『社会科学のためのベイズ統計モデリング』第 6 章　6.1–6.6

Hiroshi Hamada
Tohoku University

Jun, 2020
at Tohoku University

ハートのエースという情報の価値

私が引いた 1 枚のトランプをあなたが当てるゲームを考える（正解はハートのエース）.

- 情報 A : カードはハートだ
- 情報 B : カードはエースだ

情報 A と B を得ることは,「正解（ハートのエース）」の情報を得ることと等価

$$f(A \cap B) = f(A) + f(B)$$

ハートの情報量とエースの情報量の和は，ハートのエースの情報量と同じだと仮定する.

情報量の公理的導出

事象 A, B が独立なとき

$$f(P(A \cap B)) = f(P(A)) + f(P(B))$$

確率が小さいほど大きい

$$P(A) \leq P(B) \Longrightarrow f(P(B)) \leq f(P(A)) \tag{1}$$
$$f(1) = 0 \tag{2}$$

を満たす関数は $-\log$ しかない.

条件を満たす情報量の関数 f として

$$f(P(A)) = -K \log P(A)$$

が定まる（K は正の定数）.

「ハートのエース」の情報量

$$\text{ハートの情報量}: -\log 1/4 \approx 1.39$$
$$\text{エースの情報量}: -\log 1/13 \approx 2.56$$
$$\text{ハートのエースの情報量}: -\log 1/52 \approx 3.95$$

加法性を満たす $3.95 = 1.39 + 2.56$

ハートが選ばれる確率

$$P(X = 1) = \frac{1}{4}$$

エースが選ばれる確率

$$P(Y = 1) = \frac{1}{13}$$

ハートのエースが選ばれる確率は，それぞれの試行が独立ならば,

$$P(X = 1, Y = 1) = P(X = 1)P(Y = 1) = \frac{1}{52}$$

つまり，情報量は《確率の関数》とみなせる

（自己）情報量 (self information)

定義 (情報量)

離散確率変数 X の実現値 x が生起したときの情報量を

$$I(X = x) = -\log P(X = x)$$

と定義する．また，I を確率 $p = P(X = x)$ の関数として以下のように表記することもある．

$$I(p) = -\log p$$

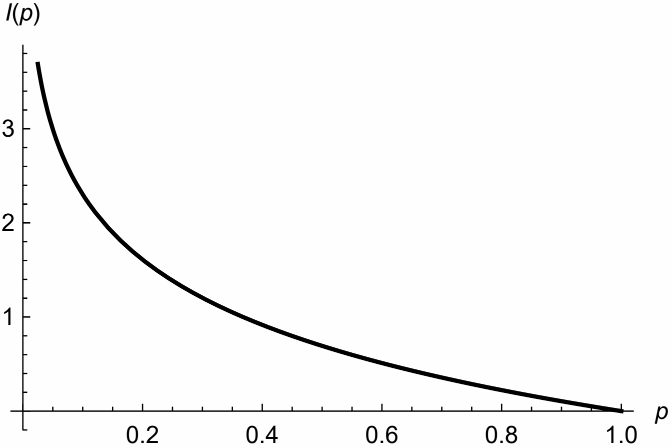

Figure: 確率 p についての情報量 $I(p)$

エントロピー

定義 (エントロピー)

離散確率変数 X のエントロピー $H(X)$ を，情報量の期待値として次のように定義する．

$$H(X) = \mathbb{E}[I(X)] = -\sum_{i=1}^{n} P(X = x_i) \log P(X = x_i)$$

$P(X = x_i) = 0$ のとき，$0 \cdot \log 0 = 0$ と約束する．また，確率質量関数 $f(x)$ を用いて，以下のように表記する．

$$H(f) = -\sum_{i=1}^{n} f(x_i) \log f(x_i)$$

ベルヌーイ確率変数 Y のエントロピー

$P(Y=1)=q, P(Y=0)=1-q$ とすると，

$$H(Y)=-q\log q-(1-q)\log(1-q).$$

$H(Y)$ を q の関数と見なして，q で微分して 0 とおくと，

$$\frac{dH}{dq}=0 \iff \log(1-q)=\log q$$

$$\frac{d^2H}{dq^2}=-\frac{1}{p}-\frac{1}{1-p}<0$$

より，$q=1/2$ でエントロピーは最大.

連続確率変数のエントロピー

定義 (連続エントロピー)

$A \subset \mathbb{R}$ 上で定義された確率密度関数 $f(x) > 0$ をもつ連続確率変数 X のエントロピーを,

$$H(X) = H(f) = -\int_A f(x) \log f(x) dx$$

と定義する.

正規分布のエントロピー

$$H(X) = \frac{1}{2}\left(1 + \log(2\pi\sigma^2)\right)$$

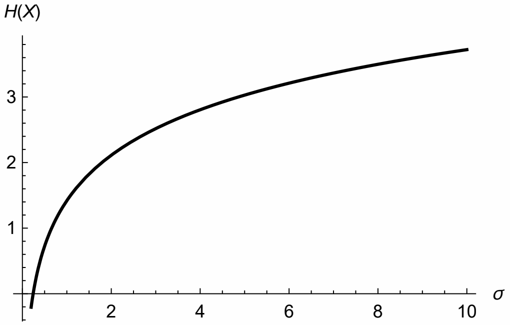

Figure: 正規確率変数 X のエントロピー $H(X)$

カルバック=ライブラー情報量

定義 (KL 情報量)

$A \subset \mathbb{R}$ 上で定義された連続確率密度関数 $q(x), p(x) > 0$ についてのカルバック=ライブラー情報量を，

$$D(q||p) = \int_A q(x) \log \frac{q(x)}{p(x)} dx$$

と定義する．

真の分布 $q(x)$ を基準にして，モデル $p(x)$ がどれくらい近いかを表す量

なぜ KL 情報量か

渾身の補足資料（KL_examples.pdf）参照！！

意味：KL 距離情報量が小さくなる予測分布 $p(x)$ を選ぶことは、データをあてはめたときに尤度関数が大きくなるような予測分布 $p(x)$ を選ぶことと一致する

関数型の導出:KL 情報量 $D(p||q)$ は《真の分布 q の経験分布がほぼ予測分布 p となる確率》を決める関数として自然に導出できる

KL 情報量の例

$$q(x) = \mathrm{Normal}(x|0,1) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}$$
$$p(x) = \mathrm{Normal}(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

2 つの正規分布間の KL 情報量は？

KL 情報量の例

$$
\begin{aligned}
D(q||p) =& \mathbb{E}_{q(X)}\left[\log \frac{q(X)}{p(X)}\right] \\
=& \mathbb{E}_{q(X)}\left[-\frac{X^2}{2} + \frac{X^2 - 2X\mu + \mu^2}{2\sigma^2} + \log\sigma\right] \\
=& -\frac{1}{2}\mathbb{E}_{q(X)}[X^2] \\
&+ \frac{1}{2\sigma^2}\left(\mathbb{E}_{q(X)}[X^2] - 2\mathbb{E}_{q(X)}[X]\mu + \mu^2\right) + \log\sigma \\
=& -\frac{1}{2} + \frac{1}{2\sigma^2}\left(1 + \mu^2\right) + \log\sigma
\end{aligned}
$$

交差エントロピー

定義 (交差エントロピー)

$A \subset \mathbb{R}$ 上で定義された確率密度関数 $q(x), p(x) > 0$ について，$q(x)$ と $p(x)$ の交差エントロピーを，

$$H_q(p) = -\int_A q(x) \log p(x) dx$$

と定義する．

$$
\begin{aligned}
D(q||p) &= \int_A q(x) \log \frac{q(x)}{p(x)} dx \\
&= -\int_A q(x) \log p(x) dx + \int_A q(x) \log q(x) dx \\
&= H_q(p) - H(q)
\end{aligned}
$$

真の分布 $q(x)$ からのモデル $p(x)$ の近さ $D(q||p)$ は，未知の定数部分 $H(q)$ をのぞくと，交差エントロピー $H_q(p)$ の大きさと完全に一致する．

交差エントロピーをデータから推定できれば，真の分布へのモデルの近さを評価できる