



链滴

系综

作者: [Roseleaves](#)

原文链接: <https://ld246.com/article/1668268146654>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

对于学过概率论的朋友们：

系综=样本空间

微观状态=样本点

宏观状态=事件

微正则系综：隔绝系统和环境的一切物质和能量交换，其中各种微观状态等概率分布。

如果总共有 g 个微态，那么每个微态的概率为 $w_i = \frac{1}{g}$ 。即 g 为其配分函数。

正则系综：允许系统和环境的能量交换（而且是任何能量的交换，排除光照这种只传递特定能量的情况），不允许系统和环境的物质交换。

记符号 $\langle F_n \rangle = \sum_{i=0}^{\infty} w_i F_i$ ，其中序列 $\{F_n\}$ 的各元素分别表示各微观状态的相应性质，特别地， $\{w_n\}$ 表示概率， $\{\epsilon_n\}$ 表示能量。

根据概率论中的熵的定义 $S = -k \ln p$ 和自由能的定义 $A = U - TS = \langle \epsilon_n + kT \ln w_n \rangle$ 和温度的定义 $T := \left(\frac{\partial S}{\partial U} \right)^{-1} \big|_{w=0}$ 使得 $dA = \delta w - S dT + \mu dN$ ，注意到无功自由换热条件的自发性判据是 $dA \leq 0$ ，故而代入概率的归一化条件 $\langle 1 \rangle = 1$ ，使用拉格朗日乘子法设任意系数 λ ，对每个微态的概率 w_i 求偏导数，分别得到的就是

$$\begin{aligned}$$

$$\frac{\partial}{\partial w_i} \left(\langle \epsilon_n + kT \ln w_n \rangle - \lambda \left(\langle 1 \rangle - 1 \right) \right) = \frac{\partial}{\partial w_i} \left(\epsilon_i + kT \ln w_i - \lambda (w_i - 1) \right) = 0$$

$$+ 0 + \dots$$

$$\epsilon_i + kT (\ln w_i - \lambda) + w_i (kT - \lambda) = 0 \end{aligned}$$

改设 $q := \exp \left(1 - \frac{\lambda}{kT} \right)$ ，得到 $w_i = \frac{1}{q} \exp \left(-\frac{\epsilon_i}{kT} \right)$ ，就是玻尔兹曼分布， q 为其配分函数。

巨正则分布：允许系统和环境的物质交换，物质传递自动携带能量。

定义巨势 $\Phi := U - TS - \mu N$ 使得 $d\Phi = \delta w - S dT - N d\mu$ ，注意到无功传质条件的自发性判据是 $d\Phi \leq 0$ 。用同样的方法得到 $w_i = \frac{1}{\Xi} \exp \left(-\frac{\epsilon_i - \mu N_i}{kT} \right)$ ， Ξ 为其配分函数。

如果设 $\{w_{gsn}\}, \{\epsilon_{gsn}\}, \{N_{gsn}\}$ 分别表示处于具有第 n 少的粒子的，第 s 低的能级的第 g 个简并态上的微观状态的概率、能量和粒子数，那么同具有第 s 少的粒子且同处于第 n 低的能的所有微观状态组成一个微正则系综，而同具有第 s 少的粒子的所有微观状态组成一个正则系综，其分函数满足

$$q_n = \sum_s g_{sn} \exp \left(-\frac{\epsilon_{sn}}{kT} \right)$$

$$\Xi = \sum_n q_n \exp\left(\frac{\mu N_n}{kT}\right)$$

彩蛋：概率论中是有熵这个概念的，而且定义方式与前文我介绍的相同。以下内容选自《大学基础数
自学丛书·概率论与数理统计基础》上科1982年版

4.2 熵

熵是刻画随机试验不确定性的数字特征。

考虑只包含两个基本事件 $A, \bar{A}$; $P(A) = \frac{1}{2}$ 的随机试验, 简单记作

$$\alpha = \begin{pmatrix} A & \bar{A} \\ 0.50 & 0.50 \end{pmatrix}.$$

这个随机试验是有不确定性的, 因为在一次试验中可能 A 出现, 也可能 $\bar{A}$ 出现. 再考虑

$$\beta = \begin{pmatrix} B & \bar{B} \\ 0.99 & 0.01 \end{pmatrix}$$

表示的随机试验, 这个试验也有不确定性, 由于在一次试验中

B “几乎必出现”, 这个随机试验的不确定性小于前一个.

$$\gamma = \begin{pmatrix} C & \bar{C} \\ 0.7 & 0.3 \end{pmatrix}$$

表示的随机试验的不确定性则介于前两个之间. 一般的可以考虑分布列

$$P(\xi = x_i) = p_i, \quad i = 1, 2, \dots, n,$$

表示的随机试验的不确定性. 记

$$A_i = \{\xi = x_i\}, \quad i = 1, 2, \dots, n,$$

这个试验可以表示为

$$\begin{pmatrix} A_1 & A_2 & \cdots & A_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}.$$

我们希望定义一个刻划这种不确定性的数值.

为了定义这个数值, 我们先分析一下这个数值应当满足哪些条件.

1) 这个数仅依赖于 $(p_1, \dots, p_n)$, 而且和 $p_1, \dots, p_n$ 的次序没有关系, 因而是 $p_1, \dots, p_n$ 的对称函数, 记为 $H(p_1, \dots, p_n)$.

2) 当 $p_1, \dots, p_n$ 中有一个是 1 时, 随机试验没有不确定性, 对应的 $H(p_1, \dots, p_n)$ 应当为 0.

3) 当 $p_i (1 \leq i \leq n)$ 只有微小变化时, H 应当也只有微小变化. 所以 $H(p_1, \dots, p_n)$ 对每个自变量来说应当是连续函数.

4) 当 $p_1 = p_2 = \dots = p_n = \frac{1}{n}$ 时, $H(p_1, \dots, p_n)$ 应当有极大值.

5) $H(p_1, \dots, p_n)$ 有可加性. H 的可加性的意义如下:

$$\text{设 } \alpha = \begin{pmatrix} A_1, \dots, A_m \\ p_1, \dots, p_m \end{pmatrix}, \quad \beta = \begin{pmatrix} B_1, \dots, B_n \\ q_1, \dots, q_n \end{pmatrix}$$

是两个随机试验, $p_i > 0, i = 1, 2, \dots, n$. 考虑乘积

$$\alpha\beta = \left(\begin{matrix} A_1B_1, \dots, A_1B_n; A_2B_1, \dots, A_2B_n; \dots; A_mB_1, \dots, A_mB_n \\ r_{11}, \dots, r_{1n}; r_{21}, \dots, r_{2n}; \dots; r_{m1}, \dots, r_{mn} \end{matrix} \right),$$

$$r_{ij} = P(A_iB_j) = P(B_j|A_i)P(A_i) = P(B_j|A_i)p_i,$$

$$i = 1, 2, \dots, m; j = 1, 2, \dots, n.$$

把 $\alpha\beta$ 表示的随机试验看成是经过两步完成的一个试验. 第一步只看 $A_1, A_2, \dots, A_m$ 中哪一个出现, 在这一步试验有不确定性 $H(p_1, \dots, p_m)$, 但还不是整个试验的不确定性 $H(r_{11}, \dots, r_{1n}, \dots, r_{m1}, \dots, r_{mn})$, 因为, 譬如在第一步试验中 A_i 出现, 但整个试验的结果 $A_iB_1, \dots, A_iB_n$ 究竟是哪一个出现还要看第二步试验

$$\left(\begin{matrix} B_1, B_2, \dots, B_n \\ \frac{r_{i1}}{p_i}, \frac{r_{i2}}{p_i}, \dots, \frac{r_{in}}{p_i} \end{matrix} \right).$$

在第一步试验中 A_i 出现的条件下, 第二步试验有不确定性

$$H\left(\frac{r_{i1}}{p_i}, \frac{r_{i2}}{p_i}, \dots, \frac{r_{in}}{p_i}\right),$$

我们取数学期望

$$\sum_{i=1}^m p_i H\left(\frac{r_{i1}}{p_i}, \frac{r_{i2}}{p_i}, \dots, \frac{r_{in}}{p_i}\right)$$

做为第二步试验的不确定性. 要求 H 有可加性, 指的是: 若试验分两步完成, 整个试验的不确定性应当是这两步试验不确定性的和, 也就是要求

$$H(r_{11}, \dots, r_{1n}, \dots, r_{m1}, \dots, r_{mn})$$

$$= H(p_1, \dots, p_m) + \sum_{i=1}^m p_i H\left(\frac{r_{i1}}{p_i}, \dots, \frac{r_{in}}{p_i}\right).$$

以上讲的是希望刻划随机试验不确定性的函数 $H(p_1, \dots, p_n)$ 应当满足的条件. 问题是满足这样多的条件的函数

有没有呢？回答是肯定的，例如函数

$$h(p_1, \dots, p_n) = -\sum_{i=1}^n p_i \log p_i$$

就满足上面的条件（当 $p_i=0$ 时约定 $p_i \log p_i=0$ ，其中的对数为常用对数）。显然 $h(p_1, \dots, p_n)$ 是 $p_1, \dots, p_n$ 的对称函数、连续函数，而且某一个 $p_i=1$ 时 $h=0$ 。其次

$$h\left(\frac{1}{n}, \dots, \frac{1}{n}\right) = -\sum_{i=1}^n \frac{1}{n} \log \frac{1}{n} = \log n. \quad (1)$$

$g(x) = x \log x$ 是（下）凸函数，因而有

$$g\left(\frac{1}{n} \sum_{i=1}^n p_i\right) \leq \frac{1}{n} \sum_{i=1}^n g(p_i).$$

注意到 $\sum_{i=1}^n p_i = 1$ ，上式即

$$\begin{aligned} \frac{1}{n} \log \frac{1}{n} &\leq \frac{1}{n} \sum_{i=1}^n p_i \log p_i, \\ \log n &\geq -\sum_{i=1}^n p_i \log p_i. \end{aligned}$$

和(1)式比较得

$$h\left(\frac{1}{n}, \dots, \frac{1}{n}\right) \geq -\sum_{i=1}^n p_i \log p_i.$$

最后，若将试验

$$\alpha\beta = \begin{pmatrix} A_1B_1, \dots, A_1B_n; \dots; A_mB_1, \dots, A_mB_n \\ r_{11}, \dots, r_{1n}; \dots; r_{m1}, \dots, r_{mn} \end{pmatrix}$$

分为两步，第一步试验为

$$\begin{aligned} \alpha &= \begin{pmatrix} A_1, A_2, \dots, A_m \\ p_1, p_2, \dots, p_m \end{pmatrix}, \\ p_i &= \sum_{j=1}^n r_{ij} > 0, \quad i=1, 2, \dots, m, \end{aligned}$$

$$\begin{aligned}
\text{则 } h(r_{11}, \dots, r_{1n}, \dots, r_{m1}, \dots, r_{mn}) &= - \sum_{i=1}^m \sum_{j=1}^n r_{ij} \log r_{ij} \\
&= - \sum_{i=1}^m \sum_{j=1}^n r_{ij} \left(\log p_i + \log \frac{r_{ij}}{p_i} \right) \\
&= - \sum_{i=1}^m \left(\sum_{j=1}^n r_{ij} \right) \log p_i - \sum_{i=1}^m p_i \sum_{j=1}^n \frac{r_{ij}}{p_i} \log \frac{r_{ij}}{p_i} \\
&= - \sum_{i=1}^m p_i \log p_i + \sum_{i=1}^m p_i h \left(\frac{r_{i1}}{p_i}, \dots, \frac{r_{in}}{p_i} \right) \\
&= h(p_1, \dots, p_m) + \sum_{i=1}^m p_i h \left(\frac{r_{i1}}{p_i}, \dots, \frac{r_{in}}{p_i} \right).
\end{aligned}$$

定义 在具有离散分布

$$\alpha = \begin{pmatrix} A_1 & A_2 & \cdots & A_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}$$

的随机试验中, 称

$$H(p_1, p_2, \dots, p_n) = - \sum_{i=1}^n p_i \log p_i$$

为 α 的熵.

其中 $\log$ 为常用对数. 又若某一 p_i 为 0, 约定 $p_i \log p_i$ 为 0.

例如,

$$\begin{aligned}
H(0.99, 0.01) &= -(0.99 \times \log 0.99 + 0.01 \times \log 0.01) \\
&= 0.0243,
\end{aligned}$$

$$\begin{aligned}
H(0.70, 0.30) &= -(0.70 \times \log 0.70 + 0.30 \times \log 0.30) \\
&= 0.2653,
\end{aligned}$$

$$H(0.50, 0.50) = 0.3010.$$

以上我们对只取有限个值的离散分布定义了熵, 并讨论了熵的性质. 可以证明: 满足条件 1~5 的 H 的形式是唯一的, 并且可以在更弱的条件下被确定. 对连续分布也可以定义熵.