2020.6.16(火)

情報統計力学 学習9

6月16日(火) 午後1時ー2時半

何回かにわたって、情報エントロピーと物理におけるエントロピーの関係について解説する。

1　情報エントロピー (続き)

定義を書いておく。X を離散型確率変数として、有限 (n) 個の値をとるとする。$p(x)$ を確率関数とすると、X の情報エントロピー H_X は、

$$H_X = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) \tag{1}$$

で与えられる。明らかに、

$$H_X \geq 0 \tag{2}$$

を満たす。

結合エントロピー
X, Y を確率変数とし、確率関数を $p(x, y)$ とする。確率変数 (X, Y) のエントロピーは、

$$H_{X,Y} = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2 p(x_i, y_j) \tag{3}$$

と定義される。これを**結合エントロピー**とよぶ。また、$p(x) > 0$ として、x が与えられたときの y の条件付き確率分布を $p(y|x)$ とする。

$$p(y|x) = p(x, y)/p(x) \tag{4}$$

となる。もし、X, Y が独立なら、$p(y|x)$ は x に依存しない。独立でないとき、その依存性を測る量があるだろうか？つまり、x の値を知ったとき、y について、どれほど情報を得ることができるだろうか？これについて考える。

条件付きエントロピー $H_{Y|X}$
$p(y|x)$ のエントロピーを x で平均したものと定義する。

$$H_{Y|X} = -\sum_{i=1}^{n} p(x_i)\sum_{j=1}^{m} p(y_j|x_i)\log_2 p(y_j|x_i). \tag{5}$$

問題　次式を示せ。

$$H_{X,Y} = H_X + H_{Y|X}, \tag{6}$$

これは、**チェーンルール**と呼ばれる。

問題　X, Y が独立なとき、以下を示せ。

$$H_{Y|X} = H_Y, \tag{7}$$
$$H_{X,Y} = H_X + H_Y. \tag{8}$$
$$\tag{9}$$

２つの変数の相関は、**相互情報量**で測ることができる。

相互情報量 $I_{X,Y}$ は次式で定義される。

$$I_{X,Y} = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2 \frac{p(x_i, y_j)}{p(x_i)P(y_j)}. \tag{10}$$

問題 次式を示せ。

$$I_{X,Y} = H_X + H_Y - H_{X,Y}, \tag{11}$$
$$I_{X,Y} = H_Y - H_{Y|X} = H_X - H_{X|Y}. \tag{12}$$

２つめの式より、相互情報量は y についての知識によって、x の不確定性がどれだけ減るかを測る量であることが分かる。また、x についての知識によって、y の不確定性がどれだけ減るかを測る量でもある。

問題　以下を示せ。

$$I_{X,Y} \geq 0, \tag{13}$$
$$I_{X,Y} = 0 \quad \text{となるのは、} X \text{ と } Y \text{ が独立であるときに限る。} \tag{14}$$

X が連続型確率変数
確率密度関数を $p(x)$ とすると、**微分エントロピー**は次式で定義される。

$$H(X) = -\int p(x)\log_2 p(x)dx \tag{15}$$

これは、必ずしも正ではない。

情報統計力学で重要な量として、**カルバックーライブラー (Kullback-Leibler, KL) 情報量**がある。離散型確率変数の場合、同じ標本空間 Ω の２つの確率関数 $p(x), q(x)$ について、

$$D(p||q) = \sum_{i=1}^{n} q(x_i \log_2 \frac{q(x_i)}{p(x_i)} \tag{16}$$

で定義される。$0\log_2 \frac{0}{0} = 0$ とする。これは、次の性質を満たす。

$$D(p||q) \text{ は、} q \text{ について、凸関数である。} \tag{17}$$
$$D(p||q) \geq 0\text{。} \tag{18}$$
$$q(x) \text{ と } p(x) \text{ が恒等的に等しいのでなければ、} D(p||q) > 0 \text{ である。} \tag{19}$$

これは、距離と似た量であるが、距離の定義を満たさない、つまり、一般に、$D(p||q) \neq D(q||p)$.

2　統計力学のエッセンス

統計力学では、アボガドロ数 $\sim 10^{22}-10^{23}$ ほどの巨視的数の分子からなる系を扱う。それを力学的に考えると、古典力学なら、ラグランジュ方程式やハミルトンの正準方程式などの時間発展方程式に従う系を扱うことになる。正準変数を $\{q_i, p_i\},\ i=1,2,\cdots,f$ とする。ここで、f は自由度。３次元の１個の質点なら、$f=6$.　巨視的系なら、$f \sim 10^{22}$.

$$\frac{d}{dt}q_i = \frac{\partial}{\partial p_i}H, \tag{20}$$

$$\frac{d}{dt}p_i = -\frac{\partial}{\partial q_i}H, \tag{21}$$

$$H = T+U,\ T: \text{運動エネルギー、}U: \text{位置エネルギー、}H: \text{ハミルトニアン。} \tag{22}$$

これを解くのは不可能。もし解いたとしても、巨視的物理量を計算できなければ、実験と比較できず意味がない。
そこで、統計的な考察を行う。系の保存量として、全エネルギー、全運動量、全角運動量がある。系が、並進運動や回転運動をしていなければ、全運動量、全角運動量は０である。一般的には、全エネルギーの他に保存量はないと考えられる。従って、軌道は、$2f$ 次元の相空間 $\{q_i, p_i\}$ で、全エネルギーが一定の面上を運動する。ここでは、エネルギーの幅を考えて、全エネルギーが $E, E+\Delta E$ の領域 R を運動すると考える。また、この領域のどの部分も、他の部分と同等と考えるのが自然である。従って、軌道はこの領域 R 内のどの部分領域 r にも到達すると考えることができる。十分長い時間 T での r への滞在時間 τ は、部分領域 r の $2f$ 次元の" 体積" を $|r|$ とすると $|r|$ に比例すると考えて良い。すなわち、同じ体積の部分領域 r' における滞在時間は同じである。
そこで、今度は、同じ系を非常にたくさん (M 個) 用意する。１つの系は相空間では１点で表される。それらの微視的状態は異なるとする。すると、それらの点は、時間とともに変化するが、十分時間がたつと、部分領域 r に入っている個数 m は、変化しないと考えられ、そこに入る確率 m/M は、１つの軌道の滞在時間の比率 τ/T に等しいと考えることができる。このような、多数の同じ系からなる集団を統計集団という。
上の考え方から、次の式が成り立つことが期待される。相空間における確率密度を $f(q,p)$ とする。ここで、(q,p) は $\{q_i, p_i\}$ を略して書いている。$A(q,p)$ は物理量。

$$\frac{1}{T}\int_0^T A(q(t), p(t))dt = \int A(q,p)f(q,p)dqdp \tag{23}$$

つまり、時間平均が統計平均に等しい。これはエルゴード仮設と呼ばれている。統計力学は、これを仮定して、実験で観測される量、すなわち上式の左辺が、右辺の統計平均で計算されるとする。次に問題なのは、$f(q,p)$ を求めることである。

2.1　ミクロカノニカル集団、ミクロカノニカル分布

簡単のため、離散的な系を考えよう。外部とエネルギーのやりとりをしない孤立系を考える。エネルギーが一定の状態が W 個あるとする。上での考察により、系はこの W 個の状態のいずれにも等しい確率で存在する。これを**等重率の原理**という。これが統計力学の基本仮定である。従って、エネルギー一定のある状態 i に存在する確率 p_i は、$p_i = 1/W$ となる。このような統計集団を**ミクロカノニカル集団**、この確率分布を**ミクロカノニカル分布**という。このとき、統計力学的エントロピー S は、

$$S = k_{\mathrm{B}} \ln W \tag{24}$$

で定義される。これは、ボルツマンによって定義され、オーストリアのウィーンの中央墓地にあるボルツマンの墓に記されている。k_{B} はボルツマン定数で、エントロピーの次元を持ち、厳密に

$k_{\mathrm{B}} = 1.380649 \times 10^{-23}\mathrm{JK}^{-1}$ である。

情報エントロピーとの関係
このときの情報エントロピーを計算してみよう。

$$H = -\sum_{i=1}^{W} \frac{1}{W} \log_2 \frac{1}{W} = \log_2 W \tag{25}$$

となる。$y = \log_2 x$ なら、$x = 2^y$. 両辺の自然対数をとると、$\ln x = y \ln 2$。従って、$\log_2 x = \frac{\ln x}{\ln 2}$。
従って、

$$H = \frac{1}{k_{\mathrm{B}} \ln 2} S \tag{26}$$

となる。つまり、情報エントロピーと統計力学的エントロピーは、ミクロカノニカル分布のとき、比例関係にある。