

信息熵与量子信息熵

吕铭

2015 年 1 月 9 日

1 Shannon 信息熵

1.1 信息熵的来源

Shannon 在论文 [Shannon, 1948] 中最早提出了对于信息量的测量方案: 信息熵. 具体来说信息熵的概念要描述一套编码所包含的信息量. 编码方案通常可以用具有一组几率分布 $\{p_i\}$ 出现的编码 Γ . 在测量实际信息量时, 具体的编码是可以变化的但对应的几率分布是不变的. 于是寻找熵的过程实际上就是如果有 n 种可能的状态, 出现的几率分别是 p_1, p_2, \dots, p_n , 希望有一个函数 $H(\mathbf{p})$ 来描述信息量.

在 Shannon 的论文中, 提出了三条这个函数应当满足的公式 [Shannon, 1948]:

- 关于 p_i 连续
- $p_i = 1/n$ 时, 关于 n 单调增的
- 拆分过程应当相当于按权重求和

$$H(p_1 \mathbf{q}_1, p_2 \mathbf{q}_2, \dots) = H(\mathbf{p}) + \sum p_i H(\mathbf{q}_i) \quad (1)$$

Shannon 证明了满足上面三条的表达式只能是:

$$H(\mathbf{p}) = -k \sum_i p_i \log p_i \quad k > 0 \quad (2)$$

证明 1 令 $A(n) = H(1/n, 1/n, \dots, 1/n)$, 据第三条公设有

$$A(mn) = A(m) + A(n) \quad (3)$$

同时 $A(n)$ 是单调函数, 从而 $A(n) = k \log n$.

对于一般的 \mathbf{p} , 不妨假定 $p_i = n_i / \sum n_i$, 据第三条公设

$$k \log \sum n_i = H(\mathbf{p}) + \sum p_i k \log n_i \quad (4)$$

于是 $H = -k \sum p_i \log p_i$, 由单调性和连续性可以推广到实数

回顾一下课堂内讲到过的熵的微观对应量 (Gibbs 熵)

$$S = -k_B \langle \ln \rho \rangle = -k_B \sum_s \rho_s \ln \rho_s$$

其中 ρ 表示态密度 (不同态的几率分布)

虽然公式的来源不同但形式完全一致, 这也是为什么将这个量称为熵¹

1.2 Shannon 熵的推广: Rényi 熵

Alfréd Rényi 将这个概念推广到一族函数 [Rényi, 1961] :

$$H_\alpha(\mathbf{p}) = \frac{1}{1-\alpha} \log \left[\sum_{i=1}^d p_i^\alpha \right] \quad \alpha \geq 0 \quad (5)$$

Shannon 熵是 $\alpha \rightarrow 1$ 时的特例. 对于所有的 Rényi 熵, 都有取值范围:

$$0 \leq H_\alpha(\mathbf{p}) \leq \log n \quad (6)$$

取到上限当且仅当 $p_i = \text{const.}$, 表示每一种可能的信号都等可能, 是信息编码效率最高的情况. (有趣的是, 纯噪音也是这种情况)

取到下限当且仅当 $p_i = \delta_{ij}$, 表示没有信息, 因为只有一种可能的信号.

1.3 信息熵与统计物理

在我们之前讲的统计力学中, 我们以等几率原理 (equal *a priori* probabilities) 为基本假设出发, 导出熵作为一个热力学函数来描述热力学关系. 信息熵的概念使得我们可能以概率论导出熵, 再根据概率论的最大熵原则导出等几率原理. 在这个意义下我们并不需要额外的物理学假设来构建统计力学 [Jaynes, 1957a, Jaynes, 1957b].

2 量子信息熵

2.1 量子态的“混合程度”

一般来说, 物理学意义下我们认为熵是描述一个状态的混乱程度的. 而对于一个量子态, 纯态通常认为是高度规则的而混合态被认为是相对混乱的. 而从信息熵的角度来看, 熵可以用来描述我们对于一个体系所知道的知识的大小 (两者单调减). 而纯态是我们对于一个量子态掌握全部信息的描述, 而混合态则是不完全信息的一个统计期望 [Jaynes, 1957b]. 因此我们可以用描述一个态作为混合态的“混合程度”的大小来试图接近于熵的概念.

如果我们用密度矩阵来表示一个态

$$\hat{\rho} = \sum_i p_i |\psi_i\rangle \langle \psi_i| \quad (7)$$

¹“The form of H will be recognized as that of entropy as defined in certain formulations of statistical mechanics...” [Shannon, 1948]

对于一般的么正过程, 纯态总是变成纯态, 是无法产生混合态的. 得到混合态, 我们可以引入随机的变换:

定义 1 态 $\hat{\rho}$ 比态 $\hat{\rho}'$ “混合程度” 更高, 当且仅当存在一组么正变换 U_i 和概率分布 p_i

$$\hat{\rho} = \sum q_k U_k \hat{\rho}' U_k^\dagger \quad (8)$$

显然对于具有相同本征值组的态, 是具有相当的“混合程度”的.

$$\hat{\rho}' = U \hat{\rho} U^\dagger \quad U = \sum_i |\psi'_i\rangle \langle \psi_i|$$

因此概念应当只针对密度矩阵的本征值组

2.2 majorization criterion 优化准则

上面的定义并不方便判定, 引入判定方法

定理 1 将密度矩阵对角化:

$$\hat{\rho} = \sum_{i=1}^d p_i |\alpha_i\rangle \langle \alpha_i| \quad (9)$$

并且假定本征值排序 $p_1 \geq p_2 \geq \dots \geq p_d \geq 0$, 则 $\hat{\rho}$ 比 $\hat{\rho}'$ 更加“混合”等价于

$$\forall t \in \{1, \dots, d-1\}, \quad \sum_{i=1}^t p_i \geq \sum_{i=1}^t p'_i \quad (10)$$

记作 $\mathbf{p} \preceq \mathbf{p}'$

证明 2

$$\begin{aligned} \sum_{i=1}^t p'_i &= \sum_{i=1}^t \langle \psi'_i | \hat{\rho}' | \psi'_i \rangle = \sum_k q_k \sum_{j=1}^d p_j \sum_{i=1}^t |\langle \psi'_i | U_k | \psi_j \rangle|^2 \\ &\leq \sum_k q_k \sum_{j=1}^t p_j \sum_{i=1}^t |\langle \psi'_i | U_k | \psi_j \rangle|^2 \\ &\quad (\text{with } \text{Span}\{U_k | \psi_i\rangle\}_{i=1}^t = \text{Span}\{|\psi'_i\rangle\}_{i=1}^t) \\ &= \sum_k q_k \sum_{j=1}^t p_j = \sum_{j=1}^t p_j \end{aligned}$$

2.3 混合程度的度量与熵

如果需要单变量来描述“混合程度”, 完全等价的描述是做不到的, 如果只关注必要性, 那么可以用任何 Schur-concave 的函数

定义 2 Schur-convex: 函数 $f: \mathbb{R}^d \mapsto \mathbb{R}$, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \mathbf{x} \preceq \mathbf{y}$, 有 $f(\mathbf{x}) \leq f(\mathbf{y})$

Schur-concave: $-f$ 是 Schur-convex 的

于是发现前面提到的 Rényi 熵是满足这个性质的

定义 3 (Rényi 熵) 一族的具有 *Schur-concave* 性质的函数

$$H_\alpha(\mathbf{p}) = \frac{1}{1-\alpha} \log \left[\sum_{i=1}^d p_i^\alpha \right] \quad \alpha \geq 0 \quad (11)$$

直接根据前面的数学定义, 用密度矩阵来表达即为:

定义 4 (量子 Rényi 熵) 定义 α 范数 $\|\hat{\rho}\|_\alpha := (\text{Tr} |\hat{\rho}^\alpha|)^{1/\alpha}$, 则量子 Rényi 熵的表达式

$$S_\alpha(\hat{\rho}) := \frac{\alpha}{1-\alpha} \log \|\hat{\rho}\|_\alpha \quad (12)$$

这样定义的熵具有广延性

$$S_\alpha(\hat{\rho} \otimes \sigma) = S_\alpha(\hat{\rho}) + S_\alpha(\sigma) \quad (13)$$

如果将 $\hat{\rho}$ 视作整个系统所有粒子的密度矩阵 $\rho \in \text{St}[\mathcal{H}^{\otimes N}]$. 则这个上限恰为 Boltzmann 熵 $S \sim \log \Omega$. 取最小时当且仅当 $\hat{\rho}$ 是纯态, 取最大值时当且仅当 $\hat{\rho} = I/d$ 是完全混合态

参照经典信息论的 Shannon 熵, 定义量子 von-Neumann 熵 [Neumann, 1932]

$$S(\hat{\rho}) := \lim_{\alpha \rightarrow 1} S_\alpha(\hat{\rho}) = -\text{Tr}[\hat{\rho} \log \hat{\rho}] \quad (14)$$

这与 Gibbs 熵用密度矩阵的表达式是一致的

如果将前面关于混合程度的定义推广, 那么

定义 5 $\hat{\rho}_1 \in \mathcal{H}_1$ 比 $\hat{\rho}_2 \in \mathcal{H}_2$ 混合程度更高, 当且仅当 $\hat{\rho}_1 \otimes |\alpha\rangle_2 \langle \alpha|_2$ 比 $|\beta\rangle_1 \langle \beta|_1 \otimes \hat{\rho}_2$ 混合程度更高

直积上一个纯态对于本征值组来说只是增加了一串 0, 从前面对于熵的定义来看, 这样的推广定义是合理的.

von-Neumann 熵相比其他 Rényi 熵具有更特殊的性质, 如定义比率 R 为 $\hat{\rho}^{\otimes N}$ 比 $\hat{\rho}'^{\otimes NR}$ 混合程度更高, 则

$$R \leq \frac{S(\hat{\rho})}{S(\hat{\rho}')} \quad (15)$$

3 量子信息熵的应用 (举例)

3.1 粒子的几率分布

对于一个 N 粒子体系, 单粒子的密度矩阵 (期望) 为 $\hat{\rho} = \sum q_m |\psi_m\rangle \langle \psi_m|$, 整个体系

$$\hat{\rho}^{\otimes N} = \sum_{\mathbf{m}} q_N(\mathbf{m}) |\psi_{\mathbf{m}}\rangle \langle \psi_{\mathbf{m}}| \quad (16)$$

其中定义

- $\mathbf{m} = (m_1, \dots, m_N) \in \{1, \dots, d\}^{\times N}$
- $q_N(\mathbf{m})$ 是相应的几率 $q_N(\mathbf{m}) = \prod_i q_{m_i}$
- $|\psi_{\mathbf{m}}\rangle$ 是相应的直积态 $|\psi_{\mathbf{m}}\rangle := |\psi_{m_1}\rangle |\psi_{m_2}\rangle \cdots |\psi_{m_N}\rangle$

另外定义分布序列

$$t_{\mathbf{m}} := (N_1/N, \dots, N_d/N) \quad (17)$$

于是有:

- 可能的分布总数量 $T_N \sim 1$
- 某一中分布 t 对应的态的数量 $S_{N,t} \sim \exp[NH(t)]$
- 某一组分布 t 出现的几率 $Q_{N,t} \sim \exp[-ND(t||q)]$

证明 3 可能的分布总数量

$$\begin{aligned} T_N &= \binom{N+d-1}{d-1} \\ &= \frac{(N+d-1)!}{(d-1)!N!} \\ &\sim O(N^{d-1}) \end{aligned}$$

某一组分布中态的数量

$$\begin{aligned} S_{N,t} &= \binom{N}{N_1} \binom{N-N_1}{N_2} \cdots \binom{N_d}{N_d} = \frac{N!}{N_1!N_2! \cdots N_d!} \\ &\sim \exp \left[N(\ln N - 1) - \sum_{i=1}^d N_i(\ln N_i - 1) \right] \\ &= \exp \left[N \left(\ln N - \sum_{i=1}^d t_i \ln N_i \right) \right] \\ &= \exp [NH(t_{\mathbf{m}})] \end{aligned}$$

其中出现了熵的表达式 $H(t_{\mathbf{m}}) := -\sum_{i=1}^d t_i \ln t_i$

某一组分布的几率

$$\begin{aligned} Q_{N,t} &= S_{N,t} \prod_{i=1}^d q_i^{N_i} \\ &\sim \exp \left[N \left(\ln N - \sum_{i=1}^d t_i \ln N_i + \sum_{i=1}^d t_i \ln q_i \right) \right] \\ &= \exp[-ND(t||q)] \end{aligned}$$

其中 $D(t||q) := \sum_{i=1}^d t_i \ln \frac{t_i}{q_i}$ 称为 *Kullback-Leibler(KL)* 散度, 也称相对熵 [Kullback and Leibler, 1951]

从上面的式子还能看出对于偏离 q_m 分布的 t 分布是随着粒子数迅速趋于 0 的

$$\begin{aligned} \sum_{t:D(t||q)\geq\epsilon_N} Q_{N,t} &\lesssim \exp[-N\epsilon_N] \sum_{t:D(t||q)\geq\epsilon_N} 1 \\ &\leq \exp[-N\epsilon_N] T_N \\ &\sim \exp[-N\epsilon_N] \end{aligned}$$

取 ϵ_N 序列使得 $\lim_{N\rightarrow\infty} N\epsilon_N \rightarrow \infty$ 即可.

3.2 量子纠缠的度量

如果定义“更纠缠”为“不可以通过局域操作 (LOCC²) 转化”, 那么可以证明前面关于“更混合”的定义与“更纠缠”有关联

定理 2 $|\Psi\rangle\langle\Psi|$ 更纠缠当且仅当 $\text{Tr}_B[|\Psi\rangle\langle\Psi|]$ 更混合

特别的, Bell 态 $|\Psi^\pm\rangle$ 是最大纠缠态, 它的部分迹是完全混合态 $\text{Tr}_B[|\Psi^\pm\rangle\langle\Psi^\pm|] = I/2$

3.3 量子信息的压缩

定义量子通道 (quantum channel) $\mathcal{C}(\rho) = \text{Tr}_B[U(\rho \otimes \sigma)U^\dagger]$, 存在一组量子量子通道编码方案 $\mathcal{E}: \mathcal{H}_d \mapsto \mathcal{H}_{d'}$ 与相应的解码方案 $\mathcal{D}: \mathcal{H}_{d'} \mapsto \mathcal{H}_d$ (其中 $d' < d$) 使得以 p_i 的几率编码为量子态 $|\psi_i\rangle$ 的量子信息 (可以描述为密度矩阵) 经过编码和解码后几乎不变

$$\|\rho - \mathcal{D}\mathcal{E}(\rho)\|_1 < \epsilon$$

定理 3 (Schumacher' s noiseless channel coding theorem) 在 $N \rightarrow \infty$ 渐进意义下, 对于 ρ^N 信号的压缩 $\mathcal{E}: \mathcal{H}_{2^N} \mapsto \mathcal{H}_{d_N}$ 以及 $\mathcal{D}: \mathcal{H}_{d_N} \mapsto \mathcal{H}_{2^N}$, 定义压缩率

$$R := \limsup_{N \rightarrow \infty} \frac{\log d_N}{N} \quad (18)$$

理论可以达到的最佳压缩率为 $R \geq S(\rho)$ [Schumacher, 1995]

4 参考文献

References

[Jaynes, 1957a] Jaynes, E. T. (1957a). Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630.

²Local operations and classical communication

- [Jaynes, 1957b] Jaynes, E. T. (1957b). Information theory and statistical mechanics. ii. *Phys. Rev.*, 108:171–190.
- [Kullback and Leibler, 1951] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, pages 79–86.
- [Neumann, 1932] Neumann, J. (1932). *Mathematische Grundlagen der Quantenmechanik*. Verlag von Julius Springer Berlin.
- [Rényi, 1961] Rényi, A. (1961). On measures of information and entropy. In *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960*, pages 547–561.
- [Schumacher, 1995] Schumacher, B. (1995). Quantum coding. *Phys. Rev. A*, 51:2738–2747.
- [Shannon, 1948] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423.