情報エントロピーと統計力学（講義ノートのページ「統計力学」講義資料に加筆：2016 年 1 月）

シャノン・エントロピー

$$I = -\sum_i P_i \log P_i = -\overline{\log P_i} \tag{1}$$

をカノニカル分布に適用する際に量子統計力学では，状態 i は縮退があっても通し番号をつけておけば問題なしに,「エネルギー固有状態 E_i」を用いて,

$$P_i = \frac{1}{Z}e^{-\beta E_i} \ , \quad Z = \sum_i \mathrm{e}^{-\beta E_i} \ , \quad -\log P_i = \beta E_i + \log Z \tag{2}$$

とすれば

$$I = -\overline{\log P_i} = \beta \overline{E_i} + \log Z = \frac{U-F}{kT} = \frac{S}{k} \qquad (U = \overline{E},\ F = -kT \log Z) \tag{3}$$

の関係が得られる。あるいは，密度行列を用いて一般的に

$$S = -k\mathrm{Tr}\ \rho \log \rho \ , \quad \rho = \frac{1}{Z}\mathrm{e}^{-\beta H} \ , \quad Z = \mathrm{Tr}\ \mathrm{e}^{-\beta H} \tag{4}$$

と書けば，縮退のことをいちいち断る必要はない。

カノニカル確率分布　しかしながら多体系を扱う場合，カノニカル分布で扱われる状態の確率としては，エネルギー状態の縮重度 $G(E)$ を用いて

$$P_E = \frac{1}{Z}\ G(E)\ \mathrm{e}^{-\beta E} \ , \quad Z = \sum_E G(E)\ \mathrm{e}^{-\beta E} \tag{5}$$

とすることが普通であり，古典統計力学では，縮重度として適当な厚さをもたせた位相空間の等エネルギー殻の体積（h^{3N} で割ったもの，必要ならギブス補正 $1/N!$）を用いる。実際，$G(E)$ が急激に増加する関数，$\mathrm{e}^{-\beta E}$ はその逆として中心極限定理が適用されるのは，確率分布をこの形に書く場合である。しかしながら，これをシャノン・エントロピーに適用すると,

$$-\log P_E = \beta E + \log Z - \log G(E) \tag{6}$$

$$\mathcal{I} = -\overline{\log P_E} = \frac{U-F}{kT} - \overline{\log G(E)} \tag{7}$$

となる。第 1 項は先ほどと同じくカノニカル分布で得られるエントロピー S_{C} である。第 2 項は，$N \gg 1$ のとき中心極限定理により，カノニカル分布がエネルギー E の関数として，平均値 $\overline{E}$ の位置の鋭いピークになる場合，これに比べれば $\log G(E)$ は緩やかな関数[1]であることから，ピーク位置での状態数 $G(\overline{E})$ で（N^{-1} 程度でほぼ正確に）置き換えることができ，ミクロカノニカル分布のエントロピー

$$\overline{\log G(E)} \simeq \log G(\overline{E}) = \frac{S_{\mathrm{MC}}}{k} \tag{8}$$

とみなすことができる。したがって，以下の関係があることになる：

$$\mathcal{I} = \frac{S_{\mathrm{C}} - S_{\mathrm{MC}}}{k} \tag{9}$$

元の「統計力学」講義資料で正規分布（中心極限定理）を用いて評価を行っているが，端的に言えば，カノニカル分布は**熱容量の相加性**から **$\sqrt{N}$ 程度の広がり**をもつ分だけエントロピーが大きくなり，$\boldsymbol{\mathcal{I} \simeq}\ \log \boldsymbol{\sqrt{N}}$ と推察できる。

[1] $\log G$ は N の程度の量であり，平均値からの「はずれ」を Nx として展開すれば，中心極限定理的には，$\overline{x} = 0,\ N\overline{x^2} \sim 1$ である。この 1 の程度の誤差は，後で得られるエントロピー差 $\log \sqrt{N}$ に比べれば十分に小さい。

連続分布　古典統計ではエネルギーを**連続変数**として扱う。連続変数 x の確率分布関数が $p(x)\mathrm{d}x$ で与えられるとき

$$-\overline{\log p(x)} = -\int p(x)\log p(x)\mathrm{d}x \tag{10}$$

をエントロピーとみなすべきではない。変数のスケール変換「$y=\lambda x,\ \mathrm{d}y=\lambda\mathrm{d}x$」を行うと，確率分布関数は確率の保存則により

$$q(y)\mathrm{d}y = p(x)\mathrm{d}x \quad \rightarrow \quad q(y)=\lambda^{-1}p(y/\lambda)\ ,\quad \log q(y)=\log p(y/\lambda)-\log\lambda \tag{11}$$

でなければならないから

$$\overline{\log q(y)} = \int q(y)\log q(y)\mathrm{d}y = \int \lambda^{-1}p(y/\lambda)\log p(y/\lambda)\mathrm{d}y - \log\lambda = \overline{\log p(x)} - \log\lambda \tag{12}$$

となる。もともとエントロピーはエネルギー準位の構造で決まる量であり，変数のスケール変換（いわば断熱変化）では不変でなければならない。この曖昧さを避けるため，変数 x を幅 Δx で離散化し

$$P_x = \int_{x-\Delta x/2}^{x+\Delta x/2} p(x)\mathrm{d}x \simeq p(x)\Delta x \tag{13}$$

に対して

$$\mathcal{I} = -\overline{\log P_x} = -\sum_x P_x\log P_x = -\overline{\log p(x)} - \log\Delta x \tag{14}$$

とする。スケール変換「$y=\lambda x$」を行った場合は,「$\Delta y=\lambda\Delta x$」ととれば，(12) により

$$\overline{\log q(y)} + \log\Delta y = \overline{\log p(x)} + \log\Delta x \tag{15}$$

となり，エントロピーは不変である。

したがって，シャノン・エントロピーの N 依存性を正しく評価しようと思えば，変数を $x=E/N$ とスケール変換して平均値を 1 の程度の量に揃えて，$\Delta x \propto 1/N$ とすればよい。

中心極限定理　N 粒子系の各粒子のエネルギー $\mathcal{E}_i$ を独立な確率変数とみなせば，その和で与えられる確率変数

$$E = \mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3 + \ \ldots\ + \mathcal{E}_N \tag{16}$$

は，$N \gg 1$ のとき中心極限定理に従い，確率分布の主要な部分は正規分布になる。これをエネルギー E についての連続な確率分布とみるとき，さきほど見たようにシャノン・エントロピーを曖昧なく定義するためには離散化の幅 ΔE — どれだけの幅にエネルギー準位が 1 つ対応するか？ — を導入しなければならない。

量子力学の場合は各粒子のエネルギー構造には代表的な単位量 ϵ があり，独立な確率変数の和 E においても，N には無関係に確率的跳躍の単位「ランダムウォークの歩幅」は ϵ とみなすことができる。つまり,「跳躍回数」N が何倍になっても「歩幅」$\Delta E \simeq \epsilon$ は共通であり，スケール変換 $x=E/N$ をした場合は，上の約束により「$\Delta x \propto N^{-1}$」とすればよい。

$\mathcal{E}_i$ を連続とみなす古典的な場合でも，その確率的な揺らぎ幅が有限な値であれば，それを確率変数 E の特徴的「酔歩幅」であるとみなすことができよう。$N \gg 1$ として得られる正規分布の分散からは，結果的に kT 程度の酔歩幅 — 多少のふらつきはあろうが — のランダムウォークと見なした構造になっており，やはり「酔歩回数」N には依存しない量 と考えることができる。カノニカル分布とミクロカノニカル分布の差を評価するだけであるから，微視的な物理的根拠を求める必要はない。その情報（エネルギー準位構造）は既に微視的状態数としてミクロカノニカル分布のエントロピーに取りこまれている。

以下，統計力学講義資料「中心極限定理と平均エントロピー」に続く。2006 年当時から 10 年の隔たりがあるため多少は記号や説明の混乱・矛盾・重複あり。

中心極限定理と平均エントロピー

2 項分布　確率 1/2 で 0, 1 の値をとる，N 個の独立な確率変数の和から成る確率変数

$$X = \sum_{i=1}^{N} X_i \tag{1}$$

を考える。$X = M\ (\leq N)$ となる状態の数は

$$G(M) = \frac{N!}{M!(N-M)!}\ , \quad \sum_{M=0}^{N} G(M) = 2^N \tag{2}$$

であり，確率は 2 項分布

$$P(M) = \frac{1}{2^N}\frac{N!}{M!(N-M)!} \tag{3}$$

に従う。$G_{\max} = G(N/2)$ であり，Stirling の公式を用いれば，中心極限定理の原形

$$\frac{\log G_{\max}}{N\log 2} = 1 - \left(\frac{\log N}{2N\log 2}\ \text{程度の量}\right) \tag{4}$$

を示すことができる。母関数の方法により

$$\overline{X} = \frac{N}{2}\ , \quad \overline{X^2} - \overline{X}^2 = \overline{(X-\overline{X})^2} = \frac{N}{4} \tag{5}$$

あるいは，確率変数を $x = (X - N/2)/N$ とすれば以下が得られる：

$$\overline{x} = 0\ , \quad \sqrt{\overline{x^2}} = \frac{1}{2\sqrt{N}} \tag{6}$$

$N \gg 1$ のとき x を連続変数とみなせば，確率分布は正規分布に近づく（**中心極限定理**）:

$$p(x)\mathrm{d}x = \sqrt{\frac{2N}{\pi}}\ \mathrm{e}^{-2Nx^2}\mathrm{d}x \tag{7}$$

（**注．**酔歩幅 ±1 のランダムウォークなら，$x/2$ が今の場合の x であり，$\sqrt{\overline{x^2}} = 1/\sqrt{N}$ となる。）

平均情報エントロピー　連続な確率分布の場合にエントロピーを定義するためには，先ず分布の特徴的な形を潰してしまわない程度に小さな幅 Δx を導入して離散化[1]する。確率を

$$P_x = \int_{x-\Delta x/2}^{x+\Delta x/2} p(x)\mathrm{d}x \simeq p(x)\Delta \mathrm{x} \tag{8}$$

とし，幅内は状態を同一視する階段状分布にした上で

$$\Delta S = -\sum_x P_x \log P_x = -\overline{\log p(x)} - \log \Delta x \tag{9}$$

としなければならず，Δx の取り方によってエントロピーの基準値が変わる。上の正規分布で計算すれば

$$\Delta S = \frac{1}{2} - \frac{1}{2}\log\frac{2N}{\pi} - \log \Delta x \tag{10}$$

[1] もともと離散的な問題であるが，このように近似的に計算するしかない。大数の法則により，こうすることで普遍的になる。

となる。Δx は，(1) 確率が激しく変化する揺らぎの幅 $N^{-1/2}$（$\sim$ **半値幅**）よりは十分小さく，(2) x を連続変数として扱う以上，N^{-1} より小さくなることは意味がなく，N^{-1} の**数倍程度**には大きくなければならない。（最初の 2 項分布の場合は，N^{-1} に 1 状態だから，$\Delta x = N^{-1}$。）すなわち

$$N^{-1/2} \gg \Delta x \sim N^{-1}\text{の数倍程度} \tag{11}$$

したがって

$$\frac{1}{2}\log N < -\log \Delta x \simeq \log N - \log a \quad (\ a \text{ は 1 程度の量}) \tag{12}$$

とみなしてよく，いずれにせよ $\log N \gg 1$ であれば以下のように見積もられる：

$$\Delta S \simeq \frac{1}{2}\log N \tag{13}$$

一方，**ミクロカノニカル分布**の考えでは，平均値 $x = 0$ を中心とする Δx の中に**全確率 1 を集中**させ，この幅の中では状態を**同一視**するから，上と同じ意味でのエントロピーは 0 である。したがって，ミクロカノニカルの場合のエントロピーを基準として，広がりをもつ正規分布では，エントロピーが上で求めた ΔS だけ多いことになる。中央に集中させた全微視状態の数から計算される N 粒子系のエントロピーは N の程度の量（今の例では $N\log 2$）であるから，エントロピーの差の相対比 $\boldsymbol{\log N/N}$ が十分に小さいなら，2 つの分布は等価であるとみなせるだろう。

カノニカル分布の中心極限定理

$$P(E)\mathrm{d}E = \frac{1}{Z(\beta)}G(E)\mathrm{e}^{-\beta E}\mathrm{d}E\ , \quad Z(\beta) = \int_0^\infty G(E)\mathrm{e}^{-\beta E}\mathrm{d}E \tag{14}$$

$$\overline{E} = -\frac{1}{Z}\frac{\partial Z}{\partial \beta} = -\frac{\partial}{\partial \beta}\log Z \tag{15}$$

より，熱容量が

$$C_V = \left(\frac{\partial \overline{E}}{\partial T}\right)_V = -\frac{1}{kT^2}\frac{\partial \overline{E}}{\partial \beta} = \frac{\overline{E^2} - \overline{E}^2}{kT^2} \tag{16}$$

で与えられる。したがって，1 分子あたりの比熱を ck（c は一般には定数ではない。臨界現象などを除き，普段は 1 程度の量，例えば固体では $c = 3$）として

$$\overline{(E - \overline{E})^2} = Nck^2T^2 \tag{17}$$

が得られる。

$$x = \frac{E - \overline{E}}{NckT} \tag{18}$$

とするとき

$$\sqrt{\overline{x^2}} = (cN)^{-1/2} \tag{19}$$

したがって，$N \gg 1$ のときカノニカル分布の**主要な部分**は，正規分布

$$P(E)\mathrm{d}E \sim p(x)\mathrm{d}x = \sqrt{\frac{cN}{2\pi}}\exp\left(-\frac{cN}{2}x^2\right)\mathrm{d}x \tag{20}$$

で近似される。このとき

$$-\overline{\log p(x)} = \frac{1}{2} - \frac{1}{2}\log\frac{cN}{2\pi} \tag{21}$$

となり，先ほどと同じ結論 $\Delta S \simeq \log N/2$ が得られる。

例：多準位系　（2 準位系は 2 項分布と変わりなく，芸がないので変更）

エネルギー $0,\ \epsilon,\ 2\epsilon,\ 3\epsilon, \ldots$ をとる原子 N 個からなる古典系を考える。簡単のため，原子の入れ替えの問題に悩まずにすむ固体としておく。エネルギーが $E = M\epsilon$ となる状態の確率は

$$P(M) = \frac{1}{Z} G(M)\, \mathrm{e}^{-\beta\epsilon M}\ , \quad G(M) = \frac{(N+M-1)!}{M!(N-1)!}\ , \quad Z = (1-\mathrm{e}^{-\beta\epsilon})^{-N} \tag{22}$$

である。Stirling の公式を用いて（簡素化のため，$N-1 \sim N$ とおく。）

$$\log G(M) \simeq (N+M)(\log(N+M)-1) - M(\log M - 1) - N(\log N - 1) \tag{23}$$

とし，$\log P(M)$ を最大とする M を求める：

$$[\log P(M)]' = \log(N+M) - \log M - \beta\epsilon = 0 \tag{24}$$

より

$$\frac{N+M}{M} = \mathrm{e}^{\beta\epsilon}\ , \quad \overline{M} = \frac{N}{\mathrm{e}^{\beta\epsilon}-1} \tag{25}$$

となる。さらに，$\log P(M)$ の 2 階微分を求めれば，

$$[\log P(M)]'' = \frac{1}{N+M} - \frac{1}{M}\ , \quad [(\log P(M))'']_{\overline{M}} = -\frac{4}{N}\sinh^2\frac{\beta\epsilon}{2} \tag{26}$$

したがって

$$P(M) \sim \exp\left[-\frac{2\sinh^2(\beta\epsilon/2)}{N}(M-\overline{M})^2\right] \tag{27}$$

となり，$\overline{M}$, $\overline{(M-\overline{M})^2}$ ともに，母関数（分配関数）から計算したものと正確に一致する。

一方，熱容量を求めると

$$C = \frac{\mathrm{d}\overline{E}}{\mathrm{d}T} = \frac{N\epsilon^2}{4kT^2\sinh^2(\beta\epsilon/2)} = Nk\left(\frac{\beta\epsilon/2}{\sinh(\beta\epsilon/2)}\right)^2 \quad (\ \propto N\ ) \tag{28}$$

となり，$C = Nck$ とおいてエネルギーについての確率分布で書けば

$$P(E) \sim \exp\left[-\frac{1}{2Nc}\left(\frac{E-\overline{E}}{kT}\right)^2\right] \tag{29}$$

が得られる（高温では $c \simeq 1$）。この場合はエネルギーの最小単位が明確なので

$$x = \frac{E-\overline{E}}{N\epsilon} \tag{30}$$

とする方が扱いやすく，

$$p(x)\mathrm{d}x \simeq \sqrt{\frac{\gamma N}{2\pi}}\exp\left[-\frac{\gamma N}{2}x^2\right]\mathrm{d}x\ , \quad \gamma = \frac{(\beta\epsilon)^2}{c} = 4\sinh^2\frac{\beta\epsilon}{2} \tag{31}$$

となる。エネルギー状態は間隔 ϵ に 1 つだから，2 項分布のときと同じく $\Delta x = N^{-1}$（あるいはその数倍でも影響はない）にとればよく，$\Delta S \simeq \log N/2$ となる。