

第 1 回

学部 3 年後期ゼミナール発表資料

青山和樹

2024 年 10 月 7 日

発表の目的

テキスト [1] の

- 2.1 Entropy
- 2.2 Joint entropy and conditional entropy

について発表する .

目次

1	Entropy	2
2	Joint entropy and conditional entropy	4

1 Entropy

まず確率変数の不確実性の尺度としてエントロピー (entropy) を導入する. X をアルファベット \mathcal{X} の離散確率変数 (discrete random variable) とし, その確率質量関数 (probability mass function) を $p(x) = \Pr(X = x), x \in \mathcal{X}$ とする.

注意 1.1 確率質量関数を $px(x)$ ではなく簡便のため $p(x)$ と表す. したがって, $p(x)$ と $p(y)$ は異なる確率変数に対して異なる確率質量関数 $px(x)$ と $py(y)$ を指す.

定義 1.2 (エントロピー)

エントロピー $H(X)$ は, 確率関数 $p(x)$ を持つ確率変数 X について次のように定義される.

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x) \quad (1)$$

また, この量を $H(p)$ と書く. 対数は 2 を底とし, エントロピーはビット (bit) で表される.

注意 1.3 $0 \log 0 = 0$ という規約を用いる, これは $x \log x$ が $x \rightarrow 0$ で 0 に収束するため, 連続性から容易に正当化される. よってゼロ確率の項を加えてもエントロピーは変わらない.

もし対数の底が b であれば, エントロピーは $H_b(X)$ と表される. 対数の底が e の場合, エントロピーはナット (nat) で測定される. 特に指定がない限り, すべての対数は底 2 であり, したがってエントロピーはビットで測定され则认为. エントロピーは X の分布の汎関数であり, 実際に確率変数 X が取る値には依存せず, 確率だけに依存する.

期待値は E によって表される. もし X が $p(x)$ に従うならば, 確率変数 $g(X)$ の期待値は以下によって表される.

$$E_p g(X) = \sum_{x \in \mathcal{X}} g(x) p(x) \quad (2)$$

あるいは, 文脈から確率関数が明らかであれば, 単に $Eg(X)$ と書く.

解説 1.4 X のエントロピーは, 確率関数 $p(x)$ に従って X を選んだ場合の確率変数 $\log \frac{1}{p(x)}$ の期待値としても解釈できる. したがって,

$$H(X) = E_p \log \frac{1}{p(X)} \quad (3)$$

と表される.

補題 1.5

$$H(X) \geq 0 \quad (4)$$

証明 1.6 $0 \leq p(x) \leq 1$ であるから, $\log \frac{1}{p(x)} \geq 0$ である. □

補題 1.7

$$H_b(X) = (\log_b a) H_a(X) \quad (5)$$

証明 1.8 $\log_b q = \log_b a \log_a p$. □

エントロピーの第 2 の性質として, 定義における対数の底を変更できることがある. エントロピーは, 適切な係数を掛けることで, 異なる底に変換できる.

例 1.9

$$X = \begin{cases} 1 & \text{確率 } p \text{ のとき,} \\ 0 & \text{確率 } 1-p \text{ のとき.} \end{cases} \quad (6)$$

のとき

$$H(X) = -p \log p - (1-p) \log(1-p) \stackrel{\text{def}}{=} H(p) \quad (7)$$

特に $p = \frac{1}{2}$ のとき $H(X) = 1$ である. 関数 $H(p)$ のグラフは図??に示した. この図はエントロピーの基本的な性質を示しており, エントロピーは分布に対して凹関数であり, $p = 0$ または 1 のとき 0 となる. これは, $p = 0$ または 1 の場合, 変数はランダムでなくなり, 不確実性がなくなるためである. 同様に, 不確実性が最大となるのは $p = \frac{1}{2}$ のときであり, このときエントロピーも最大値を取る.

例 1.10

$$X = \begin{cases} a & \text{確率 } \frac{1}{2} \text{ のとき,} \\ b & \text{確率 } \frac{1}{4} \text{ のとき,} \\ c & \text{確率 } \frac{1}{8} \text{ のとき,} \\ d & \text{確率 } \frac{1}{8} \text{ のとき.} \end{cases} \quad (8)$$

エントロピーは

$$H(X) = -\left(\frac{1}{2} \log \frac{1}{2} + \frac{1}{4} \log \frac{1}{4} + \frac{1}{8} \log \frac{1}{8} + \frac{1}{8} \log \frac{1}{8}\right) = \frac{7}{4} \text{ bits} \quad (9)$$

2 Joint entropy and conditional entropy

1 章では, 単一の確率変数のエントロピーを定義した. ここでは, 確率変数のペアに対して定義を拡張する. この定義には特に新しいことはないが, (X, Y) は単一のベクトル値確率変数と見なすことができる.

定義 2.1 (同時エントロピー) —

同時エントロピー (joint entropy) $H(X, Y)$ は, 同時分布 $p(x, y)$ を持つ 2 つの確率変数 (X, Y) について次のように定義される.

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y) \quad (10)$$

また次式のようにも表される.

$$H(X, Y) = -E \log p(X, Y) \quad (11)$$

また, ある確率変数に別の確率変数が与えられたときの条件付きエントロピーを, 条件付き確率変数で平均した条件付き分布のエントロピーの期待値として定義する.

定義 2.2 (条件付きエントロピー) —

条件付きエントロピー (conditional entropy) $H(X|Y)$ は次のように定義される.

$$H(Y|X) = \sum_{x \in \mathcal{X}} p(x) H(Y|X = x) \quad (12)$$

$$= - \sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x) \quad (13)$$

$$= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) \quad (14)$$

$$= -E \log p(Y|X). \quad (15)$$

同時エントロピーと条件付きエントロピーの定義の自然さは, 次の定理によって示される.

定理 2.3 (エントロピーのチェーン則) —

$$H(X, Y) = H(X) + H(Y|X) \quad (16)$$

証明 2.4

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y) \quad (17)$$

$$= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x) p(y|x) \quad (18)$$

$$= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x) - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) \quad (19)$$

$$= - \sum_{x \in \mathcal{X}} p(x) \log p(x) - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) \quad (20)$$

$$= H(X) + H(Y|X). \quad (21)$$

同等に次のように書くことができる。

$$\log p(X, Y) = \log p(X) + \log p(Y|X) \quad (22)$$

系 2.5

$$H(X, Y|Z) = H(X|Z) + H(Y|X, Z) \quad (23)$$

証明 2.6

$$H(X, Y|Z) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y|z) \log p(x, y|z) \quad (24)$$

$$= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y|z) \log p(x) p(y|z|x) \quad (25)$$

$$= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x) - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) \quad (26)$$

$$= - \sum_{x \in \mathcal{X}} p(x) \log p(x) - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) \quad (27)$$

$$= H(X) + H(Y|X). \quad (28)$$

参考文献

- [1] T.M.Cover and J.A.Thomas, Elements of Information Theory, Second Edition, John Wiley & Sons, 2006.