证明 decoded bit-error p_b 和信道传输噪音率 R 构成的 (R, p_b) 的可达点集和不可达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
- 描述主观的可信度. [Bayesian / Subjective View]

↓故可用于描述假设及其推断.　P26　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

$h(x) = \log_2 \frac{1}{P(x)}$　　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）

$H(X) \equiv \sum_{x \in A_x} P(x) \log \frac{1}{P(x)}$　E(Shannon information)

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

$H(X) \geqslant 0$

"=" $\Leftrightarrow \exists! i$, s.t. $P_i = 1$

Redundancy　$1 - H(X)/\log|A_x|$

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

$$\begin{aligned} H(\mathbf{p}) &= H\left[(p_1 + p_2 + \cdots + p_m), (p_{m+1} + p_{m+2} + \cdots + p_I)\right] \\ &+ (p_1 + \cdots + p_m) H\left(\frac{p_1}{(p_1 + \cdots + p_m)}, \ldots, \frac{p_m}{(p_1 + \cdots + p_m)}\right) \\ &+ (p_{m+1} + \cdots + p_I) H\left(\frac{p_{m+1}}{(p_{m+1} + \cdots + p_I)}, \ldots, \frac{p_I}{(p_{m+1} + \cdots + p_I)}\right) \end{aligned} \quad (2.44)$$

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

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

Jensen's inequality: f 是凸函数，则

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

凹函数相反

1 byte = 8 bits 可表示 0～255

ASCII 编码，只用 8 bits 中的 7 个. 表示从 0～127

Raw bit content of X

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

Compression method

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

lossless: 类一对一

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. 给定 $\varepsilon > 0$ and $0 < \delta < 1$. 存 N_0

s.t 对 $N > N_0$, $\left|\frac{1}{N} H_\delta(X^N) - H\right| < \varepsilon$

Typical set (β)

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

$$\log_2 \frac{1}{P(x)} \simeq N \sum_i p_i \log_2 \frac{1}{p_i} \simeq NH.$$

$$\therefore T_{N\beta} \equiv \left\{ x \in A_X^N : \left| \frac{1}{N} \log_2 \frac{1}{P(x)} - H \right| < \beta \right\}$$

Asymptotic equipartition principle

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

N 充分大.

outcome $x = (x_1, x_2, \cdots, x_N)$ is almost certain to belong to a subset of 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 $N \to \infty$;
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)，那么它们至少也该被压缩成 $NH(X)$ 长.

更短的表示必然会丢失信息.

Source coding theorem (symbol codes)

对ensemble X，存在一个变长encoding C，满足 encoded symbol 的平均长度 $L(C,X)$ 满足 $L(C,X) \in [H(X), H(X)+1)$ $L(C,X) = \sum_{i=1}^{I} p_i l_i$

$L(C,X) = H(X) \iff$ codelength for each outcome = Shannon information content

Binary symbol code C of X.

map the range of X, $A_X = \{a_1, \cdots, a_I\}$ to $\{0,1\}^+$

$c(x)$ denote the codeword corresponding to x
$l(x)$ denote its length $l_i = l(a_i)$

C^+ 是 mapping from $A_X^+ \to \{0,1\}^+$.

$$C^+(x_1 x_2 \cdots x_n) = c(x_1)c(x_2)\cdots c(x_N)$$ (这里是拼起来，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:
对 uniquely decodeable code C over the binary alphabet $\{0,1\}$. the codeword length 后满足

$$\sum_{i=1}^{I} 2^{-l_i} \le 1 \qquad I = |A_X|$$

Completeness.
一个 uniquely decodeable code 满足 Kraft 不等式的等号成立.

Kraft inequality and prefix codes
给定 a set of codeword lengths, 满足 Kraft 不等式. 存在
一个 uniquely decodeable prefix code with these codeword lengths.

Lower bound on E[length]

Uniquely decodeable code.

$$L(C,X) = \sum_i p_i l_i \ge H(X) = \sum_i p_i \log 1/p_i$$

对每组 $\{l_i\}$. 隐含着一组分布 $\{q_i\}$

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

Source coding theorem for symbol codes.
对 ensemble X, 存在 prefix code C

$$H(X) \le L(C,X) < H(X)+1$$

If we use a code whose lengths != optimal codelengths, Define implicit probabilities $q_i = 2^{-l_i}$

$$L(C,X) = H(X) + \sum_i p_i \log p_i/q_i.$$

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.