

2019年6月3日（川端）

データ圧縮基礎資料

9 微分エントロピー

有歪情報源符号化の対象としたいのは、連続な情報源である。連続情報源ではシャノンのエントロピーは発散するものの有歪における情報レートは有界に収まることが期待できる。本節では、連続情報源においてシャノンのエントロピーに代わる役割をもち、特にレート歪理論を構成する要素となる微分エントロピーについて述べる。

9.1 微分エントロピーの定義

実数値確率変数 $X : \Omega \rightarrow \mathcal{X}$ について、累積分布関数 $F(x) = P(X \leq x)$ が微分 $f(x) = F'(x)$ すなわち確率密度関数をもつと仮定する。 $f(x)$ の台を S で表す。

定義 1 (微分エントロピー) : 確率密度関数 $f(x)$ をもつ実数値確率変数 X について

$$\begin{aligned} h(X) &:= - \int_S f(x) \log f(x) dx \\ &= E[\log \frac{1}{f(X)}] \end{aligned}$$

を X の微分エントロピーとよぶ。

2つの例を挙げる。

例 1 (一様分布)

$$\begin{aligned} f(x) &= \frac{1}{a}, 0 \leq x \leq a \\ h(X) &= E \log a = \log a \end{aligned} \tag{1}$$

例 2 (正規分布)

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \\ h(X) &= E[\log \sqrt{2\pi}\sigma + \log e^{-\frac{X^2}{2\sigma^2}}] \\ &= \frac{1}{2} \log(2\pi e\sigma^2) \end{aligned}$$

確率変数 X を幅 $\Delta > 0$ で量子化した離散確率変数 X^Δ を以下のように定義する。平均値の定理より

$$f(x_i)\Delta := \int_{i\Delta}^{(i+1)\Delta} f(x)dx = P(i\Delta \leq X \leq (i+1)\Delta)$$

を満たす x_i が区間 $[i\Delta, (i+1)\Delta]$ に存在する。そこで X^Δ を

$$X^\Delta = x_i \quad \text{if} \quad i\Delta \leq X \leq (i+1)\Delta$$

のように定義する。 $p_i = P(X^\Delta = x_i)$ は $f(x_i)\Delta$ に等しいことに注意する。

定理 1 (微分エントロピーとシャノンエントロピーの関係)

$$H(X^\Delta) \rightarrow h(X) - \log \Delta \quad \text{as} \quad \Delta \rightarrow 0.$$

証明:

$$\begin{aligned}
H(X^\Delta) &= \sum_i p_i \log \frac{1}{p_i} \\
&= E \log \frac{1}{f(X^\Delta)\Delta} \\
&= E \log \frac{1}{f(X^\Delta)} - E \log \Delta \\
&\rightarrow h(X) - \log \Delta
\end{aligned}$$

証明終

9.2 同時微分エントロピー, 条件付き微分エントロピー

2つの実数値確率変数 (X, Y) の同時確率密度関数についても

$$f(x, y) := \frac{\partial^2}{\partial x \partial y} P(X \leq x, Y \leq y)$$

が定義される.

定義 2 同時微分エントロピーは以下のように定義される.

$$h(X, Y) := E \log \frac{1}{f(X, Y)} \quad (2)$$

X に関する Y の条件付き確率密度は

$$f(y|x) := \frac{f(x, y)}{f(x)}$$

で与えられる.

定義 3 X に関する Y の条件付き微分エントロピーを

$$h(Y|X) = E \log \frac{1}{f(Y|X)}$$

と定義する.

定理 2

$$h(Y|X) = h(X, Y) - h(X) \quad (3)$$

すなわち, 微分エントロピーに関するチェインルール

$$h(X, Y) = h(X) + h(Y|X) \quad (4)$$

が成り立つ.

以上の定義は, さらに多変数についても一般化される. 例えば, $X_1, X_2, X_3, \dots, X_n$ について同時確率密度関数 $f(x_1, x_2, x_3, \dots, x_n)$ から同時微分エントロピー $h(X_1, X_2, X_3, \dots, X_n)$ が定義される. これらについて, $X^n, \mathbf{X} = X^n = X_1^n, f(x^n), f(\mathbf{x}) = f(x^n) = f(x_1^n)$ 等の記法の利用が可能となる.

例 3 n 次元確率ベクトル $\mathbf{X}$ の多次元確率密度関数がベクトル $\mu = (\mu_i)$, および $n \times n$ 正定値行列 K を用いて

$$f(\mathbf{x}) = \frac{1}{\sqrt{2\pi}^n |K|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu)K^{-1}(\mathbf{x} - \mu)^T\right\}$$

と表されるとき, $\mathbf{X}$ は平均ベクトル μ , 分散共分散行列 K をもつ多次元正規分布に従うという. $K = (K_{ij})$ の (i, j) 要素は $K_{ij} = E(X_i - \mu_i)(X_j - \mu_j)$ で与えられる. このとき, $\mathbf{X} \sim N(\mu, K)$ と略記する.

多変数の確率密度についても微分エントロピーが定義できる. n 変数の場合, $h(\mathbf{X}) = h(X^n) = h(X_1^n)$ 等のような記法で表すことにする.

例 4 n 次元正規分布に従う $\mathbf{X}$ の微分エントロピーは次のように計算できる.

$$\begin{aligned} h(\mathbf{X}) &= h(X_1, X_2, \dots, X_n) = E \log \frac{1}{f(\mathbf{X})} \\ &= \log \sqrt{2\pi}^n |K|^{1/2} + \frac{1}{2} \{E(\mathbf{x} - \mu)K^{-1}(\mathbf{x} - \mu)^T\} \log e \\ &= \frac{1}{2} \log(2\pi)^n |K| + \frac{1}{2} \text{tr}(K^{-1}K) \log e \\ &= \frac{1}{2} \log(2\pi e)^n |K| \end{aligned} \quad (5)$$

また, チェインルールは以下のように一般化できる.

例 5

$$h(X^n) = \sum_{i=1}^n h(X_i | X^{i-1})$$

すなわち, シヤノン理論と記号の上でも対応した微分エントロピーの情報理論が成り立つことが分かる.

9.3 KL ダイバージェンス, 相互情報量

次にシャノン情報量の理論における鍵概念であった KL ダイバージェンスについて考える.

定義 4 多次元実空間 $\mathcal{X}$ 上の 2 つの連続な確率密度を $f(x), g(x)$ とする. $g(x)$ に関する $f(x)$ の KL ダイバージェンスを

$$D(f||g) = \int f(x) \log \frac{f(x)}{g(x)} dx$$

と定義する.

定理 3

$$D(f||g) \geq 0$$

等号は $f(\cdot) = g(\cdot)$ のときのみ.

証明:

$$\begin{aligned}
D(f||g) &= \lim_{\Delta \downarrow 0} E_f \log \frac{f(X_f^\Delta)\Delta}{g(X_g^\Delta)\Delta} \\
&= \lim_{\Delta \downarrow 0} D(\{f(X_f^\Delta)\Delta\} || \{g(X_g^\Delta)\Delta\}) \\
&\geq 0
\end{aligned} \tag{6}$$

最後から2番目の式は離散分布に対するKLダイバージェンスであり、最後の不等式にはその非負性を用いている。従って等号に関する事実もこのことと密度の連続性から分かる。証明終

- 同じ核文字が用いられるのは同じ値を取る理由による。
- 直接 Jensen の不等式を用いて

$$\begin{aligned}
-E_f \log \frac{g(X)}{f(X)} &\geq -\log E_f \frac{g(X)}{f(X)} \\
&\geq 0
\end{aligned}$$

と証明しても良い。

定義 5 連続な確率密度をもつ確率変数 X と Y の相互情報量は

$$I(X; Y) = D(f_{XY}(x, y) || f_X(x)f_Y(y))$$

と定義される。

- $I(X; Y) \geq 0$ であり、等号は X と Y が独立のときだけである。
-

$$\begin{aligned}
I(X; Y) &= h(X) + h(Y) - h(X, Y) \\
&= h(X) - h(X | Y) \\
&= h(Y) - h(Y | X)
\end{aligned}$$

例 6 1次元密度をもつ X について

$$h(aX) = h(X) + \log |a|$$

証明: $Y = aX$ とおく。 $f_Y(ax) = f_X(x) \left| \frac{dx}{dy} \right| = f_X(x) \frac{1}{|a|}$ に注意すると、

$$\begin{aligned}
h(aX) &= E \log \frac{1}{f(aX)} \\
&= E \log \frac{1}{f_X(X) \left| \frac{dx}{dy} \right|} \\
&= h(X) + \log |a|
\end{aligned}$$

証明終

定理 4

$$h(\mathbf{A}\mathbf{X}) = h(\mathbf{X}) + \log |\det \mathbf{A}|$$

定理 5 (最大エントロピー分布の正規性) $\mathbf{X}$ は平均が $E\mathbf{X} = \mathbf{0}$ を満たし, 分散共分散行列 $E\mathbf{X}^T\mathbf{X}$ が K で与えられる n 次元確率変数とする. このとき,

$$h(\mathbf{X}) \leq \frac{1}{2} \log(2\pi e)^n |\det K|$$

であり, 等号は多次元正規分布のときに成り立つ.

証明: $\mathbf{X}$ の密度関数を f とおく. f とおなじ平均, 分散共分散行列をもつ多次元正規分布の密度関数を ϕ すなわち

$$\phi(\mathbf{x}) = \frac{1}{\sqrt{2\pi}^n |\mathbf{K}|^{1/2}} \exp\left\{-\frac{1}{2} \mathbf{x} \mathbf{K}^{-1} \mathbf{x}^T\right\}$$

とおく. この関数の対数は $\mathbf{X}$ についての二次形式であるから, f でも ϕ でも同じ期待値をもつ, すなわち $E_f - \log \phi(\mathbf{X}) = E_\phi - \log \phi(\mathbf{X})$. これより

$$0 \leq D(f \parallel \phi) = h(\phi) - h(f)$$

証明終

○