確率・統計　第1回　～事象と確率～

定義 1　試行と事象

- 試行: 同じ条件のもとで反復可能な実験や観測のこと
- 事象: 試行の結果として起こること
- 全事象: 試行の起こりうる結果全体、U と表す
- 根元事象: それ以上分けることができない事象
- 空事象: 決して起こらない事象、$\emptyset$ と表す

事象は、集合に対応させることができる。

$$\begin{array}{rcl} \text{全事象} & \longleftrightarrow & \text{全体集合} \\ \text{事象} & \longleftrightarrow & \text{部分集合} \\ \text{根元事象} & \longleftrightarrow & 1\,\text{点集合} \\ \text{空事象} & \longleftrightarrow & \text{空集合} \end{array}$$

以後、事象と対応する集合は同一視する。

例 1. 1から5までの番号が書かれた5枚のカードから、1枚のカードを引く試行を考える。

- 全事象 $U = \{1, 2, 3, 4, 5\}$
- 根元事象: $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$
- 偶数である事象 $A = \{2, 4\}$

試行の根元事象がいずれも公平に起こると考えられるとき、これらの根元事象は同様に確からしいという。以後、根元事象が同様に確からしい試行について考える。

定義 2　確率

試行において、全事象 U に含まれる根元事象の数を $n(U)$, 事象 A に含まれる根元事象の数を $n(A)$ とするとき、$\frac{n(A)}{n(U)}$ を事象 A の確率といい、$P(A)$ で表す:

$$P(A) = \frac{n(A)}{n(U)} = \frac{\text{事象 } A \text{ が起こる場合の数}}{\text{起こりうる全ての場合の数}}.$$

確率の基本性質

試行における全事象 U, 空事象 $\emptyset$, 任意の事象 A の確率に対して、次が成り立つ:

$$P(U) = 1, \quad P(\emptyset) = 0, \quad 0 \leq P(A) \leq 1.$$

二項係数

n 個のものから r 個取る組合せの総数は、

$${}_nC_r = \frac{n!}{r!(n-r)!}.$$

ただし、$0! = 1$ とする。

例題 1　確率の計算

赤球 3 個と白球 7 個が入っている袋から、4 個の球を同時に取り出すとき、次の確率を求めよ。

(1) 4 個とも白球である確率

(2) 赤球が 3 個、白球が 1 個である確率

解答.

10 個の球から 4 個の球を取り出す場合の数は、${}_{10}C_4$ 通り。

(1) 白球 4 個を取り出す場合の数は、${}_7C_4$ 通り。よって、求める確率は、$\frac{{}_7C_4}{{}_{10}C_4} = \frac{1}{6}$.

(2) 赤球 3 個と白球 1 個を取り出す場合の数は、${}_3C_3 \cdot {}_7C_1$ 通り。よって、求める確率は、$\frac{{}_3C_3 \cdot {}_7C_1}{{}_{10}C_4} = \frac{1}{30}$.

□

定義 3　和事象と積事象

試行における事象 A, B に対して、A または B が起こる事象を A と B の和事象といい、$A \cup B$ で表す。また、A と B がともに起こる事象を A と B の積事象といい、$A \cap B$ で表す。

$$\begin{array}{ccc} \text{和事象} & \longleftrightarrow & \text{和集合} \\ \text{積事象} & \longleftrightarrow & \text{共通部分} \end{array}$$

定義 4　排反事象

試行における事象 A, B に対して、A と B が同時に起こりえないとき、A と B は互いに排反であるという。

$$A \text{ と } B \text{ は互いに排反} \quad \longleftrightarrow \quad A \cap B = \emptyset$$

根元事象は、全て互いに排反である。

定理 1　加法定理

試行における事象 A, B に対して、次が成り立つ:

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

特に、A と B が互いに排反であるとき、

$$P(A \cup B) = P(A) + P(B).$$

証明.

$n(A \cup B) = n(A) + n(B) - n(A \cap B)$ より、

$$P(A \cup B) = \frac{n(A \cup B)}{n(U)} = \frac{n(A)}{n(U)} + \frac{n(B)}{n(U)} - \frac{n(A \cap B)}{n(U)} = P(A) + P(B) - P(A \cap B).$$

□

例題 2　加法定理

1 から 100 までの番号が書かれた 100 枚のカードから、1 枚のカードを引くとき、次の確率を求めよ。

(1) カードの番号が 6 の倍数または 8 の倍数である確率

(2) カードの番号が 10 の倍数または 11 の倍数である確率

解答.

(1) 6 の倍数である事象を A, 8 の倍数である事象を B とすると、$A = \{6, 12, \cdots, 96\}$, $B = \{8, 16, \cdots, 96\}$. よって、$A \cap B = \{24, 48, 72, 96\}$. ゆえに、$P(A) = \frac{16}{100}$, $P(B) = \frac{12}{100}$, $P(A \cap B) = \frac{4}{100}$. 加法定理 1 より、求める確率は、

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{16}{100} + \frac{12}{100} - \frac{4}{100} = \frac{6}{25}.$$

(2) 10 の倍数である事象を C, 11 の倍数である事象を D とすると、$C = \{10, 20, \cdots, 100\}$, $D = \{11, 22, \cdots, 99\}$. よって、$C \cap D = \emptyset$ より、C と D は互いに排反である。また、$P(C) = \frac{10}{100}$, $P(D) = \frac{9}{100}$. 加法定理 1 より、求める確率は、

$$P(C \cup D) = P(C) + P(D) = \frac{10}{100} + \frac{9}{100} = \frac{19}{100}.$$

□

定義 5　余事象

試行における事象 A に対して、A が起こらない事象を A の余事象といい、$\overline{A}$ で表す。

余事象　$\longleftrightarrow$　補集合

事象 A と余事象 $\overline{A}$ について、$A \cup \overline{A} = U$, $A \cap \overline{A} = \emptyset$.

定理 2　余事象の確率

試行における事象 A に対して、次が成り立つ:

$$P(\overline{A}) = 1 - P(A).$$

証明.

$A \cup \overline{A} = U$, $A \cap \overline{A} = \emptyset$ より、加法定理 1 から、

$$\begin{aligned}
P(A) + P(\overline{A}) &= P(A) + P(\overline{A}) - P(\emptyset) && \leftarrow P(\emptyset) = 0 \\
&= P(A) + P(\overline{A}) - P(A \cap B) && \leftarrow A \cap \overline{A} = \emptyset \\
&= P(A \cup \overline{A}) && \leftarrow \text{加法定理 } 1 \\
&= P(U) && \leftarrow A \cup \overline{A} = U \\
&= 1 && \leftarrow P(U) = 1.
\end{aligned}$$

□

例題 3　余事象の確率

赤球 5 個と白球 3 個が入っている袋から、3 個の球を同時に取り出すとき、少なくとも 1 個は白球である確率を求めよ。

解答.

少なくとも 1 個は白球である事象は、3 個とも赤球である事象 A の余事象 $\overline{A}$ となる。8 個の球から 3 個の球を取り出す場合の数は、${}_8C_3$ 通り。3 個とも赤球である場合の数は、${}_5C_3$ 通り。よって、$P(A) = \frac{{}_5C_3}{{}_8C_3} = \frac{5}{28}$. 定理 2 より、求める確率は、

$$P(\overline{A}) = 1 - P(A) = 1 - \frac{5}{28} = \frac{23}{28}.$$

□

問題 1.

12 本のくじの中に当たりくじが 4 本ある。この中から 3 本のくじを同時に引くとき、次の確率を答えよ。

(1) 2 本当たる確率

(2) 少なくとも 1 本当たる確率

解答.

(1) 12 本のくじから 3 本のくじを引く場合の数は、${}_{12}C_3$ 通り、当たり 2 本とはずれ 1 本を引く場合の数は、${}_4C_2 \cdot {}_8C_1$ 通り。よって、求める確率は、$\frac{{}_4C_2 \cdot {}_8C_1}{{}_{12}C_3} = \frac{12}{55}$.

(2) 少なくとも 1 本当たるという事象は、3 本ともはずれであるという事象 A の余事象 $\overline{A}$ である。3 本ともはずれである場合の数は、${}_8C_3$ 通り。よって、$P(A) = \frac{{}_8C_3}{{}_{12}C_3} = \frac{14}{55}$. 定理 2 より、求める確率は、$P(\overline{A}) = 1 - P(A) = 1 - \frac{14}{55} = \frac{41}{55}$.

□

確率・統計　第2回　～独立試行と反復試行の確率～

定義 6　独立試行

いくつかの試行の結果が互いに他の試行の結果に影響を及ぼさないとき、これらの試行は互いに独立であるという。互いに独立な試行をまとめた試行を独立試行という。

例 2. 10本のくじの中に当たりくじが3本ある。この中から A, B の2人が順に1本ずつくじを引く試行を、それぞれ T_1, T_2 とする。

(1) A が引いたくじを戻してから B がくじを引く（復元抽出）とき、T_1 と T_2 は独立である。

(2) A が引いたくじを戻さないで B がくじを引く（非復元抽出）とき、T_1 と T_2 は独立でない。

定理 3　独立試行の確率

互いに独立な n 個の試行 $T_1, \cdots, T_n$ について、各 T_i で事象 A_i が起こる確率を $P(A_i)$ とする。このとき、$T_1, \cdots, T_n$ をまとめた独立試行 T において、各 T_i で A_i が起こる事象を A とすると、その確率は

$$P(A) = P(A_1) \cdot \cdots \cdot P(A_n).$$

証明.

各試行 T_i の全事象を U_i, 独立試行 T の全事象を U とおく。T において起こりうる全ての場合の数は、$n(U) = n(U_1) \cdot \cdots \cdot n(U_n)$ 通り。T において事象 A が起こる場合の数は、$n(A) = n(A_1) \cdot \cdots \cdot n(A_n)$ 通り。よって、

$$P(A) = \frac{n(A)}{n(U)} = \frac{n(A_1) \cdot \cdots \cdot n(A_n)}{n(U_1) \cdot \cdots \cdot n(U_n)} = \frac{n(A_1)}{n(U_1)} \cdot \cdots \cdot \frac{n(A_n)}{n(U_n)} = P(A_1) \cdot \cdots \cdot P(A_n).$$

□

例題 4　独立試行の確率

赤球3個と白球2個が入っている袋 A と、赤球4個と白球1個が入っている袋 B から、球を1個ずつ取り出すとき、2個の球の色が同じである確率を求めよ。

解答.

袋 A から球を1個取り出す試行を T_1, 袋 B から球を1個取り出す試行を T_2, これらををまとめた独立試行を T とおく。T において、2個の球の色が同じである事象 E は、2個の球の色が赤である事象 F と2個の球の色が白である事象 G の和事象である。T_1 において球の色が赤である事象を F_1, T_2 において球の色が赤である事象を F_2 とすると、定理3より、

$$P(F) = P(F_1) \cdot P(F_2) = \frac{3}{5} \cdot \frac{4}{5} = \frac{12}{25}.$$

また、T_1 において球の色が白である事象を G_1, T_2 において球の色が白である事象を G_2 とすると、定理 3 より、

$$P(G) = P(G_1) \cdot P(G_2) = \frac{2}{5} \cdot \frac{1}{5} = \frac{2}{25}.$$

F と G は互いに排反であるから、加法定理 1 より、求める確率は、$P(E) = P(F \cup G) = P(F) + P(G) = \frac{12}{25} + \frac{2}{25} = \frac{14}{25}$. □

定義 7　反復試行

同じ試行を何回か繰り返し行うことを、反復試行という。

反復試行における各回の試行は、互いに独立である。

定理 4　反復試行の確率

試行 T について、事象 A が起こる確率を p とする。このとき、T を n 回繰り返す反復試行において、A が r 回起こる確率 q は

$$q = {}_nC_r p^r (1-p)^{n-r}.$$

証明.

n 回の試行のうち事象 A が r 回起こる場合の数は、${}_nC_r$ 通り。反復試行における各回の試行は互いに独立であるから、各場合の確率は $p^r(1-p)^{n-r}$. 各場合は互いに排反であるから、加法定理 1 より、求める確率は、

$$q = \underbrace{p^r(1-p)^{n-r} + \cdots + p^r(1-p)^{n-r}}_{{}_nC_r} = {}_nC_r p^r (1-p)^{n-r}.$$

□

例題 5　反復試行の確率

さいころを 5 回続けて投げるとき、2 以下の目が 3 回出る確率を求めよ。

解答.

さいころを 1 回投げるとき、2 以下の目が出る確率は $p = \frac{2}{6}$. 定理 4 より、求める確率は、

$$q = {}_5C_3 (\frac{2}{6})^3 (1 - \frac{2}{6})^{5-3} = \frac{40}{243}.$$

□

問題 2.

赤球 5 個と白球 3 個が入っている袋がある。この中から 4 個の球を取り出すとき、赤球が 2 個、白球が 2 個となる確率を、次の場合について答えよ。

(1) 1回目に2個の球を同時に取り出して、色を確認して袋の中に戻し、2回目に2個の球を同時に取り出す場合

(2) 1個ずつ4回球を取り出し、取り出した球はそのつど色を確認して袋の中に戻す場合

解答.

(1) 『1回目に赤球2個を取り出し、2回目に白球2個を取り出す』という事象を A, 『1回目に赤球1個と白球1個を取り出し、2回目も赤球1個と白球1個を取り出す』という事象を B, 『1回目に白球2個を取り出し、2回目に赤球2個を取り出す』という事象を C とすると、求める確率は A, B, C の和事象 $A \cup B \cup C$ の確率 $P(A \cup B \cup C)$ である。1回目に球を取り出すとき、赤球2個を取り出す確率は、$\frac{{}_5C_2}{{}_8C_2}$, 2回目に球を取り出すとき、白球2個を取り出す確率は、$\frac{{}_3C_2}{{}_8C_2}$. 1回目に球を取り出す試行と2回目に球を取り出す試行は独立であるから、定理3より、$P(A) = \frac{{}_5C_2}{{}_8C_2} \cdot \frac{{}_3C_2}{{}_8C_2} = \frac{15}{392}$. 同様に、$P(B) = \frac{{}_5C_1 \cdot {}_3C_1}{{}_8C_2} \cdot \frac{{}_5C_1 \cdot {}_3C_1}{{}_8C_2} = \frac{225}{784}$, $P(C) = \frac{{}_3C_2}{{}_8C_2} \cdot \frac{{}_5C_2}{{}_8C_2} = \frac{15}{392}$.
事象 A, 事象 B, 事象 C は互いに排反であるから、加法定理1より、$P(A \cup B \cup C) = P(A) + P(B) + P(C) = \frac{15}{392} + \frac{225}{784} + \frac{15}{392} = \frac{285}{784}$.

(2) 1回球を取り出すとき、赤球を取り出す確率は、$\frac{5}{8}$. 定理4より、求める確率は、${}_4C_2(\frac{5}{8})^2(1-\frac{5}{8})^{4-2} = \frac{675}{2048}$.

□

確率・統計　第3回　～条件付き確率～

定義 8　条件付き確率

試行における空でない事象 A, 事象 B に対して、A が起こったという条件のもとで B が起こる確率を、条件 A のもとでの B の条件付き確率といい、$P_A(B)$ で表す:

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

条件 A のもとでの B の条件付き確率は、A を新たに全事象としたときの事象 $A \cap B$ の確率と考えられる。

例 3. 赤球3個と白球2個が入っている袋から、2個の球を1個ずつ元に戻さないで取り出すとき、1回目に赤球が出たという条件のもとで2回目に白球が出る条件付き確率について考える。

(1) 1回目に赤球が出る事象を A, 2回目に白球が出る事象を B とする。事象 A が起こる場合の数 $n(A)$ は、1回目の赤球の出方3通りの各々について2回目の球の出方4通りが考えられるから、$n(A) = 3 \cdot 4 = 12$. 事象 $A \cap B$ が起こる場合の数 $n(A \cap B)$ は、1回目の赤球の出方3通りの各々について2回目の白球の出方2通りが考えられるから、$n(A \cap B) = 3 \cdot 2 = 6$. よって、1回目に赤球が出たという条件のもとで2回目に白球が出る条件付き確率は、

$$P_A(B) = \frac{n(A \cap B)}{n(A)} = \frac{6}{12} = \frac{1}{2}.$$

(2) 1回目に赤球が出たとき、袋の中には赤球2個と白球2個が入っているから、その中から白球を取り出す確率は、

$$\frac{2}{2+2} = \frac{1}{2}.$$

1回目に赤球が出たという条件のもとで2回目に白球が出る条件付き確率は、その確率に等しいので、$P_A(B) = \frac{1}{2}$.

定理 5　乗法定理

試行における空でない事象 A, 事象 B に対して、次が成り立つ:

$$P(A \cap B) = P(A) \cdot P_A(B).$$

証明.

定義より、

$$P_A(B) = \frac{n(A \cap B)}{n(A)} = \frac{\frac{n(A \cap B)}{n(U)}}{\frac{n(A)}{n(U)}} = \frac{P(A \cap B)}{P(A)}.$$

□

例 4. くじ引きの、引く順番と当たる確率について考える。n 本のくじの中に当たりくじが m 本あるとき、1回目に引いた人が当たる事象を A. 2回目に引いた人が当たる事象を B とする。このとき、$P(A)=\frac{m}{n}$. また、$P(\overline{A})=1-\frac{m}{n}=\frac{n-m}{n}$. 1回目に当たりが出たとき、残りのくじの中には当たりくじ $m-1$ 本が入っているから、その中から当たりくじを引く確率は、$\frac{m-1}{n-1}$. よって、1回目に当たりが出たという条件のもとで2回目に当たりが出る条件付き確率は、その確率に等しいので、$P_A(B)=\frac{m-1}{n-1}$. 1回目にはずれが出たとき、残りのくじの中には当たりくじ m 本が入っているから、その中から当たりくじを引く確率は、$\frac{m}{n-1}$. ゆえに、1回目にはずれが出たという条件のもとで2回目に当たりが出る条件付き確率は、その確率に等しいので、$P_{\overline{A}}(B)=\frac{m}{n-1}$. 乗法定理 *5* より、

$$P(A\cap B)=P(A)\cdot P_A(B)=\frac{m}{n}\cdot\frac{m-1}{n-1}=\frac{m(m-1)}{n(n-1)},$$

$$P(\overline{A}\cap B)=P(\overline{A})\cdot P_{\overline{A}}(B)=\frac{n-m}{n}\cdot\frac{m}{n-1}=\frac{m(n-m)}{n(n-1)}.$$

$A\cup\overline{A}=U,\ A\cap\overline{A}=\emptyset$ であるから、加法定理 *1* より、

$$\begin{aligned}P(B)&=P(U\cap B)=P((A\cup\overline{A})\cap B) && \leftarrow A\cup\overline{A}=U\\&=P((A\cap B)\cup(\overline{A}\cap B))=P(A\cap B)+P(\overline{A}\cap B) && \leftarrow A\cap\overline{A}=\emptyset,\ \text{加法定理 } 1\\&=\frac{m(m-1)}{n(n-1)}+\frac{m(n-m)}{n(n-1)}=\frac{m}{n}.\end{aligned}$$

したがって、くじ引きの当たる確率は引く順番によらない。

例題 6 乗法定理

赤球3個と白球5個が入っている袋から、2個の球を1個ずつ元に戻さないで取り出すとき、2回目に初めて白球が出る確率を求めよ。

解答.

1回目に赤球が出る事象を A. 2回目に白球が出る事象を B とすると、2回目に初めて白球が出る事象は、A と B の積事象 $A\cap B$ となる。8個の球から2個の球を1個ずつ元に戻さないで取り出す場合の数は、$8\cdot 7=56$ 通り。事象 A が起こる場合の数は、1回目の赤球の出方3通りの各々について2回目の球の出方7通りが考えられるから、$3\cdot 7=21$ 通り。よって、$P(A)=\frac{21}{56}=\frac{3}{8}$. 1回目に赤球が出たとき、袋の中には赤球2個と白球5個が入っているから、その中から白球を取り出す確率は、$\frac{5}{2+5}=\frac{5}{7}$. 1回目に赤球が出たという条件のもとで2回目に白球が出る条件付き確率は、その確率に等しいので、$\frac{5}{7}$. 乗法定理5より、求める確率は、

$$P(A\cap B)=P(A)\cdot P_A(B)=\frac{3}{8}\cdot\frac{5}{7}=\frac{15}{56}.$$

□

定義 9　事象の独立性

試行における空でない事象 A, B に対して、次の関係を満たすとき A と B は独立であるという:

$$P_A(B) = P(B),\ P_B(A) = P(A).$$

事象 A と事象 B が独立であるとき、互いにもう一方の起こる確率に影響を及ぼさない。

定理 6　事象の独立性

試行における空でない事象 A, B に対して、A と B が独立であるための必要十分条件は、次が成り立つことである:

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

証明.

(必要性) A と B は独立であるから、$P_A(B) = P(B)$. 乗法定理 5 より、

$$P(A \cap B) = P(A) \cdot P_A(B) = P(A) \cdot P(B).$$

(十分性) 仮定から、$P(A \cap B) = P(A) \cdot P(B)$. 乗法定理 5 より、

$$P_A(B) = \frac{P(A \cap B)}{P(A)} = \frac{P(A) \cdot P(B)}{P(A)} = P(B).$$

同様にして、$P_B(A) = P(A)$. □

例 5. さいころを投げるとき、奇数の目が出る事象を A, 3 以下の目が出る事象を B, 4 または 5 の目が出る事象を C とする。

(1) A と B の積事象 $A \cap B$ は 1 または 3 の目が出ることであるから、

$$P(A \cap B) = \frac{2}{6} \neq \frac{3}{6} \cdot \frac{3}{6} = P(A) \cdot P(B).$$

よって、A と B は独立でない。

(2) A と C の積事象 $A \cap C$ は 5 の目が出ることであるから、

$$P(A \cap C) = \frac{1}{6} = \frac{3}{6} \cdot \frac{2}{6} = P(A) \cdot P(C).$$

よって、A と C は独立である。

定義 10　事前確率と事後確率

試行における空でない事象 A, B に対して、結果 B の原因として A が考えられているとする。このとき、$P_B(A)$ を事前確率といい、$P(A)$ を事後確率という。

事前確率 $P_B(A)$ は、結果 B の原因が A である確率を表している。結果 B の原因として、全ての場合を尽くし、かつ同時に起こらない原因 $A_1, \cdots, A_n$ が考えられているとする。各原因 A_j によって結果 B が生じる因果関係の確率 $P_{A_j}(B)$ がわかっているとき、事前確率 $P_B(A_i)$ は次のように計算される。

定理 7 Bayes **の定理**

試行における空でない事象 $A_1, \cdots, A_n$, B に対して、$U = A_1 \cup \cdots \cup A_n$, $A_i \cap A_j = \emptyset$, $i \neq j$ とする。このとき、任意の $i = 1, \cdots, n$ に対して、次が成り立つ:

$$P_B(A_i) = \frac{P(A_i) \cdot P_{A_i}(B)}{\sum_{j=1}^{n} P(A_j) \cdot P_{A_j}(B)}.$$

証明.

乗法定理 5 より、$P(A_i \cap B) = P(A_i) \cdot P_{A_i}(B) = P(B) \cdot P_B(A_i)$. $U = A_1 \cup \cdots \cup A_n$, $A_i \cap A_j = \emptyset$, $i \neq j$ であるから、加法定理 1 より、

$$\begin{aligned}
P(B) = P(U \cap B) &= P((A_1 \cup \cdots \cup A_n) \cap B) & \leftarrow U = A_1 \cup \cdots \cup A_n \\
&= P((A_1 \cap B) \cup \cdots \cup (A_n \cap B)) & \\
&= \sum_{j=1}^{n} P(A_j \cap B) & \leftarrow A_i \cap A_j = \emptyset, i \neq j,\ \text{加法定理 1.}
\end{aligned}$$

ここで、乗法定理 5 より、$P(A_j \cap B) = P(A_j) \cdot P_{A_j}(B)$, $j = 1, \cdots, n$. よって、乗法定理 5 より、

$$P_B(A_i) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i \cap B)}{\sum_{j=1}^{n} P(A_j \cap B)} = \frac{P(A_i) \cdot P_{A_i}(B)}{\sum_{j=1}^{n} P(A_j) \cdot P_{A_j}(B)}.$$

□

例題 7 Bayes **の定理**

ある製品を工場 a で 40%, 工場 b で 60% 生産したとき、工場 a では 1%, 工場 b では 2% の不合格品が出る。取り出した製品が不合格品であるとき、それが工場 a で生産されたものである確率を求めよ。

解答.

工場 a の製品であるという事象を A, 工場 b の製品であるという事象を B, 不合格品であるという事象を C とする。製品は工場 a で 40%, 工場 b で 60% 生産されているから、$P(A) = \frac{40}{100}$, $P(B) = \frac{60}{100}$. 工場 a の製品は 1% の割合で不合格品が出るから、工場 a の製品を取り出したとき、それが不合格品である確率は、$P_A(C) = \frac{1}{100}$. 同様に、$P_B(C) = \frac{2}{100}$. 不合格品を取り出したとき、それが工場 a の製品である確率は、Bayes の定理 7 より、

$$P_C(A) = \frac{P(A) \cdot P_A(C)}{P(A) \cdot P_A(C) + P(B) \cdot P_B(C)} = \frac{\frac{40}{100} \cdot \frac{1}{100}}{\frac{40}{100} \cdot \frac{1}{100} + \frac{60}{100} \cdot \frac{2}{100}} = \frac{1}{4}.$$

□

問題 3.

ある製品を工場 a で 60%, 工場 b で 40% 生産したとき、工場 a では 2%, 工場 b では 3% の不合格品が出る。取り出した製品が不合格品であるとき、それが工場 a で生産されたものである確率を求めよ。

解答.

工場 a の製品であるという事象を A, 工場 b の製品であるという事象を B, 不合格品であるという事象を C とする。製品は工場 a で 60%, 工場 b で 40% 生産されているから、$P(A)=\frac{60}{100}$, $P(B)=\frac{40}{100}$. 工場 a の製品は 2% の割合で不合格品が出るから、工場 a の製品を取り出したとき、それが不合格品である確率は、$P_A(C)=\frac{2}{100}$. 同様に、$P_B(C)=\frac{3}{100}$. 不合格品を取り出したとき、それが工場 a の製品である確率は、Bayes の定理 7 より、$P_C(A)=\frac{P(A)\cdot P_A(C)}{P(A)\cdot P_A(C)+P(B)\cdot P_B(C)}=\frac{\frac{60}{100}\cdot\frac{2}{100}}{\frac{60}{100}\cdot\frac{2}{100}+\frac{40}{100}\cdot\frac{3}{100}}=\frac{1}{2}$. □

確率・統計　第4回　〜データの整理〜

定義 11　母集団と標本

- 母集団: 統計調査の対象全体
- 標本: 母集団から取り出した要素の集まり
- 抽出: 母集団から標本を取り出すこと
- 標本の大きさ: 標本に含まれる要素の個数
- 無作為抽出: 母集団のどの要素も標本に抽出される確率が等しくなるような抽出方法
- 復元抽出: 母集団から標本を抽出するとき、標本を元に戻しながら抽出を行うこと
- 非復元抽出: 母集団から標本を抽出するとき、標本を元に戻さずに抽出を行うこと

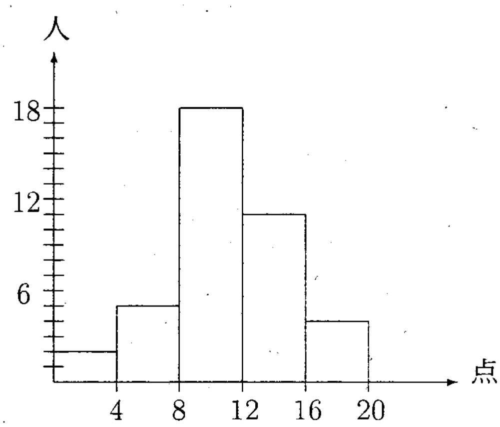
母集団から標本を抽出し、その標本から統計的手法によって母集団の特性を推し量ることを統計的推測という。このとき、抽出した標本を整理する必要がある。

定義 12　度数分布表とヒストグラム

標本がとる数値の範囲をいくつかの区間に分けて、各区間に含まれる標本の個数を数える。この区間を階級といい、階級の中央の値を階級値、階級の端点の値の差を階級の幅という。また、階級に含まれる標本の個数を度数という。このとき得られる、階級別の度数の分布を表した表を度数分布表という。度数分布表から、階級の幅を底辺、度数を高さとする長方形を並べた柱状のグラフをヒストグラムという。

例 6. あるクラスの数学のテストの結果を、度数分布表とヒストグラムにまとめた。

階級（得点）以上〜未満	度数（人数）
0〜4	2
4〜8	5
8〜12	18
12〜16	11
16〜20	4



標本の特徴を適当な1つの数値で表すことがあり、そのような値を代表値という。以下、標本が次のような度数分布表に整理されているとき、その代表値について調べる。

階級 以上〜未満	階級値	度数
$a_0 \sim a_1$	x_1	f_1
$\vdots$	$\vdots$	$\vdots$
$a_{i-1} \sim a_i$	x_i	f_i
$\vdots$	$\vdots$	$\vdots$
$a_{n-1} \sim a_n$	x_n	f_n
計		N

まず、標本の中心的な値について考えてみる。

定義 13　標本の平均値

標本の平均値 $\overline{x}$ を次のように定める:

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{n} x_i f_i.$$

度数分布表から得られる平均値は、個々の標本がとる数値をそれが含まれる階級の階級値に置き換えて、平均を計算している。標本の大きさが十分大きく、階級の幅が十分小さいとき、度数分布表から得られる平均値と、実際の個々の標本がとる数値から得られる平均値には、大きな差は生じない。

定義 14　標本の中央値

標本の中央値 Me を次のように定める: $\sum_{k=1}^{m-1} f_k \leq \frac{N}{2} < \sum_{k=1}^{m} f_k$ を満たす m に対して、

$$M_e = a_{m-1} + (a_m - a_{m-1}) \cdot \frac{\frac{N}{2} - \sum_{k=1}^{m-1} f_k}{f_m}.$$

定義 15　標本の最頻値

度数の最も大きい階級の階級値を最頻値といい、Mo と表す。

注意 1. 標本の中で、他の標本がとる数値に比べて極端に大きかったり小さかったりする少数の標本の値を、外れ値という。平均値は、全ての標本の値を使って計算しているため外れ値の影響を受けやすく、このような場合、中央値や最頻値の方が標本を特徴付ける代表値としてふさわしい。あるクラスの学生の貯金高を調べて、度数分布表に整理した。

階級（円） 以上〜未満	階級値（円）	度数（人数）
0〜1000	500	1
1000〜2000	1500	2
2000〜3000	2500	4
3000〜4000	3500	1
4000〜5000	4500	1
5000〜6000	5500	0
⋮	⋮	⋮
998000〜999000	998500	0
999000〜1000000	999500	1
計		10

このとき、平均値は、

$$\begin{aligned}\overline{x} = \frac{1}{10}(500 \cdot 1 + 1500 \cdot 2 + 2500 \cdot 4 + 3500 \cdot 1 + 4500 \cdot 1 + 5500 \cdot 0 + \cdots \\ \cdots + 998500 \cdot 0 + 999500 \cdot 1) \\ = 102100\end{aligned}$$

であり、貯金高が 999000〜1000000 円（外れ値）の学生 1 人が全体の平均を上げてしまい、標本の特徴を表す代表値に適しているとは言えない。一方、中央値は、

$$M_e = 2000 + (3000 - 2000) \cdot \frac{\frac{10}{2} - (1+2)}{4} = 2500$$

となり、こちらの方が標本の代表値としてふさわしい。

次に、標本のばらつきを表す値について考えてみる。個々の標本がとる数値と平均値の差を偏差という。偏差の平均を計算すると、

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{n}(x_i - \overline{x})f_i &= \frac{1}{N}\sum_{i=1}^{n} x_i f_i - \frac{\overline{x}}{N}\sum_{i=1}^{n} f_i = \overline{x} - \frac{\overline{x}}{N} \cdot N \qquad \leftarrow \frac{1}{N}\sum_{i=1}^{n} x_i f_i = \overline{x}, \sum_{i=1}^{n} f_i = N \\ &= \overline{x} - \overline{x} = 0\end{aligned}$$

となり、これを用いて標本のばらつきを表すことはできない。そこで、偏差の 2 乗の平均を使って、標本のばらつきを表すことを考える。

定義 16 標本の分散と標準偏差

標本の分散 s^2 と標準偏差 s を次のように定める:

$$s^2 = \frac{1}{N}\sum_{i=1}^{n}(x_i - \overline{x})^2 f_i, \qquad s = \sqrt{s^2}.$$

分散は計算の過程で2乗するため、標本がとる数値と同じ単位を持つばらつきを表す指標として、分散の正の平方根である標準偏差を用いることが多い。分散は、標本の値の2乗の平均値から平均値の2乗を引いた値に等しい。

定理 8　標本の分散

標本の分散 s^2 について、次が成り立つ:

$$s^2 = \overline{x^2} - \overline{x}^2.$$

証明.

定義より、

$$\begin{aligned}
s^2 &= \frac{1}{N}\sum_{i=1}^{n}(x_i - \overline{x})^2 f_i = \frac{1}{N}\sum_{i=1}^{n}(x_i^2 - 2x_i\overline{x} + \overline{x}^2) f_i = \frac{1}{N}\sum_{i=1}^{n} x_i^2 f_i - \frac{2\overline{x}}{N}\sum_{i=1}^{n} x_i f_i + \frac{\overline{x}^2}{N}\sum_{i=1}^{n} f_i \\
&= \overline{x^2} - 2\overline{x}\cdot\overline{x} + \frac{\overline{x}^2}{N}\cdot N \qquad \leftarrow \frac{1}{N}\sum_{i=1}^{n} x_i^2 f_i = \overline{x^2}, \frac{1}{N}\sum_{i=1}^{n} x_i f_i = \overline{x}. \sum_{i=1}^{n} f_i = N \\
&= \overline{x^2} - \overline{x}^2.
\end{aligned}$$

□

例題 8　標本の平均値と標準偏差

例 6 の度数分布表から、平均値、中央値、最頻値、分散、標準偏差を求めよ。

解答.

階級	階級値 x	度数 f
0〜4	2	2
4〜8	6	5
8〜12	10	18
12〜16	14	11
16〜20	18	4
計		40

平均値は、$\overline{x} = \frac{1}{40}(2\cdot 2 + 6\cdot 5 + 10\cdot 18 + 14\cdot 11 + 18\cdot 4) = 11.$

中央値は、$M_e = 8 + (12-8)\cdot\frac{\frac{40}{2}-(2+5)}{18} = \frac{98}{9}.$

最頻値は、$M_o = 10.$

x^2 の平均値は、$\overline{x^2} = \frac{1}{40}(2^2\cdot 2 + 6^2\cdot 5 + 10^2\cdot 18 + 14^2\cdot 11 + 18^2\cdot 4) = 136.$

定理 8 より、分散は、$s^2 = \overline{x^2} - (\overline{x})^2 = 136 - 11^2 = 15.$

よって、標準偏差は、$s = \sqrt{s^2} = \sqrt{15}.$ □

問題 4.

次の度数分布表は、あるクラスの英語の試験の採点結果である。平均値と標準偏差を求めよ。

得点	$0 \sim 20$	$20 \sim 40$	$40 \sim 60$	$60 \sim 80$	$80 \sim 100$
人数	2	4	6	5	3

解答.

階級	階級値 x	度数 f
$0 \sim 20$	10	2
$20 \sim 40$	30	4
$40 \sim 60$	50	6
$60 \sim 80$	70	5
$80 \sim 100$	90	3
計		20

平均値は、$\bar{x} = \frac{1}{20}(10 \times 2 + 30 \times 4 + 50 \times 6 + 70 \times 5 + 90 \times 3) = 53$. x^2 の平均値は、$\bar{x^2} = \frac{1}{20}(10^2 \times 2 + 30^2 \times 4 + 50^2 \times 6 + 70^2 \times 5 + 90^2 \times 3) = 3380$. 定理 8 より、分散は、$s^2 = \bar{x^2} - (\bar{x})^2 = 3380 - 53^2 = 571$. よって、標準偏差は、$\sqrt{s^2} = \sqrt{571} = 23.8\cdots$. □