

前回の復習

- 講義の概要
- chapter 1: 情報を測る... エントロピーの定義

- 確率変数 X の(一次)エントロピー

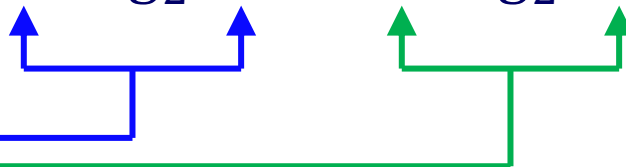
$$H_1(X) = \sum_{i=1}^M -p_i \log_2 p_i \quad (\text{bit})$$



- ◆ M は実現値の個数, p_i は i 番目の実現値が取られる確率

実現値	表	裏
確率	0.5	0.5

$$H_1(X) = -0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1 \text{ bit}$$



練習問題の解答

- 講義 webページにあるデータを使い, エントロピーを計算せよ
 - ◆ 英語の文字出現頻度 ... 約4.18 bit (値の解釈には要注意)
 - ◆ 野球の優勝チーム
 - セ... 2.01, パ... 2.36, 高(春)... 4.42, 高(夏)... 4.17
 - 相撲... 1.98, 単位はいずれも bit

表計算ソフトを使えば簡単に計算できる

	A	B	C	D	E
1	文字	出現頻度(百分率)			
2	a	8.16		0.0816	0.295007
3	b	1.49		0.0149	0.090421
4	c	2.78		0.0278	0.143692

$$-p_i \log_2 p_i$$

	C	D	E
5	1.97	0.0197	0.111614
6	7	0.007	0.007336
7			4.175863

どういう文脈で話が進んでいるか

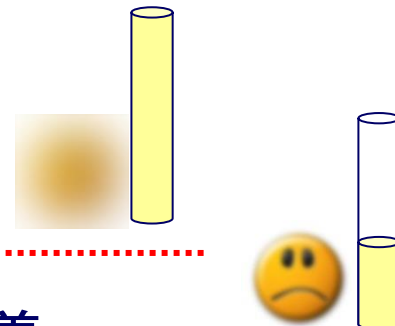
最終目標: 「情報」の量を測る定量的指標を導入する

■ step 1: 確率変数の「エントロピー」を定義

◆ エントロピー大 $\Leftrightarrow$ 不確実さ大

前回

今回

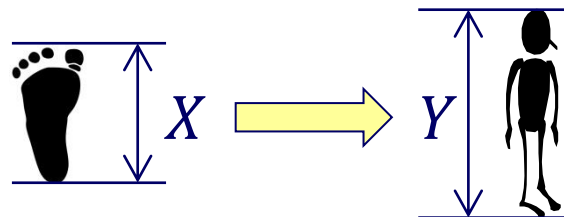


■ step 2: 一つのニュースが持つ情報量を定義

◆ 情報量 = (BEFORE エントロピー) - (AFTER エントロピー)

■ step 3: 確率変数の間の相互情報量を定義

◆ ある確率変数の値が, 他の確率変数について何を語るか



本日の講義(詳細)

- エントロピーの諸性質
- (確率計算に関する復習)
- 関連する概念・量の定義
 - ◆ 結合エントロピー ... 結合確率に対応
 - ◆ 条件付きエントロピー ... 条件付き確率に対応
 - ◆ 相互情報量
- 各種エントロピー・情報量の性質



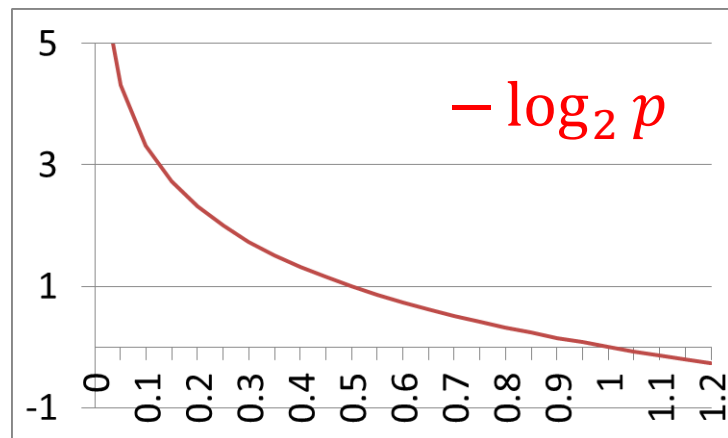
エントロピーの性質(1)

[性質1] $H_1(X) \geq 0$

証明:

$$H_1(X) = \sum_{i=1}^M \boxed{-p_i \log_2 p_i}$$

↑
 $0 \leq p_i \leq 1$ では、
ここは非負

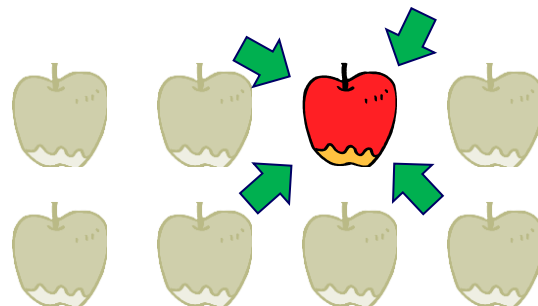


[性質2] $H_1(X) = 0 \Leftrightarrow$ ある i に対し $p_i = 1$, それ以外は $p_j = 0$

証明:

($\Leftarrow$) 定義より明らか

($\Rightarrow$) 背理法を用いて証明



エントロピーの性質(2)

[性質3] X の取り得る値が M 通りならば...

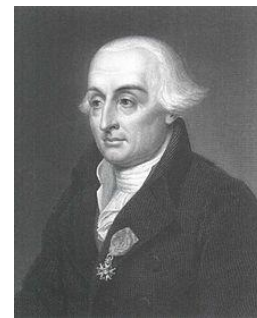
$p_1 = \cdots = p_M = \frac{1}{M}$ のとき $H_1(X)$ は最大, その値は $\log_2 M$ となる

証明: ラグランジュの未定乗数法を用いる

- 目的関数: $H_1(X)$ (M 変数 $p_1, \dots, p_M$ の式と考える)
- 束縛関数: $p_1 + \cdots + p_M - 1 = 0$
- 最大化条件

$$\frac{\partial}{\partial p_i} (H_1(X) + \lambda(p_1 + \cdots + p_M - 1)) = 0$$

これから $p_1 = \cdots = p_M = \frac{1}{M}$ が得られ, そのとき $H_1(X) = \log_2 M$



J. L. Lagrange
1736-1813

エントロピー = 不確実さ

■ $\min H_1(X) = 0$

- ◆ ある i に対し $p_i = 1$, それ以外は $p_j = 0$
- ◆ 何が発生するのか, あらかじめわかっている
- ◆ 不確実な要素が, まったくない

■ $\max H_1(X) = \log_2 M$

- ◆ $p_1 = \dots = p_M = 1/M$
- ◆ どの可能性も, 等しく考えられる
- ◆ 非常に不確実で, 振る舞いの予測がつかない

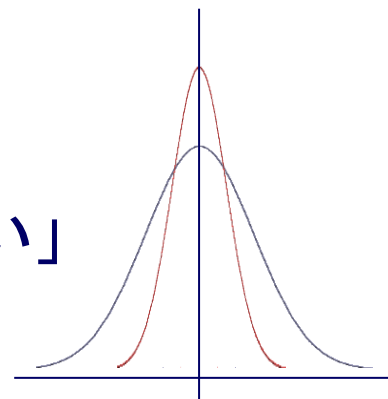


エントロピー = 不確実さ = 予測の難しさ

エントロピー vs. 分散

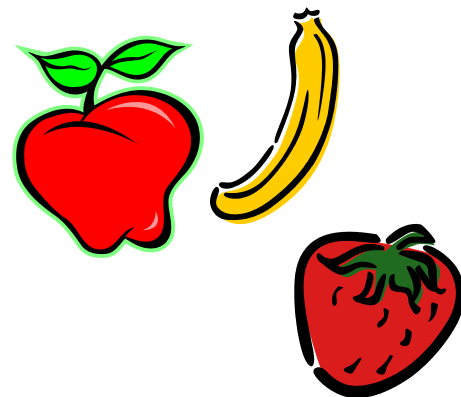
不確実さの指標として、「分散」ではダメなのか？

- ◆ 確率変数 X の分散 $V(X) = E[(X - E[X])^2]$
- ◆ 直観的には「分散が大きい＝ばらつきが大きい」



エントロピーの利点 (vs. 分散)

- ◆ 実現値の上で「演算」ができなくても良い
 - $X = \{\text{りんご, バナナ, いちご}\} \dots$
- ◆ 「工学的な量」と密接に関係
 - 「符号化」の性能の限界を与える



情報理論は, エントロピーの概念を中心に組み立てられている

エントロピーに関するまとめ

確率変数 X のエントロピー:

$$H_1(X) = \sum_{i=1}^M -p_i \log_2 p_i \quad (\text{bit})$$

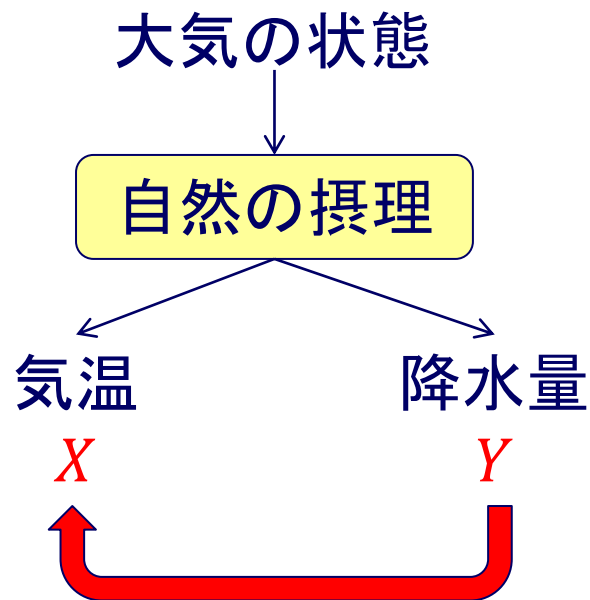


- $H_1(X) \geq 0$...【エントロピーの非負性】
- $\min H_1(X) = 0$...【エントロピーの最小値】
 - ◆ 1個の実現値に対して $p_i = 1$
- $\max H_1(X) = \log_2 M$...【エントロピーの最大値】
 - ◆ $p_1 = \dots = p_M = 1/M$

直観的には...エントロピー大 $\Leftrightarrow$ 不確実さ大

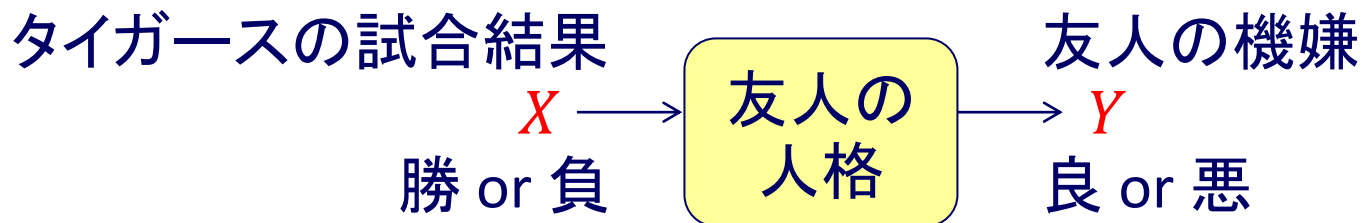
複数の確率変数

- ここまで...確率変数1個に限定
- 「情報の伝達」を考えるには、複数の確率変数が必要



Y の値を知れば,
 X の値に関する情報が得られる
= X の不確かさが減少する

議論すべき「情報の量」



「友人の機嫌が良い」⇒「タイガースは勝った？」
...ここに潜む「情報の伝達」を、数理的に考える

1. 確率論に関する復習
2. **Yの個別の値**が、Xの値について与える**情報量**
 - 「Y = 良」が、Xの値について与える情報量
3. Yの値が、Xの値について与える**情報量の期待値**

同時確率・結合確率

$P_{X,Y}(x,y)$: $X = x$ と $Y = y$ とが同時に発生する確率

例: 過去 100 日間の, 試合結果(X) と友人の機嫌 (Y) の統計

$X \backslash Y$	良	悪
勝	45	12
負	15	28

- 勝って, 機嫌が良かった... 45日
- 勝ったのに, 機嫌が悪かった... 12日
- ...

$P_{X,Y}(\text{勝}, \text{良}) = 0.45$ $P_{X,Y}(\text{勝}, \text{悪}) = 0.12$
 $P_{X,Y}(\text{負}, \text{良}) = 0.15$ $P_{X,Y}(\text{負}, \text{悪}) = 0.28$

同時確率, 結合確率, と呼ばれる

確率の周辺化

同時確率からは、他の様々な確率を導き出せる

$X \backslash Y$	良	悪	
勝	45	12	57
負	15	28	33
	60	40	100

■ 勝ったのは $45+12=57$ 日

$$\begin{aligned} P_X(\text{勝}) &= P_{X,Y}(\text{勝}, \text{良}) + P_{X,Y}(\text{勝}, \text{悪}) \\ &= 0.45 + 0.12 = 0.57 \end{aligned}$$

■ 一般には、
$$P_X(x) = \sum_{y \in D(Y)} P_{X,Y}(x, y)$$

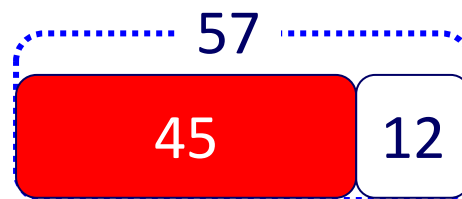
$$P_Y(y) = \sum_{x \in D(X)} P_{X,Y}(x, y)$$

...確率の**周辺化** (marginalize) と呼ばれる操作

条件付き確率

$P_{Y|X}(y|x)$: $X = x$ の条件のもとで, $Y = y$ となる確率

$X \backslash Y$	良	悪	
勝	45	12	57
負	15	28	33
	60	40	100



$$P_{Y|X}(\text{良}|\text{勝})$$

= 57の中での45の割合

試合に勝った日は (条件) ▶ $45/57 = 0.79$ の確率で機嫌が良い
▶ $12/57 = 0.21$ の確率で機嫌が悪い

■ 一般には, $P_{Y|X}(y|x) = \frac{P_{X,Y}(x,y)}{P_X(x)}$... **ベイズの定理**

$P_{Y|X}(y|x)$ と $P_{X|Y}(x|y)$ を混同しないこと

条件付き確率に関する注意

$P_{Y|X}(y|x)$ と $P_{X|Y}(x|y)$ を混同しないこと

- 試合に勝つ確率

$$P_X(\text{勝}) = 0.45 + 0.12 = 0.57$$

- 機嫌が良い確率

$$P_Y(\text{良}) = 0.45 + 0.15 = 0.60$$

X \ Y	良	悪	
勝	45	12	57
負	15	28	43
	60	40	100

$$\blacksquare P_{Y|X}(\text{良}|\text{勝}) = \frac{P_{X,Y}(\text{勝}, \text{良})}{P_X(\text{勝})} = \frac{0.45}{0.57} = 0.79$$

$$\blacksquare P_{X|Y}(\text{勝}|\text{良}) = \frac{P_{X,Y}(\text{勝}, \text{良})}{P_Y(\text{良})} = \frac{0.45}{0.60} = 0.75$$

確率変数の独立性

確率変数 X, Y が独立

$\Leftrightarrow$ 任意の x, y に対し $P_{X,Y}(x, y) = P_X(x)P_Y(y)$

$\Leftrightarrow$ 任意の x, y に対し $P_{Y|X}(y|x) = P_Y(y)$

$\Leftrightarrow$ 任意の x, y に対し $P_{X|Y}(x|y) = P_X(x)$



独立でない $\Rightarrow$ 従属関係にある

(どちらかが主で, どちらかが従, というわけでは**ない**点に注意)

同時エントロピー・結合エントロピー

X と Y の同時エントロピー, 結合エントロピー;

$$H_1(X, Y) = \sum_{x \in D(X)} \sum_{y \in D(Y)} -P_{X,Y}(x, y) \log_2 P_{X,Y}(x, y).$$

$X \backslash Y$	良	悪
勝	45	12
負	15	28

$$\begin{aligned} H_1(X, Y) = & -0.45 \log_2 0.45 \\ & -0.12 \log_2 0.12 \\ & -0.15 \log_2 0.15 \\ & -0.28 \log_2 0.28 = 1.81 \text{bit} \end{aligned}$$

X の値と Y の値とを同時に予測する「難しさ」に相当

結合エントロピーの性質

補題: $H_1(X, Y) \leq H_1(X) + H_1(Y)$

証明:

$$H_1(X) = \sum_{x \in D(X)} -P_X(x) \log_2 P_X(x) = \sum_{x \in D(X)} \sum_{y \in D(Y)} -P_{X,Y}(x, y) \log_2 P_X(x)$$

$$H_1(Y) = \sum_{y \in D(Y)} -P_Y(y) \log_2 P_Y(y) = \sum_{x \in D(X)} \sum_{y \in D(Y)} -P_{X,Y}(x, y) \log_2 P_Y(y)$$

右辺

$$H_1(X) + H_1(Y) = \sum_{x \in D(X)} \sum_{y \in D(Y)} -P_{X,Y}(x, y) (\log_2 P_X(x) + \log_2 P_Y(y))$$

$$= \sum_{x \in D(X)} \sum_{y \in D(Y)} -P_{X,Y}(x, y) \log_2 P_X(x) P_Y(y)$$

左辺

$$H_1(X, Y) = \sum_{x \in D(X)} \sum_{y \in D(Y)} -P_{X,Y}(x, y) \log_2 P_{X,Y}(x, y)$$

微妙に
違う

シャノンの補助定理

シャノンの補助定理, *Shannon's lemma* を導入

[補題]

$p_1 + \cdots + p_M = 1, q_1 + \cdots + q_M \leq 1$ を満たす非負数 p_i, q_i に対し,

$$\sum_{i=1}^M -p_i \log_2 q_i \geq \sum_{i=1}^M -p_i \log_2 p_i$$

等号成立は, すべての i に対して $p_i = q_i$ のとき

補助定理の証明(概略)

左辺 - 右辺 =

$$\sum_{i=1}^M -p_i \log_2 q_i + \sum_{i=1}^M p_i \log_2 p_i = \sum_{i=1}^M -p_i \log_2 \frac{q_i}{p_i} = \sum_{i=1}^M \frac{p_i}{\log_e 2} (-\log_e \frac{q_i}{p_i})$$

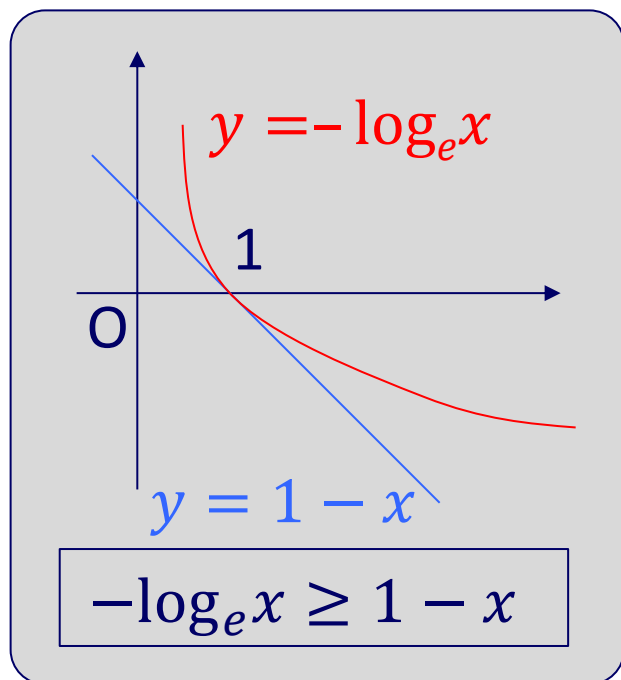
$$\geq \sum_{i=1}^M \frac{p_i}{\log_e 2} \left(1 - \frac{q_i}{p_i}\right) = \frac{1}{\log_e 2} \sum_{i=1}^M (p_i - q_i)$$

$$= \frac{1}{\log_e 2} \left(\sum_{i=1}^M p_i - \sum_{i=1}^M q_i\right) = \frac{1}{\log_e 2} \left(1 - \sum_{i=1}^M q_i\right)$$

$$\geq 0$$

等号成立

$\Leftrightarrow$ 全ての i に対し $q_i/p_i = 1$ のとき



結合エントロピーの性質

補題: $H_1(X, Y) \leq H_1(X) + H_1(Y)$

証明:

右 $H_1(X) + H_1(Y) = \sum_{x \in D(X)} \sum_{y \in D(Y)} -P_{X,Y}(x, y) \log_2 P_X(x) P_Y(y)$

左 $H_1(X, Y) = \sum_{x \in D(X)} \sum_{y \in D(Y)} -P_{X,Y}(x, y) \log_2 P_{X,Y}(x, y)$

シャノンの補助定理

(証明終了)

系: 確率変数 X, Y が独立なら $H_1(X, Y) = H_1(X) + H_1(Y)$

例で確かめてみる

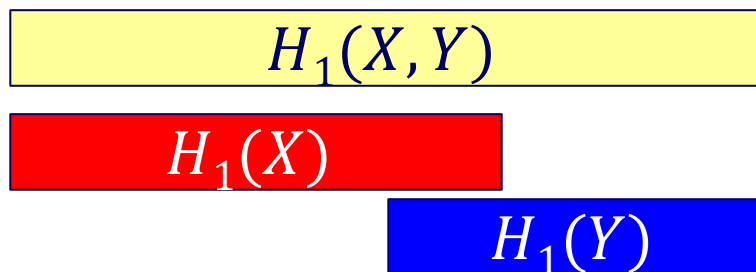
$X \backslash Y$	良	悪	
勝	45	12	57
負	15	28	33
	60	40	100

$$\begin{aligned} H_1(X, Y) &= -0.45 \log_2 0.45 \\ &\quad -0.12 \log_2 0.12 \\ &\quad -0.15 \log_2 0.15 \\ &\quad -0.28 \log_2 0.28 = 1.81 \text{ bit} \end{aligned}$$

- $P_X(\text{勝}) = 0.57, P_X(\text{負}) = 0.43$
 $\Rightarrow H_1(X) = -0.57 \log_2 0.57 - 0.43 \log_2 0.43 = 0.99 \text{ bit}$
- $P_Y(\text{良}) = 0.60, P_Y(\text{悪}) = 0.40$
 $\Rightarrow H_1(Y) = -0.60 \log_2 0.60 - 0.40 \log_2 0.40 = 0.97 \text{ bit}$
- $H_1(X, Y) = 1.81 < 1.96 = H_1(X) + H_1(Y)$

$H_1(X, Y) \leq H_1(X) + H_1(Y)$ の意味

- $H_1(X, Y)$... X の値と Y の値を同時に予測する難しさ
- $H_1(X) + H_1(Y)$... X の値と Y の値を別々に予測する難しさ
- $H_1(X, Y) \leq H_1(X) + H_1(Y)$
 - ◆ 同時に予測するほうが, 別々に予測するよりも簡単
 - ◆ Y の値を決めれば, X の値もある程度限定される
 $\Rightarrow X$ の不確実さが, 少し減少する
 - ◆ Y の値の中には, X の値に関する情報が含まれている



友人の機嫌と情報量

$X \backslash Y$	良	悪
勝	45	12
負	15	28

- 友人の機嫌が良い $\Rightarrow$ 試合に勝った？
... 友人の機嫌が、試合結果に関する
情報を与えてくれる

- 友人の機嫌を知る前...

$$P_X(\text{勝}) = 0.57, P_X(\text{負}) = 0.43$$

- 友人の機嫌が良いのを見た後...

- ◆ $P_Y(\text{良}) = 0.45 + 0.15 = 0.60$

- ◆ $P_{X|Y}(\text{勝} \mid \text{良}) = 0.45/0.60 = 0.75$: 「 $X = \text{勝}$ 」の確率 up

- ◆ $P_{X|Y}(\text{負} \mid \text{良}) = 0.15/0.60 = 0.25$

エントロピーは？

個別値による条件付きエントロピー

- $Y = y$ のときのエントロピーを以下で定義

$$H_1(X|Y = y) = \sum_{x \in D(X)} -P_{X|Y}(x|y) \log_2 P_{X|Y}(x|y)$$

- 前ページの例では

$$H_1(X | Y = \text{良}) = -0.75 \log_2 0.75 - 0.25 \log_2 0.25 = 0.81$$

before $P_X(\text{勝}) = 0.57$

$P_X(\text{負}) = 0.43$

after $P_{X|Y}(\text{勝} | \text{良}) = 0.75$

$P_{X|Y}(\text{負} | \text{良}) = 0.25$

$H_1(X) = 0.99 \longrightarrow H_1(X | Y = \text{良}) = 0.81$

$$0.99 - 0.81 = 0.18 \text{ bit}$$

...「友人の機嫌が良い」ことを知って解消された不確かさ

...「友人の機嫌が良い」ことから得られる情報量

友人の機嫌が悪いときは...

■ $P_Y(\text{悪}) = 0.12 + 0.28 = 0.40$

◆ $P_{X|Y}(\text{勝} | \text{悪}) = 0.12/0.40 = 0.30$

◆ $P_{X|Y}(\text{負} | \text{悪}) = 0.28/0.40 = 0.70$

◆ $H_1(X | Y = \text{悪}) = -0.30 \log_2 0.30 - 0.70 \log_2 0.70 = 0.88$

X \ Y	良	悪
勝	45	12
負	15	28

before $P_X(\text{勝}) = 0.57$
 $P_X(\text{負}) = 0.43$

 $H_1(X) = 0.99$

after $P_{X|Y}(\text{勝} | \text{悪}) = 0.30$
 $P_{X|Y}(\text{負} | \text{悪}) = 0.70$

 $H_1(X | Y = \text{悪}) = 0.88$

$0.99 - 0.88 = 0.11 \text{ bit}$

...「友人の機嫌が悪い」ことを知って解消された不確かさ

...「友人の機嫌が悪い」ことから得られる情報量

「平均的」な情報量

■ 「友人の機嫌が良い」

- ◆ 確率 $P_Y(\text{良}) = 0.60$ で発生する事象
- ◆ $H_1(X) - H_1(X | Y = \text{良}) = 0.99 - 0.81 = 0.18 \text{ bit}$
- ◆ 情報量は 0.18bit

■ 「友人の機嫌が悪い」

- ◆ 確率 $P_Y(\text{悪}) = 0.40$ で発生する事象
- ◆ $H_1(X) - H_1(X | Y = \text{悪}) = 0.99 - 0.88 = 0.11 \text{ bit}$
- ◆ 情報量は 0.11bit

■ Y の値がもたらす, X に関する情報量の期待値は

$$0.60 \times 0.18 + 0.40 \times 0.11 = 0.152 \text{ bit} \dots X \text{と} Y \text{の相互情報量}$$

相互情報量, 条件付きエントロピー

■ X と Y の相互情報量

$$\begin{aligned} I(X; Y) &= \sum_{y \in D(Y)} P_Y(y) (H(X) - H(X | Y = y)) \\ &= H(X) - \sum_{y \in D(Y)} P_Y(y) H(X | Y = y) \end{aligned}$$

■ X の Y による条件付きエントロピー

$$H_1(X|Y) = \sum_{y \in D(Y)} P_Y(y) H_1(X | Y = y) \quad \leftarrow$$

「個別値による条件付きエントロピー」の期待値

例で確認

- 「友人の機嫌が良い」

- ◆ 確率 $P_Y(\text{良}) = 0.60$

- ◆ $H_1(X | Y = \text{良}) = 0.81$

- 「友人の機嫌が悪い」

- ◆ 確率 $P_Y(\text{悪}) = 0.40$

- ◆ $H_1(X | Y = \text{悪}) = 0.88$

- 条件付きエントロピー

$$H_1(X|Y) = 0.60 \times 0.81 + 0.40 \times 0.88 = 0.838 \text{ bit}$$

- $H_1(X) = 0.99 \text{ bit}$

- 相互情報量

$$I(X; Y) = H(X) - H(X|Y)$$

$$= 0.99 - 0.838 = 0.152 \text{ bit}$$

条件付きエントロピーの性質(1)

補題: $H_1(X|Y) = H_1(X, Y) - H_1(Y)$

証明:

$$\begin{aligned} H(X|Y) &= \sum_{y \in D(Y)} P_Y(y) \sum_{x \in D(X)} -P_{X|Y}(x|y) \log_2 P_{X|Y}(x|y) \\ &= \sum_{y \in D(Y)} \sum_{x \in D(X)} -P_{X,Y}(x, y) (\log_2 P_{X,Y}(x, y) - \log_2 P_Y(y)) \\ &= H_1(X, Y) - \sum_{y \in D(Y)} \left(\sum_{x \in D(X)} -P_{X,Y}(x, y) \right) \log_2 P_Y(y) \\ &= H_1(X, Y) - \sum_{y \in D(Y)} -P_Y(y) \log_2 P_Y(y) \\ &= H_1(X, Y) - H_1(Y) \end{aligned}$$

$P_{X|Y}(x|y) = P_{X,Y}(x, y)/P_Y(y)$

周辺化計算

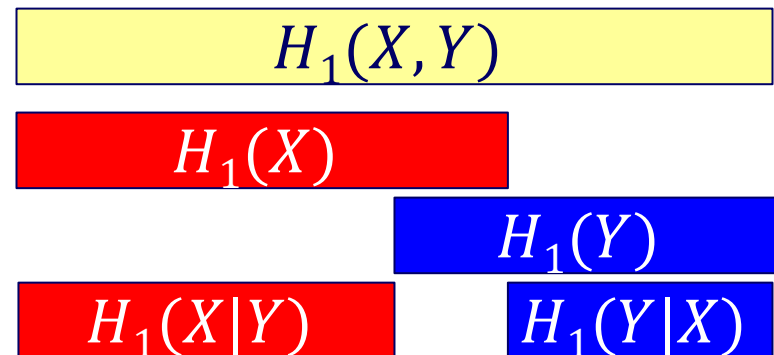
条件付きエントロピーの性質(2)

前ページの補題: $H_1(X|Y) = H_1(X, Y) - H_1(Y)$

系: $H_1(X, Y) = H_1(Y) + H_1(X|Y)$
 $= H_1(X) + H_1(Y|X)$

証明:

$H_1(X, Y)$ は, 変数 X, Y について
対称であるため



相互情報量の性質(1)

系: $I(X; Y) = I(Y; X)$

証明: $H_1(X, Y) = H_1(Y) + H_1(X|Y) = H_1(X) + H_1(Y|X)$ より

$$\begin{aligned} I(X; Y) &= H_1(X) - H_1(X|Y) \\ &= H_1(Y) - H_1(Y|X) \\ &= I(Y; X) \end{aligned}$$

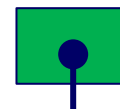
$$H_1(X, Y)$$

$$H_1(X)$$

$$H_1(Y)$$

$$H_1(X|Y)$$

$$H_1(Y|X)$$



$$I(X; Y) = I(Y; X)$$

Y が X について教えてくれる情報量

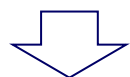
||

X が Y について教えてくれる情報量

相互情報量の性質(2)

p.18 の補題: $H_1(X, Y) \leq H_1(X) + H_1(Y)$

p.30の補題: $H_1(X|Y) = H_1(X, Y) - H_1(Y)$



$$H_1(X|Y) = H_1(X, Y) - H_1(Y) \leq H_1(X) + H_1(Y) - H_1(Y) = H_1(X)$$

系: $I(X; Y) \geq 0$, 等号成立は X, Y が独立のとき

証明: $I(X; Y) = H_1(X) - H_1(X|Y) \geq 0$

- ▶ Y の値を知ることで、失うものは何もない
- ▶ X と Y が独立なら、 Y の値を知っても得るものはない

相互情報量について, まとめ

右図で表現されていることが全て

$$H_1(X, Y)$$

$$H_1(X)$$

$$H_1(Y)$$

$$H_1(X|Y)$$

$$H_1(Y|X)$$

たとえば...

■ 相互情報量の計算法は3通りある

$$1. \quad I(X; Y) = H_1(X) + H_1(Y) - H_1(X, Y)$$

$$2. \quad I(X; Y) = H_1(X) - H_1(X|Y)$$

$$3. \quad I(X; Y) = H_1(Y) - H_1(Y|X)$$



$$I(X; Y) = I(Y; X)$$

本日のまとめ

- エントロピーと, それに関連する概念
 - ◆ 結合, 条件付きエントロピー
 - ◆ 相互情報量

練習問題

- X : タイガースの試合結果, $P_X(w) = P_X(d) = P_X(l) = 1/3$
- Y : 阪神ファンの友人のtweet

X	Y
w	やったー
d	くやしー
l	くやしー



- ▶ p.13 のように同時確率の表を書き, 周辺確率も求めよ
- ▶ p.34 に示した3つの異なる方法で, 相互情報量 $I(X; Y)$ を求めよ