第一章　線形代数

１．　線形代数学（行列）

（1）固有値・固有ベクトルの求め方を確認する。

（2）固有値分解について理解を深める。

（3）特異値・特異ベクトルの概要を知る。

（4）特異値分解の概要を知る。

１．１　スカラーとベクトルの違い

（1）スカラー

　普通の数字

　四則演算ができるもの

（2）ベクトル

　大きさ、向きを表すために用いられる。

　数字の組み合わせ

１．２　行列

スカラーを表のようにしてまとめたもの

（1）ベクトルの変換

１．３　連立１次方程式

ｘ１＋２ｘ２＝３

ｘ１、ｘ２：未知のもの

関係性はわかる。下記直線の上の値だけを取りうる。

ｘ２＝（－１／２ｘ１）＋（３／２）

ｘ２＝（－２／５ｘ１）＋１

１．４　連立１次方程式を行列で表す（ａｘ＝ｂ）

ｘ１＋２ｘ２＝３　　　　　　　Ａｘ＝ｂ

２ｘ１＋５ｘ２＝５

$$\begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} x1 \\ x2 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

１．５　行列の積

$$\begin{pmatrix} 6 & 4 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 6*1+4*2 \\ 3*1+5*2 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 1 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 2*1+1*3 & 2*3+1*1 \\ 4*1+1*3 & 4*3+1*1 \end{pmatrix}$$

１．６　連立方程式の解き方

下限法

$$\begin{cases} x_1 + 4x_2 = 7 \\ 2x_1 + 6x_2 = 10 \end{cases}$$

(1)　２行目を１／２する

$$\begin{cases} x_1 + 4x_2 = 7 \\ x_1 + 3x_2 = 5 \end{cases}$$

(2)　１行目に２行目の－１倍を加える

$$\begin{cases} x_2 = 2 \\ x_1 + 3x_2 = 5 \end{cases}$$

(3)　２行目に１行目の－３倍を加える

$$\begin{cases} x_2 = 2 \\ x_1 = -1 \end{cases}$$

(4)　１行目と２行目を入れ替える

$$\begin{cases} x_1 = -1 \\ x_2 = 2 \end{cases}$$

注）必要な技術（行基本変形）

①.　ｉ行目をｃ倍する

②.　ｓ行目にｔ行目のｃ倍を加える

③.　ｐ行目とｑ行目を入れ替える。

$$\begin{pmatrix} 1 & 4 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 7 \\ 10 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 4 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 7 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

１．７　単位行列と逆数

(1)　単位行列

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(2)　逆行列

$$A(A^{-1}) = (A^{-1})A = I$$

１．８　逆行列の求め方

$$\begin{pmatrix} 1 & 4 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} x1 \\ x2 \end{pmatrix} = \begin{pmatrix} 7 \\ 10 \end{pmatrix}$$

まず、左右同じ形にする

$$\begin{pmatrix} 1 & 4 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} x1 \\ x2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 7 \\ 10 \end{pmatrix}$$

$$\left[\begin{array}{cc|cc} 1 & 4 & 1 & 0 \\ 2 & 6 & 0 & 1 \end{array}\right]$$

２行目を１／２倍する

$$\left[\begin{array}{cc|cc} 1 & 4 & 1 & 0 \\ 1 & 3 & 0 & 1/2 \end{array}\right]$$

１行目に２行目の－１倍を加える

$$\left[\begin{array}{cc|cc} 0 & 1 & 1 & -1/2 \\ 1 & 3 & 0 & 1/2 \end{array}\right]$$

２行目に１行目の－３倍を加える

$$\left[\begin{array}{cc|cc} 0 & 1 & 1 & -1/2 \\ 1 & 0 & -3 & 2 \end{array}\right]$$

１行目と２行目を入れ替える

$$\left[\begin{array}{cc|cc} 1 & 0 & -3 & 2 \\ 0 & 1 & 1 & -1/2 \end{array}\right]$$

ガウスの掃出し法で逆行列が求まる

１．１１　逆行列が存在しない条件

解が無い連立方程式

$$\begin{matrix} a & b \\ c & d \end{matrix}$$

ａ：ｂ　≠　ｃ：ｄ

ａ：ｂ　＝　ｃ：ｄ　　　　　逆行列を持たない

ａｄ－ｂｃ＝０

平行四辺形の面積が０となる

１．９　行列式の特徴

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} v1 \\ v2 \end{vmatrix}$$

(1) 同じものを含んでいる場合０となる

$$\begin{vmatrix} v1 \\ v2 \\ w \\ v4 \\ \vdots \\ w \\ \vdots \\ vn \end{vmatrix} = 0$$

(2) １つのベクトルがλ倍されると、行列式はλ倍される

$$\begin{vmatrix} v1 \\ v2 \\ \vdots \\ \vdots \\ \lambda vi \\ \vdots \\ \vdots \\ vn \end{vmatrix} = \lambda \begin{vmatrix} v1 \\ v2 \\ \vdots \\ vi \\ \vdots \\ \vdots \\ \vdots \\ vn \end{vmatrix}$$

(3) 他の部分が全部同じでｉ番目のベクトルだけが違った場合、行列式の足し合わせになる。

$$\begin{vmatrix} v1 \\ v2 \\ \vdots \\ \vdots \\ vi+w \\ \vdots \\ \vdots \\ vn \end{vmatrix} = \begin{vmatrix} v1 \\ v2 \\ \vdots \\ \vdots \\ vi \\ \vdots \\ \vdots \\ vn \end{vmatrix} + \begin{vmatrix} v1 \\ v2 \\ \vdots \\ \vdots \\ w \\ \vdots \\ \vdots \\ vn \end{vmatrix}$$

(4) 行を入れ替えると符号が変わる

$$\begin{vmatrix} v1 \\ v2 \\ \vdots \\ vs \\ vt \\ \vdots \\ \vdots \\ vn \end{vmatrix} = - \begin{vmatrix} v1 \\ v2 \\ \vdots \\ vt \\ vs \\ \vdots \\ \vdots \\ vn \end{vmatrix}$$

①．同じものがあったら０になる。

②．他が同じで、１つだけ足したものは行列の足し合わせになる。

$$\begin{vmatrix} v1 \\ v2 \\ \vdots \\ vs \\ vt \\ \vdots \\ \vdots \\ vn \end{vmatrix} + \begin{vmatrix} v1 \\ v2 \\ \vdots \\ vt \\ vs \\ \vdots \\ \vdots \\ vn \end{vmatrix} = \begin{vmatrix} v1 \\ v2 \\ \vdots \\ vs+vt \\ vt+vs \\ \vdots \\ \vdots \\ vn \end{vmatrix} = 0$$

(5) ３つ以上のベクトルからできている行列式は、展開できる。

$v1 = (a、b、c)$

$v2 = (d、e、f)$

$v3 = (g、h、i)$

$$\begin{vmatrix} v1 \\ v2 \\ v3 \end{vmatrix} = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = \begin{vmatrix} a & b & c \\ 0 & e & f \\ 0 & h & i \end{vmatrix} + \begin{vmatrix} 0 & b & c \\ d & e & f \\ 0 & h & i \end{vmatrix} + \begin{vmatrix} 0 & b & c \\ 0 & e & f \\ g & h & i \end{vmatrix}$$

$$= a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - d\begin{vmatrix} b & c \\ h & i \end{vmatrix} + g\begin{vmatrix} b & c \\ e & f \end{vmatrix}$$

１．１０　行列式の求め方

ある一つの正方行列に、ある一つの数値が対応する。

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

$$\begin{vmatrix} a11 & a12 & a13 \\ a21 & a22 & a23 \\ a31 & a32 & a33 \end{vmatrix} = a11\begin{vmatrix} a22 & a23 \\ a32 & a33 \end{vmatrix} - a21\begin{vmatrix} a12 & a13 \\ a32 & a33 \end{vmatrix} + a31\begin{vmatrix} a12 & a13 \\ a22 & a23 \end{vmatrix}$$

２．線形代数学（固有値）

２．１　固有値と固有ベクトル

$Ax = \lambda x$

行列とベクトルをかけると、同じベクトル×スカラになる。

$$\begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

固有値 $\lambda = 5$

固有値ベクトル（うちの一つ）$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ 特定の比率になっている（定数倍）

２．２　固有値と固有ベクトルの求め方

$Ax = \lambda x$

$(A - \lambda I)x = 0$（Aとλはベクトルとスカラは引き算できない）

$|A - \lambda I| = 0$　　行列が０（逆行列を持たない）

$$\begin{vmatrix} 1-\lambda & 4 \\ 2 & 3-\lambda \end{vmatrix} = ((1-\lambda) * (3-\lambda)) - 4 * 2 = 0$$

$1 * 3 + (-\lambda) + (-3\lambda) + \lambda * \lambda - 8 = 0$

$\lambda * \lambda - 4\lambda - 5 = 0$

二次方程式の解

$$x = \frac{-b \mp \sqrt{b * b - 4ac}}{2a}$$

$(4 \mp \sqrt{16 + 20}) / 2 = 4 \mp 6 / 2 = 5, -1$

（1）固有値５の時

$$\begin{matrix} 1 & 4 \\ 2 & 3 \end{matrix} \; \begin{matrix} x1 \\ x2 \end{matrix} = 5 \; \begin{matrix} x1 \\ x2 \end{matrix}$$

$1x1 + 4x2 = 5x1$ ―――＞ $4x2 = 4x1$ ―――＞ $x1 = x2$

$2x1 + 3x2 = 5x2$ ―――＞ $2x1 = 2x2$

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

（2）固有値－１の時

$$\begin{matrix} 1 & 4 \\ 2 & 3 \end{matrix} \; \begin{matrix} x1 \\ x2 \end{matrix} = -1 \; \begin{matrix} x1 \\ x2 \end{matrix}$$

$1x1 + 4x2 = -x1$ ―――＞ $4x2 = -2x1$ ―――＞ $x1 = -2x2$

$2x1 + 3x2 = -x2$ ―――＞ $2x1 = -4x2$

$$\begin{bmatrix} 1 \\ -1/2 \end{bmatrix}$$

２．３　固有値分解

固有値はｎ＊ｎの場合ｎ個存在する

ＡＶ＝ＶＡ

Ａ＝ＶＡＶinver

$$\begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1/2 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1/3 & 2/3 \\ 2/3 & -2/3 \end{bmatrix}$$

２．４　特異値分解

正方行列以外の固有値分解

Ｍｖ＝σｕ

M(T)ｕ＝σｖ

このような特殊な単位ベクトルがあるならば特異値分解できる。

Ｍ＝ＵＳＶ(T)　　　　　　U,V は直行行列

２．５　特異値の求め方

ＭＶ＝ＵＳ　　　M(T)Ｕ＝ＶＳ(T)

Ｍ＝ＵＳＶ(T)　M(T)＝ＶＳ(T)Ｕ(T)

積は

MM(T)＝ＵＳＶ(T)ＶＳ(T)Ｕ(T)＝ＵＳＳ(T)Ｕ(T)

２．６　特異値の求め方（具体例）

$$M = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \qquad MM(T) = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 1+4+9 & 3+4+3 \\ 3+4+3 & 9+4+1 \end{bmatrix} = \begin{bmatrix} 14 & 10 \\ 10 & 14 \end{bmatrix}$$

固有値分解

$$\begin{bmatrix} 14 & 10 \\ 10 & 14 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 24 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} (T)$$

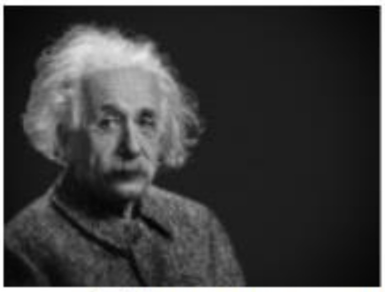
１にするため$\sqrt{2}$にしている

$$M(T)M = \begin{bmatrix} 1 & 2 \\ 3 & 2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 8 & 6 \\ 8 & 8 & 8 \\ 6 & 8 & 10 \end{bmatrix}$$

これを固有値分解して

$$\begin{bmatrix} 10 & 8 & 6 \\ 8 & 8 & 8 \\ 6 & 8 & 10 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3}\ - & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} 24 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix} \text{(T)}$$

２．７　特異値分解の利用例


元画像は960×720

特異値分解した画像データの行列から成分の小さい部分を
取り除いていくと画像はぼやけて行が画像の特徴は維持できる。
このことから画像のデータ量を少なくすることができる。
すなわち機械学習における計算コストを小さくすることができる。

第二章　確率・統計

１．　統計学１

１．１　統計学 1-1（プロローグ）
機械学習で扱うデータは、大きいため統計学が必要になる。
専門用語を理解し、読めるようになることが目的である。

１．２　統計学 1-2（集合とは何か）
要素の記述方法を学ぶ
集合とはものの集まりのことである。
Ｓ＝｛ａ，ｂ，ｃ，ｄ，ｅ，ｆ，ｇ｝
中に入っているものは

$$a \in S$$

ａはＳの要素のうちの一つ（分解できないもの）で、要素は隣と明確に区別できる。
内部に、M＝｛ｃ，ｄ，ｇ｝が有ったとすると

$$M \subset S$$

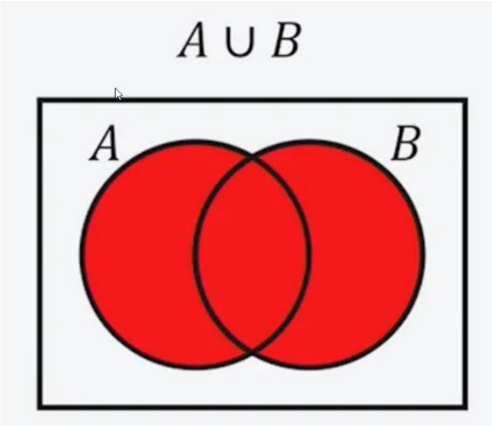
集合に含まれない場合

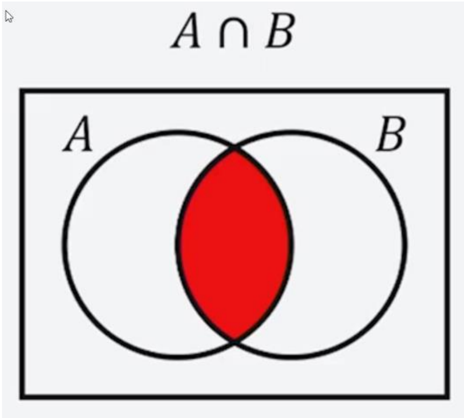
$$h \notin S$$

と表す。

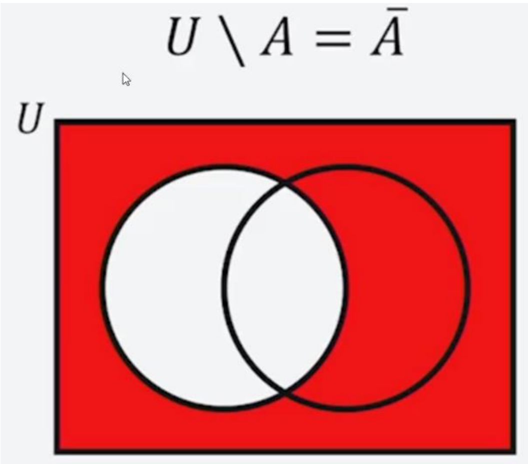
１．３　統計学 1-3（和集合と共通部分）

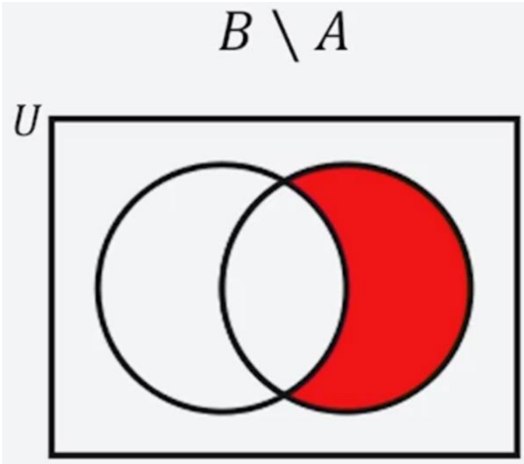
１）和集合Ａ∪Ｂ・共通部分Ａ∩Ｂ





２）絶対補Ｕ＼Ａ（$\overline{\text{Ａ}}$）・相対補Ｂ＼Ａ





１．４　統計学 1-5（確率）

確率には２つの考え方がある。(確かさの率)

１）頻度確率（客観確率）

・発生する頻度

１０本の内１本だけ当たりのクジを引いて当選する確率

１／１０＝０．１（１０％）

２）ベイズ確率（主観確率）

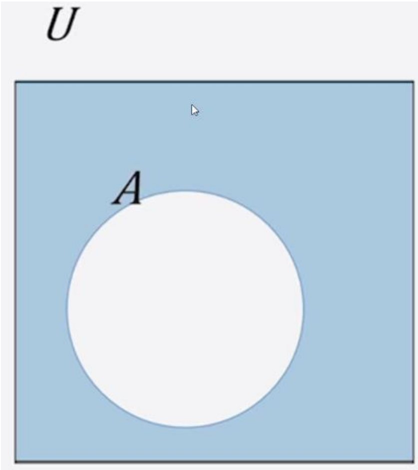
・信念の度合い

あなたは４０％の確率でインフルエンザです。

１．５　統計学 1-6

確率の定義　$$P(A) = \frac{n(A)}{n(U)} = \frac{\text{事象Aが起こる数}}{\text{全ての事象の数}}$$　確率は０~１の値

$$P(\underline{A}) = 1 - P(A)$$



$$P(\overline{A}) = \frac{\text{事象Aが起こらない数}}{\text{全ての事象の数}}$$

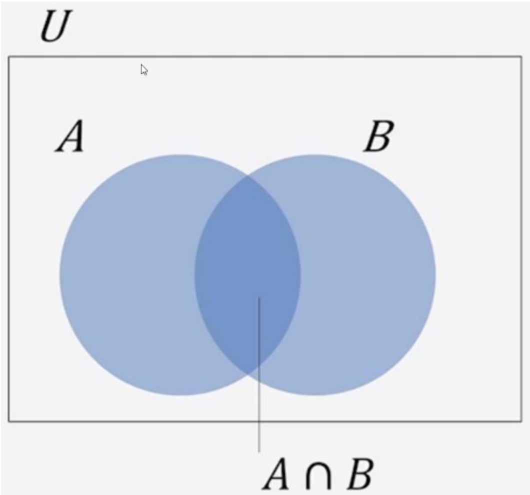
$$= \frac{\text{全ての事象の数—事象Aの起こる数}}{\text{全ての事象の数}}$$

$$= \frac{n(U) - n(A)}{n(U)}$$

$$= \frac{n(U)}{n(U)} - \frac{n(A)}{n(U)} = 1 - P(A)$$

１．６　統計学 1-7

同時確立



Ｐ（Ａ∩Ｂ）＝Ｐ（Ａ）Ｐ（Ｂ｜Ａ）

Ａの条件下でＢであるもの

Ｐ（Ａ∩Ｂ）＝Ｐ（Ｂ∩Ａ）

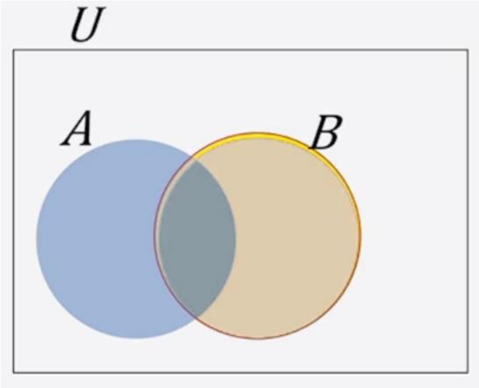
Ｐ（Ａ）Ｐ（Ｂ｜Ａ）＝Ｐ（Ｂ）Ｐ（Ａ｜Ｂ）

１．７　統計学 1-8（条件付き確率）

ある事象Ｂが与えられた下でＡとなる確率

１）雨が降っている条件下で交通事故に合う確率（条件付き）

２）雨が降っていて、事故にあった確率（同時）



$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{n(A \cap B)}{n(B)}$$

１．８　統計学 1-9（独立な事象の同時確立）

同時確立と条件付き確率の関係

１）お互いの発生には因果関係のない事象Ａと事象Ｂが同時に発生する確率

$$P(A \cap B) = P(A)\ P(B \mid A)$$
$$= P(A)\ P(B)$$

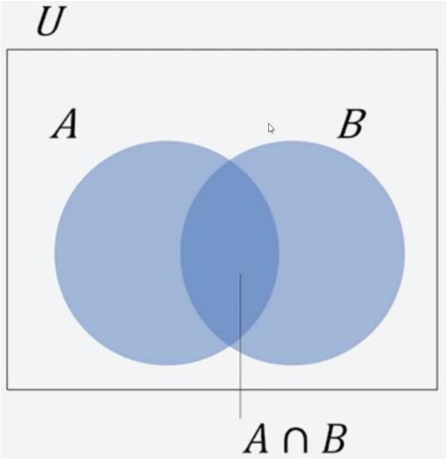
猫を見ると風邪をひく

猫を見る

風邪をひく

独立なものが同時に発生する。

１．９　統計学 1-10（Ｐ（Ａ∪Ｂ）



$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

重なり分を引く

１．１０　統計学 1-11（ベイズ則）

　ある街の子供たちは毎日１／４の確率で飴玉をもらうことができ、飴玉をもらうと１／２の確率で笑顔になるという。その街の、笑顔な子供が飴玉をもらっている確率を求めよ。（ただし、この街の子供たちが笑顔でいる確率は１／３である）

・笑顔の条件で飴玉をもらっている確率

Ｐ（Ａ）Ｐ（Ｂ｜Ａ）＝Ｐ（Ｂ）Ｐ（Ａ｜Ｂ）が成り立つ　―＞ベイズ則

１）問題整理

Ｐ（飴玉）＝１／４

Ｐ（笑顔｜飴玉）＝１／２

Ｐ（笑顔）＝１／３

２）解答

Ｐ（笑顔｜飴玉）×Ｐ（飴玉）＝Ｐ（笑顔∩飴玉）＝１／２×１／４＝１／８

Ｐ（笑顔∩飴玉）＝Ｐ（飴玉∩笑顔）

Ｐ（飴玉∩笑顔）＝Ｐ（飴玉｜笑顔）×Ｐ（笑顔）　　１／８＝Ｐ（飴玉｜笑顔）×１／３

Ｐ（飴玉｜笑顔）＝３／８

第三章　情報理論

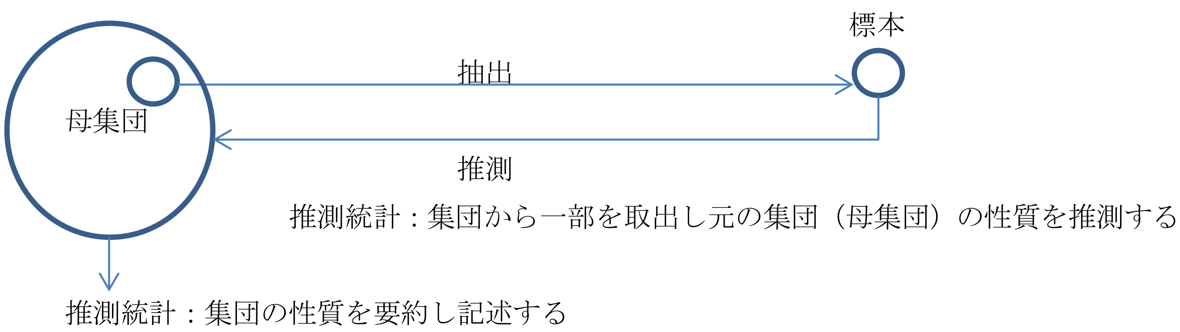
１．　統計学２

１．１　統計学 2-1（プロローグ）

(1) 記述統計

・集団の性質を要約し記述する。

(2) 推測統計



１．　２　統計学 2-2（確率変数と確率分布）

（1）確率変数

- 確率的な変数
- 事象と結び付けられた数値
- 当たりくじの金額
- 事象そのものを指すと解釈する場合も多い

（2）確率分布

- 事象の発生する確率分布
- 離散値であれば表に表せる。

事象	裏が０枚 表が４枚	裏が１枚 表が３枚	裏が２枚 表が２枚	裏が３枚 表が１枚	裏が４枚 表が０枚
確率変数（裏を０、表を１と対応させ和を取った）	4	3	2	1	0
事象が発生した回数	７５	３００	４５０	３００	７５
事象と対応する確率	１／１６	４／１６	６／１６	４／１６	１／１６

１．３　統計学 2-3（期待値）

期待値はおおむね平均値

・平均は足し合わせて個数で割る

・期待値はその分布における、確率変数の平均の値又は「ありそう」な値

事象X	ｘ１	ｘｎ
確率変数ｆ（X）　ｆ（ｘ１）		ｆ（ｘｎ）
確率Ｐ（X）	Ｐ（ｘ１）	Ｐ（ｘｎ）

$$\text{期待値}E(f)=\sum_{k=1}^{n} P(X=x_k)\, f(X=x_k)$$

連続する値ならば

$$\text{期待値}E(f)=\int P(X=x_k)\, f(X=x_k)\, dx$$

１．４　統計学 2-4（分散と共分散）

（1）分散＝平均との差の二乗

・データの散らばり具合

・データの各々の値が、期待値からどれだけズレているのか平均したもの

分散Ｖａｒ（ｆ）＝Ｅ（(（ｆ（Ｘ＝ｘ）－Ｅ（ｆ））2乗）

＝Ｅ（ｆ（Ｘ－ｘ）2乗）－（Ｅ（ｆ））2乗

もともとは絶対値であったが、計算が面倒なので２乗した。

$$
\begin{aligned}
&\text{分散}\mathrm{Var}(f)\\
&= \mathrm{E}\Big(\big(f_{(X=x)} - \mathrm{E}_{(f)}\big)^2\Big)\\
&= \mathrm{E}\big(f_{(X=x)}^2\big) - \big(\mathrm{E}_{(f)}\big)^2
\end{aligned}
$$

（2）共分散

・２つのデータ系列の傾向の違い

・正の値を取れば似た傾向

・負の値を取れば逆の傾向

・ゼロを取れば関係性に乏しい

$$
\begin{aligned}
&\text{共分散}\mathrm{Cov}(f, g)\\
&= \mathrm{E}\Big(\big(f_{(X=x)} - \mathrm{E}(f)\big)\big(g_{(Y=y)} - \mathrm{E}(g)\big)\Big)\\
&= \mathrm{E}(fg) - \mathrm{E}(f)\mathrm{E}(g)
\end{aligned}
$$

１．５　統計学 2-5（分散と標準偏差）

分散は２乗してしまっているので元のデータと単位が違う
２乗することの逆演算（つまり平行根を求める）をすれば元の単に戻る
標準偏差

$$
\begin{aligned}
\sigma &= \sqrt{\mathrm{Var}(f)} \\
&= \sqrt{\mathrm{E}\left(\left(f_{(X=x)} - \mathrm{E}_{(f)} \right)^2 \right)}
\end{aligned}
$$

１．６　統計学 2-6（様々な確率分布Ｉ・ベルヌーイ分布）

確率分布
(1) ベルヌーイ分布
- コイントスのイメージ
- 裏と表で出る割合が等しくなくとも扱える

μ：平均の値－＞ｘが１の時（表の確率）

$$P(x|\mu) = \mu^x (1-\mu)^{1-x}$$

ｘ＝０：裏の確率

(2) マルチヌーイ（カテゴリカル）分布
- サイコロを転がすイメージ
- 各面の出る割合が等しくなくても扱える

１．７　統計学 2-7（様々な確率分布Ⅱ）

（1）二項分布

・ベルヌーイ分布の多試行版

$$P(x|\lambda, n) = \frac{n!}{x!\,(n-x)!}\lambda^{x}(1-\lambda)^{n-x}$$

２項係数　　　　　　　表　　　　裏

全てが表、裏の場合は係数はいらない。

パターン分係数をかける。

ｘとｎ－ｘが同じぐらいになった時一番大きい

（2）ガウス分布

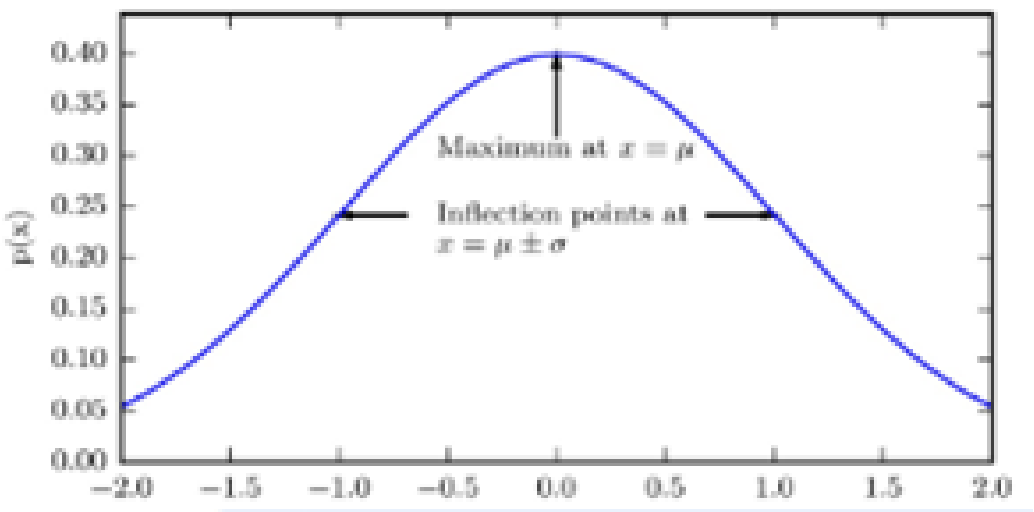
・釣鐘型の連続分布

・人工的に２項分布を作ったもの

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

・真の分布がわからなくてもサンプルが多ければ正規分布に近づく

・全て足し合わせると１になるように係数を付けている



１．８　統計学 2-8（推定）

推定：母集団を特徴づける母数（パラメーター：平均など）を統計的に推測すること。
　母数：母集団が持っている数

（1）点推定
　　平均値などを１つの値に推定すること。

（2）区間推定
　　平均値などが存在する範囲（区間）を推定すること

１．９　統計学 2-9（推定量と推定値）

（1）推定量（ｅｓｔｉｍａｔｏｒ）
　　パラメータを推定するために利用する数値の計算方法や計算式のこと（関数)。
　　推定関数ともいう。

２）推定値（ｅｓｔｉｍａｔｅ）
　　実際に試行を行った結果から計算した値。

　　真の値を θ とすると、推定値は θ ﾊｯﾄ（$\hat{\theta}$）のように表す

１．１０　統計学 2-10（標本平均）

標本平均

　母集団から取り出した標本の平均値

（1）サンプル数が大きくなれば、母集団に近づく。―――＞一致性

（2）サンプル数がいくらであっても、その期待値は母集団の値と同様。―――＞不遍性

$$E\left(\hat{\theta}\right) = \theta$$

１．１１　統計学 2-11（標本分散）

（1）標本分散

サンプルサイズをｎとすると

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

一致性は満たすが、不偏性は満たさない。

たくさんのデータのバラつき具合は、少数のデータのバラつきより大きくなる。

（2）不偏分散

$$s^2 = \frac{n}{n-1} \times \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

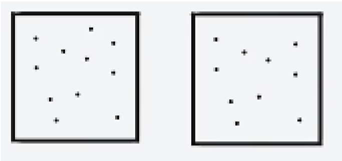
値を修正している

母集団の分散に近づける

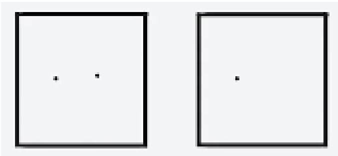
１．１２　統計学 2-12（増えた量は同じなのに、何が違う）

情報に関する考え方・情報科学

（1）　個数の違い



１０，１１（１個変化）　Δｗ＝１　　１／１０



１，２（１個変化）　　　Δｗ＝１　　１／２０

どちらもΔｗ＝１だが基の量に対しての増えた量の「比率」が異なる

（2）情報の感じ方
　１／ｗの積分ｌｏｇ（ｗ）

１．１３　統計学 2-13（自己情報量）

自己情報量

Ｉ（ｘ）＝－ｌｏｇ（Ｐ（ｘ））＝ｌｏｇ（Ｗ（ｘ））
　Ｗ（）＝１／Ｐ（）

低により単位を変える

（1）　対数の底が２の時、単位はビット（bit）

（2）対数の底がネイピアｅの時、単位はナット（nat）

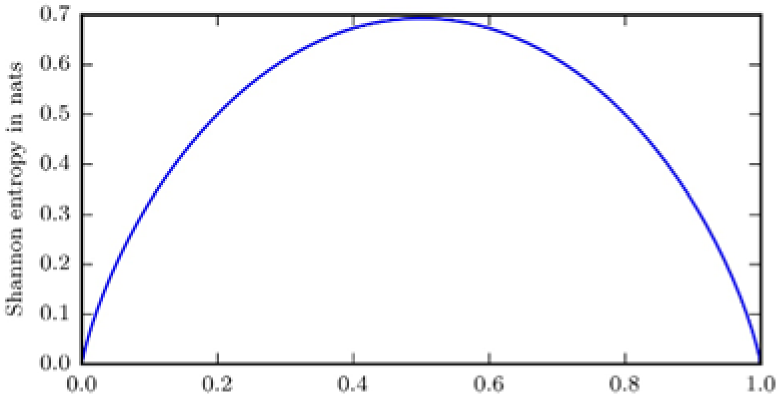
１．１４　統計学 2-14（シャノンエントロピー）

シャノンエントロピー

・微分エントロピーともいうが、部分しているわけでもない
・自己情報量の期待値

$$H(x) = E\big(I(x)\big)$$
$$= -E\Big(\log\big(P(x)\big)\Big)$$
$$= -\sum\Big(P(x)\log\big(P(x)\big) \Big)$$

コイン投げの一例



必ず裏　　　　　　必ず表

・情報量が最大になるようなところ

１．　１５　統計学 2-15（カルバック・ライブラー）

カルバック・ライブラー　ダイバージェンス

・同じ事象・確率変数における異なる確率分布 P,Q の違いを表す。

$$D_{\mathrm{KL}}(P\|Q) = \mathbb{E}_{\mathrm{x}\sim P}\left[\log \frac{P(x)}{Q(x)}\right] = \mathbb{E}_{\mathrm{x}\sim P}\left[\log P(x) - \log Q(x)\right]$$

平均

コインを投げる、裏と表は同じ確率であるが、実際確率が違う場合に使用する。

ダイバージェンス：距離に近い概念

$$I\big(Q(x)\big) - I\big(P(x)\big)$$
$$= \big(-\log\big(Q(x)\big)\big) - \big(-\log\big(P(x)\big)\big) = \log \frac{P(x)}{Q(x)}$$

I(Q)　当選最初の情報

I(P)　後になってわかった情報

期待値

$$E\big(f(x)\big) = \sum_{x} P(x) f(x)$$

情報利得

$$D_{\mathrm{KL}}(P\|Q) = \sum_{x} P(x)\,\big(-\log\big(Q(x)\big)\big) - \big(-\log\big(P(x)\big)\big) = \sum_{x} P(x) \log \frac{P(x)}{Q(x)}$$

自己情報量の期待値に似ている。

１．１６　統計学 2-16（交差エントロピー）

交差エントロピー

・KL ダイバージェンスの一部分を取り出したもの

・Q についての自己情報量を P の分布で平均している

$$D_{\mathrm{KL}}(P\|Q) = \sum_x P(x)\left(-\log(Q(x))\right) - \left(-\log(P(x))\right)$$

$$\boxed{H(P,Q)} = \boxed{H(P)} + D_{\mathrm{KL}}(P\|Q)$$

$$H(P,Q) = -\mathbb{E}_{\mathrm{x}\sim P}\log Q(x) = -\sum_x P(x)\log Q(x)$$