補足資料　確率・統計

松尾・岩澤研究室
MATSUO-IWASAWA LAB UTOKYO

許諾なく撮影や第三者
への開示を禁止します

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

今回の目的と目標

- **目的**
  - ニューラルネットワークで重要となる確率・統計の基礎を理解する.

- **目標**
  - 確率分布の基本的な性質および代表的な分布について説明できる.
  - ナイーブベイズの考え方について説明できる.
  - ベイズ推定と最尤推定の違いを説明できる.
  - 自己情報量，相互情報量について説明できる.
  - 結合エントロピーの計算ができる.

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

確率変数・確率分布

- **確率変数**
  - 無作為に異なる値をとることができる変数.
  - より正確には確率変数は関数であり，標本空間から実数への写像として定義される.
  - 離散値の場合と連続値の場合がある.

- **確率分布**
  - 確率変数やその集合が取りうる値それぞれの起こりやすさを記述したもの.

例：

コインを2回投げたとき，表が出る回数の
確率分布は以下のようになる.

表が出る回数 （確率変数X）	0	1	2
確率$P(X)$	0.25	0.5	0.25

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

同時分布・周辺分布

- **同時確率分布**
  - 多変数の確率分布.
  - $X = x$ かつ $Y = y$ である確率は,
    $P(X = x, Y = y) = P(x, y)$ で表せる.

- **確率周辺分布**
  - 変数の集合の確率分布がわかっているとき,
    その部分集合の確率分布.
  - 離散変数 X, Y について, $P(X, Y)$ がわかっているとき,
    $P(x)$ は確率の加法定理で求められる.

$$\forall x \epsilon X, P(x) = \sum_{y} P(x, y)$$

  - 連続変数の場合,

$$p(x) = \int p(x, y) dy$$

同時分布

X \ Y	y_1	y_2	計
x_1	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{3}{5}$
x_2	$\frac{2}{5}$	0	$\frac{2}{5}$
計	$\frac{4}{5}$	$\frac{1}{5}$	1

周辺分布

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

確率分布の期待値・分散

- 確率分布の期待値
  - 確率分布 $P(X)$ から x が抽出されたときの，関数 $f(x)$ がとる値の平均（値）．
  - 離散変数の場合

$$\mathbb{E}_{X\sim P}[f(x)] = \sum_{x} P(x)f(x)$$

  - 連続変数の場合

$$\mathbb{E}_{X\sim p}[f(x)] = \int p(x)f(x)dx$$

- 確率分布の分散
  - 確率分布からさまざまな x を抽出したときの，関数 $f(x)$ のばらつきの度合い.

$$\mathrm{Var}(f(x)) = \mathbb{E}[(f(x) - \mathbb{E}[f(x)])^2]$$

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

確率質量関数

- **確率質量関数**
  - 離散変数の確率分布.

- 確率質量関数 p が満たすべき性質
  - p の定義域は，X が取りうる状態全ての集合でなければならない.
  - $\forall x \in X, 0 \leq p(x) \leq 1$.
  - $\sum_{x \epsilon X} p(x) = 1$.

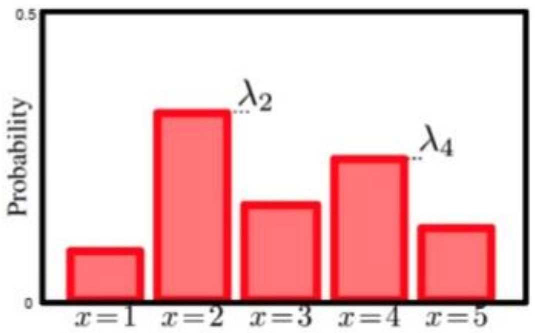
多項分布・カテゴリ分布

- **多項分布**
  - K個のカテゴリから1つ選ぶ試行を M回繰り返したとき， k番目が選ばれた回数 m_k の組み合わせに対する離散確率分布.
  - 確率質量関数（π_k は k番目のカテゴリが選ばれる確率）

$$p_{\pi,M}(m) = M!\prod_{k=1}^{K}\frac{{\pi_k}^{m_k}}{m_k\,!} \equiv Mult(x;\pi,M)$$

- **二項分布**
  - カテゴリを2個（$K=2$）にした多項分布.

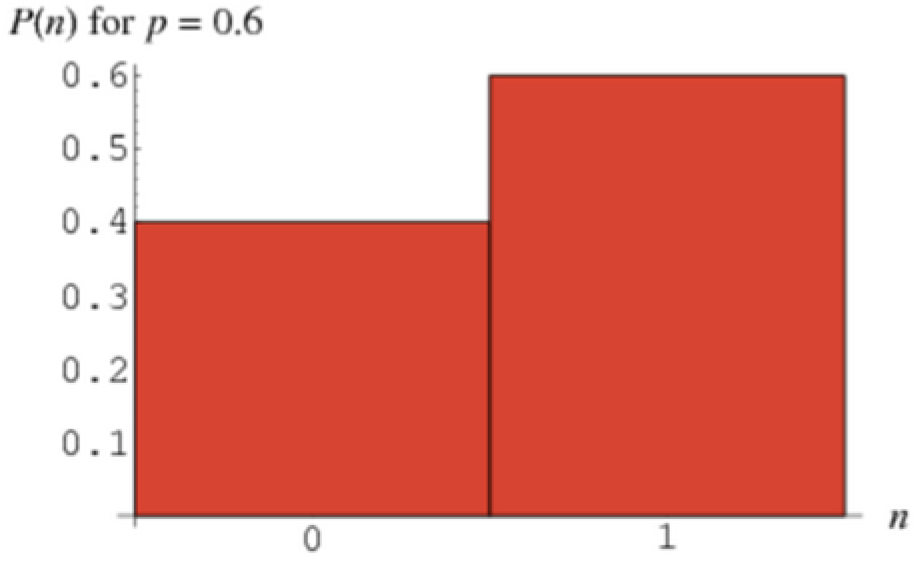
- **カテゴリ分布**
  - 試行を1回（$M=1$）にした多項分布.



ベルヌーイ分布

- **ベルヌーイ分布**
  - カテゴリを2個（$K = 2$），試行を1回（$M = 1$）にした多項分布.
  - コインを1回だけ投げたときの確率分布に相当し，確率変数 x が $\{0,1\}$ の二値のみをとりうる離散確率分布.
  - 1をとる確率である μ をパラメータとして持つ.
  - 確率質量関数

$$p_{\mu}(x) = \mu^{x}(1-\mu)^{1-x} \equiv Bern(x;\mu)$$



©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

確率密度関数

- **確率密度関数**
  - 連続変数の確率分布.

- 確率密度関数 p が満たすべき性質
  - p の定義域は，X が取りうる状態全ての集合でなければならない
  - $\forall x \in X, p(x) \geq 0$．ただし，$p(x) \leq 1$ は必要条件でない.
  - $\int p(x)\, dx = 1.$

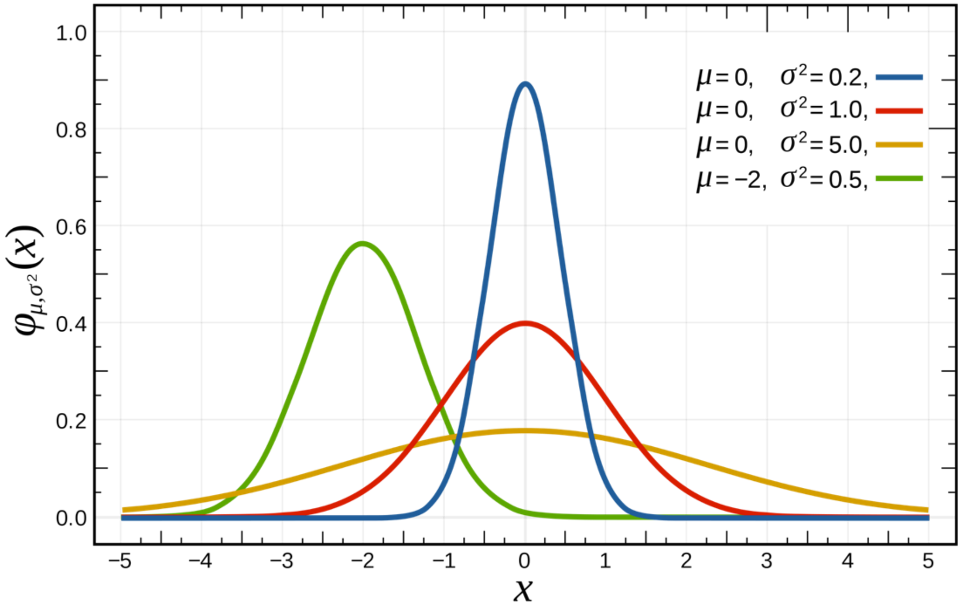
- 確率密度関数 p からは特定の状態の確率は直接的に得られない.
  - 容積が δx の微小領域にある確率は $p(x)\delta x$ で与えられる.

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

ガウス分布（正規分布）

- **ガウス分布（正規分布）**
  - 平均 μ，標準偏差 σ をパラメータとして持つ連続確率分布.
  - 台（確率分布が取りうる値の範囲）は実数全体.
  - 確率密度関数

$$p_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) \equiv \mathcal{N}(x;\mu,\sigma^2)$$



©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

ベイズの定理

- **条件付き確率**
  - $P(Y|X)$：事象 X が起きたもとで，事象 Y が起こる（条件付き）確率.
    $P(Y|X) = \frac{P(X,Y)}{P(X)}$ を満たす.

- **ベイズの定理**
  - 結果Dに対する原因の推定Hが正しい確率を求める定理.

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

$P(H|D)$：結果 D に対する原因の推定 H が正しい確率（事後確率）.
$P(D|H)$：推定 H が正しいと仮定した上での結果 D の発生確率（尤度）.
$P(H)$：結果 D によらず，推定 H が正しい確率（推定 H の事前確率）.
$P(D)$：推定 H によらない，結果 D の発生確率（結果 D の事前確率）.

  - 分母は $P(y) = \sum_x P(x,y) = \sum_x P(x)P(y \mid x)$ より，分子の x に関する総和
    - 事後分布を確率分布にするためのnormalizerと捉えられる

ナイーブベイズ

- **ナイーブベイズ**
  - 確率に基づいた予測手法であり，ベイズの定理の考えを元にしたアルゴリズム.
  - ベイズの定理において，結果 D の要素 $(d_1, d_2, d_3, \ldots, d_n)$ は,
    互いに独立である（お互いの確率に影響しない）.
    → この仮定が単純（ナイーブ）なので，ナイーブベイズと呼ばれる.
  - 自然言語の分類に多用される.

- 特徴
  - 他手法に比べて計算が早い.
  - 実装が容易であるにも関わらず，精度が良い.

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

分類問題におけるナイーブベイズの考え方

- **ナイーブベイズ**
  - 結果 D の要素 $(d_1, d_2, d_3, \ldots, d_n)$ は，互いに独立である（お互いの確率に影響しない）．

$$P(H|d_1, d_2\ d_3, \ldots, d_n) = \frac{P(d_1, d_2\ d_3, \ldots, d_n|H)P(H)}{P(D)}$$

$$= \frac{P(d_1|H)P(d_2|H)\ldots P(d_n|H)P(H)}{P(d_1, d_2, d_3, \ldots, d_n)}$$

$$= \frac{P(H)\prod_{i=1}^{n} P(d_i|H)}{P(d_1, d_2, d_3, \ldots, d_n)} - ①$$

$H \in \{A, B, C \ldots\}$：分類カテゴリ
$D = (d_1, d_2, d_3, \ldots d_n)$：結果データ
$P(H|d_1, d_2, d_3, \ldots d_n)$：$d_1, d_2$ …からカテゴリが Hである確率.
$P(H)$：全データに対してのカテゴリHの確率.

①の値が最も大きくなるカテゴリ $H \in \{A, B, C \ldots\}$ を，
そのデータのカテゴリとする.

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

ベイズ推定

- **ベイズ推定**
  - とある条件における事象の発生確率を，既知の確率と観測データから求める手法.

事後確率分布 ∝ **事前確率分布** × **尤度関数**

- **ベイズの定理**
  - 結果 D に対する原因の推定 H が正しい確率を求める定理.

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

結果 $D = (d_1, d_2, d_3 \ldots, d_n)$ は互いに独立と仮定.

確率変数 θ, x に関する確率密度関数 $f(\theta, x)$ を導入.

尤度関数

$$\underbrace{\boxed{f(\theta|x)}}_{\text{事後確率分布}} = \frac{\boxed{f(x|\theta)}\boxed{f(\theta)}}{f(x)}$$

事前確率分布

ベイズ推定

- 確率分布におけるベイズ推定

確率分布におけるベイズの定理

$$f(\theta|x)=\frac{f(x|\theta)f(\theta)}{f(x)} \longrightarrow f(\theta|x)=\frac{f(x|\theta)f(\theta)}{\int_{-\infty}^{\infty}f(x|\theta)f(\theta)\mathrm{d}\theta}$$

- $f(x)=\int_{-\infty}^{\infty}f(x|\theta)f(\theta)\mathrm{d}\theta$

（参考）条件付き確率に関する式： $\mathrm{P}(B_j)=\sum_{i=1}^{n}p\left(B_j|A_i\right)p(A_i)$

$$\longrightarrow f\left(\theta\middle|x^{(n)}\right)=\frac{f(x_n|\theta)f\left(\theta|x^{(n-1)}\right)}{\int_{-\infty}^{\infty}f(x_n|\theta)f(\theta)\mathrm{d}\theta}$$

- 互いに独立な

$$x^{(n)}=x_1,x_2,\dots,x_n$$

を導入.

（参考）確率におけるベイズ推定式： 事象B_jと事象C_kが互いに独立であるならば,

$$p(A_i|B_j,C_k)=\frac{p(C_k|A_i)p(A_i|B_j)}{\sum_{i=1}^{n}p(C_k|A_i)p(A_i|B_j)}$$

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

ベイズ推定

- 確率分布におけるベイズ推定

（続き）

- $k_n = \frac{1}{\int_{-\infty}^{\infty} f(x_n|\theta)f(\theta)d\theta}$
  として，

確率分布におけるベイズの定理の一般式

$$\boxed{f\left(\theta\middle|x^{(n)}\right) = k_n f(x_n|\theta)f(\theta|x^{(n-1)})}$$

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

ベイズ推定

例題：

表が出る確率がθであるコインを3回投げた時，裏→表→裏の順に出た．
この時，θに関する確率分布$f(\theta|x^{(3)})$と$f(\theta|x^{(3)})$の極大値を取るθを求めよ．
事前分布は$f(\theta)=1$とする．

解答：

1回目は裏が出たことから，$f(x_1|\theta)=1-\theta$.
よって，$f(\theta|x_1)=k_1 f(x_1|\theta)f(\theta)=k_1(1-\theta)$.
したがって，

$$\int_0^1 f(\theta|x_1)\mathrm{d}\theta=\int_0^1 k_1(1-\theta)\mathrm{d}\theta=k_1\left[\theta-\frac{1}{2}\theta^2\right]_0^1=\frac{k_1}{2}=1.$$

$k_1=2$より， $f(\theta|x_1)=2-2\theta$.

2回目は表が出たことから，$f(x_2|\theta)=\theta$.
よって，$f\left(\theta\middle|x^{(2)}\right)=k_2 f(x_2|\theta)f(\theta|x_1)=k_2\theta(2-2\theta)=2k_2(\theta-\theta^2)$.
したがって，

$$\int_0^1 f\left(\theta\middle|x^{(2)}\right)\mathrm{d}\theta=\int_0^1 2k_2(\theta-\theta^2)\mathrm{d}\theta=2k_2\left[\frac{1}{2}\theta^2-\frac{1}{3}\theta^3\right]_0^1=\frac{k_2}{3}=1.$$

$k_2=3$ より， $f\left(\theta\middle|x^{(2)}\right)=6\theta(1-\theta)$.

3回目は裏が出たことから，$f(x_3|\theta)=1-\theta$.
よって，

$$\begin{aligned} f\left(\theta\middle|x^{(2)}\right)&=k_3 f(x_3|\theta)f\left(\theta\middle|x^{(2)}\right)\\ &=k_3(1-\theta)6\theta(1-\theta)\\ &=6k_3(\theta-2\theta^2+\theta^3). \end{aligned}$$

したがって，

$$\begin{aligned} \int_0^1 f\left(\theta\middle|x^{(3)}\right)\mathrm{d}\theta&=\int_0^1 6k_3(\theta-2\theta^2+\theta^3)\mathrm{d}\theta\\ &=6k_3\left[\frac{1}{2}\theta^2-\frac{2}{3}\theta^3+\frac{1}{4}\theta^4\right]_0^1=\frac{k_3}{2}=1. \end{aligned}$$

$k_3=2$より， $\boldsymbol{f\left(\theta\middle|x^{(3)}\right)=12\theta(1-\theta)^2}$.
また， $f'\left(\theta\middle|x^{(3)}\right)=12(1-\theta)(1-3\theta)=0$ より，
$f\left(\theta\middle|x^{(3)}\right)$ の極大値をとる $\boldsymbol{\theta=1/3}$.

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

最尤推定

- **最尤推定**
  - とある条件における事象の発生確率を，観測データが最も発生しやすい（尤もらしい）パラメータを探すことで求める手法.
    ＝尤度関数を最大化するようなパラメータを求める.

例題：

表が出る確率が θ であるコインを3回投げた時，裏→表→裏の順に出た.
この時，尤もらしい θ を求めよ.

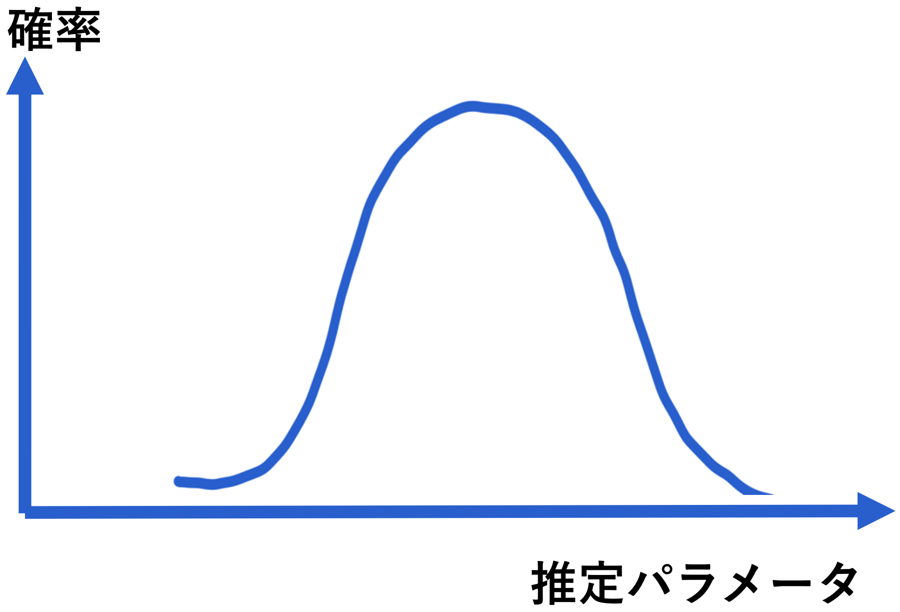
解答：

観測に関する尤度関数を L とすると $L(\theta)=\theta(1-\theta)^2$． これを最大化する θ を求める.
$\frac{dL}{d\theta}=1-4\theta+3\theta^2=0$ を解いて $\theta=\frac{1}{3},1$ であり，
$L(0)=0, L\left(\frac{1}{3}\right)=\frac{4}{27}, L(1)=0$ より， $\boldsymbol{\theta}=\boldsymbol{\frac{1}{3}}$.

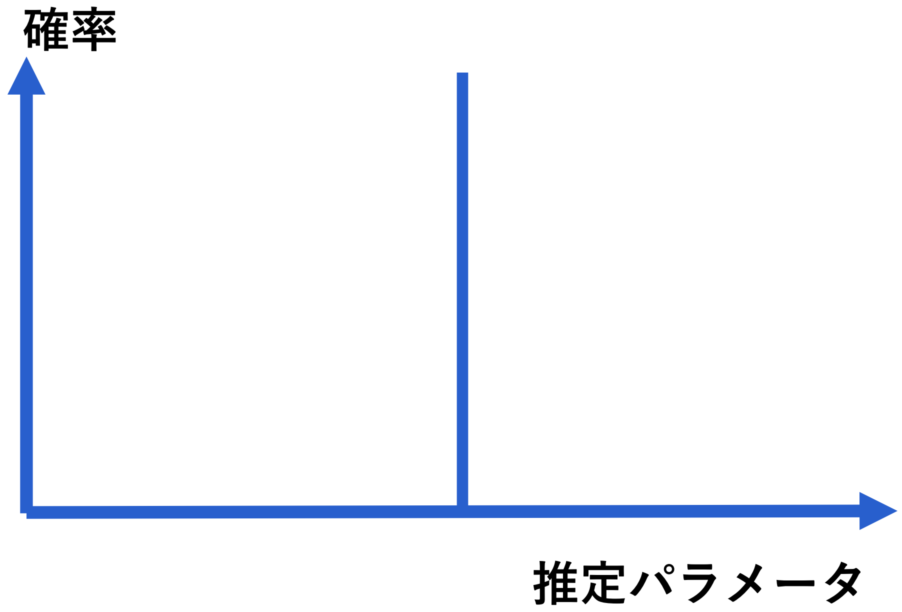
©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

ベイズ推定と最尤推定の違い

- **ベイズ推定**
  - とある条件における事象の発生確率を，既知の確率と観測データから求める手法.
  - 事前確率分布を用いる.
  - 推定するパラメータは確率分布として得られる.



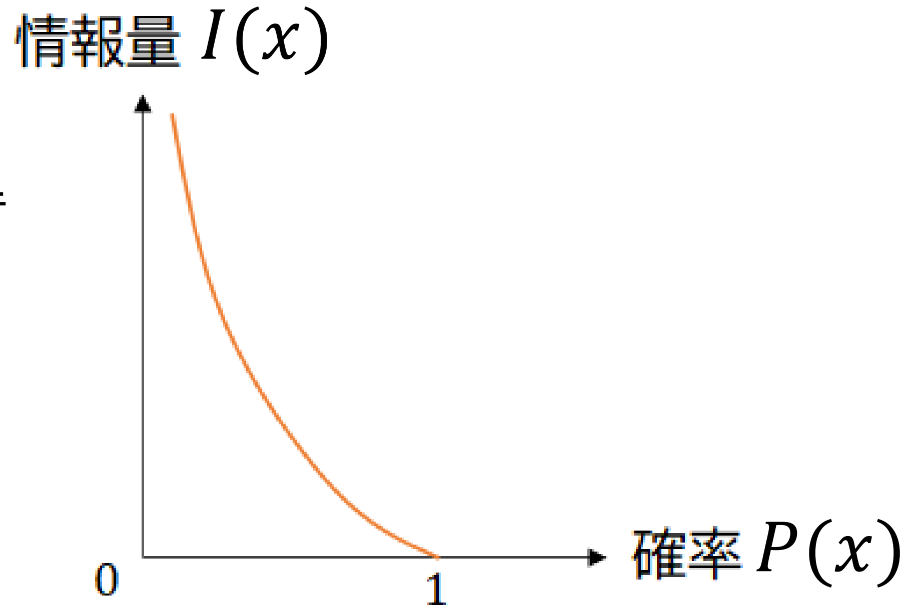
- **最尤推定**
  - とある条件における事象の発生確率を，観測データが最も発生しやすい（尤もらしい）パラメータを探すことで求める手法.
  - 事前確率分布を用いない.
  - 推定するパラメータは値として得られる.



©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

自己情報量

- **自己情報量**
  - 情報量の大きさを定量化した数値のこと.
    $I(x) \coloneqq -\log P(x)$



- 特徴
  - 直感的には，「ありふれたこと（P が大きい）が起きたとわかったときよりも，稀なこと（P が小さい）が起きたとわかったときの方が，情報量は大きい」ということ.
  - 対数の底が2のとき，単位はビット（bit）になる.
  - 加法性がある.
    発生確率が $P(x)$ の事象が起こり，
    かつ発生確率が $Q(x)$ の事象が起こる時の
    自己情報量は，
    $I(x) = -\log P(x)Q(x) = -\big(\log P(x) + \log Q(x)\big)$.

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

相互情報量

- **相互情報量**
  - 一方の事象を知ることでもう一方の事象の情報がどれくらい得られるかを表す量.

$$I(X;Y) := \sum_{x \in D_X} \sum_{y \in D_Y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)}$$

X, Y：確率変数
D_X, D_Y：それぞれ X, Y が取りうる値の全体の集合

- 特徴
  - $I(X;Y) \geq 0.$
  - X と Y が独立ならば $I(X;Y) = 0.$
  - $I(X;Y)$ が最大値をとるのは, X の分布と Y の分布が一致するとき.

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

相互情報量の特徴

- 特徴
  - $I(X;Y) \geq 0.$
  - X と Y が独立ならば $I(X;Y) = 0.$
  - $I(X;Y)$ が最大値をとるのは，X の分布と Y の分布が一致するとき．

- 証明
  ギブスの不等式：

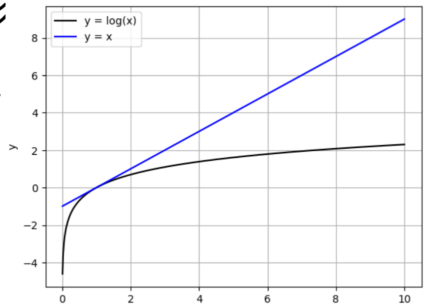
  $\sum p_k = 1, \sum q_k = 1$，かつ $p_k \geq 0, q_k \geq 0$ のとき

  $$\sum p_k \log \frac{p_k}{q_k} \geq 0.$$

  を利用する．

参考：ギブスの不等式の証明

任意の正の数 xについて成立する不等式 $\log(x) \leq x - 1$を $x = \frac{q_k}{p_k}$で置き換える．その式を変形して $p_k \log\left(\frac{p_k}{q_k}\right) \geq p_k - q_k$ を導出し両辺の全ての k に関する総和Σ_kを求めると，ギブスの不等式になる．



$\sum P(x,y) = 1$, $\sum P(x)P(y) = 1$，かつ $P(x,y) \geq 0, P(x)P(y) \geq 0$であるから，

$$I(X;Y) := \sum_{x \in D_X} \sum_{y \in D_Y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)} \geq 0. (Q.E.D.)$$

相互情報量の特徴

- 特徴
  - $I(X;Y) \geq 0.$
  - X と Y が独立ならば $I(X;Y) = 0.$
  - $I(X;Y)$ が最大値をとるのは，X の分布と Y の分布が一致するとき.

- 証明
  X と Y が独立の時，$P(x,y) = P(x)P(y)$ であるから，

$$
\begin{aligned}
I(X;Y) &:= \sum_{x \in D_X} \sum_{y \in D_Y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)} \\
&= \sum_{x \in D_X} \sum_{y \in D_Y} P(x,y) \log \frac{P(x)P(y)}{P(x)P(y)} \\
&= \sum_{x \in D_X} \sum_{y \in D_Y} P(x,y) \log 1 \\
&= 0. (Q.E.D.)
\end{aligned}
$$

相互情報量の特徴

- 特徴
  - $I(X;Y) \geq 0.$
  - X と Y が独立ならば $I(X;Y) = 0.$
  - $I(X;Y)$ が最大値をとるのは， X の分布と Y の分布が一致するとき.

- 証明

X と Y が一致するとき，$P(x,y) = P(x) = P(y)$ であるから，

$$I(X;Y) := \sum_{x \in D_X} \sum_{y \in D_Y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)} = \sum_{x \in D_X} P(x) \log \frac{P(x)}{P(x)P(x)} = -\sum_{x \in D_X} P(x) \log P(x). \ -①$$

①を「エントロピー（平均情報量）」といい，$H(X)$で表す.
また，「条件付きエントロピー $H(X|Y)$」は以下で示される.

$$H(X|Y) = -\sum_{y \in D_Y} P(y) \sum_{x \in D_X} P(x|y) \log P(x|y) \geq 0$$

↓ 続く

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

相互情報量の特徴

- 特徴
  - $I(X;Y) \geq 0.$
  - X と Y が独立ならば $I(X;Y) = 0.$
  - $I(X;Y)$ が最大値をとるのは，X の分布と Y の分布が一致するとき.

- 証明続き
このとき，

$$H(X) = -\sum_{x \in D_X} P(x) \log P(x) = -\sum_{y \in D_y} P(y|x) \sum_{x \in D_X} P(x) \log P(x) = -\sum_{x \in D_x} \sum_{y \in D_y} P(x,y) \log P(x)$$

$$-H(X|Y) = \sum_{y \in D_Y} P(y) \sum_{x \in D_X} P(x|y) \log P(x|y) = \sum_{x \in D_x} \sum_{y \in D_y} P(x,y) \log \frac{P(x,y)}{P(y)}$$

$$\therefore\ H(X) - H(X|Y) = \sum_{x \in D_X} \sum_{y \in D_Y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)} = I(X;Y)$$

よって，$I(X;Y) = H(X) - H(X|Y), H(X|Y) \geq 0$ より，$I(X;Y) \leq H(X). (Q.E.D.)$

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

エントロピー（平均情報量）

• エントロピー（平均情報量）

- 確率変数 X のエントロピーは以下で表される.

$$H(X) = -\sum_{x \in D_X} P(x) \log P(x)$$

- X がとりうる事象の情報量の平均.
- X が連続確率変数ならば，総和は積分になる.

条件付きエントロピー

- **条件付きエントロピー**
  - 一方の事象を知ることでもう一方の事象の情報がどれくらい不確かさを持つかを表す量.

$$H(X|Y) = -\sum_{y \in D_Y} P(y) \sum_{x \in D_X} P(x|y) \log P(x|y)$$

X, Y：確率変数.
D_X, D_Y：それぞれ X, Y が取りうる値の全体の集合.
$P(x|y)$：Y が特定の値 y をとるという条件下において X が x という値をとる条件付き確率分布.

条件付きエントロピーに関する例題

例題：

52枚のトランプからカードを1枚引くという確率事象系 A を考える.
この時，事前に「引いたカードはハートである．」という事象 B が与えられた時に
得られる B の条件付きエントロピー $H(A|B)$ を求めよ.

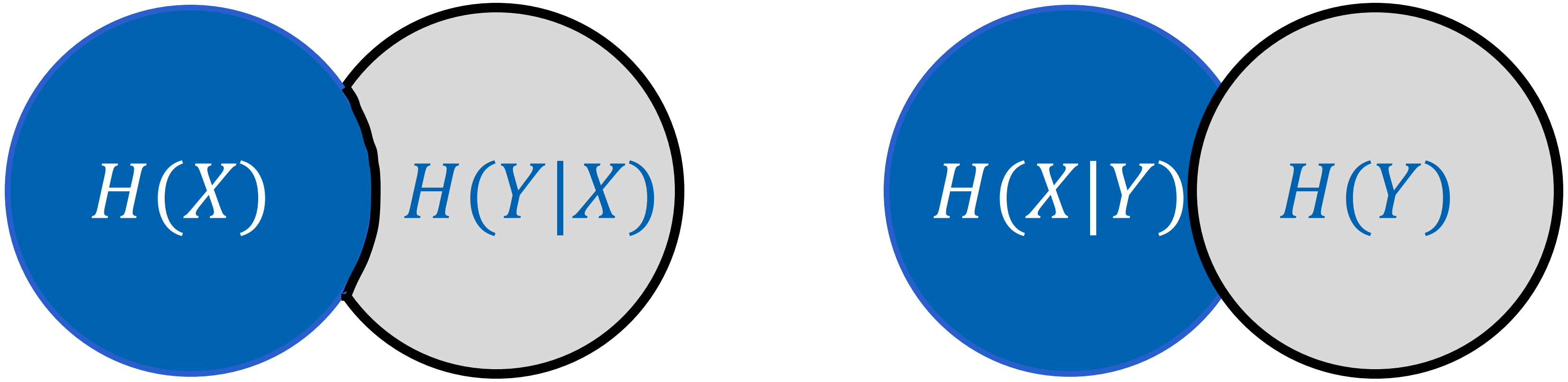
解答：

条件付きエントロピー：$H(A|B) = -\sum_{b\in D_B} P(b) \sum_{a\in D_A} P(a|b)\log P(a|b)$.
ハートのカードはトランプ内に$52 \div 4 = 13$枚含まれるから，$P(a|b) = \frac{1}{13}$.
よって，条件付きエントロピー $H(A|B) = -1 \cdot \sum_{i=1}^{13} \frac{1}{13} \log_2 \frac{1}{13} = \log_2 13 \approx 3.70$ ビット.

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

結合エントロピー

- **結合エントロピー**
  - 二つの事象が同時に起こった時に得られる平均情報量（エントロピー）のこと.



結合エントロピー：$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

結合エントロピーに関する例題

例題：

サイコロを1回振る時，目が「3以下か4以上か」という事象系$A=\{a_1,a_2\}$と，
目が「偶数か奇数か」という事象系$B=\{b_1,b_2\}$ があるとする.
AとBの結合エントロピー $H(A,B)$を求めよ.

解答：

結合エントロピー：$H(A,B)=H(A)+H(B|A)=H(B)+H(A|B)$
事象系Bのエントロピー $H(B)=-\sum_{b\in D_B}P(b)\log P(b)=-\frac{1}{2}\log_2\frac{1}{2}-\frac{1}{2}\log_2\frac{1}{2}=\log_2 2=1$.

A と $B=b_1$ の条件付きエントロピー $H(A|b_1)=-\sum_{a\in D_A}P(a|b_1)\log P(a|b_1)$

$$=-\frac{1}{3}\log_2\frac{1}{3}-\frac{2}{3}\log_2\frac{2}{3}=\log_2 3-\frac{2}{3}.$$

A と $B=b_2$ の条件付きエントロピー $H(A|b_2)=-\sum_{a\in D_A}P(a|b_2)\log P(a|b_2)$

$$=-\frac{2}{3}\log_2\frac{2}{3}-\frac{1}{3}\log_2\frac{1}{3}=\log_2 3-\frac{2}{3}.$$

↓ 続く

結合エントロピーに関する例題

解答続き：

よって，条件付きエントロピー $H(A|B) = H(A|b_1)P(b_1) + H(A|b_2)P(b_2)$

$$= \left(\log_2 3 - \frac{2}{3}\right) \times \frac{1}{2} + \left(\log_2 3 - \frac{2}{3}\right) \times \frac{1}{2}$$

$$= \log_2 3 - \frac{2}{3}$$

$$\approx 0.918$$

結合エントロピー $H(A, B) = H(B) + H(A|B) \approx 1 + 0.918 \approx 1.918.$

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO

参考文献

[1]　C.M.ビショップ (2012), ”パターン認識と機械学習 上”, 丸善出版

[2]　縄田和満 (2013), “東京大学工学教程 基礎系 数学 確率・統計I ”, 丸善出版

[3]　Ian Goodfellow et al. (2016), “Deep Learning”, The MIT Press

[4]　@kazuya_minakuchi(Kazuya Minakuchi) (2020), ”ナイーブベイズ（単純ベイズ）”, Qiita, アクセス日：2024/01/09

[5]　Avinton (2022), “機械学習入門者向け Naive Bayse(単純ベイズ)に触れてみる”, AVINTON, アクセス日：2024/01/09

[6]　津山工業高等専門学校 (2024), “ベイズ推定（基本編）”, 津山工業高等専門学校, アクセス日：2024/01/09

[7]　高鳥 (2022), “第 8 章 ベイズ推定: データを元に「確信」を高める手法”, Cloud Ace, アクセス日：2024/01/09

[8]　kr (2024), “自己情報量とは？分かりやすく解説します！”, 「なんとなくわかる」大学の数学・物理・情報, アクセス日：2024/01/16

[9]　高校数学の美しい物語 (2022), “相互情報量の意味とエントロピーとの関係”,高校数学の美しい物語, アクセス日：2024/01/16

[10] kr (2024), “条件付きエントロピーとは？簡単に解説！”, 「なんとなくわかる」大学の数学・物理・情報, アクセス日：2024/01/16

[11] kr (2024), “結合エントロピーって何？分かりやすく解説しました！”, 「なんとなくわかる」大学の数学・物理・情報, アクセス日：2024/01/16

松尾・岩澤研究室

MATSUO-IWASAWA LAB UTOKYO

©MATSUO-IWASAWA LAB, THE UNIVERSITY OF TOKYO