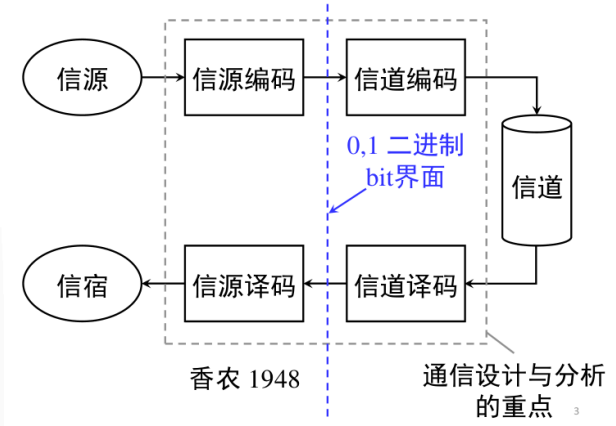
信息论

信息论的基本模型



信源

信源

信源的概念

> ㊀ 信源
>
> 信息的产生者称为信源 (source of information)，产生一个随机过程发出信息。

离散无记忆信源

离散信源 (discrete source) 是时间、取值上都离散的信源，可以表示为

$$X[k] \in \{x_1, x_2, \cdots, x_N\}$$

为简化讨论，可以假设 $X[k]$ **是独立同分布的随机过程**，即：

> ㊀ 离散无记忆信源
>
> 持续产生独立同分布的符号 $X \in \{x_1, x_2, \cdots, x_N\}$ 的信源，称为离散无记忆信源 (discrete memoryless source, DMS)，记为
>
> $$X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_N \\ p_1 & p_2 & \cdots & p_N \end{pmatrix}, \quad p_i = \Pr\{X = x_i\}$$

信源编码的基本要求

将信源产生的符号 X 映射为 0, 1 比特串 $f(X)$ 的过程称为**信源编码 (source coding)**，产生的比特串称为**码字 (codeword)**，其长度 l 称为**码长 (code length)**。

- 若不同 x_i 映射出的码长 l_i 相同，称为**定长码 (fixed-length code)**；
- 若不同 x_i 映射出的码长 l_i 不同，称为**变长码 (variable-length code)**。

定长码的可解码条件

对**定长码**，要求

$$f(x_i) \neq f(x_j), \quad \forall i \neq j$$

因此固定码长 l 应满足

$$N \leq 2^l \implies l \geq \lceil \log N \rceil$$

此处 log 均为以 2 为底的对数。

变长码的可解码条件

对**变长码**，要求任意码字不能是另一个码字的前缀，因此又称为**前缀码 (prefix code)**。

引入**平均码长 (average code length)**

$$\bar{l} = \sum_{i=1}^{N} p_i l_i$$

我们希望在可解码的条件下，使 $\bar{l}$ 尽可能小。

可以证明，信源 $X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_N \\ p_1 & p_2 & \cdots & p_N \end{pmatrix}$ 的最小平均码长为

$$\bar{l}_{\min} = -\sum_{i=1}^{N} p_i \log p_i$$

信源的熵

离散信源的熵

> ㊀ 离散信源的熵
>
> 离散无记忆信源 $X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_N \\ p_1 & p_2 & \cdots & p_N \end{pmatrix}$ 的熵 (entropy) 定义为
>
> $$H(X) = -\sum_{i=1}^{N} p_i \log p_i$$

$H(X)$ 可以写为

$$H(X) = \mathbb{E}_X[-\log p(X)]$$

因而 $-\log \Pr\{X = x_i\}$ 刻画事件 $\{X = x_i\}$ 所包含的信息量，**概率越小，信息量越大**。

熵的性质

考虑 $H(X) = \bar{l}_{\min}$ 的物理含义，应有 $0 \leq H(X) \leq \log N$。

- 当且仅当
  $$\exists i \in \{1, 2, \cdots, N\}, \quad p_i = 1$$
  时，$H(X) = 0$，此时信源没有不确定性，不包含信息；
- 当且仅当
  $$p_i = \frac{1}{N}, \quad i = 1, 2, \cdots, N$$
  时，$H(X) = \log N$，此时信源不确定性最大，包含信息量最大。
  对离散信源，**均匀分布的不确定度最大**。

联合熵、条件熵

> ㊀ 联合熵
>
> 考虑两个离散无记忆信源 X 和 Y，事件 $\{X = x_i, Y = y_j\}$ 的概率为 $p_{ij} = \Pr\{X = x_i, Y = y_j\}$，则 X 和 Y 的联合熵 (joint entropy) 定义为
>
> $$H(X,Y) = \mathbb{E}_{XY}[-\log p(X,Y)] = -\sum_{i=1}^{N}\sum_{j=1}^{M} p_{ij} \log p_{ij}$$
>
> 在没有歧义的情况下，联合熵 $H(X,Y)$ 可记为 $H(XY)$。

联合熵刻画了将 X 和 Y 一起编码所需的最小平均码长，即综合考虑 X 和 Y 的信息量。

> ㊀ 条件熵
>
> 考虑 X 和 Y 的联合分布 $p_{ij} = \Pr\{X = x_i, Y = y_j\}$，在 Y 已知的条件下，事件 $\{X = x_i\}$ 发生的概率为
>
> $$p_{i|j} = \Pr\{X = x_i \mid Y = y_j\} = \frac{\Pr\{X = x_i, Y = y_j\}}{\Pr\{Y = y_j\}} = \frac{p_{ij}}{p_j}$$
>
> 则以 Y 为条件的 X 的条件熵 (conditional entropy) 定义为
>
> $$H(X \mid Y) = \mathbb{E}_{XY}[-\log p(X \mid Y)] = -\sum_{i=1}^{N}\sum_{j=1}^{M} p_{ij} \log p_{i|j}$$

条件熵 $H(X \mid Y)$ 刻画了在已知 Y 的条件下，X 所包含的信息量，即在观测 Y 后 X 残存的不确定度。

熵的通信意义

信源的熵 $H(X)$ 刻画了信源 X 所包含的信息量，其值即对 X 进行无失真编码时所需的最小平均码长。

- 考虑对一个离散无记忆信源 X 编码传输，当信道速率（每传输一个信源符号所传输的平均比特数）$R \geq H(X)$ 时，可以实现无失真传输，即信源译码环节可无失真恢复 X；
- 考虑对两个离散无记忆信源 X 和 Y 做联合信源编码传输，当信道速率 $R \geq H(X,Y)$ 时，在信源译码环节可无失真恢复 (X,Y)；
- 考虑对离散无记忆信源 X 编码传输，编译码器可以共同观测另一个信源，当信道速率 $R \geq H(X \mid Y)$ 时，在信源译码环节可无失真恢复。

▶ 熵的链式法则

熵的链式法则

对于任意两个离散随机变量 X 和 Y，有

$$H(X,Y) = H(Y) + H(X \mid Y) = H(X) + H(Y \mid X)$$

进一步地，

- 若 X 与 Y 独立（记为 $X \perp Y$），则 $p_{ij} = p_i p_j$，$p_{i|j} = \dfrac{p_i p_j}{p_j} = p_i$，于是

$$\begin{aligned} H(X \mid Y) &= -\sum_{i=1}^{N}\sum_{j=1}^{M} p_{ij} \log p_{i|j} = -\sum_{i=1}^{N}\sum_{j=1}^{M} p_i p_j \log p_i \\ &= -\sum_{i=1}^{N} p_i \log p_i = H(X) \end{aligned}$$

$$H(XY) = H(X \mid Y) + H(Y) = H(X) + H(Y)$$

即**独立随机变量的联合熵等于各自熵之和**。观测 Y 不会减少 X 的不确定性。

- 若 X 是 Y 的**确定性映射**（记为 $X = f(Y)$），则 $p_{i|j}$ 要么为 0，要么为 1，因此 $H(X \mid Y) = 0$，于是

$$H(X,Y) = H(Y)$$

即**确定性映射不会增加不确定性**。观测 Y 会完全消除 X 的不确定性。

▶ 互信息

$H(X \mid Y)$ 表征通过观测 Y 后 X 残存的不确定度，这种观测所消除不确定度的程度，即称为**互信息 (mutual information)**。

显然，有

$$\begin{aligned} H(X) - H(X \mid Y) &= H(X) - \big(H(X,Y) - H(Y)\big) \\ &= H(X) + H(Y) - H(X,Y) \\ &= H(Y) - H(Y \mid X) \end{aligned}$$

即，通过观测 Y 消除的 X 的不确定度，等于通过观测 X 消除的 Y 的不确定度。X 与 Y 的**互信息是对称的**。

互信息

离散随机变量 X 和 Y 的互信息 (mutual information) 定义为

$$I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$$

一般地，

$$I(X;Y) = H(X) + H(Y) - H(X,Y) = \sum_{i=1}^{N}\sum_{j=1}^{M} p_{ij} \log \frac{p_{ij}}{p_i p_j}$$

考虑两个离散无记忆信源 X 和 Y，

- 若 $X \perp Y$，则

$$I(X;Y) = H(X) - H(X \mid Y) = H(X) - H(X) = 0$$

即**独立信源间没有互信息**；

- 若 $X = f(Y)$，则

$$I(X;Y) = H(X) - H(X \mid Y) = H(X) - 0 = H(X)$$

即**确定性映射的信源间互信息是其全部不确定度**。

- $0 \leq H(X \mid Y) \leq H(X)$，因此 $I(X;Y) \geq 0$，**信源之间不存在欺骗**。

■ 连续信源的微分熵

对于连续随机变量 X，其概率分布由**概率密度函数 (probability density function, PDF)** $p_X(x)$ 描述，满足

$$\int_{-\infty}^{+\infty} p_X(x)\,\mathrm{d}x = 1, \quad p_X(x) \geq 0$$

微分熵

微分熵 (differential entropy) 定义为

$$h(X) = \mathbb{E}_X[-\log p_X(X)] = -\int_{-\infty}^{+\infty} p_X(x) \log p_X(x)\,\mathrm{d}x$$

微分熵 $h(X)$ **刻画了信源 X 的相对不确定度**，其值可以为负数。

Gaussian 分布的微分熵

对 $X \sim \mathcal{N}(\mu, \sigma^2)$，有 $p_X(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)$，因此

$$\begin{aligned} h(X) &= -\int_{-\infty}^{+\infty} p_X(x) \log p_X(x)\,\mathrm{d}x \\ &= -\int_{-\infty}^{+\infty} p_X(x) \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right)\mathrm{d}x \\ &= -\int_{-\infty}^{+\infty} p_X(x) \left(-\frac{(x-\mu)^2}{2\sigma^2}\log \mathrm{e} - \log\sqrt{2\pi\sigma^2}\right)\mathrm{d}x \\ &= \frac{\log \mathrm{e}}{2\sigma^2}\int_{-\infty}^{+\infty} (x-\mu)^2 p_X(x)\,\mathrm{d}x + \log\sqrt{2\pi\sigma^2}\int_{-\infty}^{+\infty} p_X(x)\,\mathrm{d}x \\ &= \frac{\log \mathrm{e}}{2\sigma^2}\mathbb{E}[(X-\mu)^2] + \log\sqrt{2\pi\sigma^2} \\ &= \frac{\log \mathrm{e}}{2\sigma^2}\left(\mathbb{E}^2[(X-\mu)] + \mathrm{Var}[(X-\mu)]\right) + \log\sqrt{2\pi\sigma^2} \\ &= \frac{\log \mathrm{e}}{2\sigma^2}\cdot\sigma^2 + \log\sqrt{2\pi\sigma^2} = \boxed{\log\sqrt{2\pi\mathrm{e}\sigma^2}} \end{aligned}$$

▶ 联合微分熵、条件微分熵

联合微分熵

考虑两个连续随机变量 X 和 Y，其联合概率密度函数为 $p_{X,Y}(x,y)$，则 X 和 Y 的联合微分熵 (joint differential entropy) 定义为

$$h(X,Y) = \mathbb{E}_{X,Y}[-\log p_{X,Y}(X,Y)] = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{X,Y}(x,y) \log p_{X,Y}(x,y)\,\mathrm{d}$$

在没有歧义的情况下，联合微分熵 $h(X,Y)$ 可记为 $h(XY)$。

条件微分熵

考虑 X 和 Y 的联合概率密度函数 $p_{X,Y}(x,y)$，条件随机变量 $X \mid Y$ 的概率密度函数为

$$p_{X|Y}(x,y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

则以 Y 为条件的 X 的条件微分熵 (conditional differential entropy) 定义为

$$h(X \mid Y) = \mathbb{E}_{X,Y}[-\log p_{X|Y}(X,Y)] = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{X,Y}(x,y) \log p_{X|Y}(x,y)$$

类似地，微分熵也有**链式法则**

$$h(X,Y) = h(Y) + h(X \mid Y) = h(X) + h(Y \mid X)$$

▶ 微分熵的互信息

微分熵的互信息

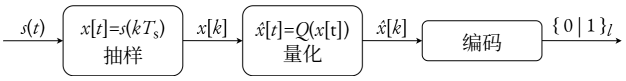
连续随机变量 X 和 Y 的互信息 (mutual information) 定义为

$$I(X;Y) = h(X) - h(X \mid Y) = h(Y) - h(Y \mid X)$$

一般地，

$$I(X;Y) = h(X) - h(X \mid Y) = h(X) + h(Y) - h(X,Y)$$
$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{X,Y}(x,y)\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\,\mathrm{d}x\mathrm{d}y$$

模拟信源的数字编码



模拟信源的数字编码

信道

离散幅值信道

离散无记忆信道模型

离散无记忆信道 (Discrete Memoryless Channel, DMC) 模型假设信道在每次传输时的行为独立且不依赖于之前的传输结果。DMC 由一个条件概率来描述，即**给定信源输入符号 $X = x_i$ 时信宿输出符号 $Y = y_j$ 的概率分布** $p_{j|i}$。

当给定 X 的分布 p_i 时，可以得到联合分布

$$p_{ij} = \Pr\{X = x_i, Y = y_j\} = \Pr\{Y = y_j \mid X = x_i\}\Pr\{X = x_i\} = p_i p_{j|i}$$

信宿输出 Y 的边缘分布为

$$p_j = \Pr\{Y = y_j\} = \sum_i p_{ij} = \sum_i p_i p_{j|i}$$

这样，**互信息 $I(X;Y)$ 就可以给定**为

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) \\ &= -\sum_i p_i \log p_i - \sum_j p_j \log p_j + \sum_i\sum_j p_{ij}\log p_{ij} \\ &= -\sum_i\sum_j p_{ij}\log p_i - \sum_i\sum_j p_{ij}\log p_j + \sum_i\sum_j p_{ij}\log p_{ij} \\ &= \sum_i\sum_j p_{ij}\log\frac{p_{ij}}{p_i p_j} = \sum_i\sum_j p_{ij}\log\frac{p_{j|i}}{p_j} = \sum_i\sum_j p_i p_{j|i}\log\frac{p_{j|i}}{\sum_i p_i p_{j|i}} \end{aligned}$$

若允许通过映射**调整输入分布 p_i**，则可以在此意义下**最大化互信息 $I(X;Y)$**，这一最大值称为**信道容量 (channel capacity)**，记为

$$C = \max_{p_i} I(X;Y) = \max_{p_i}\sum_i\sum_j p_i p_{j|i}\log\frac{p_{j|i}}{\sum_i p_i p_{j|i}}$$

这一优化问题还有两个约束条件，即概率归一化条件 $\sum_i p_i = 1$ 和非负性条件 $p_i \geq 0$。这是一个**非凸**的带约束优化问题，有一定的求解难度。一般 DMC 的容量需要通过数值方法求解，如 Blahut-Arimoto 算法。

BSC 信道

我们希望得到容量的解析闭式解，因此需要在**数字通信**背景下简化 DMC 模型。

> ㊀ 对称二进制信道
>
> **对称二进制信道 (Binary Symmetric Channel, BSC)** 是这样的 DMC：
>
> - 输入和输出符号均为二进制（0 和 1），
> - 每个比特在传输过程中都有一个固定的**差错概率 (cross-over probability)** ε，即输出比特有 ε 的概率变为输入比特的反码。

BSC 信道可以等效为

$$Y = X \oplus Z, \quad Z \sim \text{Bernoulli}(\varepsilon) = \begin{pmatrix} 0 & 1 \\ 1-\varepsilon & \varepsilon \end{pmatrix}$$

其中 Z 是与输入 X 独立的噪声变量，$\oplus$ 表示**按位异或**运算。

使用这一等效，BSC 信道的容量为

$$\begin{aligned} C &= \max_{p_i} I(X;Y) = \max_{p_i} I(X; X\oplus Z) \\ &= \max_{p_i}\big(H(X\oplus Z) - H(X\oplus Z \mid X)\big) \\ &= \max_{p_i}\big(H(X\oplus Z) - H(Z)\big) \\ &= \max_{p_i} H(X\oplus Z) + \underbrace{\varepsilon\log\varepsilon + (1-\varepsilon)\log(1-\varepsilon)}_{\text{const}} \end{aligned}$$

因此，最大化互信息等价于**最大化** $H(X\oplus Z)$。由于 $X\oplus Z \in \{0,1\}$，其熵的最大值为 1，**比特均匀分布**时取得，即

$$C = 1 + \varepsilon\log\varepsilon + (1-\varepsilon)\log(1-\varepsilon) \quad \text{iff} \quad Y \sim \text{Bernoulli}(0.5)$$
$$\iff X \sim \text{Bernoulli}(0.5)$$

连续幅值信道

类似于离散幅值信道，**连续幅值信道 (continuous amplitude channel)** 也可以用条件概率密度函数 $p_{Y|X}(y,x)$ 来描述，所传递的信息量仍然用互信息 $I(X;Y)$ 来衡量。

若允许通过映射**调整输入分布** $p_X(x)$，则可以在此意义下**最大化互信息** $I(X;Y)$，这一最大值仍然称为**信道容量 (channel capacity)**，记为

$$C = \max_{p_X(x)} I(X;Y)$$

Gauss 信道

Gauss 信道 (Gaussian channel) 是一种特殊的连续幅值信道，假设信道输出 Y 是输入 X 与高斯白噪声 Z 的和，即

$$Y = X + Z, \qquad \text{where} \quad Z \sim \mathcal{N}(0,\sigma^2)$$

其中 X 有功率约束 $\mathbb{E}\left[X^2\right] = E_s$。

Gauss 信道的容量为

$$\begin{aligned} C &= \max_{p_X(x)} I(X;Y) = \max_{p_X(x)} I(X; X+Z) = \max_{p_X(x)}\big(h(X+Z) - h(X+Z\mid X)\big) \\ &= \max_{p_X(x)}\big(h(X+Z) - h(Z)\big) = \max_{p_X(x)} h(X+Z) - \log\sqrt{2\pi e\sigma^2} \end{aligned}$$

因此，最大化互信息等价于**最大化** $h(X+Z)$。由于 X、Z 独立，

$$\mathbb{E}\left[(X+Z)^2\right] = \mathbb{E}\left[X^2\right] + \mathbb{E}\left[Z^2\right] + 2\mathbb{E}\left[X\right]\underbrace{\mathbb{E}\left[Z\right]}_{0} = E_s + \sigma^2$$

当且仅当 $X \sim \mathcal{N}(0, E_s)$ 时，$X+Z \sim \mathcal{N}(0, E_s+\sigma^2)$，$h(X+Z)$ 最大化为

$$h(X+Z) = \log\sqrt{2\pi e(E_s+\sigma^2)}$$

从而

$$C = \log\sqrt{2\pi e(E_s+\sigma^2)} - \log\sqrt{2\pi e\sigma^2} = \boxed{\log\sqrt{1+\frac{E_s}{\sigma^2}}}$$

这里 C 无量纲，单位为 bit/次。

带宽受限的 Gauss 信道

在实际通信系统中，信道通常是**带宽受限 (bandwidth-limited)** 的，即信号只能在有限的频率范围内传输。设信道的带宽为 W，则根据 **Nyquist 采样定理**，单位时间最多可以传输 $2W$ 个独立符号。

考虑带宽受限为 W 的 Gauss 信道，由于这一信道中的噪声是与信号**时域相加**的、功率谱在 $[-W, W]$ 均匀的 Gauss 噪声，因此称为**加性白 Gauss 噪声 (Additive White Gaussian Noise, AWGN) 信道**。设其中噪声 $N(t)$ 的方差为 σ^2，则

$$R_N(t,s) = \begin{cases} \mathbb{E}\left[N^2(t)\right] = \sigma^2, & t = s \\ 0, & t \neq s \end{cases} = \sigma^2\delta(t-s)$$

因此 $N(t)$ 是**宽平稳**的，$R_N(\tau) = \sigma^2\delta(\tau)$，其功率谱密度为

$$S_N(f) = \int_{-\infty}^{+\infty} R_N(\tau)\mathrm{e}^{-\mathrm{j}2\pi f\tau}\,\mathrm{d}\tau = \sigma^2, \quad |f| \leq W$$

将正负频部分合并，得到单边功率谱密度为

$$S_N^{(\text{single})}(f) = 2\sigma^2 =: n_0, \quad 0 \leq f \leq W$$

同时，记通信功率 P 为单位时间内传输的平均能量，则每个符号的平均能量为 $E_s = \dfrac{P}{2W}$。则 AWGN 信道的容量为

$$C = 2W\log\sqrt{1+\frac{P}{2W\sigma^2}} = W\log\left(1+\frac{P}{Wn_0}\right)$$

Shannon 公式

带宽受限于 $|f| \leq W$、噪声功率谱密度为 $S_N(f) = \frac{n_0}{2}$ 的 **AWGN 信道**在功率 P 下的容量为

$$C = W\log\left(1 + \frac{P}{Wn_0}\right)$$

其中 C 的单位为 bit/s。

信号功率与噪声功率的比值 $\frac{P}{Wn_0}$ 称为**信噪比 (signal-to-noise ratio, SNR)**，可记为 SNR。因此，Shannon 公式也可以写为

$$C = W\log(1 + \mathrm{SNR})$$

- 当 $\mathrm{SNR} \to 0$ 时，

  $$C \to W\log \mathrm{e} \cdot \mathrm{SNR} = \frac{P}{n_0}\log \mathrm{e}$$

  容量 C 与带宽 W 无关，而与功率 P 呈线性关系；
- 当 $\mathrm{SNR} \to +\infty$ 时，$C \to W\log \mathrm{SNR}$，引入信噪比的分贝值 $\mathrm{SNR}_{\mathrm{dB}} = 10\log_{10}\mathrm{SNR} = \frac{10}{\log 10}\log \mathrm{SNR}$，则

  $$C \to W \cdot \frac{\log 10}{10}\mathrm{SNR}_{\mathrm{dB}} \approx 0.3322W \cdot \mathrm{SNR}_{\mathrm{dB}}$$