

情報量と確率の関係

事象	情報量	確率
(a) 沖縄に雪が降った。	$i(a)$	$P(a)$
(b) 北海道に雪が降った。	$i(b)$	$P(b)$

$i(a) > i(b) \quad P(a) < P(b)$

1. 情報量は、確率の関数
 2. 情報量は、確率に関する減少関数
 (確率が増加すれば、情報量は減少する。)

7

独立事象の確率と情報量(情報量の性質2)

事象「宝くじが当たって、しかも事故にあった。」の情報量を考えよう。

(宝a) 買った宝くじが外れた。
 (宝b) 買った宝くじが1等当たった。

↔

(事a) 今日事故にあった。
 (事b) 今日事故にあわなかった

(互いに無関係な事象を独立な事象と言う。)

”独立”な事象の積事象の確率
 $P(\text{宝}c) \times P(\text{事}b)$

”独立”な事象の積事象の情報量
 $i(\text{宝}c) + i(\text{事}b)$

確率統計の復習
 積事象の確率が、各事象の確率の和であるとき、独立な事象という。

3. 独立な事象の積事象の情報量は、個々の情報量の和

8

情報量の性質から情報量の数値化へ

事象 x がおきる確率を $P(x)$ と表し、
 事象 x がおきたことを知った情報量を $i(x)$ と表す。
 このとき、以下を満たす関数 $f(x)$ で情報量を定義する。

1. 情報量は、確率の関数である。

$$i(x) = f(P(x))$$
2. 情報量は、確率に対する減少関数である。

$$P(x_1) < P(x_2) \Leftrightarrow i(x_1) > i(x_2)$$
3. 独立な積事象を知ったときの情報量は、個々の情報量の和である。

$$\begin{aligned} i(x_1 \wedge x_2) &= f(P(x_1 \wedge x_2)) = f(P(x_1) \cdot P(x_2)) \\ &= i(x_1) + i(x_2) \end{aligned}$$

9

(自己)情報量と情報量の単位

先のスライドを満たすように、
 ある事象 x を知る情報量 $i(x)$ は以下の関数で定義される。

$$i(x) = -\log_s P(x)$$

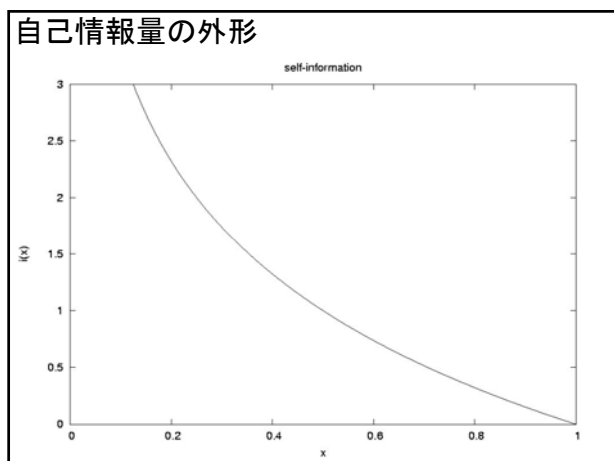
ただし、 $s > 1$

情報量は、確率の逆数の対数。

情報量の単位

$(s = 2) \rightarrow$ [bit](ビット)
 $(s = e) \rightarrow$ [nat](ナット)
 $(s = 10) \rightarrow$ [decit](デシット)

1ビットは確率0.5の事象が起きたことを知る情報量。以後は、ビットだけを扱う。



練習

次の事象の自己情報量を求めよ。

- (1)
a : 2枚のコインを投げて両方とも裏がでる。
- (2)
b : 52枚のトランプから絵札を1枚引く
- (3)
c : アルファベットの書いてある26個の玉から、cの玉を取り出す。

12

事象系

単独の事象ではなくて、事象の集合を考える。
事象の集合とそれが生じる確率を以下のように表す。

$$X = \left\{ \begin{array}{cccc} x_1 & , & x_2 & , & \cdots & , & x_n \\ P(x_1) & , & P(x_2) & , & \cdots & , & P(x_n) \end{array} \right\}$$

ここで、 $0 \leq P(x_i) \leq 1 \quad (1 \leq i \leq n)$

$$\sum_{i=1}^n P(x_i) = 1$$

このように、確率が定められた事象の集合を事象系と言う。

13

事象系例

(1) コイントスの事象系

$$C = \left\{ \begin{array}{cc} \text{賭} & , & \text{趙} \\ \frac{1}{2} & , & \frac{1}{2} \end{array} \right\}$$

(2) サイコロの事象系

$$D = \left\{ \begin{array}{cccccc} 1 & , & 2 & , & 3 & , & 4 & , & 5 & , & 6 \\ \frac{1}{6} & , & \frac{1}{6} & , & \frac{1}{6} & , & \frac{1}{6} & , & \frac{1}{6} & , & \frac{1}{6} \end{array} \right\}$$

(3) トランプを引いたときの数の事象系

$$T = \left\{ \begin{array}{cccc} 1 & , & 2 & , & \cdots & , & K \\ \frac{1}{13} & , & \frac{1}{13} & , & \cdots & , & \frac{1}{13} \end{array} \right\}$$

14

練習

次の事象系を形式的に示せ。

(1) トランプを引いたときのカードのマークの事象系

(2) 26文字のアルファベットと空白文字が書かれた、27個の玉が袋に入っている。その袋から1つの玉をとりだす事象系。

15

事象系の平均情報量

定義(エントロピー)

事象系 X のすべての事象に対して、その情報量の平均をとったものを平均情報量またはエントロピーという。

ある事象系 X が以下のように与えられるとする。

$$X = \left\{ \begin{array}{cccc} x_1 & , & x_2 & , & \cdots & , & x_n \\ P(x_1) & , & P(x_2) & , & \cdots & , & P(x_n) \end{array} \right\}$$

このとき、平均情報量 $H(X)$ は自己情報量を元に以下のように表される。

$$\begin{aligned} H(X) &= \sum_{x \in X} P(x) i(x) \\ &= - \sum_{x \in X} P(x) \log P(x) \end{aligned}$$

16

自己情報量と平均情報量

自己情報量は事象に対して定義され、
平均情報量は事象系(事象の集合)に対して定義される。

事象 $\longrightarrow$ **自己情報量**

事象系 $\longrightarrow$ **平均情報量**
(情報源) (エントロピー)

17

平均情報量の計算例

$$\begin{aligned} (1) \quad C = \left\{ \begin{array}{cc} \text{賭} & , & \text{趙} \\ \frac{1}{2} & , & \frac{1}{2} \end{array} \right\} \quad H(C) &= -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} \\ &= \frac{1}{2} \log_2 2 + \frac{1}{2} \log_2 2 \\ &= 1 \end{aligned}$$

$$(2) \quad D = \left\{ \begin{array}{cccccc} 1 & , & 2 & , & 3 & , & 4 & , & 5 & , & 6 \\ \frac{1}{6} & , & \frac{1}{6} & , & \frac{1}{6} & , & \frac{1}{6} & , & \frac{1}{6} & , & \frac{1}{6} \end{array} \right\}$$

$$\begin{aligned} H(D) &= -\frac{1}{6} \log_2 \frac{1}{6} - \frac{1}{6} \log_2 \frac{1}{6} - \frac{1}{6} \log_2 \frac{1}{6} - \frac{1}{6} \log_2 \frac{1}{6} - \frac{1}{6} \log_2 \frac{1}{6} - \frac{1}{6} \log_2 \frac{1}{6} \\ &= 6 \times \frac{1}{6} \log_2 6 \\ &= \log_2 6 \\ &= 2.585 \dots \end{aligned}$$

18

練習1

次の確率事象系(情報源)の平均情報量(エントロピー)を求めよ。

$$(1) \quad A = \left\{ \begin{matrix} a_1 & a_2 \\ \frac{1}{4} & \frac{3}{4} \end{matrix} \right\} \quad (2) \quad B = \left\{ \begin{matrix} b_1 & b_2 & b_3 \\ \frac{1}{10} & \frac{3}{10} & \frac{6}{10} \end{matrix} \right\}$$

$$(3) \quad C = \left\{ \begin{matrix} d_2 & \text{ブヅピ(土)(名)肝ヤヨ※ わ(一)ムキ< ワ蚪} \\ d_3 & \text{ブヅピ(土)(名)肝ヤヨ※ わ< 偽婁ワ蚪} \end{matrix} \right\}$$

19

練習2

次の事象系のエントロピーを求めよ。

(1) トランプを引いたときのカードのマークの事象系

(2) 26文字のアルファベットと空白文字が書かれた、27個の玉が袋に入っている。その袋から1つの玉をとりだす事象系。

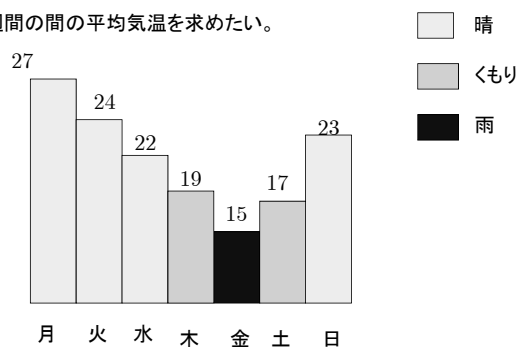
20

補足: 平均(期待値)の話

21

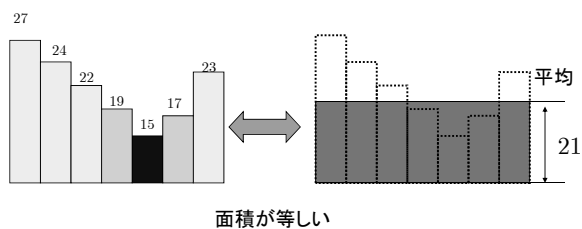
期待値の式

1週間の間の平均気温を求めたい。



22

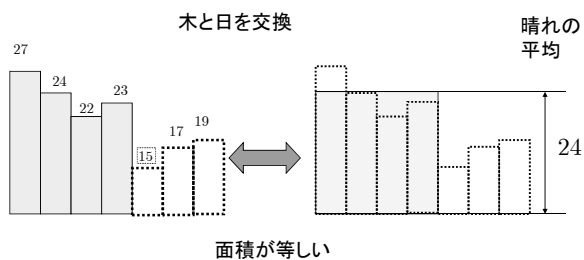
単純平均



$$\bar{T} = \frac{27 + 24 + 22 + 19 + 15 + 17 + 23}{7} = 21$$

23

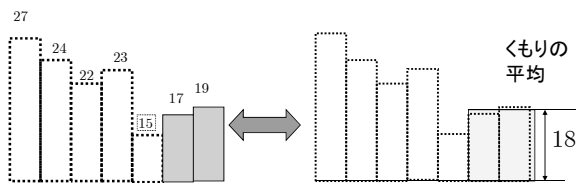
晴れの平均気温



$$\bar{T}_f = \frac{27 + 24 + 22 + 23}{4} = 24$$

24

曇りの平均気温

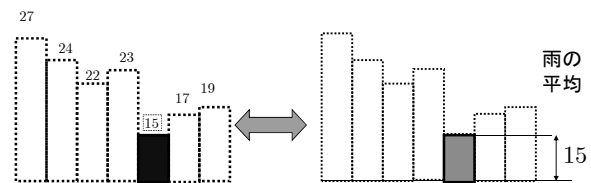


面積が等しい

$$\begin{aligned}\bar{T}_c &= \frac{17 + 19}{2} \\ &= 18\end{aligned}$$

25

雨の平均気温



面積が等しい

$$\begin{aligned}\bar{T}_r &= \frac{15}{1} \\ &= 15\end{aligned}$$

26

よくある間違い

1週間のすべての日は、「晴れ」か「曇り」か「雨」であった。

「晴れ」の日の平均気温は $\bar{T}_f = 24$ 度、

「曇り」の日の平均気温は $\bar{T}_c = 18$ 度、

「雨」の日の平均気温は $\bar{T}_r = 15$ 度であった。

一週間の平均気温を求めよ。

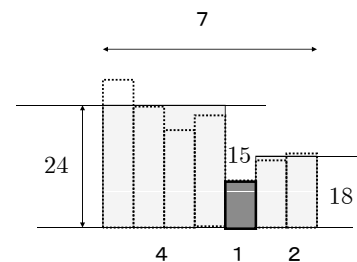
$$\begin{aligned}\bar{T}_{\text{wrong}} &= \frac{\bar{T}_f + \bar{T}_c + \bar{T}_r}{3} \\ &= \frac{24 + 18 + 15}{3} \\ &= 19\end{aligned}$$

✗

この計算は間違い。
(平均の平均を求めるときには、
注意が必要。)
このような計算ができるためには、
どのような条件が必要か？

27

確率



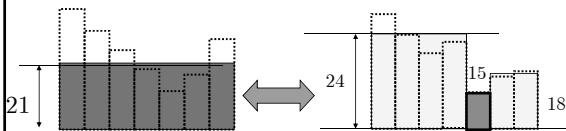
晴れの確率: $P_f = \frac{4}{7}$

雨の確率: $P_r = \frac{1}{7}$

曇りの確率: $P_c = \frac{2}{7}$

28

単純平均と期待値の関係



面積が等しい

$$\begin{aligned}\bar{T} &= P_f \bar{T}_f + P_r \bar{T}_r + P_c \bar{T}_c \\ &= \frac{4}{7} \cdot \frac{27 + 24 + 22 + 23}{4} + \frac{1}{7} \cdot \frac{15}{1} + \frac{2}{7} \cdot \frac{19 + 17}{2} \\ &= \frac{27 + 24 + 22 + 19 + 15 + 17 + 23}{7} \\ &= 21\end{aligned}$$

29

2事象の事象系の平均情報量(重要)

ある事象系 X が以下のように与えられるとする。

$$X = \left\{ \begin{matrix} x & , & \bar{x} \\ p & , & 1-p \end{matrix} \right\}$$

引数が、事象系

このとき、エントロピーは次式となる。

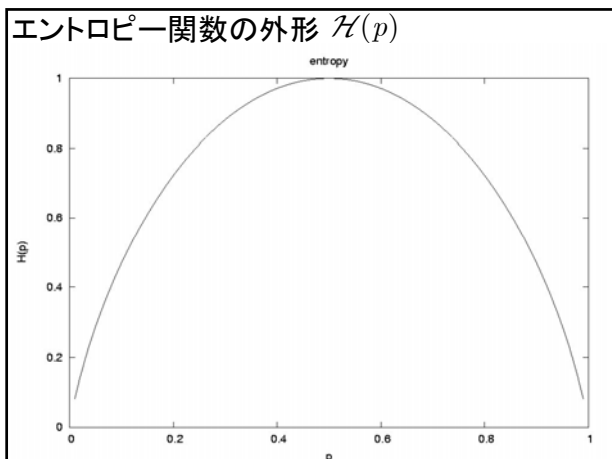
$$H(X) = -p \log p - (1-p) \log(1-p)$$

この形の事象系は非常によく用いられ、この右辺の形の関数をエントロピー関数といい以下のように表す。

$$\mathcal{H}(p) \equiv -p \log p - (1-p) \log(1-p)$$

引数が、数

30



具体例1: 文書の情報量

(1 記号あたりの平均情報量の意味)

アルファベット $\mathcal{A} = \{a, b, c, \dots, z\}$ に対して、
個々の記号の出現確率を考えよう。

記号 $\alpha \in \mathcal{A}$ の出現する確率を $P(\alpha)$ と書く。

例えば、 $P(a)$: a の現れる確率

$P(b)$: b の現れる確率

辞書



このとき、一般の英文において、各記号の出現
確率は異なる。すなわち、以下である。

$$P(a) \neq P(b) \neq \dots \neq P(z) \neq \frac{1}{26}$$

32

まず、アルファベット $\mathcal{A} = \{a, b, c, \dots, z\}$ 中の文字を
用いた文書の情報量を考える。

今、 $\mathcal{A}$ の n 個の文字からなる英文

$$S = s_1 s_2 \dots s_n$$

の情報量を $I(S)$ とする。

$I(S)$ は各記号が S に出現する自己情報量の総和である。

$$I(S) = \sum_{i=1}^n i(s_i)$$

事象に対する
確率の逆数の対数

出現確率に基づく
事象とみなせること
に注意する。

33

今度は逆に、情報量 $I(S)$ を持つ事象を文書 S で表すことを考える。

まず、アルファベットの (1 記号あたりの) 平均情報量 (エントロピー)
が以下の式で表される。

$$\begin{aligned} H(\mathcal{A}) &= \sum_{\alpha=a}^z \{P(\alpha) \times i(\alpha)\} \\ &= P(a)i(a) + P(b)i(b) + \dots + P(z)i(z) \\ &= -P(a) \log P(a) - P(b) \log P(b) - \dots - P(z) \log P(z) \\ &= -\sum_{\alpha=a}^z P(\alpha) \log P(\alpha) \quad [\text{bit/記号}] \end{aligned}$$

34

よって、文書 S 中に含まれる文字数の期待値 $\bar{n}$ は次式で求
められる。

$$\bar{n}[\text{軋噴}] = \frac{I(S)[\text{bit}]}{H(\mathcal{A})[\text{bit/軋噴}]}$$

したがって、文書の長さは情報源 $\mathcal{A}$ の 1 記号あたりの平均
情報量 (エントロピー) $H(\mathcal{A})$ を“単位”とした情報量で
表される。

また、情報源 $\mathcal{A}$ の長さ n の文書 (事象) の平均情報
量 $\bar{I}(S)$ に関して次式が成り立つ。

$$\bar{I}(S)[\text{bit}] = n[\text{軋噴}] \times H(\mathcal{A})[\text{bit/軋噴}]$$

35

練習1

次の事象系のエントロピー (平均情報量) を求めよ。

$$(1) \quad A = \left\{ \begin{matrix} a_1 & a_2 & a_3 \\ \frac{1}{5} & \frac{2}{5} & \frac{2}{5} \end{matrix} \right\}$$

$$(2) \quad B = \left\{ \begin{matrix} b_1 & b_2 & b_3 & b_4 \\ \frac{1}{9} & \frac{2}{9} & \frac{2}{9} & \frac{5}{9} \end{matrix} \right\}$$

36

練習2

情報源(事象系)から生成された文字列の自己情報量を求めよ。

(1)

$$A = \left\{ \begin{matrix} a_1 & a_2 & a_3 \\ \frac{1}{5} & \frac{2}{5} & \frac{2}{5} \end{matrix} \right\}$$

から生成された、文字列 $s_1 = a_3 a_2 a_1 a_3$ の自己情報量 $I(s_1)$

(2)

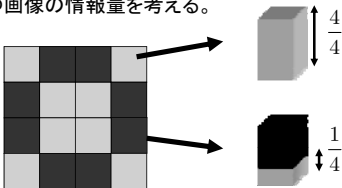
$$B = \left\{ \begin{matrix} b_1 & b_2 & b_3 & b_4 \\ \frac{1}{9} & \frac{2}{9} & \frac{2}{9} & \frac{5}{9} \end{matrix} \right\}$$

から生成された、文字列 $s_2 = b_3 b_4 b_1 b_2$ の自己情報量 $I(s_2)$

37

具体例2: 画像の情報量

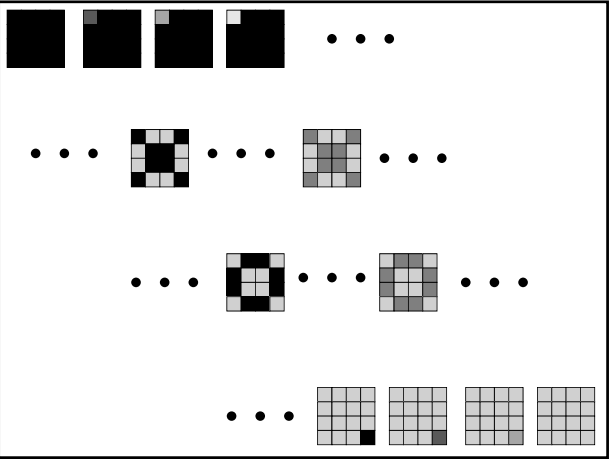
4 × 4 の画素数を持ち、個々の画素が 4 階調の輝度値を持つ画像の情報量を考える。



各画素がある輝度値を持つ確率を、 $\frac{1}{4}$ とする。このとき、この画像が持つ情報量 $I(\text{画})$ は次式で表される。

$$I(\text{画}) = \sum_{j=1}^{16} i(\text{画}_{j,k}) = 16 \times \log 4 = 32[\text{bit}] = 4[\text{byte}]$$

38



前のスライドでは、輝度値が全ての画素で確率が均等であると仮定してあった。
実は、風景画、人物画等では、画素における輝度値の出現確率は一様ではない。
このような場合、より少ない情報量しか持たない。
すなわち、確率が一様でない場合には、1画素あたりの平均情報量は、一様な場合より小さくなる。

40

練習

下の画像の自己情報量を求めよ。ただし、各画素の諧調値はすべて均等に現れるものとする。

(1)

256 × 256 の画素で、各画素が 16 階調の白黒画像の自己情報量を求めよ。

(1)

256 × 256 の画素のカラー画像の自己情報量を求めよ。ただし、1画素は赤、緑、青(RGB)の各色で表され、各色はそれぞれ16階調で表されるものとする。

41