信息熵与量子信息熵

吕铭

清华大学物理系

统计力学课堂报告, 2015

Shannon 信息熵

- 动机: 描述一套编码对应的几率分布所包含的信息量
  有 n 种可能的状态, 出现的几率分别是 $p_1, p_2, \cdots, p_n$, 希望有一个函数 $H(\boldsymbol{p})$ 来描述信息量

Shannon 信息熵

- 动机: 描述一套编码对应的几率分布所包含的信息量
  有 n 种可能的状态, 出现的几率分别是 $p_1, p_2, \cdots, p_n$, 希望有一个函数 $H(\boldsymbol{p})$ 来描述信息量
- 要求满足的条件:
  三条公理 [Shannon, 1948]

Shannon 信息熵 I

三条公理 [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

$$H(p_1\boldsymbol{q}_1, p_2\boldsymbol{q}_2, \cdots) = H(\boldsymbol{p}) + \sum p_i H(\boldsymbol{q}_i) \tag{1}$$

Shannon 信息熵 II

满足上面三条的表达式只能是:

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

Shannon 信息熵 III

Proof.

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

$$A(mn) = A(m) + A(n) \tag{3}$$

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

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

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

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 □

为什么叫 “熵”

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)

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

为什么叫 “熵”

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

Shannon 熵的推广: Rényi 熵

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

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

Shannon 熵的推广: Rényi 熵

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

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

Shannon 熵是 $\alpha \to 1$ 时的特例

信息熵的取值范围

$$0 \leq H_\alpha(\boldsymbol{p}) \leq \log n \tag{6}$$

上限时 $p_i = \text{const.}$ 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)

下限时 $p_i = \delta_{ij}$ 表示没有信息, 因为只有一种可能的信号

信息熵的概念与统计力学

等几率原理 $\Rightarrow$ 热力学关系 $\Rightarrow$ 熵

v.s. 熵 $\Rightarrow$ 统计学关系 $\Rightarrow$ 等几率原理

(Maximum entropy thermodynamics)

信息熵的概念与统计力学

等几率原理 $\Rightarrow$ 热力学关系 $\Rightarrow$ 熵

v.s. 熵 $\Rightarrow$ 统计学关系 $\Rightarrow$ 等几率原理

(Maximum entropy thermodynamics)

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].

图: Edwin Thompson Jaynes (1922–1998)

对于一个量子态 “混合程度” 的度量

用密度矩阵来表示一个态

$$\hat{\rho} = \sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \tag{7}$$

对于一个量子态 “混合程度” 的度量

用密度矩阵来表示一个态

$$\hat{\rho} = \sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \tag{7}$$

Definition

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

$$\hat{\rho} = \sum q_k U_k \hat{\rho}' U_k^\dagger \tag{8}$$

对于一个量子态 “混合程度” 的度量

用密度矩阵来表示一个态

$$\hat{\rho} = \sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \tag{7}$$

Definition

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

$$\hat{\rho} = \sum q_k U_k \hat{\rho}' U_k^\dagger \tag{8}$$

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.

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

majorization criterion 优化准则 I

Theorem

将密度矩阵对角化:

$$\hat{\rho} = \sum_{i=1}^{d} p_i \left|\alpha_i\right\rangle \left\langle\alpha_i\right| \tag{9}$$

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

$$\forall t \in \{1, \cdots, d-1\}, \qquad \sum_{i=1}^{t} p_i \geq \sum_{i=1}^{t} p_i' \tag{10}$$

记作 $\boldsymbol{p} \preceq \boldsymbol{p}'$

majorization criterion 优化准则 II

Proof.

$$\sum_{i=1}^{t} p'_i = \sum_{i=1}^{t} \langle \psi'_i | \hat{\rho}' | \psi'_i \rangle = \sum_k q_k \sum_{j=1}^{d} p_j \sum_{i=1}^{t} | \langle \psi'_i | U_k | \psi_j \rangle |^2$$

$$\leq \sum_k q_k \sum_{j=1}^{t} p_j \sum_{i=1}^{t} | \langle \psi'_i | U_k | \psi_j \rangle |^2$$

$$\left(\text{with } \text{Span}\{U_k |\psi_i\rangle\}_{i=1}^{t} = \text{Span}\{|\psi'_i\rangle\}_{i=1}^{t}\right)$$

$$= \sum_k q_k \sum_{j=1}^{t} p_j = \sum_{j=1}^{t} p_j$$

majorization 的度量

Definition

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

Schur-concave: $-f$ 是 Schur-convex 的

majorization 的度量

Definition

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

Schur-concave: $-f$ 是 Schur-convex 的

Definition (Rényi 熵)

一族的具有 Schur-concave 性质的函数

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

混合程度的度量

直接根据前面的数学定义, 用密度矩阵来表达即为:

Definition (量子 Rényi 熵)

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

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

Rényi 熵的性质

取值范围

$$0 \leq S_\alpha \leq \log d \tag{13}$$

Rényi 熵的性质

取值范围

$$0 \leq S_\alpha \leq \log d \tag{13}$$

广延性

$$S_\alpha(\hat{\rho} \otimes \sigma) = S_\alpha(\hat{\rho}) + S_\alpha(\sigma) \tag{14}$$

Rényi 熵的性质

取值范围

$$0 \leq S_\alpha \leq \log d \tag{13}$$

广延性

$$S_\alpha(\hat{\rho} \otimes \sigma) = S_\alpha(\hat{\rho}) + S_\alpha(\sigma) \tag{14}$$

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

von-Neumann 熵

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

$$S(\hat{\rho}) := \lim_{\alpha \to 1} S_\alpha(\hat{\rho}) = -\operatorname{Tr}[\hat{\rho} \log \hat{\rho}] \tag{15}$$

这与 Gibbs 熵用密度矩阵的表达式是一致的

更“混合”定义的推广

Definition

$\hat{\rho}_1 \in \mathcal{H}_1$ 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

更“混合”定义的推广

Definition

$\hat{\rho}_1 \in \mathcal{H}_1$ 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

补上纯态对于本征值组来说只是增加了一串 0

von-Neumann 熵的特殊性质

特殊的性质, 定义比率 R 为 $\hat{\rho}^{\otimes N}$ 比 $\hat{\rho}'^{\otimes NR}$ 混合程度更高, 则

$$R \leq \frac{S(\hat{\rho})}{S(\hat{\rho}')} \tag{16}$$

从信息熵的观点看粒子分布 I

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

$$\hat{\rho}^{\otimes N} = \sum_{\boldsymbol{m}} q_N(\boldsymbol{m}) \left|\psi_{\boldsymbol{m}}\right\rangle \left\langle\psi_{\boldsymbol{m}}\right| \tag{17}$$

其中定义

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

从信息熵的观点看粒子分布 II

另外定义分布序列

$$t_{\boldsymbol{m}} := (N_1/N, \cdots, N_d/N) \tag{18}$$

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 $Q_{N,t} \sim \exp[-ND(t\|q)]$

Proof.

可能的分布总数量

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

Proof.

某一组分布中态的数量

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

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

Proof.

某一组分布的几率

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

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 $\lim_{N\to\infty} N\epsilon_N \to \infty$ 即可

量子纠缠的度量 (略)

如果定义 “更纠缠” 为 “不可以通过局域操作 (LOCC[2]) 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

Theorem

$|\Psi\rangle\langle\Psi|$ 更纠缠当且仅当 $\mathrm{Tr}_B[|\Psi\rangle\langle\Psi|]$ 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

$\mathrm{Tr}_B[|\Psi^{\pm}\rangle\langle\Psi^{\pm}|] = I/2$

[2] Local operations and classical communication

量子信息的压缩 (略) I

定义量子通道 (quantum channel) $\mathscr{C}(\rho) = \mathrm{Tr}_B[U(\rho \otimes \sigma)U^\dagger]$, 存在一组量子量子通道编码方案 $\mathscr{E} : \mathcal{H}_d \mapsto \mathcal{H}_{d'}$ 与相应的解码方案 $\mathscr{D} : \mathcal{H}_{d'} \mapsto \mathcal{H}_d$ (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

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

量子信息的压缩 (略) II

Theorem (Schumacher's noiseless channel coding theorem)

在 $N \to \infty$ 渐进意义下, 对于 ρ^N 信号的压缩 $\mathscr{E} : \mathcal{H}_{2^N} \mapsto \mathcal{H}_{d_N}$ 以及 $\mathscr{D} : \mathcal{H}_{d_N} \mapsto \mathcal{H}_{2^N}$, 定义压缩率

$$R := \limsup_{N\to\infty} \frac{\log d_N}{N} \tag{19}$$

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

参考文献 I

Jaynes, E. T. (1957a).

Information theory and statistical mechanics.

Phys. Rev., 106:620–630.

Jaynes, E. T. (1957b).

Information theory and statistical mechanics. ii.

Phys. Rev., 108:171–190.

Kullback, S. and Leibler, R. A. (1951).

On information and sufficiency.

The Annals of Mathematical Statistics, pages 79–86.

参考文献 II

Neumann, J. (1932).

Mathematische Grundlagen der Quantenmechanik.

Verlag von Julius Springer Berlin.

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, B. (1995).

Quantum coding.

Phys. Rev. A, 51:2738–2747.

参考文献 III

Shannon, C. E. (1948).

A mathematical theory of communication.

Bell System Technical Journal, 27:379–423.