

情報エントロピーとの関係

各アンサンブルでの分布関数が情報エントロピーと関係していることを見ます。情報エントロピーは情報理論で使われるものなので、通常は統計力学で使われるものではないです。しかし、ここで見るように各アンサンブルと関係しており、情報エントロピーから分布関数を構築することもできます。

まず、各アンサンブルでの分布関数を見直しておきます。位相空間における状態の分布 (確率密度) を表す分布関数を $f(p, q, t)$ とします。 q は一般化座標、 p はその共役量、 t は時間です。 p と q は $p = (p_1, p_2, \dots, p_N)$ 、 $q = (q_1, q_2, \dots, q_N)$ と略したものです。 N は粒子数です。これは、位相空間全体にわたる積分で規格化して

$$\frac{1}{N!(2\pi\hbar)^{3N}} \int f(p, q, t) dpdq = \int f(p, q, t) d\Gamma = 1$$

$d^3p_1 d^3p_2 \dots d^3p_N, d^3q_1 d^3q_2 \dots d^3q_N$ を省略して dp, dq と書いています。 $N!(2\pi\hbar)^{3N}$ は「ミクロカノニカルアンサンブル」での話と同じ理由でつけています。この分布関数の対数 $\log f$ の平均値

$$S = - \int f(p, q, t) \log f(p, q, t) d\Gamma$$

をギブスのエントロピーと言います。このエントロピーは、この式からはすぐには分かりませんが、時間独立であることに注意してください。なぜなら、熱力学によって孤立系ではエントロピーが増加するとされているからです。

時間独立であることは、リウヴィルの定理から位相空間での分布関数が

$$f(p, q, t) = f(p', q', t') \tag{1}$$

という関係を持っているのを利用すれば分かります。リウヴィルの定理は、正準方程式 (運動方程式) に従って運動しているとき、位相空間においてある時間 t での微小体積 $dpdq$ とそこから時間経過した時間 t' での $dp'dq'$ は等しいという定理で (それぞれの領域の形が変わったとしても)

$$dpdq = dp'dq' \tag{2}$$

という関係を持っていることです (証明は省きます)。つまり、リウヴィルの定理は、時間 t での位相空間の領域 R 内の点は、時間 t' で R' 内に移り (運動方程式に従って) これらの領域の体積は等しいと言っています。このことから、時間 t の微小体積を $dpdq$ 、時間 t' の微小体積を $dp'dq'$ としたとき、 $dpdq$ に含まれる点は、 $dp'dq'$ に移ります。なので、点の数は変わらないことから分布関数によって

$$f(p, q, t) dpdq = f(p', q', t') dp'dq'$$

となり、(1) が出てきます。ちなみに、 f の時間変化を考えると

$$\frac{\partial f}{\partial t} = \{H, f\}$$

となっていて、この式は同じ意味を持っていてリウヴィル方程式と呼ばれます。 $\{ \}$ はポアソン括弧、 H はハミルトニアンです。(1) で $t' = t + dt$ として $f(p + \dot{p}dt, q + \dot{q}dt, t + dt)$ を展開すれば出てきます (詳しいことは「リウヴィル方程式」参照)。この式をリウヴィルの定理と言う場合もあります。リウヴィルの定理は解析力学の範疇で証明できますが、統計力学の方が出会いやすいです。(2) と (??) を合わせることでギブスのエントロピーは

$$\begin{aligned} S &= - \frac{1}{N!(2\pi\hbar)^{3N}} \int f(p, q, t) \log f(p, q, t) dpdq \\ &= - \frac{1}{N!(2\pi\hbar)^{3N}} \int f(p', q', t') \log f(p', q', t') dp' dq' \\ &= - \frac{1}{N!(2\pi\hbar)^{3N}} \int f(p', q', t') \log f(p', q', t') dpdq \end{aligned}$$

このときに $t' = 0$ とすれば、 S は時間に依存しなくなるので、ギブスのエントロピーは時間に対して不変です。しかし、熱平衡状態ではエントロピーは一定であるために熱平衡では何の問題もないので、これについての問題は無視することにします。

熱平衡でのミクロカノニカルアンサンブルを考えてみます。ミクロカノニカルアンサンブルはエネルギーが E と $E + \Delta E$ の間の状態数 W がエネルギー E 、粒子数 N 、体積 V によって決められ、位相積分によって状態数は

$$W(E, N, V) = \int_{E \sim E + \Delta E} d\Gamma$$

で与えられます。そして、等確率の原理から分布関数 $f(p, q)$ は

$$f(p, q) = \frac{1}{W(E, N, V)} \quad (E \sim E + \Delta E)$$

$E \sim E + \Delta E$ の外では $f(p, q) = 0$ です。時間変化はしないので t を省いています。これをギブスのエントロピーに入れると

$$\begin{aligned} S &= - \int f(p, q) \log f(p, q) d\Gamma \\ &= \int_{E \sim E + \Delta E} \frac{1}{W(E, N, V)} \log W(E, N, V) d\Gamma \\ &= \frac{1}{W(E, N, V)} \log W(E, N, V) \int_{E \sim E + \Delta E} d\Gamma \\ &= \log W(E, N, V) \end{aligned}$$

このようにボルツマン定数 k_B はいませんが、状態数によるエントロピーの式になります。カノニカルアンサンブルでの分布関数は「カノニカルアンサンブル」での P_n で

$$P_n = Z^{-1} \exp[-\beta E_n], \quad Z = \sum_i \exp[-\beta E_i]$$

と与えられています。これを位相空間で書き直せば

$$f(p, q) = Z^{-1} \exp[-\beta H(p, q)], \quad Z = \int \exp[-\beta H(p, q)] d\Gamma$$

$\beta = 1/k_B T$ で、 $H(p, q)$ はハミルトニアンです。位相空間上であることと、ミクロカノニカルでの与えられたエネルギー E と区別するために $H(p, q)$ を使います。

グランドカノニカルでは

$$P_n = \frac{\exp[-\beta(E_n - \mu N_n)]}{\Xi(T, V, \mu)}, \quad \Xi(T, V, \mu) = \sum_m \sum_n \exp[-\beta(E_{mn} - \mu N_n)]$$

E_{mn} は粒子数 N_n の状態における位相空間上でのエネルギー E_m という意味です。 E_m が位相空間上で $H(p, q)$ としたとき、分布関数は

$$f_n(p, q) = \frac{\exp[-\beta(H(p, q) - \mu N_n)]}{\Xi}, \quad \Xi = \sum_n \int \exp[-\beta(H(p, q) - \mu N_n)] d\Gamma_n$$

となります。エネルギーと粒子がいるので分布関数の規格化は

$$\sum_n \int f_n(p, q) d\Gamma_n = \sum_n \frac{1}{n!(2\pi\hbar)^{3n}} \int f_n d^3 p_1 d^3 q_1 \cdots d^3 p_n d^3 q_n = 1$$

となり、平均は

$$\langle A \rangle = \sum_n \int A(p, q) f_n(p, q) d\Gamma_n$$

と与えられます。当たり前ですが、 $n \geq 0$ です。

ここで情報エントロピー (information entropy) と呼ばれるものを持ち出します。情報エントロピーは何かの確率 p_i (離散的) によって

$$S_I = - \sum_i p_i \log p_i, \quad \sum_i p_i = 1$$

と定義されます。確率は1以下なので S_I が正の値になるようにマイナスを付けています。情報エントロピーは H で表されることが多いですが、ハミルトニアンと区別するために S_I としています。情報エントロピーはシャノン (Shannon) エントロピーとも言い、特に対数の底を2としたもの (\log_2) をさすときにシャノンエントロピーと言うことが多いです。情報理論はシャノン1人によってほぼ作られました。

ちなみに、 $-\log p_i$ は情報量、シャノンエントロピー $-p_i \log p_i$ は情報量の平均の形になっているために平均情報量と呼ばれ区別されています。この2つは等確率 $p_1 = p_2 = \cdots$ なら同じになります。

式を見てすぐに分かるように、 p_i の内のどれかが1で他が0というときに $S_I = 0$ となっています。最大となるときは、 $-p \log p$ ($0 \leq p \leq 1$) が上に凸の関数 (凹関数) であることから、凹関数 $F(x_i)$ の平均の取り方に対する性質

$$\frac{1}{N} \sum_{i=1}^N F(x_i) \leq F\left(\frac{1}{N} \sum_{i=1}^N x_i\right)$$

を利用すれば分かります。これから

$$S_I(p_1, \dots, p_N) = - \sum_{i=1}^N p_i \log p_i = \sum_{i=1}^N F(p_i) \leq NF\left(\frac{1}{N} \sum_{i=1}^N p_i\right) = NF\left(\frac{1}{N}\right) = -N\left(\frac{1}{N} \log \frac{1}{N}\right) = S_I\left(\frac{1}{N}, \frac{1}{N}, \dots\right)$$

この関係は p_i が全て同じときに情報エントロピーは最大になることを示しています。というわけで、情報エントロピーは得られる情報が確定しているときに最小になり、すべての確率が同じときに最大になります。これは感覚的に分かりやすい性質で、確率が低いほうが情報量は多くなります（よく起きることよりたまにしか起きないことの方が情報量が多い）。

不確かさを表すのが情報エントロピーと言うのも分かりやすい言い回しです。これは全ての確率が同じときが最も不確かで、確率に偏りがあると不確かさが減るというのに対応します（表裏が確率 $1/2$ で出るコインではどちらが出るか分からないが、表が $3/4$ 、裏が $1/4$ で出るコインでは表が出やすくなる）。また、情報エントロピーをエントロピーと対応させて乱雑さを表すとするのは、意味合いが不鮮明になるので止めたほうがいいと思います。

情報エントロピーの単位についても簡単に触れておきます。情報エントロピーは対数で与えられているので、その底をどうするかで単位が変更されます。底が 2 のときをビット (bit)、 e のときをナット (nat) と呼びます。底の変更は、情報エントロピー S_I の底を 2 にしたものを S_2 とすると、底の変換式

$$\log_a b = \frac{\log_c b}{\log_c a}$$

から

$$S_2 = - \sum_i p_i \log_2 p_i = - \sum_i p_i \frac{\log p_i}{\log 2} = \frac{S_I}{\log 2}$$

ここで、起こる事象は 2 個しかなく、その確率は等しいとします。そうすると $p_1 = p_2 = 1/2$ なので

$$S_I = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} = \log 2$$

となり、これは全ての確率が等しいことから S_I の最大値なので、 S_I の範囲は $0 \leq S_I \leq \log 2$ となります。そうすると、これを S_2 の式に入れれば S_2 の範囲は

$$0 \leq S_2 \leq 1$$

となります。この話から、 $0, 1$ の 2 通りによるデータ量の単位がビットと呼ばれることとの対応が分かると思います。

話が逸れますがついでなので、例として、使える色が白黒のみの画像ファイル (ビットマップ形式) のファイルサイズを計算してみます。画像は 100×100 (ピクセル) のサイズとします。今使える色は白か黒なので、 1 ピクセルに入る色の確率は $1/2$ です。このため、 1 ピクセルが持つ情報量は

$$-\log_2 2^{-1} = 1 \text{ (ビット)}$$

となります (等確率なので平均情報量と情報量で同じになる)。後はこれが 100×100 あるので、ファイルサイズは 10000 ビットとなります。 1 バイトは 8 ビットのことなので 8 で割れば、 1250 バイトとなります。実際にはこ

れにヘッダサイズとかが加わるので、1250 バイトよりは大きくなります。このように、画像を作るのに必要な要素の確率の対数を取ることでファイルサイズは決まります (画像に何が描いてあるかは関係ない)。また、1 ピクセルが持つ情報量は、16 色なら確率は $1/16 = 2^{-4}$ なので 4 ビット、256 色なら $1/256 = 2^{-8}$ なので 8 ビットとなります。これから、色の選択に幅があるときのほうが確率が低くなり情報量が多くなるのが分かると思います。情報エントロピーについてはこれで終わりにします。

情報エントロピーは位相空間における連続分布に対して

$$S_I = - \int f \log f d\Gamma$$

となっているので、ギブスのエントロピーは情報エントロピーです。このように対応しているので、各アンサンブルの分布関数に対して情報エントロピーがどうなっているのか求めてみます。

ミクロカノニカルアンサンブルでの分布関数を f 、任意の分布関数を f' とします (両方とも同じエネルギー面で定義されている)。単純な不等式

$$\log x + \frac{1}{x} \geq 1 \quad (x > 0)$$

を使います。これは、 $x = 1$ から $x = 0$ に行くとき $1/x$ のほうが $\log x$ よりも急激に発散し、 $x \sim 1$ のときには

$$\log x = \log(1 - \epsilon) \simeq -\epsilon - \frac{1}{2}\epsilon^2 - \dots, \quad \frac{1-x}{x} = \frac{1-(1-\epsilon)}{1-\epsilon} = \epsilon(1+\epsilon-\dots) \quad (\epsilon \ll 1)$$

であることを踏まえれば $0 < x < 1$ のとき $\log x + 1/x - 1 > 0$ となり、 $x = 1$ のときには等号になり、 $x > 1$ では $\log x + 1/x - 1$ が増加関数であるために $\log x + 1/x - 1 > 0$ となるからです。これに $x = f'/f$ (確率分布なので $f, f' > 0$) とし入れると

$$\begin{aligned} \log \frac{f'}{f} + \frac{f}{f'} &\geq 1 \\ f' \log \frac{f'}{f} &\geq f' - f \\ \int f' \log \frac{f'}{f} d\Gamma &\geq \int f' d\Gamma - \int f d\Gamma \\ \int f' \log \frac{f'}{f} d\Gamma &\geq 0 \end{aligned} \tag{3}$$

という形が出てきます (f と f' は全位相空間積分で 1 に規格化)。これに f がミクロカノニカルであることを入れると

$$- \int f' \log f' d\Gamma \leq - \int f' \log f d\Gamma = - \int f \log f d\Gamma$$

最右辺に行くときに、ミクロカノニカルの分布関数はエネルギー $E \sim E + \Delta E$ の範囲内での状態数 W の逆数であるために位相積分に引っかけられないことと、 f, f' の規格化が同じであることを使っています。 f' は任意の分布関数なので、ミクロカノニカルの分布関数による情報エントロピーが最大になっています。これは等確率の原理から予想できる結果です。

次に情報エントロピーに対して平均エネルギーが分布関数 $f(p, q)$ によって

$$\langle E \rangle = \int H(p, q) f(p, q) d\Gamma$$

と与えられている場合を考えてみます。このような条件付きでの極値はラグランジュの未定乗数法 (解析力学の「拘束条件」参照) で求められるので、情報エントロピーの極値は

$$S'_I = S_I + \lambda\phi, \quad dS'_I = dS_I + \lambda d\phi = 0$$

によって求められます。条件 ϕ は今の場合は

$$\lambda\phi = \lambda_1 \left(\langle E \rangle - \int H f d\Gamma \right) + \lambda_2 \left(1 - \int f d\Gamma \right)$$

となっています。規格化の式も拘束条件になっています。変分を計算すると

$$dS_I = \frac{dS_I}{df} df = -df \int d\Gamma (\log f + 1)$$

$$\lambda d\phi = -\lambda_1 df \int H d\Gamma - \lambda_2 df \int d\Gamma$$

なので

$$\left(- \int d\Gamma (\log f + 1) - \lambda_1 \int H d\Gamma - \lambda_2 \int d\Gamma \right) df = 0$$

$$- \int d\Gamma \log f - \lambda_1 \int H d\Gamma - \int d\Gamma - \lambda_2 \int d\Gamma =$$

$$\log f + \lambda_1 H + 1 + \lambda_2 =$$

$$f = \exp[-\lambda_1 H - 1 - \lambda_2]$$

そして、規格化の条件

$$\int f d\Gamma = \exp[-1 - \lambda_2] \int \exp[-\lambda_1 H] d\Gamma = 1$$

から

$$f = Z^{-1} \exp[-\lambda_1 H], \quad Z = \int \exp[-\lambda_1 H] d\Gamma$$

ここで $\lambda_1 = \beta$ とすれば、これはカノニカルでの分布関数になっています。というわけで、情報エントロピーに平均エネルギーを与えた時の極値はカノニカルの分布関数で与えられます。これが最大値になっていることはミクロカノニカルと同じようにすれば分かります。不等式 (3) に入れて f がカノニカルでの分布関数とすれば

$$- \int f' \log f' d\Gamma \leq - \int f' \log f d\Gamma = - \log Z^{-1} \int f' d\Gamma + \beta \int f' H d\Gamma = \log Z + \beta \int f' H d\Gamma$$

最右辺にいくときに f' と f が同じ規格化で、同じ平均エネルギーを与える

$$\int f' H d\Gamma = \int f H d\Gamma \quad (4)$$

という条件を入れています (平均エネルギー $\langle E \rangle$ を与える分布関数を考えているから)。そして、

$$\begin{aligned} \int f \log f d\Gamma &= \int Z^{-1} e^{-\beta H} (-\log Z - \beta H) d\Gamma \\ &= -Z^{-1} \log Z \int e^{-\beta H} d\Gamma - \beta \int Z^{-1} e^{-\beta H} H d\Gamma \\ &= -\log Z - \beta \int f H d\Gamma \end{aligned}$$

であることから

$$-\int f' \log f' d\Gamma \leq -\int f \log f d\Gamma$$

となるので、カノニカルでの分布関数は平均エネルギーを与えた時の情報エントロピーの最大値を与えます。同様に平均エネルギー $\langle E \rangle$ と粒子数の平均 $\langle N \rangle$

$$\langle E \rangle = \sum_n \int H f_n d\Gamma, \quad \langle N \rangle = \sum_n \int n f_n d\Gamma \quad \left(\sum_n \int f_n d\Gamma = 1 \right)$$

を与えると、ラグランジュの未定乗数法によってグランドカノニカルの分布関数

$$f_n = \Xi^{-1} \exp[-\beta(H - \mu N_n)]$$

が情報エントロピーの極値と分かり、最大値であることは、カノニカルと同じように任意の分布関数 f' に

$$\sum_n \int H f_n d\Gamma = \sum_n \int H f'_n d\Gamma, \quad \sum_n \int n f_n d\Gamma = \sum_n \int n f'_n d\Gamma$$

という条件をかけたときに

$$-\sum_n \int f'_n \log f'_n d\Gamma \leq -\sum_n \int f'_n \log f_n d\Gamma = -\sum_n \int f_n (-\log \Xi - \beta(H - \mu N_n)) d\Gamma = -\sum_n \int f_n \log f_n d\Gamma$$

となることから分かります。よって、平均エネルギーと平均粒子数が与えられた情報エントロピーの最大値はグランドカノニカルでの分布関数によって与えられます。

このように統計力学と情報理論の間には関連性があります。歴史的には統計力学の方が古いので、情報理論は統計力学を参考にしている部分が結構あります。また、情報エントロピーは統計力学でも必要になる場面があ

ります。熱平衡状態におけるエントロピーの定義は明確に与えていますが、非平衡状態でのエントロピーは明確に定義されていなく、そのため確率で定義できる情報エントロピーが非平衡状態の話で使われます。

・補足

情報理論に関連する確率についての小話をします。

確率には主観確率 (subject probability) と客観確率 (object probability) という分類があります。大雑把に言えば、主観確率は個人の判断で決められる確率、客観確率は実験結果から作られる確率です。言い方を合わせれば、個人による確率が主観確率、共通認識として使える確率が客観確率です。例えば、サイコロの目の確率は $1/6$ というのは実験結果としてあるので、一般常識として共有されています (実験結果を信じるという暗黙の了解がある)。主観確率は、例えばすれ違った人がそのまま帰宅する確率は何かというものです。これは直感やすれ違った人の見た目に対する個人的な印象に依存しています。

主観、客観確率の境界はかなり曖昧です。例として、条件付き確率 (conditional probability) の問題として、サイコロを振ると偶数が出るという条件において、2 が出る確率はいくつかというのを持ってきます。答えはすぐ分かるように $1/3$ です。この条件を情報と読み替えます。つまり、サイコロを振ると偶数しか出ないことを知っている人にとって、2 が出る確率は $1/3$ ということです。これは主観確率です (知っているかどうかによって依存するから)。しかし、偶数しか出ないということを知っていさえすれば、誰でも確率は $1/3$ になるという点から見れば、客観確率です。