# データサイエンスのための統計学

# 確率

by 尚晋　(名古屋大学経済学研究科助教)
on 2025 年 5 月 20 日

# 目次

# 目次

# » 確率論の基本概念：試行・事象・確率

**試行** 偶然に左右され、事前に結果が分からない行動を試行（確率的試行）と呼ぶ。例：「サイコロを振る。」

**事象** 試行の結果として起こり得る事柄を、事象と呼ぶ。例：「サイコロを振って 1 が出る。」

**確率** 事象それぞれの起こりやすさを測った数量を、確率と呼ぶ。例：「サイコロを振って 1 が出る確率は 1/6.」

**ランダム性** 「確率」と関係の深い概念に「ランダム性」がある. ランダム性とは, 次に何が起こるか確定的に予測できないこと. 例：コインを投げて表か裏かを事前に言い当てるのは困難. ただし, ランダム性には法則性があり, それに基づいて確率論が成り立つ.

**確率論とランダム性** 確率論は「ランダム性そのもの」ではなく, 「ランダム性の法則」を扱い, 統計学はそれを利用する.

# 目次

* 確率論の基本概念
* 標本空間と事象
* 確率
* 条件付き確率
* 事象の独立性
* ベイズの定理

# » 標本空間・標本点

**標本点** ある実験において起こりうる結果のこと. 一般的に, $\omega$ で表す. 例えば, サイコロを一回投げた可能な結果は, $1, 2, 3, 4, 5, 6$.

**標本空間 (全事象)** 標本点全体の集合のこと. 一般的に, $\Omega$ で表す. 例 (1), サイコロを一回投げた例では, 標本空間は $\{1, 2, 3, 4, 5, 6\}$. 一般化したもの:$\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$

**事象** 標本点 $\omega$ を組み合わせてできる集合のこと, 即ち, 事象は標本空間の部分集合. 例えば, サイコロを一回投げた場合,「出た目が奇数である」という事象 $A = \{1, 3, 5\}, A \subset \Omega$

**根元事象** ただ一つの標本点からなり分解できない事象のこと. 例えば, サイコロ 2 回投げた場合, $(1, 2)$ は標本点の一つとなり, さらに分解すると, $\{1\}, \{2\}$ は根本事象.

## » 事象の演算 例

排反 例えば, サイコロを一回投げた可能な結果は, 事象 $A$ は奇数出る, 排反事象となる事象 $B$ は____ 出る.

和事象 例, さいころを 1 回投げる場合, A を奇数の目が出るという事象, B を 3 以下 (3 を含め) の目が出るという事象とすると,
$A \cup B = \{\_\_\_\_\_\_\_\}$

積事象 例, さいころを 1 回投げる場合, A を奇数の目が出るという事象, B を 3 以下の目が出るという事象とすると,
$A \cap B = \{\_\_\_\}$

補事象 例, サイコロ投げに関する事象 $A = 1, 3, 5$（奇数）の補事象
$A^c = \{\_\_\_\_\_\}$

# » 事象の演算 例

**排反** 例えば, サイコロを一回投げた可能な結果は, 事象 $A$ は奇数出る, 排反事象となる事象 $B$ は偶数 出る.

**和事象** 例, さいころを 1 回投げる場合, A を奇数の目が出るという事象, B を 3 以下 (3 を含め) の目が出るという事象とすると,
$A \cup B = \{\underline{1, 2, 3, 5}\}$

**積事象** 例, さいころを 1 回投げる場合, A を奇数の目が出るという事象, B を 3 以下の目が出るという事象とすると,
$A \cap B = \{\underline{1, 3}\}$

**補事象** 例, サイコロ投げに関する事象 $A = 1, 3, 5$（奇数）の補事象
$A^c = \{\underline{2, 4, 6}\}$

# 目次

# » 確率の公理

**確率論の公理は次の三つからなる**

$$(a)\text{すべての事象 } A \text{ に対して} 0 \leq P(A) \leq 1 \tag{1a}$$
$$(b) P(\Omega) = 1 \tag{1b}$$
$$(c)\text{排反事象 } A_1, A_2, \ldots \text{ に対して} \quad P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots \tag{1c}$$

式1cは $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ とも表記できる.

例えば, コインを投げた場合, 公理主義的定義によると:

$$P(A) := \begin{cases} 0 & for \quad A = \phi \\ 1/2 & for \quad A = \{\text{表}\}, \{\text{裏}\} \\ 1 & for \quad A = \Omega \end{cases}$$

# » 加法定理

**確率論の公理は次の三つからなる**

$$(a) A \text{ と } B \text{ が排反事象の場合：} P(A \cup B) = P(A) + P(B) \tag{2a}$$

$$(b) A \text{ と } B \text{ が共通部分持つ場合：} \quad P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2b}$$

加法定理計算例：

さいころ投げにおいて $A$ を奇数の目が出る事象, $B$ を 3 以下の目が出る事象とする.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 3/6 + 3/6 - 2/6 = 2/3$$

# 目次

* 確率論の基本概念
* 標本空間と事象
* 確率
* 条件付き確率
* 事象の独立性
* ベイズの定理

# » 条件付き確率

**条件付き確率** ある事象 B が起こったという条件のもとでの事象 A の起こる確率. $P(A|B)$ で表す (P A ギブン B と読む).「B を条件とする A の条件付き確率」と言っても良い。定義:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{3}$$

例, さいころを 1 回投げる場合, A を奇数の目が出るという事象, B を 3 以下 (3 を含め) の目が出るという事象とすると,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\text{奇数かつ 3 以下})}{P(\text{3 以下})} = \frac{2/6}{3/6} = \frac{2}{3}$$

**乗法定理**

条件付き確率の定理の式3を積の形に変形すれば:

$$P(A \cap B) = P(B) \cdot P(A|B) \tag{4}$$

# » 乗法定理の応用例

例：下記の情報からある学生が経済学部 ($P(EC)$) の学生でしかも数学好きと答えた確率は？

| | $P(EC)$ | $P(EN)$ | $P(SC)$ |
|---|---|---|---|
| 在籍人数 | 200 | 500 | 300 |
| $P(suki)$ 数学が好き | 30 | 60 | 70 |
| $P(kirai)$ 数学が嫌い | 70 | 40 | 30 |
| 調査対象の人数の合計 | 100 | 100 | 100 |

表: 数学が好きかどうかの調査データ

まず、答えた学生は経済学部に所属する確率を計算し、次に、経済学部の学生の中に数学が好きと答えた確率を計算し、乗法定理を使うと、

________________________________________ となる。

# » 乗法定理の応用例

例：下記の情報からある学生が経済学部 ($P(EC)$) の学生でしかも数学好きと答えた確率は？

| | $P(EC)$ | $P(EN)$ | $P(SC)$ |
|---|---|---|---|
| 在籍人数 | 200 | 500 | 300 |
| $P(suki)$ 数学が好き<br>$P(kirai)$ 数学が嫌い | 30<br>70 | 60<br>40 | 70<br>30 |
| 調査対象の人数の合計 | 100 | 100 | 100 |

表: 数学が好きかどうかの調査データ

まず、答えた学生は経済学部に所属する確率を計算し、次に、経済学部の学生の中に数学が好きと答えた確率を計算し、乗法定理を使うと、
$\underline{P(suki \cap EC) = P(EC) \times P(suki|EC) = 0.2 \times 0.3 = 0.06}$ となる。

# 目次

# » 事象の独立性

| | | | |
|---|---|---|---|
| 黒 | ❶❶❷ | 「1」が 2 個 | 「2」が 1 個 |
| 白 | ①①② | 「1」が 2 個 | 「2」が 1 個 |

ボールの数字が 1 である確率とボールの色に基づく条件付き確率をそれぞれ計算すれば: ________________ となる。

即ち, 数字に関する事象の確率は, 色に関する事象に影響されない.

事象の独立性 事象 $A$ の起る確率が他の事象 $B$ に影響されない場合,

$$P(A) = P(A|B) \tag{5}$$

である場合, この二つ事象は独立であるという.

この関係を乗法定理の式4に代入すると, 事象 $A$ と事象 $B$ が独立である場合, 下記の式が成り立つ:

$$P(A \cap B) = P(A) \cdot P(B) \tag{6}$$

即ち, 積事象の確率を ________________ として表すことができるとき二つの事象は独立である.

# » 事象の独立性

| | | | |
|---|---|---|---|
| 黒 | ❶❶❷ | 「1」が 2 個 | 「2」が 1 個 |
| 白 | ①①② | 「1」が 2 個 | 「2」が 1 個 |

ボールの数字が 1 である確率とボールの色に基づく条件付き確率をそれぞれ計算すれば：$P(1) = 4/6, P(1|Black) = \frac{2/6}{3/6}, P(1|White) = \frac{2/6}{3/6}$ となる。

即ち, 数字に関する事象の確率は, 色に関する事象に影響されない.

事象の独立性　事象 $A$ の起る確率が他の事象 $B$ に影響されない場合,

$$P(A) = P(A|B) \tag{5}$$

である場合, このニつ事象は独立であるという.

この関係を乗法定理の式4に代入すると, 事象 $A$ と事象 $B$ が独立である場合, 下記の式が成り立つ：

$$P(A \cap B) = P(A) \cdot P(B) \tag{6}$$

即ち, 積事象の確率を二つの事象の確率の積 として表すことができるとき二つの事象は独立である.

# 目次

* 確率論の基本概念
* 標本空間と事象
* 確率
* 条件付き確率
* 事象の独立性
* ベイズの定理

# » ベイズの定理

**ベイズの定理**

いま, 原因 $R_1, R_2, \ldots, R_k$ は互いに排反で, かつ $R_1 \cup R_2 \cup \ldots \cup R_k$ のごとくすべての場合をつくしているとする. このとき, 以下の規則が成り立つ:

$$P(R_i|A) = \frac{P(R_i) \cdot P(A|R_i)}{P(A)} = \frac{P(R_i) \cdot P(A|R_i)}{\Sigma P(R_j) \cdot P(A|R_j)} \tag{7}$$

* ここに $P(R_i)$ は事前確率, $P(R_i|A)$ は事後確率と呼ばれる.「事前」「事後」は事象 $A$ が起こったことを基準としている. 事前確率はしばしば主観確率が用いられる.

# » ベイズの定理 応用例 (1)

下記の条件で、無作為に選んだメールが『キャンペーン』という単語を含んでいたという条件のもとで、それが迷惑メールである確率は?(spam は迷惑メール;ham は一般メール)

$$
\begin{cases}
\text{事象 A: メールが迷惑メールの確率は} P(spam) = 0.2 \\
\text{事象 B: キャンペーンという単語を含める確率は二つの場合ある:} \\
\begin{cases}
\text{迷惑メール (0.2) 条件の下で『キャンペーン』を含める確率は 0.3:} P(camp|spam) = 0.3 \\
\text{一般メール (0.8) 条件の下で『キャンペーン』を含める確率は 0.04:} P(camp|ham) = 0.04
\end{cases}
\end{cases}
$$

$$
\begin{aligned}
P(camp) &= P(camp|spam) \times P(spam) + P(camp|ham) \times P(ham) \\
&= 0.3 \times 0.2 + 0.04 \times 0.8 \\
&= 0.092
\end{aligned}
$$

$$
\begin{aligned}
P(spam|camp) &= \frac{P(camp|spam) \times P(spam)}{P(camp)} \\
&= \frac{0.3 \times 0.2}{0.092} \approx 0.652
\end{aligned}
$$

## » ベイズの定理 応用例 (2)

祐介さんは桜さんと同じクラスで、祐介さんは桜さんに片思いが、桜さんの気持ちは分からず、告白するかを悩んでいる。この場合、ベイズの定理は使えるかも。
具体的には、まず、「一緒にゲームをやりましょうか」を誘ってみる。

いま、事件 $S$ を桜さんも祐介さんのことが好き ($P(S) = 0.5, P(S^c) = 0.5$)、事件 $Y$ を「一緒にゲームをやりましょうか」の答えは「Yes」とする。

事象 Y: 答えは Yes の確率は二つの場合ある:

$$\begin{cases} \text{好き (0.5) 条件の下で『Yes』の確率は 0.7:} \rightarrow P(Y|S) = 0.7 \\ \text{好きではない (0.5) 条件の下で『Yes』の確率は 0.2:} \rightarrow P(Y|S^c) = 0.2 \end{cases}$$

したがって、$P(Y^c|S) = 0.3, P(Y^c|S^c) = 0.8$ も分かる。

## » ベイズの定理 応用例 (2) 続き

もし一回目で、桜さんの答えは「No」であれば、桜さんは祐介さんのことが好きの確率は 0.5 から 0.27 に下がる：

$$
\begin{aligned}
P(S|Y^c) &= \frac{P(S) \times P(Y^c|S)}{P(Y^c)} \\
&= \frac{P(S) \times P(Y^c|S)}{P(Y^c|S) \times P(S) + P(Y^c|S^c) \times P(S^c)} \\
&= \frac{0.5 \times 0.3}{0.3 \times 0.5 + 0.8 \times 0.5} \approx 0.27
\end{aligned}
$$

もし祐介さんは諦めが悪い方で、再度誘った、二回目で桜さんの答えも「No」であれば、桜さんは祐介さんのことが好きの確率は 0.27 から、さらに 0.12 に下がる。(この時、$P(S) = 0.27$, $P(S^c) = 0.73$ となることを注意。)

$$
\begin{aligned}
P(S|Y^c) &= \frac{P(S) \times P(Y^c|S)}{P(Y^c|S) \times P(S) + P(Y^c|S^c) \times P(S^c)} \\
&= \frac{0.27 \times 0.3}{0.3 \times 0.27 + 0.8 \times 0.73} \approx 0.12
\end{aligned}
$$