機械学習・統計検定

のための 数学

〜 確率 〜

確率とは、ある事象が起こる起こりやすさ。

↳ サイコロ（あまり出したくない）

非常に重要な一文

どの面も出る確率が等しいものとする。

事象

ある事象が起こったときの値

非常に重要

サイコロは1〜6の目が出る。

事象	1が出る	2が出る	3が出る	4が出る	5が出る	6が出る
確率変数	1	2	3	4	5	6
確率	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

全て足すと必ず1

事象：1の目が出る

確率変数：1

これは自分で定義できる。

確率の特徴

1. 各事象の確率は0以上1以下
2. 全ての事象の確率を足すと、必ず1になる。

確率変数と事象

例：コイン.

コインの表/裏が出る確率はそれぞれ $\frac{1}{2}$ とする。

事象	表	裏
確率変数?	1	0
確率	$\frac{1}{2}$	$\frac{1}{2}$

← 自分で決められる.

← 仮定・観測によって求められるため自分では決められない.

確率は P(事象) で表現する。すなわち

P(表) $= \frac{1}{2}$, P(裏) $= \frac{1}{2}$

厳密

全ての事象を X とする、

$X = \underbrace{\{表, 裏\}}_{集合の記号}$

このとき X の事象 x に対して

$P(x) = P(X = x)$

として確率を表現する。

先程の性質

1. 各事象の確率は0以上1以下

2. 全ての事象の確率の和は1

全ての事象 $\Longleftrightarrow$ $X := \{x_1, x_2, \dots, x_n\}$ とすると

サイコロの「1が出る」
コインの「表が出る」

1. $0 \leq p(x_i) \leq 1 \quad (i = 1, 2, \dots, n)$

2. $p(x_1) + p(x_2) + \dots + p(x_n) = \sum_{i=1}^{n} p(x_i) = 1$

総和、$i = 1, 2, \dots, n$ を順に足す。

期待値：全ての確率×確率変数を足したもの

← "平均"

あまりにも誤用が多い

・「期待~~できる度~~合い」

期待値：全ての確率×確率変数を足したもの　③

①　②

サイコロ

事象	1が出る	2が出る	3が出る	4が出る	5が出る	6が出る
確率変数	1	2	3	4	5	6
確率	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

← 事象と確率変数は1セット

→ xを確率変数ともみなせる。

確率　$P(x)$

事象，確率変数

$$\underbrace{\left(\frac{1}{6} \times 1\right)}_{\text{1が出る場合}} + \underbrace{\left(\frac{1}{6} \times 2\right)}_{\text{2が出る場合}} + \cdots + \underbrace{\left(\frac{1}{6} \times 6\right)}_{\text{6が出る場合}}$$

①　②　①　②　①　②

期待値

$$= \frac{1}{6} + \frac{2}{6} + \cdots + \frac{6}{6} = \frac{21}{6} = 3.5$$

$X = \{x_1, x_2, \ldots, x_n\}$: 全ての事象(確率変数)

$p(x_i)$ $(i=1,2,\cdots,n)$: x_i が起こる確率

コイン

$\underbrace{\frac{1}{2}\times 1}_{表} + \underbrace{\frac{1}{2}\times 0}_{裏} = \frac{1}{2}$

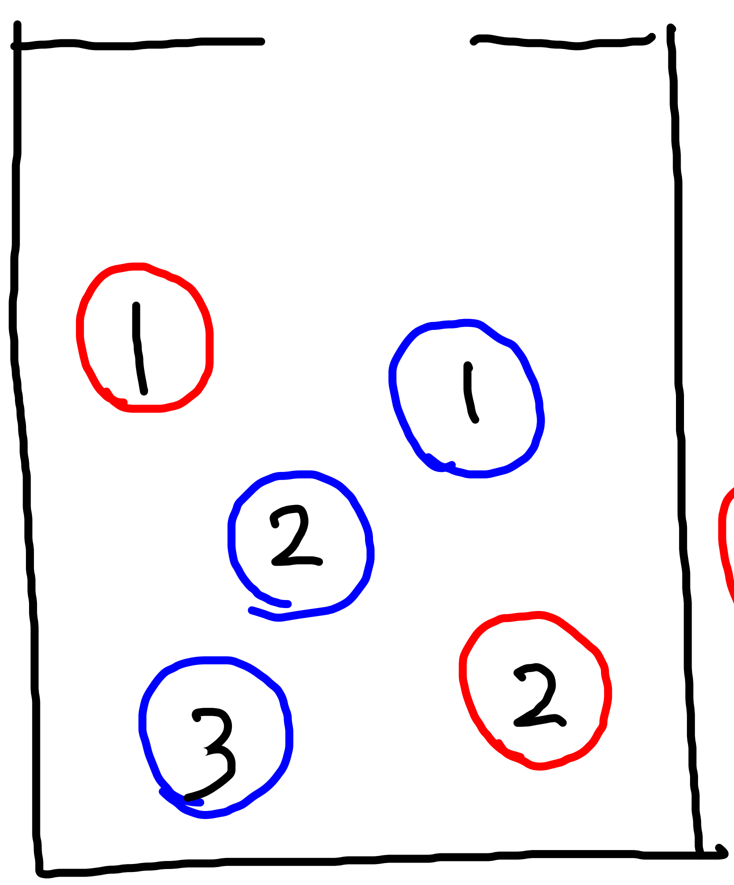
このときの期待値 μ (mu ミュー) は

$$\mu = \underbrace{\sum_{i=1}^{n}}_{総和} \underbrace{p(x_i)}_{確率} \underbrace{x_i}_{確率変数}$$

と定義する。

次に

確率は全て足すと1 (しつこいですが)



・各ボールを取る確率は $\frac{1}{5}$ → $\frac{1}{5}+\frac{1}{5}+\frac{1}{5}+\frac{1}{5}+\frac{1}{5}=1$

色

・赤のボールを取る確率は $\frac{2}{5}$

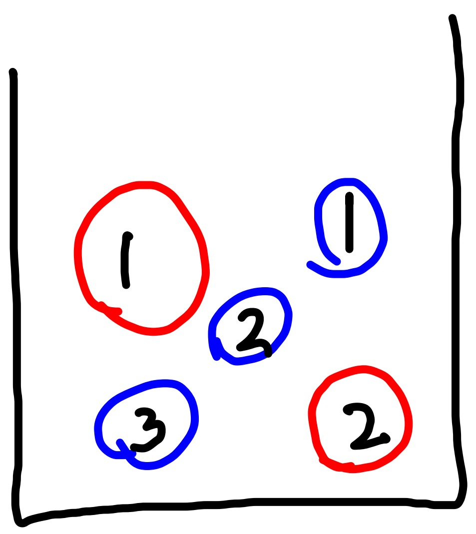
青のボールを取る確率は $\frac{3}{5}$ → $\frac{2}{5}+\frac{3}{5}=1$

数

1のボールを取る確率は $\frac{2}{5}$

2のボールを取る確率は $\frac{2}{5}$ → $\frac{2}{5}+\frac{2}{5}+\frac{1}{5}=1$

3のボールを取る確率は $\frac{1}{5}$



各ボール … $\frac{1}{5}$ ずつ

色 $\begin{cases} \frac{2}{5} & (赤) \\ \frac{3}{5} & (青) \end{cases}$

数 $\begin{cases} \frac{2}{5} & (1) \\ \frac{2}{5} & (2) \\ \frac{1}{5} & (3). \end{cases}$

赤かつ1

赤かつ3 → ない

	1	2	3	
赤	$\frac{1}{5}$	$\frac{1}{5}$	0	$\frac{2}{5}$
青	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{3}{5}$
	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$ → 1	↓ 1

青かつ2

ここで

	1	2	3
赤	$\frac{1}{5}$	$\frac{1}{5}$	0
青	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

ボールを取った後に ボールの色が青であることがわかりました。

書かれている数字が 2 である確率は？

$\cancel{\frac{1}{5}}$

答えは $\frac{1}{3}$ です。

解説 (但し直感的でないため難しいです。)

	$\frac{1}{5}$	$\frac{1}{5}$	0
	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

青とわかったときに、$\boxed{}$ のいずれかということになる

↓

	0	0	0
	?	?	?

その場合赤である確率は 0 になる。

しかし、①②③の確率が $\frac{1}{5}$ のままだと、

$\frac{1}{5}+\frac{1}{5}+\frac{1}{5}=\frac{3}{5}$ で確率を全て足すと1ということと矛盾。

↓

	0	0	0
	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

青の玉は3つのため、各確率は $\frac{1}{3}$ ずつとなる。

?

このような とある条件を元に計算される π 条件付き確率 という.

Y を条件に X を起こる確率を

X かつ Y が起こる確率

$$P(X \mid Y) = \frac{P(X \cap Y)}{P(Y)}$$

Y を条件とした X

Y が起こる確率

と定義する。

青を条件として 2 が出るとき

	1	2	3	
赤	$\frac{1}{5}$	$\frac{1}{5}$	0	$\frac{2}{5}$
青	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{3}{5}$

P(2 かつ 青)　　P(青)

$$P(2 \mid 青) = \frac{P(2 かつ 青)}{P(青)}$$

$$= \frac{\frac{1}{\cancel{5}}}{\frac{3}{\cancel{5}}} = \frac{1}{3}$$

	1	2	3
赤	$\frac{1}{5}$	$\frac{1}{5}$	0
青	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$

P(青∩3)

P(3)

3を条件としたときの青が出る確率

$$P(青|3) = \frac{P(青かつ3)}{P(3)}$$

$$= \frac{\frac{1}{5}}{\frac{1}{5}}$$

$$= 1$$

「3とわかったら青しかない.」という直感と合う.

Yを条件としたときの

Xが出る確率

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$$

ここで「かつ」を確認。 ← 本だと必ず条件付き確率の前にやる。

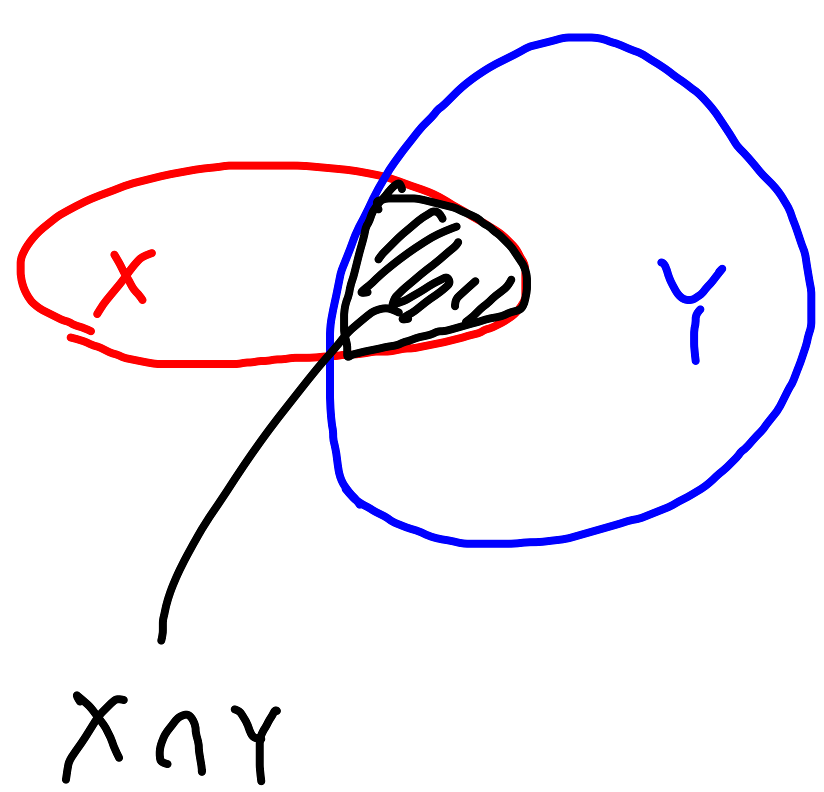
X と Y が"同時に"起こるとき、「X かつ Y」といい、

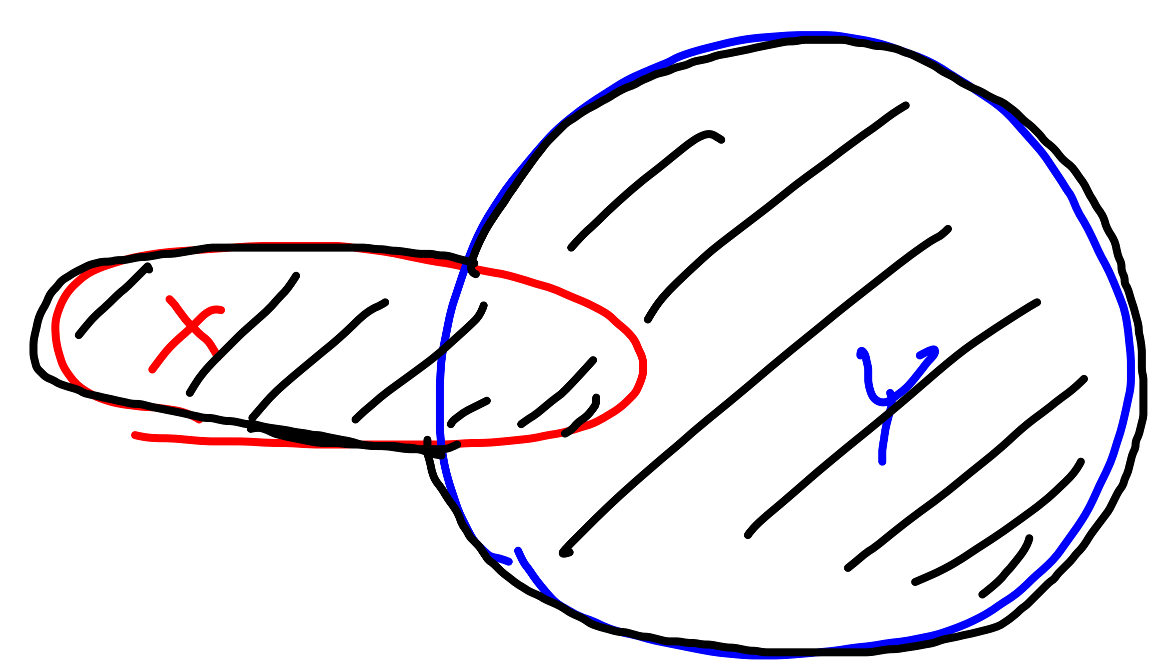
$$X \cap Y \quad \cap : \text{cap}$$

と表します。また、X と Y のいずれかが起こるとき「X または Y」といい、

$$X \cup Y \quad \cup : \text{cup}$$

と表します。





X が起こる確率を P(X), Y が起こる確率を P(Y) とする.

また

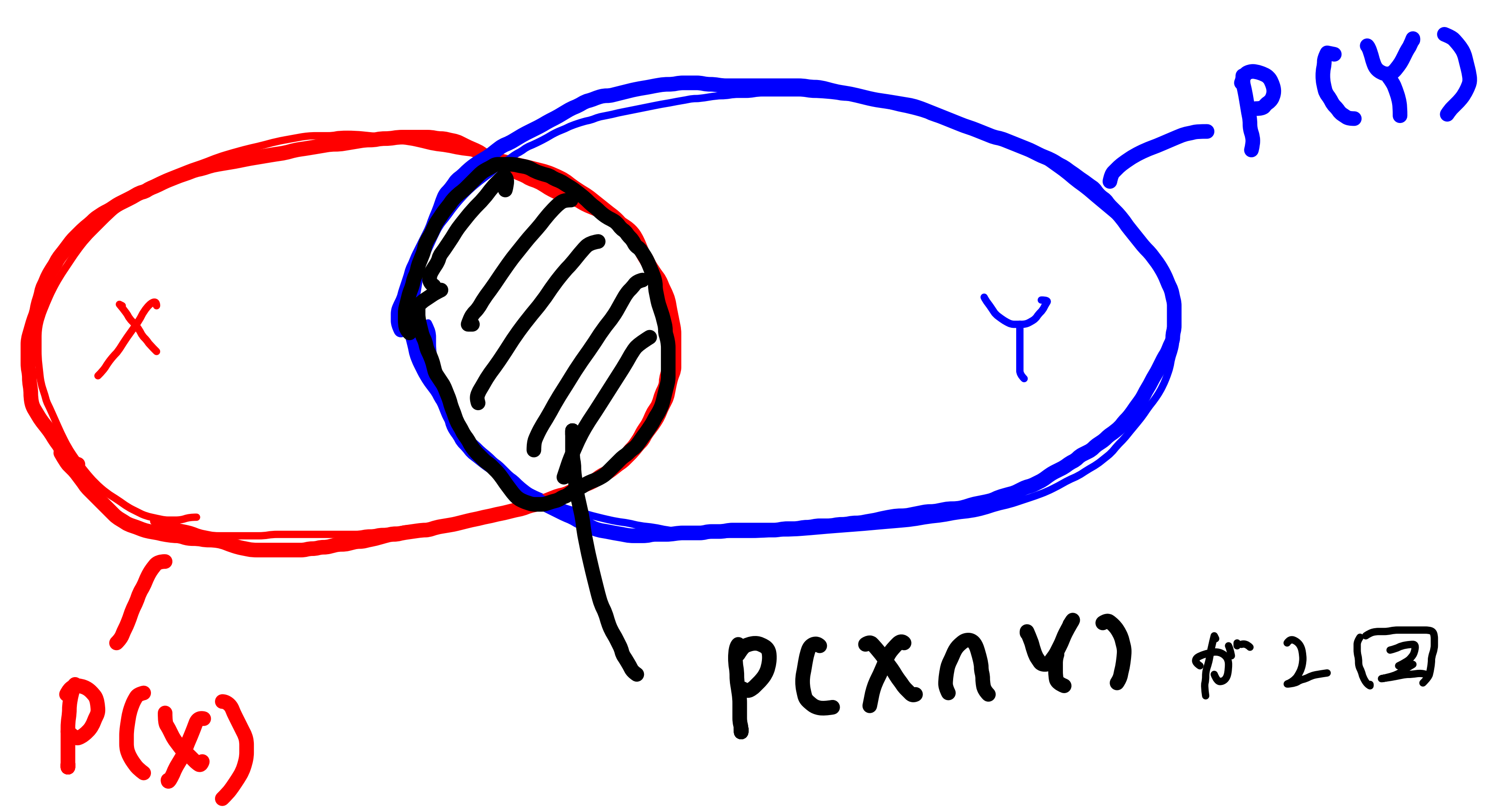
$$\begin{cases} P(X\cap Y) : X \text{ かつ } Y \text{ が起こる確率} \\ P(X\cup Y) : X \text{ または } Y \text{ が起こる確率} \end{cases}$$

とする。

このとき

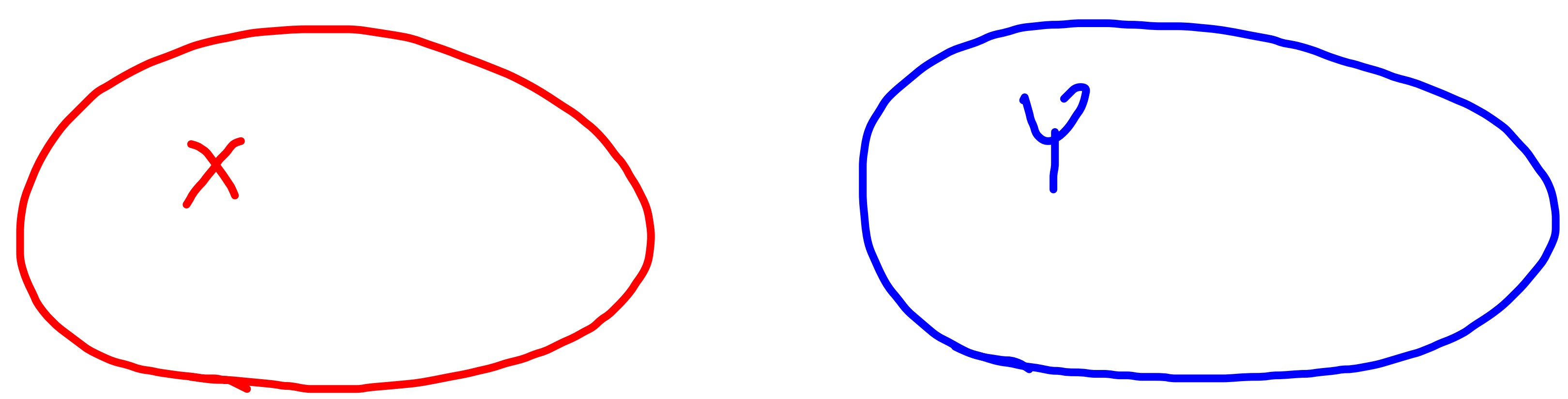
$$P(X\cup Y) = P(X) + P(Y) - \underline{P(X\cap Y)}$$

が成り立つ。



P(X∩Y) が2回足されている。

では X と Y が離れているときは？



X∩Y がない

$\Longleftrightarrow P(X \cap Y) = 0$

$$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$$

$$= P(X) + P(Y)$$

どちらか一方が起こる確率

$P(X \cap Y)$

XとYが両方同時に起こる確率

例：感染症が流行する確率と感染する確率

$\underbrace{\text{感染症が流行する確率}}_{X}$ と $\underbrace{\text{感染する確率}}_{Y}$

$P(X \cap Y)$

感染症が流行したときに感染する確率 ← 条件付き確率

この計算で使う

$$P(Y=O \mid X=O) = \frac{P(Y=O, X=O)}{P(X=O)}$$

$$= \frac{0.06}{0.6}$$

$$= \frac{1}{10}$$

感染症が流行したときに感染する確率

$P(Y=O, X=O)$

流行

○	60%
×	40%

感染

○	10%
×	90%

$P(X)$

感染 / 流行	○	×	
○	0.06	0.54	0.6
×	0.04	0.36	0.4
	0.1	0.9	

流	感	
○	○	$0.6 \times 0.1 = 0.06$
○	×	$0.6 \times 0.9 = 0.54$
×	○	$0.4 \times 0.1 = 0.04$
×	×	$0.4 \times 0.9 = 0.36$

問題

計算して下さい(STOP!)

	○	×
流行	60%	40%
感染	10%	90%

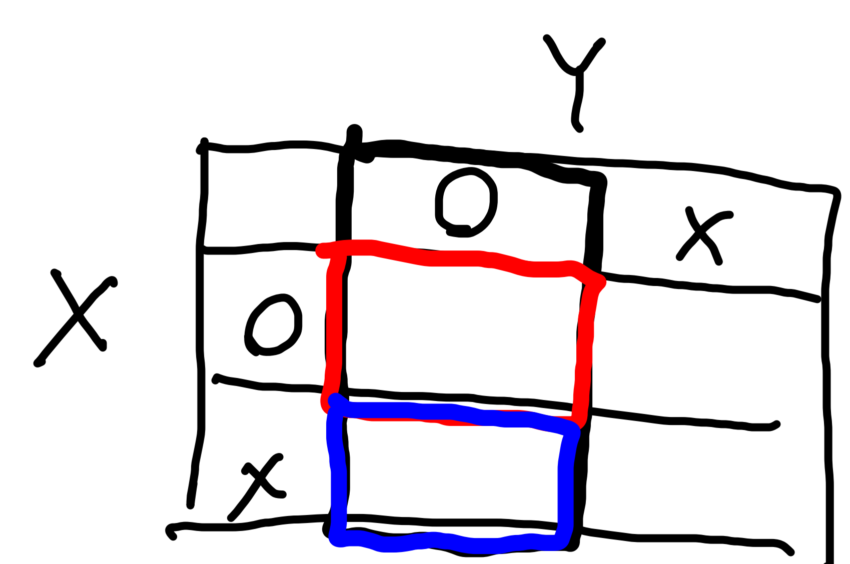
$$P(Y=\bigcirc \mid X=\bigcirc) = \frac{1}{10}$$
$$P(Y=\bigcirc \mid X=\times) = \frac{1}{10}$$
$$P(Y=\times \mid X=\bigcirc) = \frac{9}{10}$$
$$P(Y=\times \mid X=\times) = \frac{9}{10}$$

足すと1になる.

$$P(X=\bigcirc \mid Y=\bigcirc) = \frac{3}{5}$$
$$P(X=\bigcirc \mid Y=\times) = \frac{3}{5}$$
$$P(X=\times \mid Y=\bigcirc) = \frac{2}{5}$$
$$P(X=\times \mid Y=\times) = \frac{2}{5}$$

→ 条件を固定した上で起こりうる事象の確率を足すと1になる.

流 \ 感	○	×	
○	0.06	0.54	0.6
×	0.04	0.36	0.4
	0.1	0.9	



Y=○を条件

$P(X=\bigcirc \mid Y=\bigcirc)$

$P(X=\times \mid Y=\bigcirc)$

$$P(X=\bigcirc \mid Y=\bigcirc) + P(X=\times \mid Y=\bigcirc) = 1$$

$X = \{x_1, x_2, \dots, x_n\}$, $Y = \{y_1, y_2, \dots, y_m\}$

n個の事象　　　　m個の事象

あるy_iを条件として起こるXの全ての事象の確率の和、すなわち

$$\sum_{x_j \in X} p(X = x_j \mid Y = y_i) = 1$$

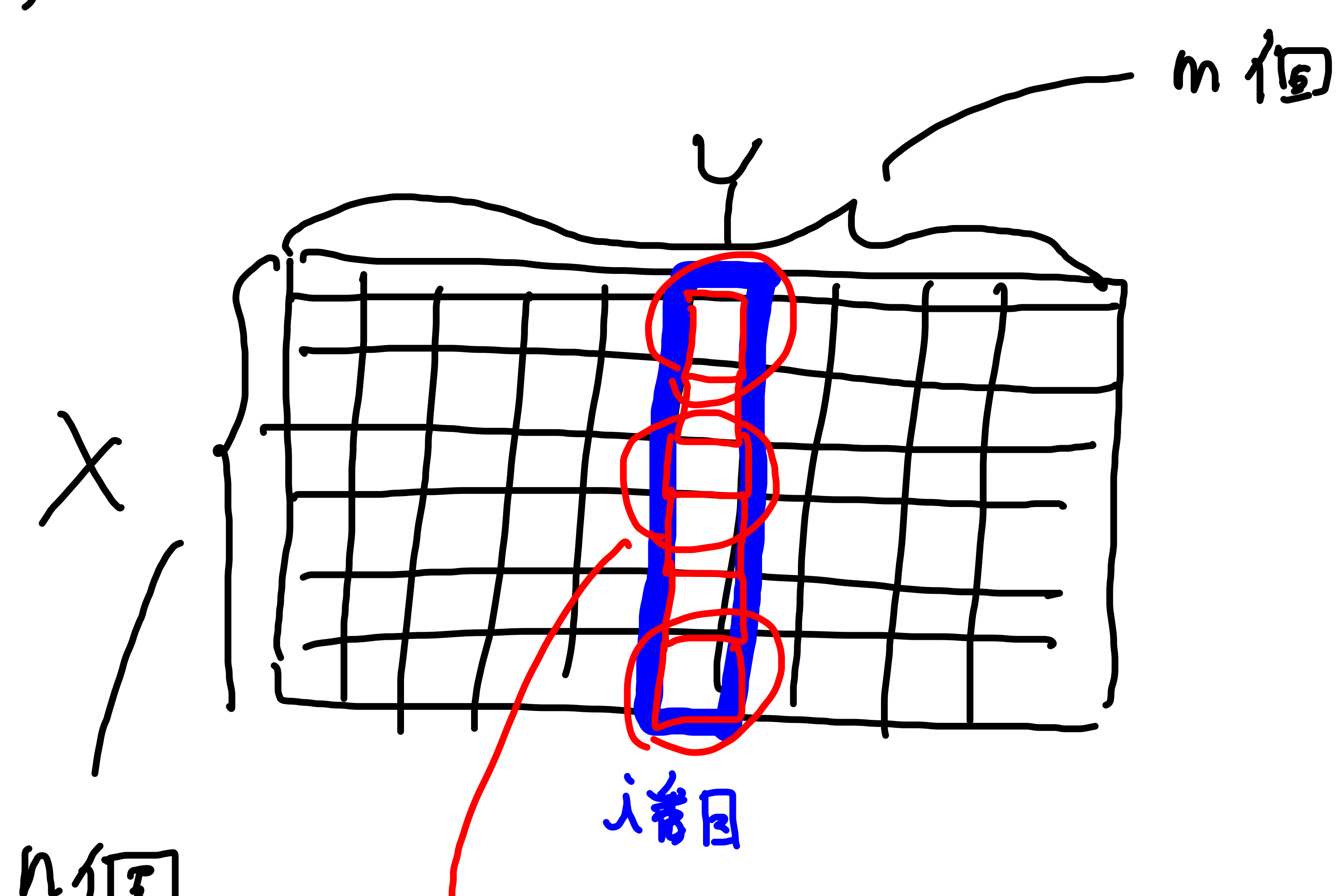
jを使っているのは

iを使っているから。

$x_j \in X$

in

Xからx_jを順に取り出す。



$$\sum_{x_j \in X} P(x_j \mid y_i)$$

独立性

サイコロを2個転がしたときに出るそれぞれの目は、それぞれ独立していて変わらない。

一方、先の問題のように、感染症の流行と感染リスクは影響する。

サイコロのように独立している事象を独立事象と呼び、

感染症のように一方が他方に影響するものを従属事象という。

AとBが独立事象のとき

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

が成り立つ

サイコロAが1、サイコロBが6であるとき、その確率は

$$P(A \cap B) = \underbrace{\frac{1}{6}}_{= P(A)} \times \underbrace{\frac{1}{6}}_{= P(B)} = \frac{1}{36}$$

さらにその条件つき確率 $P(A|B)$, $P(B|A)$ は

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

$$= P(A)$$

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

$$= P(B)$$

まとめ

・見返して何度も
　計算してみよう.

次回予告

連続と確率