スケジュール

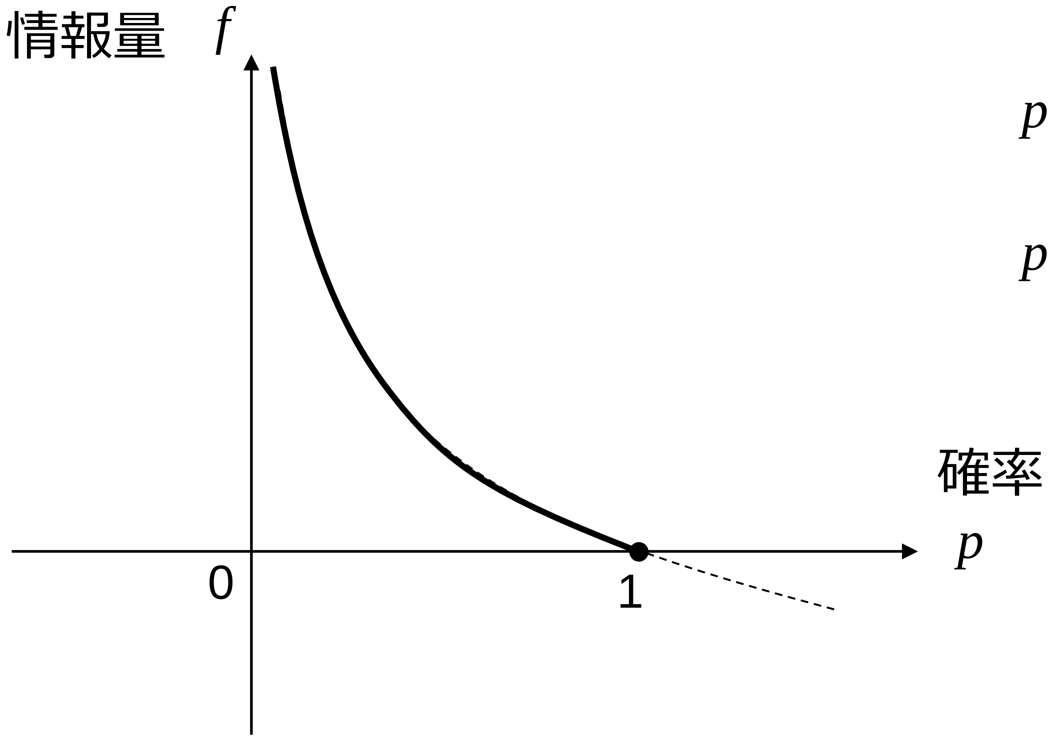
1. 情報理論の概要・情報の表現　4/14 done
2. 確率の基礎　4/21 一部持ち越し
3. 情報量（エントロピー、ダイバージェンス、相互情報量） 4/28
4. 情報量の性質　5/2　(5/4祝日)
5. 演習　5/11　（羽石不在）
6. 情報源のモデルとエントロピーレート 5/19
7. 情報源の符号化　5/26
8. 中間テスト 6/2　（羽石不在）
9. 相互情報量（１）基礎　6/9
10. 相互情報量（２）応用　6/16
11. 演習　6/23　（羽石不在？）
12. 情報量統計学（１）最尤法 6/30
13. 情報量統計学（２）ＡＩＣ 7/7
14. 情報理論の応用（１）ベイズ推定 7/14
15. 情報理論の応用（２）パターン認識　7/21
16. 期末テスト 7/28

情報量

情報量の定義：

$$f = -\log_2 p$$

⇔確率が低ければ、情報量は大きい。



確率	情報量
$p = \frac{1}{2} \Rightarrow -\log_2 p = 1$	
$p = \frac{1}{4} \Rightarrow -\log_2 p = 2$	
$p = \frac{1}{8} \Rightarrow -\log_2 p = 3$	
$p = \frac{1}{16} \Rightarrow -\log_2 p = 4$	

エントロピー：平均情報量

エントロピー（平均情報量）：

$$H = \sum_{k=1}^{n} f_k p_k = \sum_{k=1}^{n} (-\log_2 p_k) \cdot p_k = -\sum_{k=1}^{n} p_k \log_2 p_k$$

単位：ビット(bit)

演習3-1

次の２つの確率分布についてそれぞれエントロピーを求めよ。

1）
$$p_1 = \frac{1}{4}, p_2 = \frac{1}{4}, p_3 = \frac{1}{4}, p_4 = \frac{1}{4}$$

2）
$$p_1 = \frac{1}{2}, p_2 = \frac{1}{4}, p_3 = \frac{1}{8}, p_4 = \frac{1}{8}$$

$0\log_2 0 = 0$ の導出

$v = \lim_{x \to +0} x \log x \quad \cdots (1)$ が0になることを以下のように導く（logの底は2）

$u := \frac{1}{x} \Leftrightarrow x = \frac{1}{u}$ とおく。このとき $x \to +0 \Leftrightarrow u \to \infty$

となる。このとき式（１）は

$$v = \lim_{u \to \infty} \frac{1}{u} \log \frac{1}{u} = \lim_{u \to \infty} \frac{1}{u} (-\log u) = -\lim_{u \to \infty} \frac{\log u}{u}$$

ここで $y = \log u \Leftrightarrow 2^y = u$ とおくと $u \to \infty \Leftrightarrow y \to \infty$

$$v = -\lim_{u \to \infty} \frac{\log u}{u} = -\lim_{u \to \infty} \frac{y}{2^y} = 0$$

$f(y) = y$と$f(y) = 2^y$とを比べると
後者の方が増加が速いから。

したがって

$$v = \lim_{x \to +0} x \log x = -\lim_{u \to \infty} \frac{\log u}{u} = -\lim_{u \to \infty} \frac{y}{2^y} = 0$$

エントロピーの非負性

$0 \leq p_k \leq 1, k = 1,2,\ldots,n$ より $-p_k \log_2 p_k \geq 0$ が成り立つ。

よって、ただちに

$$H(X) == \sum_{k=1}^{n} (-p_k \log_2 p_k) \geq 0$$

が成り立つ。

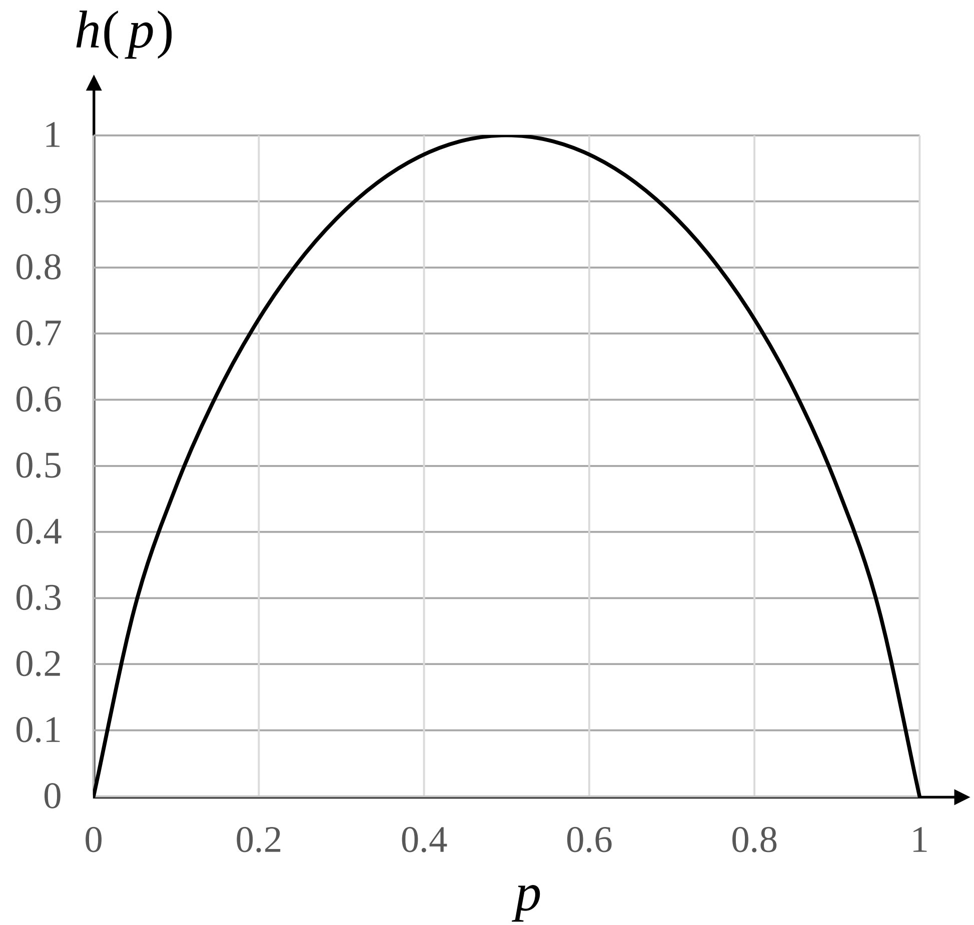
２元エントロピー関数

標本空間$A = \{0,1\}$において、
　値０の生起確率：p
　値１の生起確率：$1 - p$

このときエントロピーは

$$h(p) = -p \log p - (1-p)\log(1-p)$$

と書ける。
これを２元エントロピーという。



同時確率

確率変数が２つある場合を考える。

例１）サイコロを２回振る。１回めの目をx, ２回めの目をyとする。

このとき、(x, y)が発生する確率$p(x, y)$を同時確率という。この場合の同時確率は以下のように行列で表すことができる。

$$
\begin{array}{cc|c}
 & & y \\
 & & \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\
\hline
x & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} & \begin{bmatrix}
1/36 & 1/36 & 1/36 & 1/36 & 1/36 & 1/36 \\
1/36 & 1/36 & 1/36 & 1/36 & 1/36 & 1/36 \\
1/36 & 1/36 & 1/36 & 1/36 & 1/36 & 1/36 \\
1/36 & 1/36 & 1/36 & 1/36 & 1/36 & 1/36 \\
1/36 & 1/36 & 1/36 & 1/36 & 1/36 & 1/36 \\
1/36 & 1/36 & 1/36 & 1/36 & 1/36 & 1/36
\end{bmatrix}
\end{array}
$$

同時確率

例２）サイコロを２回振る。１回めに出る目を確率変数xとする。
一方２つめの確率変数は以下のように与える。すなわち

１回めの目が奇数だった場合は２回めの目をそのまま確率変数yとする
１回めの目が偶数mだった場合は、

２回めの目が３以下のとき　$y = m - 1$

２回めの目が３を超えるとき　$y = m$

演習3-2

		y					
		1	2	3	4	5	6
	1						
	2						
	3						
x	4						
	5						
	6						

6×６の同時確率の配列を埋めよ

同時確率と周辺確率の関係

周辺確率 $\boxed{P_X(x)} = -\sum_{y \in B} \boxed{P(x,y)}, \quad x \in A$ 同時確率

		y					
		1	2	3	4	5	6
	1	p_{11}	p_{12}	p_{13}	p_{14}	p_{15}	p_{16}
	2	p_{21}	p_{22}	p_{23}	p_{24}	p_{25}	p_{26}
	3	p_{31}	p_{31}	$\ddots$			$\vdots$
x	4	p_{41}	p_{31}		同時確率		
	5	p_{51}	p_{31}				
	6	p_{61}	p_{31}	$\cdots$			p_{66}

周辺確率

$$P_X(X=1) = \sum_{y=1}^{6} P(X=1, Y=y) = \sum_{j=1}^{6} p_{1j}$$

行内の和をとる

$$P_X(X=1) = \sum_{y=1}^{6} P(X=1, Y=y) = \sum_{j=1}^{6} p_{1j}$$

列内の和をとる

$$P_Y(Y=2) = \sum_{x=1}^{6} P(X=x, Y=2) = \sum_{i=1}^{6} p_{i2}$$

周辺確率：演習

演習3-3

先の演習3-2の周辺確率を求めよ。

条件付確率

同時確率

条件付き確率 $P(y|x)=\dfrac{P(x,y)}{P(x)}=\dfrac{P(x,y)}{\sum_y p(x,y)}$

周辺確率

$$
\begin{array}{cc|cccccc}
 & & & & y & & & \\
 & & 1 & 2 & 3 & 4 & 5 & 6 \\
\hline
 & 1 & & & & & & \\
 & 2 & & & & & & \\
 & 3 & & & & & & \\
x & 4 & & & & & & \\
 & 5 & & & & & & \\
 & 6 & & & & & &
\end{array}
\left[\begin{array}{cccccc}
\boxed{p_{11} \; \boxed{p_{12}} \; p_{13} \; p_{14} \; p_{15} \; p_{16}} \\
p_{21} & p_{22} & p_{23} & p_{24} & p_{25} & p_{26} \\
p_{31} & p_{32} & \ddots & & & \vdots \\
p_{41} & p_{42} & & & & \\
p_{51} & p_{52} & & & & \\
p_{61} & p_{62} & \cdots & & & p_{66}
\end{array}\right]
$$

周辺確率

同時確率

$$\boxed{P(X=1)=\sum_{j=1}^{6} p_{1j}}$$

条件 y

$$
P(y|x) \quad
\begin{array}{cc|cccccc}
 & & 1 & 2 & 3 & 4 & 5 & 6 \\
\hline
 & 1 & & & & & & \\
 & 2 & & & & & & \\
 & 3 & & & & & & \\
x & 4 & & & & & & \\
 & 5 & & & & & & \\
 & 6 & & & & & &
\end{array}
\left[\begin{array}{cccccc}
p_{C11} & \boxed{p_{C12}} & p_{C13} & p_{C14} & p_{C15} & p_{C16} \\
p_{C21} & p_{C22} & p_{C23} & p_{C24} & p_{C25} & p_{C26} \\
p_{C31} & p_{C32} & \ddots & & & \vdots \\
p_{C41} & p_{C42} & & & & \\
p_{C51} & p_{C52} & & & & \\
pC_{61} & p_{C62} & \cdots & & & pC_{66}
\end{array}\right]
$$

条件付き確率

$$p_{C12}=\frac{p_{12}}{P(X=1)}$$

条件付き確率:演習

演習3-4　先の演習3-2の条件付き確率$P(x|y)$を求めよ

		y					
		1	2	3	4	5	6
	1						
	2						
	3						
x	4						
	5						
	6						

6× 6 の条件付き確率の配列を埋めよ

同時エントロピー

直積$A \times B$を標本空間とする確率変数(X, Y)が同時確率$P(x, y)$を有するとき、この確率変数の同時エントロピーを以下の式で定義する。

$$H(X,Y) = -\sum_{x \in A} \sum_{y \in B} P(x,y) \log P(x,y)$$

演習3-5

直積演習３-2の同時確率について、同時エントロピーを計算せよ。
計算手段は何でもよい。たとえばExcelでもよい。

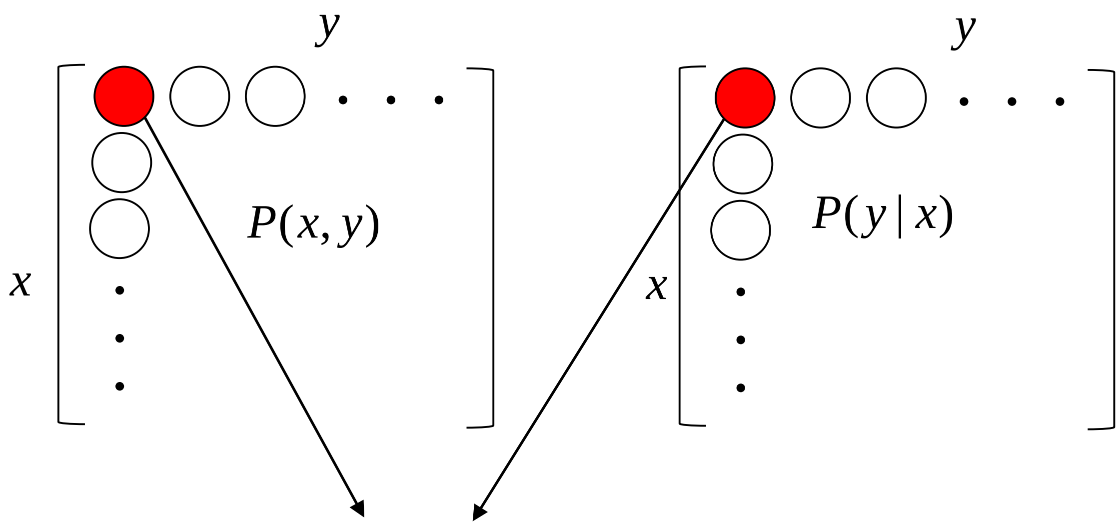
条件付エントロピー

直積$A \times B$を標本空間とする確率変数(X, Y)が同時確率$P(x, y)$を有するとき、Xに対するYの条件付エントロピー$H(Y|X)$を以下の式で定義する。

$$H(Y \mid X) = -\sum_{x \in A} \sum_{y \in B} P(x, y) \log P(y \mid x)$$

Ｘを知ったときの
Ｙのあいまいさを表す

計算方法



対応する２つの確率の積和演算

演習3-6

直積演習３-2の同時確率について、条件付きエントロピーを計算せよ。
計算手段は何でもよい。たとえばExcelでもよい。

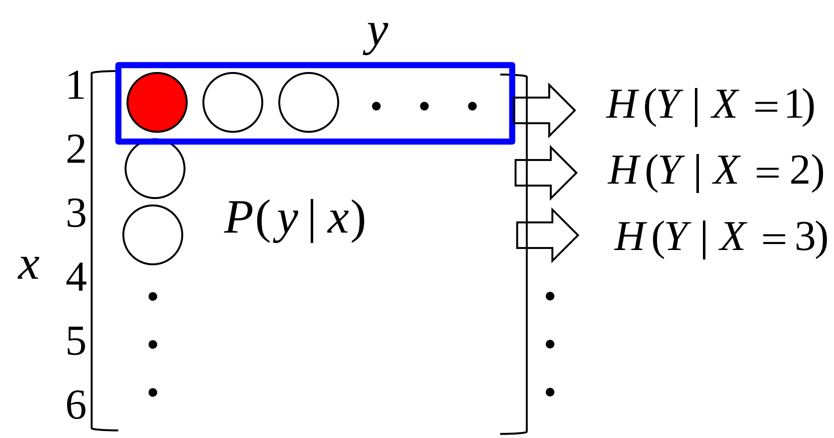
条件付エントロピー

$X = x$が与えられたときの条件付き確率$P(y|x)$のエントロピーを
$H(Y|X = x)$と書けば、

$$H(Y \mid X = x) = -\sum_{y \in B} P(y \mid x) \log P(y \mid x)$$

と書ける。一方、前ページに定義された
条件付きエントロピーは
以下のように式変形できる。

$$\begin{aligned}
H(Y \mid X) &= -\sum_{x \in A} \sum_{y \in B} P(x, y) \log P(y \mid x) \\
&= -\sum_{x \in A} \sum_{y \in B} P(x) P(y \mid x) \log P(y \mid x) \\
&= \sum_{x \in A} P(x) \left(-\sum_{y \in B} P(y \mid x) \log P(y \mid x) \right) \\
&= \sum_{x \in A} P(x) H(Y \mid X = x)
\end{aligned}$$



すなわち、Xに対するYの条件付エントロピー$H(Y|X)$は、$X = x$を知ったときのＹの条件付きエントロピー$H(Y|X = x)$のXに関する平均とも考えられる。

条件付エントロピー

Xに対するYの条件付エントロピー$H(Y|X)$

$$H(Y \mid X) = -\sum_{x \in A} \sum_{y \in B} P(x, y) \log P(y \mid x)$$

XとYを入れ替えれば以下も表現もできる。

Yに対するXの条件付エントロピー$H(X|Y)$

$$H(X \mid Y) = -\sum_{x \in A} \sum_{y \in B} P(x, y) \log P(x \mid y)$$

Kullback-Leibler(KL)ダイバージェンス

標本空間Ａ上に値をとる２つの確率分布Ｐ、Ｑに対して、ＫＬダイバージェンス（またはＫＬ情報量、または相対エントロピー）$D(P \parallel Q)$を以下の式で定義する。

$$D(P \| Q) = -\sum_{x \in A} P(x) \log \frac{P(p)}{Q(x)}$$

ＫＬ情報量は以下のように書くこともできる。

$$D(P \| Q) = -\sum_{x \in A} P(x)(\log P(x) - \log Q(x))$$

これより、２つの情報量の差の平均値と解釈できる。

相互情報量

同時確率$P(x,y)$を有する２つの確率変数X,Yを考える。このとき相互情報量$I(X;Y)$を以下の式で定義する。

$$I(X;Y) = -\sum_{x\in A}\sum_{y\in B} P(x,y)\log\frac{P(x,y)}{P_X(x)P_Y(y)}$$

ただし、$P_X(x)$と$P_Y(y)$は周辺確率を表す。

$$D(P\,\|\,Q) = -\sum_{x\in A} P(x)(\log P(x) - \log Q(x))$$